LIMITING SPECTRAL DISTRIBUTION FOR NON-HERMITIAN RANDOM MATRICES WITH A VARIANCE PROFILE

NICHOLAS COOK, WALID HACHEM, JAMAL NAJIM AND DAVID RENFREW

Abstract. For each $n$, let $A_n = (\sigma_{ij})$ be an $n \times n$ deterministic matrix and let $X_n = (X_{ij})$ be an $n \times n$ random matrix with i.i.d. centered entries of unit variance. We study the asymptotic behavior of the empirical spectral distribution $\mu^Y_n$ of the rescaled entry-wise product

$$Y_n = \left( \frac{1}{\sqrt{n}} \sigma_{ij} X_{ij} \right).$$

For our main result we provide a deterministic sequence of probability measures $\mu_n$, each described by a family of Master Equations, such that the difference $\mu^Y_n - \mu_n$ converges weakly in probability to the zero measure. A key feature of our results is to allow some of the entries $\sigma_{ij}$ to vanish, provided that the standard deviation profiles $A_n$ satisfy a certain quantitative irreducibility property. Examples are also provided where $\mu^Y_n$ converges to a genuine limit. As an application, we extend the circular law to a class of random matrices whose variance profile is doubly stochastic.

Contents

1. Introduction 1
2. Presentation of the results 5
3. Important special cases, remarks and open questions 17
4. Asymptotics of singular values distributions 28
5. Proof of Theorem 2.2: Analysis of the Master Equations 36
6. Proof of Theorem 2.3-(1): Tail estimates and asymptotics of the logarithmic potential 41
7. Conclusion of proofs of Theorems 2.3 and 2.4 49
8. Boundedness of solutions to the Regularized Master Equations 53
Appendix A. Remaining proofs for Section 3 64
Appendix B. Remaining proofs for Section 4 71
Appendix C. Remaining proofs for Section 6 82
References 83

1. Introduction

For an $n \times n$ matrix $M$ with complex entries and eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ (counted with multiplicity and labeled in some arbitrary fashion), the empirical spectral distribution (ESD) is given by

$$\mu_M = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}.$$ (1.1)
A seminal result in non-Hermitian random matrix theory is the circular law, which describes the asymptotic global distribution of the spectrum for matrices with i.i.d. entries of finite variance. The following strong form of the circular law was established by Tao and Vu [65], and is the culmination of the work of many authors [33, 50, 34, 17, 14, 36, 53, 66] – see the survey [22] for a detailed historical account.

**Theorem 1.1** (Circular law). Let $\xi$ be a complex random variable of zero mean and unit variance, and for each $n$ let $X_n = (X_{ij}^{(n)})$ be an $n \times n$ matrix whose entries are i.i.d. copies of $\xi$. Then almost surely, the ESDs $\mu_{1/n} X_n$ converge weakly to the circular measure

$$
\mu_{\text{circ}}(dx \, dy) := \frac{1}{\pi} 1_{\{x^2 + y^2 \leq 1\}} \, dx \, dy.
$$

One of the remarkable features of the circular law is that the asymptotic behavior of ESDs is insensitive to specific details of the entry distributions, apart from the first two moments. This is an instance of the universality phenomenon in random matrix theory.

The circular law has been an important tool for understanding the stability of dynamical systems on complex networks, going back to work of May in ecology [49], and later work of Sompolinski et al. in neuroscience [63]. May used an i.i.d. matrix $X_n$ to model the community matrix for a food network of $n$ species, where the entry $X_{ij}$ determines the rate of growth (or decay) of the population of species $i$ due to species $j$. The stability of the system is determined by the spectrum of $X_n$ – specifically by whether it has eigenvalues with sufficiently large real part – and May used the circular law to derive a criterion for stability.

Recently there has been increasing interest in extending the arguments of [49, 63] to matrix models with more structured distributions. In neural networks, where random matrices are used to model the synaptic matrix, the work [56] considered perturbed i.i.d. matrices of the form $X_n + M_n$, where $M_n$ is a fixed matrix with all entries within a fixed proportion of columns taking a fixed positive value $\mu^+$, and all remaining entries taking a fixed negative value $\mu^-$. Their motivation was to conform to Dale’s Law, stating that neurons are either inhibitory or excitatory. In this case $M_n$ is a rank-one perturbation; as was later shown rigorously in work of Tao [64], low rank perturbations do not affect the limiting spectral distribution, but may lead to the creation of outlier eigenvalues.

Several recent works have studied the limiting spectral distribution for random matrices of the form

$$
A_n \odot X_n = (\sigma_{ij} X_{ij})
$$

(suitably rescaled) where $A_n = (\sigma_{ij})$ is a fixed (deterministic) standard deviation profile. From a modeling perspective the $\sigma_{ij}$ can reflect the varying degrees of interaction between species/neurons. In theoretical ecology the works [7, 9] considered asymmetric standard deviation profiles, i.e. taking $\sigma_{ij} \neq \sigma_{ji}$, in order to create more realistic predator-prey cascading relationships. In neuroscience the works [6, 5] considered matrices $A_n$ partitioned into a bounded number of block submatrices having constant entries within each block, in order to model networks with a bounded number of cell types. We also note that predating these works, Girko [35, Chapter 25, 26] (see also the references therein) studied non-Hermitian matrices with standard deviation and mean profiles and provided canonical equations to describe the limiting spectral densities.

Some works have also gone beyond matrices with independent entries of specified mean and variance, for instance considering products and sums of deterministic matrices with a random matrix having i.i.d. entries [1], or allowing correlations between entries $X_{ij}, X_{ji}$ [8]. We also mention that parallel to the study of non-Hermitian matrices there have been many works devoted to the study of

\[ 1 \] At the time the circular law was only known to hold in the complex Gaussian case thanks to work of Ginibre and Mehta [33, 50]. Strictly speaking, May’s argument assumes that there are asymptotically no eigenvalues outside the limiting support, which is now known to hold under some moment hypotheses [18, 21].
Hermitian random matrices with a variance profile, both Wigner and Gram-type – see for instance Girko [35, Chapter 7, 8], Shlyakhtenko [61], Guionnet [37], Anderson and Zeitouni [12], Hachem et al. [39], Ajanki et al. [2].

As has been pointed out in the ecology literature [9], a key feature that is missing from the literature on models of the form (1.2) is to allow $A_n$ to have zero entries. Indeed, the nodes in large real-world ecological or neural networks do not interact with all other nodes. One fix has been to take $A_n$ to have i.i.d. Bernoulli($p$) indicator entries, independent of $X_n$, i.e. to model the support of the network by a sparse Erdős–Rényi digraph. As was shown by Wood [70] the circular law still holds for $A_n \odot X_n$ (after rescaling by $(pn)^{-1/2}$ if $p \geq n^{\alpha-1}$ for any fixed $\alpha \in (0,1]$. However, the valence of the nodes in the resulting network is highly concentrated around $pn$, while the valence distribution for real-world networks is highly non-uniform [9]. With the ability to set $A_n$ deterministically one can reflect some known underlying geometry of the network.

Our aim in the present work is to rigorously establish the limiting spectral distribution for a broad class of random matrices of the form (1.2). Key features of our results, which also present significant obstacles for the proofs, are to allow a large number of the entries $\sigma_{ij}$ to be zero, as well as to allow asymmetry (i.e. $\sigma_{ij} \neq \sigma_{ji}$).

As compared with the proof of the circular law, the identification and description of a limiting measure is significantly more involved. In this article we prove the existence of a sequence of deterministic measures – called deterministic equivalents – which asymptotically approximate the random ESDs. In particular we obtain non-trivial information even when the ESDs themselves do not converge to a limit. The identification of the deterministic equivalents involves analysis of a (cubic) polynomial system of Master Equations determined by the variance profile. A relative of the Master Equations known as the Quadratic Vector Equation was studied in recent work of Ajanki, Erdős and Krüger on the spectrum of Hermitian matrices with a variance profile [4, 10].

Since the initial release of this paper, a local law version of our main statement (Theorem 2.3) was proved in [11] under the restriction that the standard deviation profile $\sigma_{ij}$ is uniformly strictly positive and that the distribution of the matrix entries possesses a bounded density and has all its moments finite. In this case, it is also proved that the density of the deterministic equivalents is positive and bounded on its support. As will be shown hereafter, these properties might not be true anymore if the standard deviation profile has zero entries or is not uniformly lower bounded. In these cases, the limiting distribution may offer a wider variety of behavior, such as a blowup or vanishing density at zero, or a point mass at zero.

It is by now well-known that the study of ESDs for non-Hermitian random matrices is intimately connected with proving the quantitative invertibility of such matrices – that is, establishing lower tail estimates for small singular values. The possible sparsity of the matrices considered here gives rise to significant challenges for this task. Bounds on the smallest singular value sufficient for our purposes were established by the first author in [24]. In the present work we obtain control on the remaining small singular values from Wegner-type bounds, which are established by a quantitative analysis of the Master Equations. As in the analysis of the smallest singular value in [24], our argument for the Wegner estimates makes heavy use of expansion properties of a directed graph naturally associated to the variance profile. We discuss these aspects of the proof in more detail after presenting the results in Section 2.

1.1. The model. In this article we study the following general class of random matrices with non-identically distributed entries.

**Definition 1.2 (Random matrix with a variance profile).** For each $n \geq 1$, let $A_n$ be a (deterministic) $n \times n$ matrix with entries $\sigma_{ij}^{(n)} \geq 0$, let $X_n$ be a random matrix with i.i.d. entries $X_{ij}^{(n)} \in \mathbb{C}$ satisfying

$$\mathbb{E}X_{11}^{(n)} = 0, \quad \mathbb{E}|X_{11}^{(n)}|^2 = 1$$

(1.3)
and set
\[ Y_n = \frac{1}{\sqrt{n}} A_n \odot X_n \] (1.4)
where \( \odot \) is the matrix Hadamard product, i.e. \( Y_n \) has entries \( Y_{ij}^{(n)} = \frac{1}{\sqrt{n}} \sigma_{ij}^{(n)} X_{ij}^{(n)} \). The empirical spectral distribution of \( Y_n \) is denoted by \( \mu_n^Y \). We refer to \( A_n \) as the standard deviation profile and to \( A_n \odot A_n = (\sigma_{ij}^{(n)})^2 \) as the variance profile. We additionally define the normalized variance profile as
\[ V_n = \frac{1}{n} A_n \odot A_n. \]
When no ambiguity occurs, we will drop the index \( n \) and simply write \( \sigma_{ij}, X_{ij}, V, \) etc.

**Remark 1.1.** Note we do not assume the variables \( X_{ij}^{(n)} \) are independent or identically distributed for different values of \( n \).

Our goal is to describe the asymptotic behavior of the ESDs \( \mu_n^Y \) given a sequence of standard deviation profiles \( A_n \). In general the matrices \( A_n \) can be sparse, and may not converge in any sense to a limiting variance profile.

1.2. Master equations and deterministic equivalents. The main result of this article states that under certain assumptions on the sequence of standard deviation profiles \( A_n \) and the distribution of the entries of \( X_{ij} \), there exists a sequence of deterministic probability measures \( \mu_n \) that are deterministic equivalents of the spectral measures \( \mu_n^Y \) in the sense that for every test function, e.g. a continuous and bounded function \( f : \mathbb{C} \to \mathbb{C} \) (in fact a compactly supported function will suffice),
\[
\int f \, d\mu_n^Y - \int f \, d\mu_n \longrightarrow 0 \quad \text{in probability}.
\]
In other words, the signed measures \( \mu_n^Y - \mu_n \) converge weakly in probability to zero. In the sequel this convergence will be simply denoted by
\[
\mu_n^Y \sim \mu_n \quad \text{in probability} \quad (n \to \infty).
\]

The measures \( \mu_n \) are described by a polynomial system of Master Equations. Denote by \( V_n^T \) the transpose matrix of \( V_n \) and by \( \rho(V_n) \) its spectral radius. For a parameter \( s \geq 0 \), the Master Equations are the following system of \( 2n + 1 \) equations in \( 2n \) unknowns \( q_1, \ldots, q_n, \tilde{q}_1, \ldots, \tilde{q}_n \):
\[
\begin{align*}
q_i &= \frac{(V_n^T q)_i}{s^2 + (V_n q)_i (V_n^T q)_i} \\
\tilde{q}_i &= \frac{(V_n \tilde{q})_i}{s^2 + (V_n q)_i (V_n^T q)_i} \quad q_i, \tilde{q}_i \geq 0, \quad i \in [n],
\end{align*}
\]
(1.5)
where \( q, \tilde{q} \) are the \( n \times 1 \) column vectors with components \( q_i, \tilde{q}_i \), respectively.

If \( s \geq \sqrt{\rho(V_n)} \), it can be proved that the Master Equations admit the unique trivial solution \((q, \tilde{q}) = 0\). Provided that \( 0 < s < \sqrt{\rho(V_n)} \) and that the matrix \( V_n \) is irreducible, it can be shown that the Master Equations admit a unique positive solution \((q, \tilde{q})\) which depends only on \( s \). This solution \( s \mapsto (q(s), \tilde{q}(s)) \) is continuous on \((0, \infty)\). With this definition of \( q \) and \( \tilde{q} \), the deterministic equivalent \( \mu_n \) is defined as the radially symmetric probability distribution over \( \mathbb{C} \) satisfying
\[
\mu_n \{ z \in \mathbb{C} , \ |z| \leq s \} = 1 - \frac{1}{n} q^T(s) V_n \tilde{q}(s), \quad s > 0.
\]
It readily follows that the support of \( \mu_n \) is contained in the disk of radius \( \sqrt{\rho(V_n)} \).
1.3. **Specialization to specific variance profiles.** We now briefly describe some specific cases of interest. These examples will be expanded upon in Sections 2 and 3.

**Doubly stochastic variance profile.** In the case where the variance profile is doubly stochastic, i.e.

\[
\frac{1}{n} \sum_i \sigma^2_{ij} = \frac{1}{n} \sum_j \sigma^2_{ij} = 1 \quad \text{for all } 1 \leq i, j \leq n,
\]

with \(\sigma_{ij}\) uniformly bounded from above, our main theorem yields the convergence of \(\mu^Y_n\) toward the circular law \(\mu_{\text{circ}}\); see Theorem 2.4. This parallels results of Girko [35, §7.11, §8.2] and Anderson and Zeitouni [12] (see also [61], [37] for the Gaussian case) for random Hermitian matrices with a circular law \(\mu\) with \(\sigma\).

**Separable variance profile.** Consider two arrays of positive real numbers \((d_i^{(n)}, i \leq n, n \geq 1)\) and \((\tilde{d}_i^{(n)}, i \leq n, n \geq 1)\). Denote by \(D_n = \text{diag}(d_i^{(n)}, 1 \leq i \leq n)\) and \(\tilde{D}_n = \text{diag}(\tilde{d}_i^{(n)}, 1 \leq i \leq n)\) two \(n \times n\) diagonal matrices and by \(d_n = (d_i^{(n)})_{1 \leq i \leq n}, \tilde{d}_n = (\tilde{d}_i^{(n)})_{1 \leq i \leq n}\) two associated \(n \times 1\) vectors. Then the matrix model

\[
Y_n = \frac{1}{\sqrt{n}} D_n X_n \tilde{D}_n
\]

admits a separable variance profile in the sense that \(\text{var}(Y_{ij}^{(n)}) = \frac{1}{n} d_i^{(n)} \tilde{d}_i^{(n)}\). This model falls into Definition 1.2 with \(A_n \odot A_n = d_n \tilde{d}_n^T\). In this case the \(2n\) Master Equations (??) simplify to a single equation with unknown \(u_n(s)\):

\[
\frac{1}{n} \sum_{i \in [n]} \frac{d_i \tilde{d}_i}{s^2 + d_i \tilde{d}_i u_n(s)} = 1, \quad \text{and} \quad \mu_n\{\xi \in \mathbb{C}, |\xi| \leq s\} = 1 - u_n(s)
\]

for \(s \geq 0\); see Theorems 3.1 and 3.2. As a particular instance of this model, we recover Girko's sombrero probability distribution [35, Section 26.12].

**Sampled variance profile.** One may also like to consider sequences of standard deviation profiles which are converging in some sense to a limiting profile. A natural way to do this is to obtain the matrices \(A_n\) by evaluating a fixed continuous function \(\sigma(x, y)\) on the unit square at the grid points \(\{(i/n, j/n): 1 \leq i, j \leq n\}\).

Here, in the large \(n\) limit the Master Equations (1.5) turn into an integral equation defining a genuine limit for the ESDs:

\[
\mu^Y_n \xrightarrow{n \to \infty} \mu^\sigma
\]

weakly in probability; see Theorem 3.3.

**Acknowledgements.** The work of NC was partially supported by NSF grants DMS-1266164 and DMS-1606310. The work of WH and JN was partially supported by the program “modèles numériques” of the French Agence Nationale de la Recherche under the grant ANR-12-MONU-0003 (project DIONISOS). Support of the Labex BEZOUT from the Université Paris Est is also acknowledged.

## 2. Presentation of the results

2.1. **Notational preliminaries.** Denote by \([n]\) the set \(\{1, \ldots, n\}\) and let \(\mathbb{C}_+ = \{z \in \mathbb{C}, \text{Im}(z) > 0\}\). For \(\mathcal{X} = \mathbb{C}\) or \(\mathbb{R}\), let \(C_c(\mathcal{X})\) (resp. \(C_c^\infty(\mathcal{X})\)) the set of \(\mathcal{X} \to \mathbb{R}\) continuous (resp. smooth) and compactly supported functions. Let \(B(z, r)\) be the open ball of \(\mathbb{C}\) with center \(z\) and radius \(r\). If \(z \in \mathbb{C}\), then \(\bar{z}\) is its complex conjugate; let \(i^2 = -1\). The Lebesgue measure on \(\mathbb{C}\) will be either denoted by \(\ell(dz)\) or \(dxdy\). For \(x, y \in \mathbb{R}\) we write \(\max(x, y) = x \vee y\) and \(\min(x, y) = x \wedge y\). The
cardinality of a finite set \( S \) is denoted by \(|S|\). For \( S \subset \mathbb{N} \) and when clear from the context we will abbreviate \( S^c = \mathbb{N} \setminus S \).

### 2.1.1. Matrices

We denote by \( \mathbf{1}_n \) the \( n \times 1 \) vector of 1’s. Given two \( n \times 1 \) vectors \( \mathbf{u}, \mathbf{v} \), we denote their scalar product \( \langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i \in [n]} \bar{u}_i v_i \).

For a given matrix \( A \), denote by \( A^\top \) its transpose and by \( A^* \) its conjugate transpose. Denote by \( I_n \) the \( n \times n \) identity matrix. If clear from the context, we omit the dimension. For \( a \in \mathbb{C} \) and when clear from the context, we sometimes write \( a \) instead of \( a I \) and similarly write \( a^* \) instead of \((aI)^* = \bar{a}I\). For matrices \( A, B \) of the same dimensions we denote by \( A \odot B \) their Hadamard, or entrywise, product (i.e. \((A \odot B)_{ij} = A_{ij} B_{ij}\)).

Given two Hermitian matrices \( A \) and \( B \), the notations \( A \succeq B \) and \( A > B \) refer to the usual positive semidefinite ordering. Notations \( \succcurlyeq \) and \( \succcurlyeq \) refer to the elementwise inequalities for real matrices or vectors. Namely, if \( A \) and \( B \) are real matrices,

\[
A \succeq B \quad \iff \quad A_{ij} \geq B_{ij} \quad \forall i, j \quad \mathrm{and} \quad A \succcurlyeq B \quad \iff \quad A_{ij} \geq B_{ij} \quad \forall i, j.
\]

The notation \( A \succcurlyeq 0 \) stands for \( A \succeq 0 \) and \( A \neq 0 \). Given a matrix \( A \), \( \|A\| \) refers to its spectral norm, and \( \|A\|_\infty \) to its max-row norm, defined as:

\[
\|A\|_\infty := \max_{i \in [n]} \sum_{j=1}^n |A_{ij}|.
\]

We denote the spectral radius of an \( n \times n \) matrix \( A \) by

\[
\rho(A) := \max \big\{ |\lambda| : \lambda \text{ is an eigenvalue of } A \big\}.
\]

If \( A \) is a square matrix, we write \( \text{Im}(A) = (A - A^*)/(2i) \). \( \text{diag}(a_i : 1 \leq a_i \leq n) \) denotes the \( n \times n \) diagonal matrix with the \( a_i \)'s as its diagonal elements.

### 2.1.2. Convergence of measures

Given probability distributions \( \nu_n, \nu \) over some set \( \mathcal{X} (= \mathbb{R} \text{ or } \mathbb{C}) \), we will denote the weak convergence of \( \nu_n \) to \( \nu \) by \( \nu_n \xrightarrow{w} \nu \). If \( \nu_n \) is random, \( \nu_n \xrightarrow{w} \nu \) almost surely (resp. in probability) stands for the fact that for all \( f \in C_c(\mathcal{X}) \),

\[
\int f \, d\nu_n \xrightarrow{n \to \infty} \int f \, d\nu \quad \text{almost surely (resp. in probability)}.
\]

Let \((\mu_n)\) and \((\nu_n)\) be deterministic sequences of probability distributions over \( \mathcal{X} \), and let \((\nu_n)\) be tight, i.e. for all \( \varepsilon > 0 \), one can find a compact set \( K_\varepsilon \) such that

\[
\sup_n \nu_n(\mathcal{X} \setminus K_\varepsilon) \leq \varepsilon.
\]

We will denote by

\[
\mu_n \sim \nu_n \quad \text{as} \quad n \to \infty
\]

the fact that the signed measure \( \mu_n - \nu_n \) weakly converges to zero, i.e. \( \int f \, d\mu_n - \int f \, d\nu_n \to 0 \) for all \( f \in C_c(\mathcal{X}) \). If the sequence \((\mu_n)\) is random while \((\nu_n)\) is deterministic and tight, then

\[
\mu_n \sim \nu_n \quad \text{almost surely (resp. in probability)}
\]

stands for

\[
\int f \, d\nu_n - \int f \, d\nu \xrightarrow{n \to \infty} 0 \quad \text{almost surely (resp. in probability)},
\]

for all \( f \in C_c(\mathcal{X}) \).

Notice that \( \mu_n \sim \nu_\infty \) is equivalent to \( \mu_n \xrightarrow{w} \nu_\infty \).
2.1.3. Stieltjes transforms. Let $\mu$ be a nonnegative finite measure on $\mathbb{R}$ and
\[ g_\mu(\eta) = \int \frac{\mu(d\lambda)}{\lambda - \eta}, \quad \eta \in \mathbb{C}^+ \] (2.2)
its Stieltjes transform. Then the following properties are standard

(i) $g_\mu(\eta) \in \mathbb{C}^+$,
(ii) $|g_\mu(\eta)| \leq \mu(\mathbb{R}) \cdot \text{Im}(\eta)$,
(iii) $\lim_{y \to +\infty} -iyg_\mu(iy) = \mu(\mathbb{R})$.

Moreover, $-(z + g_\mu(z))^{-1}$ is the Stieltjes transform of a probability measure, see for instance [69, Theorem B.3]. In particular
\[ \left| \frac{1}{z + g_\mu(z)} \right| \leq \frac{1}{\text{Im}(z)}, \quad z \in \mathbb{C}^+. \] (2.3)

2.1.4. Graph theoretic notation. Given an $n \times n$ non-negative matrix $A = (\sigma_{ij})$ we form a directed graph $\Gamma = \Gamma(A)$ on the vertex set $[n]$ that puts an edge $i \to j$ whenever $\sigma_{ij} > 0$. We denote the out-neighborhood of a vertex $i \in [n]$ in the graph $\Gamma$ by
\[ \mathcal{N}_A(i) := \{ j \in [n] : \sigma_{ij} > 0 \}. \] (2.4)
Consequently, the in-neighborhood is denoted $\mathcal{N}_A^T(i)$. For a set $S \subset [n]$ we write
\[ \mathcal{N}_A(S) := \bigcup_{i \in S} \mathcal{N}_A(i) = \{ j \in [n] : \mathcal{N}_A^T(j) \cap S \neq \emptyset \}. \] (2.5)

For $\delta \in (0, 1)$ we denote the associated densely-connected out-neighbors of a set $S \subset [n]$ by
\[ \mathcal{N}_A^{(\delta)}(S) = \{ j \in [n] : |\mathcal{N}_A^T(j) \cap S| \geq \delta|S| \}. \] (2.6)

To obtain quantitative results we will generally work with the graph associated to the matrix
\[ A(\sigma_0) = (\sigma_{ij}1_{\sigma_{ij} \geq \sigma_0}) \] (2.7)
which only keeps the entries exceeding a fixed cutoff parameter $\sigma_0 > 0$, setting the remaining entries to zero.

2.2. Model assumptions. We will establish results concerning sequences of matrices $Y_n$ as in Definition 1.2 under various additional assumptions on $A_n$ and $X_n$, which we now summarize. We note that many of our results only require a subset of these assumptions.

For our main result we will need the following additional assumption on the distribution of the entries of $X_n$.

**A0 (Moments).** We have $\mathbb{E}|X_{11}^{(n)}|^{4+\varepsilon} \leq M_0$ for all $n \geq 1$ and some fixed $\varepsilon > 0$, $M_0 < \infty$.

We will also assume the entries of $A_n$ are bounded uniformly in $i, j \in [n], n \geq 1$:

**A1 (Bounded variances).** There exists $\sigma_{\max} \in (0, \infty)$ such that
\[ \sup_n \max_{1 \leq i,j \leq n} \sigma_{ij}^{(n)} \leq \sigma_{\max}. \]

**Remark 2.1 (Convention).** While we will keep the value $\sigma_{\max}$ generic in the statements, we will always set it to 1 in the proofs, with no loss of generality.
Remark 2.2 (Sparse matrices). While our assumptions allow any fixed proportion of the entries \( \sigma_{ij} \) to be zero, the above assumption precludes us from making nontrivial statements about matrices with a vanishing proportion of nonzero entries. Indeed, by the Weyl comparison inequality (cf. e.g. [44, Theorem 3.3.13]) we have in this case that

\[
\mathbb{E} \int_{\mathbb{C}} |z|^2 \mu_n^Y(dz) \leq \mathbb{E} \|Y_n\|_{\text{HS}}^2 = \frac{1}{n^2} \sum_{i,j=1}^n \sigma_{ij}^2 \leq \sigma_{\max} \frac{1}{n^2} \{ (i,j) : \sigma_{ij} > 0 \} \rightarrow 0.
\]

Consequently, the empirical spectral distributions \( \mu_n^Y \) converge weakly in probability to \( \delta_0 \), the point mass at the origin. To obtain a nontrivial limit would require a rescaling of the matrices \( Y_n \), which amounts to rescaling \( A_n \) to have entries of growing size, violating \( \text{A1} \). It would thus be interesting to relax this assumption – see Section 3.5 for further discussion.

In order to express the next key assumption, we need to introduce the following Regularized Master Equations which are a specialization of the Schwinger–Dyson equations of Girko’s Hermitized model associated to \( Y_n \) (see Section 2.5 for further discussion). The following is proved in Section 5.1.

**Proposition 2.1 (Regularized Master Equations).** Let \( n \geq 1 \) be fixed, let \( A_n \) be an \( n \times n \) nonnegative matrix and write \( V_n = \frac{1}{n} A_n \odot A_n \). Let \( s, t > 0 \) be fixed, and consider the following system of equations

\[
\begin{align*}
    r_i &= \frac{(V_n^T r)_i + t}{s^2 + ((V_n r)_i + t)((V_n^T r)_i + t)}, \\
    \tilde{r}_i &= \frac{(V_n \tilde{r})_i + t}{s^2 + ((V_n r)_i + t)((V_n^T r)_i + t)}
\end{align*}
\]

(2.8)

where \( r = (r_i) \) and \( \tilde{r} = (\tilde{r}_i) \) are \( n \times 1 \) vectors. Denote by \( \tilde{r} = \begin{pmatrix} r \\ \tilde{r} \end{pmatrix} \). Then this system admits a unique solution \( \tilde{r} = \tilde{r}(s, t) > 0 \). This solution satisfies the identity

\[
\sum_{i \in [n]} r_i = \sum_{i \in [n]} \tilde{r}_i.
\]

(2.9)

**A2 (Admissible variance profile).** Let \( \tilde{r}(s, t) = \tilde{r}_n(s, t) > 0 \) be the solution of the Regularized Master Equations for given \( n \geq 1 \). For all \( s > 0 \), there exists a constant \( C = C(s) > 0 \) such that

\[
\sup_{n \geq 1} \sup_{t \in (0, 1]} \frac{1}{n} \sum_{i \in [n]} r_i(s, t) \leq C.
\]

A variance profile \( V_n \) for which the previous estimate holds is called admissible.

Remark 2.3. Assumption \( \text{A2} \) may seem obscure at first sight as it necessitates to solve the regularized master equations to check whether a variance profile is admissible or not. In particular, it is not clear if this assumption is compatible with some sparsity. In section 2.4, we provide sufficient conditions over the variance profile \( V_n \) which imply \( \text{A2} \), namely \( \text{A3} \) (lower bound on \( V_n \)), \( \text{A4} \) (symmetric \( V_n \)) and \( \text{A5} \) (robust irreducibility for \( V_n \)).

2.3. **Statement of the results.** Recall the Master equations (1.5), and notice that these equations are obtained from the Regularized Master Equations (2.8) by letting the parameter \( t \) go to zero. Notice however that condition \( \sum q_i = \sum \tilde{q}_i \) is now required for uniqueness and not a consequence as in (2.8).
The following compact notation for the Master Equations will sometimes be more convenient.

For two \( n \times 1 \) vectors \( \mathbf{a} \) and \( \tilde{\mathbf{a}} \) with nonnegative components, let \( \tilde{\mathbf{a}}^T = (\mathbf{a}^T \tilde{\mathbf{a}}^T) \) and define

\[
\Psi(\tilde{\mathbf{a}}) = \Psi_{A_n,s}(\tilde{\mathbf{a}}) := \text{diag} \left( \frac{1}{s^2 + (V_n \tilde{\mathbf{a}})_i(V_n^T \mathbf{a})_i} ; i \in [n] \right) = \text{diag}(\psi_i(\tilde{\mathbf{a}}) ; i \in [n])
\]

and

\[
\mathcal{I}(\tilde{\mathbf{a}}) := \begin{pmatrix} \Psi(\tilde{\mathbf{a}}) V_n^T & 0 \\ 0 & \Psi(\tilde{\mathbf{a}}) V_n \end{pmatrix} \tilde{\mathbf{a}}.
\] (2.10)

Then (1.5) can alternatively be expressed

\[
\begin{aligned}
\begin{cases}
\tilde{\mathbf{q}} & = \mathcal{I}(\tilde{\mathbf{q}}) \\
\langle 1_n, \mathbf{q} \rangle & = \langle 1_n, \tilde{\mathbf{q}} \rangle , \quad \tilde{\mathbf{q}} \succ 0.
\end{cases}
\end{aligned}
\] (2.11)

In order to prove existence of solutions \( \mathbf{q}, \tilde{\mathbf{q}} \) to the Master Equations we will need to assume the standard deviation profile \( A_n \) is irreducible. This is equivalent to assuming the associated digraph \( \Gamma(A_n) \) is strongly connected (recall the graph theoretic notation from Section 2.1). This will cause no loss of generality, as we can conjugate \( V_n \) by an appropriate permutation matrix to put \( A_n \) in block-upper-triangular form with irreducible blocks on the diagonal – the spectrum of \( Y_n \) is then the union of the spectra of the block diagonal submatrices.

**Theorem 2.2** (Master equations). Let \( n \geq 1 \) be fixed, let \( A_n \) be an \( n \times n \) nonnegative matrix and write \( V_n = \frac{1}{n} A_n \odot A_n \). Assume that \( A_n \) is irreducible. Then the following hold:

1. For \( s \geq \sqrt{\rho(V_n)} \) the system (2.11) has the unique solution \( \tilde{\mathbf{q}}(s) = 0 \).
2. For \( s \in (0, \sqrt{\rho(V)}) \) the system (2.11) has a unique non-trivial solution \( \tilde{\mathbf{q}}(s) \succ neq 0 \). Moreover, this solution satisfies \( \tilde{\mathbf{q}}(s) \succ 0 \).
3. The function \( s \mapsto \tilde{\mathbf{q}}(s) \) defined in parts (1) and (2) is continuous on \((0, \infty)\) and is continuously differentiable on \((0, \sqrt{\rho(V)}) \cup (\sqrt{\rho(V)}, \infty)\).

**Remark 2.4** (Convention). Above and in the sequel we abuse notation and write \( \tilde{\mathbf{q}} = \tilde{\mathbf{q}}(s) \) to mean a solution of the equation (2.11), understood to be the nontrivial solution for \( s \in (0, \sqrt{\rho(V)}) \).

**Remark 2.5** (Solving the Master Equations numerically). The proof of Theorem 2.2 involves the study of the solution \( \tilde{\mathbf{r}} = \mathcal{I}(\tilde{\mathbf{r}}, t) \) to the Regularized Master Equations – see Proposition 2.1, where \( t > 0 \) is a regularization parameter. These equations also provide a numerical means of obtaining an approximate value of the solution of (1.5) via the iterative procedure \( \tilde{\mathbf{r}}_{k+1} = \mathcal{I}(\tilde{\mathbf{r}}_k, t) \), obtained for a small value of \( t \). However, the convergence of this procedure becomes slower as \( t \downarrow 0 \). To circumvent this issue, one can solve the system for relatively large \( t \) and then increment \( t \) down to zero, using the previous solution as the new initial vector \( \tilde{\mathbf{r}}_0 \). In Section 3.4 we present numerical solutions obtained in this way for some specific examples of variance profiles.

The main result of this paper is the following.

**Theorem 2.3** (Main result). Let \( (Y_n)_{n \geq 1} \) be a sequence of random matrices as in Definition 1.2, and assume A0, A1 and A2 hold. Assume moreover that \( A_n \) is irreducible for all \( n \geq 1 \).

1. There exists a sequence of deterministic measures \( (\mu_n)_{n \geq 1} \) on \( \mathbb{C} \) such that
   \[
   \mu_n^Y \sim \mu_n \quad \text{in probability}.
   \]
2. Let \( \tilde{\mathbf{q}}(s)^T = (\mathbf{q}(s)^T \tilde{\mathbf{q}}(s)^T) \) be as in Theorem 2.2, and for \( s \in (0, \infty) \) put
   \[
   F_n(s) = 1 - \frac{1}{n} \langle \mathbf{q}(s), V_n \tilde{\mathbf{q}}(s) \rangle.
   \] (2.12)

Then \( F_n \) extends to an absolutely continuous function on \([0, \infty)\) which is the CDF of a probability measure with support contained in \([0, \sqrt{\rho(V_n)}]\) and continuous density on \((0, \sqrt{\rho(V_n)})\).
(3) For each $n \geq 1$ the measure $\mu_n$ from part (1) is the unique radially symmetric probability measure on $\mathbb{C}$ with $\mu_n(\{z : |z| \leq s\}) = F_n(s)$ for all $s \in (0, \infty)$.

Remark 2.6 (Almost sure convergence under different hypotheses). As in works on the circular law, a key component of the proof is a lower tail estimate for the smallest singular value of scalar shifts $Y_n - zI_n$ of the form

$$P(s_n(Y_n - zI_n) \leq n^{-\beta}) = O(n^{-\alpha})$$

(2.13)

holding for a.e. fixed $z \in \mathbb{C}$. (Crucially, we do not need such an estimate for every $z \in \mathbb{C}$, as our assumption A2 only requires $s = |z| > 0$ and allows variance profiles for which (2.13) is false when $z = 0$; see Proposition 3.5) Such a bound is available for arbitrary fixed $z \neq 0$ and $\alpha > 0$ a small constant by recent work of the first author [24]; see Proposition 6.1. Obtaining (2.13) with $\alpha > 1$ would immediately improve conclusion (1) to almost sure convergence by an application of the Borel–Cantelli lemma. It is possible that such an improvement could be obtained by incorporating tools of the Inverse Littlewood–Offord theory developed by Tao–Vu [67, 66], Rudelson–Vershynin [58] and Nguyen–Vu [52] for matrices with i.i.d. entries. Also, such improvements to the bound (2.13) are already available under stronger assumptions on $A_n$ and $X_n$ than we make here. For instance, under A3 and replacing A0 with a bounded density assumption, an easy argument gives (2.13) for any fixed $\alpha > 0$ and some $\beta = \beta(\alpha) > 0$; see [22, Section 4.4].

Remark 2.7 (Moment assumptions). The moment assumption A0 is needed in order to apply the results from [24] to bound the smallest singular value as in (2.13). It is also used to Section 4 to quantitatively bound the difference between our random measures and their deterministic equivalents, which is crucial for obtaining logarithmic integrability of singular value distributions. This latter step can likely be accomplished with fewer moments, but we do not pursue this.

Remark 2.8 (Density of $\mu_n$ versus density of $F_n$). In the previous theorem, the density $\varphi_n$ of $\mu_n$ for $0 < |z| < \sqrt{\rho \nu}$ is given by the formula:

$$\varphi_n(|z|) = \frac{1}{2\pi|z|} \frac{d}{ds} F_n(s) \bigg|_{s=|z|} = -\frac{1}{2\pi n|z|} \frac{d}{ds} \langle q(s), V \overline{q}(s) \rangle \bigg|_{s=|z|} .$$

In fact, let $\mu$ be a rotationally invariant probability measure on $\mathbb{C}$, with density $\varphi(|z|)$, i.e. $\mu(A) = \int_A \varphi(|z|) \ell(dz)$. Let $F$ be the cumulative distribution function of a probability measure with support in $\mathbb{R}^+$, such that

$$\mu\{z : |z| \leq s\} = F(s) = \int_0^s f(u) \, du .$$

By classical polar change of coordinates, one obtains

$$\mu\{z : |z| \leq s\} = \int_{\{z : |z| \leq s\}} \varphi(|z|) \ell(dz) = 2\pi \int_0^s \varphi(u) u \, du = \int_0^s f(u) \, du .$$

Identifying both integrals yields the formula

$$\varphi(|z|) = \frac{1}{2\pi|z|} f(|z|) .$$

As an illustration we show how our results recover the circular law for matrices with i.i.d. entries.

Example 2.1 (The circular law). Consider a standard deviation profile $A_n$ with all elements equal to 1 and assume that A0 holds. It is well-known in this case that $\mu_n \xrightarrow{w} \mu_{\text{circ}}$ in probability (and even almost surely), where $\mu_{\text{circ}}$ stands for the circular law with density $\pi^{-1} I_{\{1 \leq |z| \leq 1\}}$. We can recover this result with Theorem 2.3. In this case, both systems (2.8) and (1.5) simplify into a single equation:

$$r_i \equiv r = \frac{r + t}{s^2 + (r + t)^2}, \quad r > 0 \quad \text{and} \quad q_i \equiv q = \frac{q}{s^2 + q^2}, \quad q \geq 0 .$$

(2.14)
From the first equation, one can prove that \( r(s, t) \leq 1 \) for \( t \in (0, 1] \). In fact,

\[
r = \frac{r + t}{s^2 + (r + t)^2} \leq \frac{1}{r + t} \implies r^2 + rt \leq 1 \implies r \leq 1.
\]

Hence A2 is fulfilled. The second equation has the unique nontrivial solution

\[
q(s) = \begin{cases} \sqrt{1 - s^2} & 0 \leq s \leq 1 \\ 0 & s \geq 1 \end{cases}
\]  

(2.15)

Consequently, \( F_n(s) = s^2 \) for \( s \leq 1 \). From Remark 2.8, we conclude the desired convergence.

In the next example, we prove that a doubly stochastic normalized variance profile is admissible and that the associated deterministic equivalent \( \mu_n \) is the circular law.

**Example 2.2 (Doubly stochastic variance profile).** Assume that the normalized variance profiles \( V_n \) is doubly stochastic, i.e.

\[
\frac{1}{n} \sum_{i} \sigma_{ij}^2 = \frac{1}{n} \sum_{j} \sigma_{ij}^2 = 1 \quad \text{for all} \quad 1 \leq i, j \leq n.
\]

Then, one quickly verifies that the vectors \( \vec{r} = r1 \) and \( \vec{q} = q1 \) with \( r, q \) as in (2.14) respectively satisfy the Regularized Master Equations and the Master Equations. As a consequence A2 can be established as in Example 2.1. Let now A1 hold and assume that the variance profile \( V_n \) is irreducible for all \( n \geq 1 \) then one can apply Theorem 2.3 with \( \mu_n \) equal to the circular law.

**Remark 2.9.** Note that under A1 the doubly stochastic condition implies that the number of non-zero entries in each row and column is linear in \( n \).

In the following theorem, we relax the irreducibility assumption, which requires some additional argument.

**Theorem 2.4 (The circular law for doubly stochastic variance profiles).** Let \( (Y_n)_{n \geq 1} \) be a sequence of random matrices as in Definition 1.2, and assume A0 and A1 hold. Suppose also that the normalized variance profiles \( V_n \) are doubly stochastic, i.e.

\[
\frac{1}{n} \sum_{i} \sigma_{ij}^2 = \frac{1}{n} \sum_{j} \sigma_{ij}^2 = 1 \quad \text{for all} \quad 1 \leq i, j \leq n.
\]

Then \( \mu_{Y_n} \xrightarrow{w} \mu_{\text{circ}} \) in probability.

More elaborate applications of our results are provided in Section 3.

### 2.4. Sufficient conditions for admissibility.

Hereafter we introduce a series of assumptions directly checkable over the variance profiles \( (V_n) \) without solving a priori the regularized master equations. These assumptions enforce A2.

The simplest such assumption is to enforce uniform positivity of the variances, which allows one to bypass some of the most technical portions of our argument. This assumption was also made in the recent work [11] on the local law.

**A3** (Lower bound on variances). There exists \( \sigma_{\min} > 0 \) such that

\[
\inf_n \min_{1 \leq i, j \leq n} \sigma_{ij}^{(n)} \geq \sigma_{\min}.
\]

We generalize A3 below with the expansion-type condition A5.

**Proposition 2.5.** Let \( A = (\sigma_{ij}) \) be an \( n \times n \) matrix with entries \( \sigma_{ij} \geq \sigma_{\min} > 0 \) for some \( \sigma > 0 \). Let \( \vec{r} \succ 0 \) be the unique solution of the Regularized Master Equations (2.8). Then

\[
\frac{1}{n} \sum_{i=1}^{n} r_i \leq \frac{1}{\sigma_{\min}}.
\]

In particular, if \( A_n = (\sigma_{ij}^{(n)}) \) is a sequence of standard deviation profiles as in Definition 1.2 for which A3 holds, then A2 is satisfied, i.e. \( V_n \) is admissible.
A4 (Symmetric variance profile). For all \( n \geq 1 \), the normalized variance profile (or equivalently the standard deviation profile) is symmetric:
\[
V_n = V_n^T.
\]

Proposition 2.6. Let \( A = (\sigma_{ij}) \) be a symmetric matrix with nonnegative entries, and let \( \bar{r} > 0 \) be the unique solution of the Regularized Master Equations (2.8). Then
\[
\frac{1}{n} \sum_{i=1}^{n} r_i \leq \frac{1}{2s}.
\]

In particular, if \( A_n = (\sigma_{ij}(n)) \) is a sequence of standard deviation profiles as in Definition 1.2 for which A4 holds, then A2 is satisfied.

We now introduce the following strengthening of the irreducibility assumption, which can be understood as a kind of expansion condition on an associated directed graph.

Definition 2.7 (Robust irreducibility). For \( \delta, \kappa \in (0, 1) \) we say that a nonnegative \( n \times n \) matrix \( A \) is \((\delta, \kappa)\)-robustly irreducible if the following hold:

1. For all \( i \in [n] \),
\[
|\mathcal{N}_A(i)|, |\mathcal{N}_A^T(i)| \geq \delta n.
\]

2. For all \( S \subset [n] \) with \( 1 \leq |S| \leq n - 1 \),
\[
|\mathcal{N}^{(\delta)}_A(S) \cap S^c| \geq \min(\kappa|S|, |S^c|).\]

For comparison, a nonnegative \( n \times n \) matrix \( A \) is irreducible if and only if \( \mathcal{N}_A^T(S) \cap S^c \neq \emptyset \) for all \( S \subset [n] \) with \( 1 \leq |S| \leq n - 1 \). Thus, a matrix \( A \) satisfying the conditions of Definition 2.7 is "robustly irreducible" in the sense that \( A \) remains irreducible even after setting a small linear proportion of entries equal to zero.

Remark 2.10 (Relation to broad connectivity). In their work on permanent estimators, Rudelson and Zeitouni assume a stronger expansion-type condition on \( A(\sigma_{ij}) \) which they call broad connectivity [59]. The conditions for a nonnegative square matrix \( A \) to be \((\delta, \kappa)\)-broadly connected are the same as in Definition 2.7, except (2.17) is replaced by the stronger condition
\[
|\mathcal{N}^{(\delta)}_A(S)| \geq \min(n, (1 + \kappa)|S|)
\]
for all nonempty \( S \subset [n] \).

Recall the definition (2.7) of \( A(\sigma_0) \).

A5 (Robust irreducibility). There exist constants \( \sigma_0, \delta, \kappa \in (0, 1) \) such that for all \( n \geq 1 \), \( A_n(\sigma_0) \) is \((\delta, \kappa)\)-robustly irreducible.

Notice that A3 implies A5 but A5 enables variance profiles with vanishing entries.

Theorem 2.8. Consider a sequence of standard deviation profiles \( A_n = (\sigma_{ij}(n)) \) as in Definition 1.2, and assume that A1 and A5 hold. Then A2 holds, i.e. \( V_n \) is admissible.

Notice that the mere irreducibility of \( V_n \) provides a weaker form of A2.

Proposition 2.9. Let \( V_n \) be an irreducible variance profile and let \( \bar{r} = \bar{r}(s, t) \) be the solution of the associated Regularized Master Equations (2.8). Then there exists \( C = C(s, n) \) such that
\[
\sup_{[0,1]} \frac{1}{n} \sum_{i \in [n]} r_i(s, t) \leq C.
\]
The main difference here is that constant $C$ depends on $n$ and may explode with $n$. Depending on the variance profile, this proposition is sometimes sufficient to verify \textbf{A2}.

\textbf{Example 2.3} (Variance profile with a block structure). Let $k \geq 1$ be a fixed integer, and $M = (m_{ij})_{i,j \in [k]}$ be a $k \times k$ irreducible matrix with nonnegative elements. Let $J_m = 1_m 1_m^T$. Assume that $n = km$ ($m \geq 1$) and consider the $n \times n$ matrix

$$V_n = \frac{1}{n} \begin{pmatrix} m_{11} J_{m} & \cdots & m_{1k} J_{m} \\ \vdots & \ddots & \vdots \\ m_{k1} J_{m} & \cdots & m_{kk} J_{m} \end{pmatrix}.$$ \hfill (2.19)

Then the variance profile $V_n$ is admissible, i.e. \textbf{A2} is fulfilled. In fact, $V_n$ is irreducible and its block structure implies that

$$r^T = (\rho_1, \ldots, \rho_1, \ldots, \rho_k, \ldots, \rho_k), \quad \tilde{r}^T = (\tilde{\rho}_1, \ldots, \tilde{\rho}_1, \ldots, \tilde{\rho}_k, \ldots, \tilde{\rho}_k)$$

where $\rho = (\rho_i)$ and $\tilde{\rho} = (\tilde{\rho}_i)$ satisfy the $2k$ equations

$$\rho_i = \frac{(M_k \rho)_i + t}{s^2 + ((M_k \rho)_i + t)(M_k \rho)_i + t)} \quad \text{and} \quad \tilde{\rho}_i = \frac{(M_k \tilde{\rho})_i + t}{s^2 + ((M_k \tilde{\rho})_i + t)(M_k \tilde{\rho})_i + t}, \quad i \in [k]$$

with $M_k = \frac{1}{k} M$. In particular,

$$\sup_{t \in (0,1]} \frac{1}{n} \sum_{i \in [n]} r_i(s, t) = \sup_{t \in (0,1]} \frac{1}{k} \sum_{i \in [k]} \rho_i(s, t),$$

where the latter is finite by Proposition 2.9 and does not depend on $n$, hence \textbf{A2}.

\textbf{2.5. Outline of the proof}. As is well-known in the literature devoted to large non-Hermitian random matrices, the spectral behavior of such matrices can be studied with the help of the so-called Girko’s Hermitization procedure, which is intimately related with the logarithmic potential of their spectral measure [34]. By definition, the logarithmic potential $U_\mu$ of a probability measure $\mu$ on $\mathbb{C}$ which integrates $\log | \cdot |$ near infinity is the function from $\mathbb{C}$ to $(-\infty, \infty]$ defined by

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

Writing $z = x + iy$, the Laplace operator is $\Delta = \partial^2_{xx} + \partial^2_{yy} = 4 \partial_x \partial_y$, where

$$\partial_x = \frac{1}{2} (\partial_x - i \partial_y) \quad \text{and} \quad \partial_y = \frac{1}{2} (\partial_x + i \partial_y).$$

The probability measure $\mu$ can be recovered by the formula

$$\mu = - \frac{1}{2\pi} \Delta U_\mu,$$

valid in the set $\mathcal{D}'(\mathbb{C})$ of the Schwartz distributions, which means that

$$\int_{\mathbb{C}} \psi(z) d\mu(dz) = - \frac{1}{2\pi} \int_{\mathbb{C}} \Delta \psi(z) U_\mu(z) dz \quad \text{for all} \quad \psi \in C^\infty_c(\mathbb{C}).$$

Now, setting $\mu = \mu_n^Y$, the logarithmic potential can be written as

$$U_{\mu_n^Y}(z) = - \frac{1}{n} \sum_{i=1}^{n} \log | \lambda_i - z | = - \frac{1}{n} \log | \det (Y_n - z) | = - \frac{1}{n} \log \sqrt{\det (Y_n - z)(Y_n - z)^*}$$

$$= - \int_{0}^{\infty} \log(x) L_{n,z}(dx),$$
where $L_{n,z} := \frac{1}{n} \sum_{i=1}^{n} \delta_{s_{i,z}}$ is the empirical distribution of the singular values $s_{1,z} \geq \cdots \geq s_{n,z} \geq 0$ of $Y_n - z$. For technical reasons, it will be easier to consider the symmetrized empirical distribution of the singular values

$$\tilde{L}_{n,z} := \frac{1}{2n} \sum_{i=1}^{n} \delta_{-s_{i,z}} + \frac{1}{2n} \sum_{i=1}^{n} \delta_{s_{i,z}}$$

for which a similar identity holds:

$$U_{\mu_n}(z) = - \int_{\mathbb{R}} \log |x| \tilde{L}_{n,z}(dx).$$

This identity is at the heart of Girko’s strategy. In a word, it shows that in order to evaluate the asymptotic behavior the spectral distribution $\mu_{Y_n}$, we can focus on the asymptotic behavior of $\tilde{L}_{n,z}$ for almost all $z \in \mathbb{C}$. By considering $\tilde{L}_{n,z}$, we are in the more familiar world of Hermitian matrices.

Informally, for all $z \in \mathbb{C}$, we will find a sequence $(\tilde{\nu}_{n,z})_{n \in \mathbb{N}}$ of deterministic probability measures such that $\tilde{L}_{n,z} \sim \tilde{\nu}_{n,z}$, and

$$\int_{\mathbb{R}} \log |x| \tilde{L}_{n,z}(dx) \approx \int_{\mathbb{R}} \log |x| \tilde{\nu}_{n,z}(dx) \quad \text{for large } n.$$

Setting

$$h_n(z) := - \int_{\mathbb{R}} \log |x| \tilde{\nu}_{n,z}(dx),$$

we will then get that for all $\psi \in C_\infty_c(\mathbb{C})$,

$$\int_{\mathbb{C}} \psi(z) \mu_n^Y(dz) \approx -\frac{1}{2\pi} \int_{\mathbb{C}} \Delta \psi(z) h_n(z) dz \quad \text{for large } n,$$

showing that $h_n(z)$ is the logarithmic potential of a probability measure $\mu_n$, and in particular that $\mu_n^Y \sim \mu_n$. By studying the smoothness properties of $h_n(z)$, we will finally retrieve the properties of $\mu_n$ which are specified in the statement of Theorem 2.3.

We provide more details hereafter with precise pointers to the article’s results.

1. Study of the associated Hermitian model. This topic is covered in Section 4.

Given $z \in \mathbb{C}$, we establish the existence of a sequence $(\tilde{\nu}_{n,z})_{n \in \mathbb{N}}$ of deterministic probability measures such that $\tilde{L}_{n,z} \sim \tilde{\nu}_{n,z}$ almost surely. To that end, we introduce the $2n \times 2n$ Hermitian matrix

$$Y_n^z := \begin{pmatrix} 0 & Y - z \\ Y^* - z^* & 0 \end{pmatrix}, \quad z \in \mathbb{C},$$

(2.20)

whose spectral measure is $\tilde{L}_{n,z}$ [45, Theorem 7.3.7]. The asymptotic analysis of $\tilde{L}_{n,z}$ is classically done by considering the resolvent

$$R_n(z, \eta) = \begin{pmatrix} -\eta & Y - z \\ Y^* - z^* & -\eta \end{pmatrix}^{-1}, \quad \eta \in \mathbb{C}_+, \quad (2.21)$$

of $Y_n^z$. Recalling the definition (2.2) of the Stieltjes transform $g_\mu$ of a probability measure $\mu$ on $\mathbb{R}$, it is clear that

$$g_{L_{n,z}}(\eta) = \frac{1}{2n} \text{tr} R_n(z, \eta).$$

The rigorous use of the Stieltjes transform for the study of ESDs of Hermitian random matrices goes back to Pastur [55], and was further developed by Bai to obtain quantitative results [15, 16]. Beginning with the seminal works [28, 29] of Erdős, Schlein and Yau this approach has been used to show that the semicircle law governs the spectral distribution for Wigner matrices down to near-optimal scales. In these works, the basic strategy is to use resolvent identities to show that the Stieltjes transform approximately satisfies a fixed-point equation, sometimes called the Schwinger–Dyson (or master-loop) equation. This approach was extended to Hermitian matrices with doubly
stochastic variance profile [30]. However, for Hermitian matrices with more general variance profiles and non-zero mean it becomes necessary to consider a system of equations that are approximately satisfied by individual diagonal entries of the resolvent.

In Section 4 we derive the following deterministic system of equations (cf. Proposition 4.1), which are the Schwinger–Dyson equations for our setting:

\[
\begin{align*}
    p_i &= \frac{(V_n^T p)_i + \eta}{|z|^2 - ((V_n^T p)_i + \eta)((V_n^T p)_i + \eta)} \\
    \tilde{p}_i &= \frac{|z|^2 - ((V_n^T \tilde{p})_i + \eta)((V_n^T \tilde{p})_i + \eta)}{|z|^2 - ((V_n^T \tilde{p})_i + \eta)((V_n^T \tilde{p})_i + \eta)}
\end{align*}
\]

for \( \eta \in \mathbb{C}^+ \), with unique solution the \( 2n \times 1 \) vector \( \tilde{p} = (p \ \tilde{p}) \) satisfying \( \text{Im} \tilde{p} > 0 \). We prove that there exists a probability distribution \( \nu_{n,z} \) whose Stieltjes transform \( g_{\nu_{n,z}} \) is defined as

\[ g_{\nu_{n,z}} = \frac{1}{n} \sum_{i \in [n]} p_i . \]

In Theorem 4.2, it is established that for all \( \eta \in \mathbb{C}_+, g_{L_{n,z}}(\eta) - g_{\nu_{n,z}}(\eta) \to 0 \) almost surely, which in particular implies that \( \tilde{L}_{n,z} \sim \tilde{\nu}_{n,z} \) a.s. In Proposition 4.7 we obtain a quantitative estimate of the form

\[ E g_{\tilde{L}_{n,z}}(\eta) - g_{\nu_{n,z}}(\eta) = O\left( \frac{1}{|\eta|^c \sqrt{n}} \right) , \]

for some integer \( c \).

We note that recently there has been much work analyzing the Schwinger–Dyson equations corresponding to Hermitian random matrices with mean and variance profiles satisfying a range of assumptions. For the centered case one is led to the so-called Quadratic Vector Equation, which has been thoroughly analyzed in works of Ajanki, Erdős and Krüger [4, 10]; the application to universality for local spectral statistics was carried out in [2]. Very recently they have made the extension to matrices with correlated entries, which involves the study of the so-called Matrix Dyson Equation [3]. In another recent work, He, Knowles and Rosenthal prove an approximate (matrix-valued) self-consistent equation for resolvents of Hermitian random matrices with arbitrary mean and variance profile, which covers the structure of the model (2.20) [42]. However, they assume the entries have all moments finite, and their aim is to obtain a local law at the optimal scale. In the present work, a sub-optimal quantitative analysis of the system (2.22) under few moments will suffice for our purposes of understanding the spectrum of the associated non-Hermitian model \( Y_n \) at global scale.

2. From the spectral measures \( L_{n,z} \sim \tilde{\nu}_{n,z} \) to the spectral measures \( \mu_{n,z}^Y \sim \mu_n \) via the associated logarithmic potentials. This topic is covered in Sections 4 (partly) and in 6.

The fact that \( \tilde{L}_{n,z} \sim \tilde{\nu}_{n,z} \) a.s. does not ensure that the random logarithmic potential \( U_{\mu_n^Y}(z) \) becomes close to the deterministic logarithmic potential \( h_n(z) \) (assuming the latter is well defined). Essentially, this is due to the fact that \( x \mapsto \log |x| \) is unbounded near zero and infinity. While the singularity at infinity is easily handled using the almost sure tightness of the measures \( \tilde{L}_{n,z} \), the singularity at zero presents a major technical challenge (indeed, this hurdle was the reason it took decades to establish the circular law under the optimal hypotheses). We show that under the admissibility assumption A2, \( x \mapsto \log |x| \) is \( \nu_{n,z} \)-integrable, and that for all \( \tau, \tau' > 0 \), there is \( \varepsilon > 0 \) small enough such that

\[ P \left\{ \left| \int_0^\varepsilon \log |x| \, \tilde{L}_{n,z}(dx) \right| > \tau \right\} < \tau' \text{ for all large } n. \]
Together with the almost sure tightness and weak convergence \( \hat{L}_{n,z} \sim \hat{\nu}_{n,z} \), we can show that \( U_{\mu_n}^Y(z) - h_n(z) \to \gamma_0 0 \) in probability. The almost sure convergence is an open problem not covered in this article.

The proof of (2.24) is based on two ingredients. The first one is a result from [24] by the first author giving a lower tail estimate for the smallest singular value of \( Y_n - z \) for arbitrary fixed \( z \in \mathbb{C} \setminus \{0\} \) under the sole assumption \( A1 \) on the standard deviation profile \( A_n \). The second result, established in Section 6, provides a so-called Wegner-type estimate on \( \hat{L}_{n,z} \), leading to a control of the other small singular values. Namely, we show that there exist two constants \( C, \gamma_0 > 0 \) such that for all \( x > 0 \),

\[
\mathbb{E} \hat{L}_{n,z}((-x, x)) \leq C(x \vee n^{-\gamma_0}).
\]

Such an estimate follows from control on \( \text{Im} \, \mathbb{E} g_{L_{n,z}}(it) \) for small \( t > 0 \), obtained in two steps. First, sufficient control of \( \mathbb{E} g_{L_{n,z}}(it) - g_{\hat{\nu}_{n,z}}(it) \) is already provided by the estimate (2.23). Second, we rely on \( A2 \) to state that \( g_{\hat{\nu}_{n,z}}(it) \) is bounded independent of \( n \) and \( t \).

For this task, a variation of the Schwinger–Dyson equations, namely the Regularized Master Equations, comes into play (see Proposition 2.1):

\[
\begin{aligned}
\left\{ \begin{array}{l}
\tilde{r}_i(t) = \frac{(V_n^T \mathbf{r})_i + t}{|z|^2 + ((V_n \mathbf{r})_i + t)((V_n^T \mathbf{r})_i + t)} , \quad \mathbf{r}(t) = (r_i(t)) , \quad \tilde{\mathbf{r}}(t) = (\tilde{r}_i(t)). \\
\end{array} \right.
\end{aligned}
\]  

(2.25)

This system is obtained simply by setting \( \eta = it \) in the Schwinger–Dyson equations (2.22) – in this case, \( \tilde{\mathbf{p}} \in i\mathbb{R}^{2n} \) and \( \tilde{\mathbf{r}} = (\mathbf{r}^T \mathbf{r}^T)^T \) is defined as \( \tilde{\mathbf{r}}(t) = \text{Im} \tilde{\mathbf{p}}(it) \). Hence by \( A2 \),

\[
\text{Im} g_{\hat{\nu}_{n,z}}(it) = \frac{1}{n} \sum_{i \in [n]} r_i(t) \leq C
\]

(2.26)

for some \( C < \infty \) independent of \( n \) and \( t \) (depending only on \( |z| \) and the parameters in our assumptions).

3. Description of the deterministic probability measure \( \mu_n \). This is covered in Sections 5 and 7.

We have proved so far that \( \mu_n^Y \sim \mu_n \) in probability, where \( \mu_n \) is the probability measure whose logarithmic potential \( U_{\mu_n}(z) \) coincides with \( h_n(z) \). It remains to establish the properties of \( \mu_n \) that are stated by Theorem 2.3.

In Section 5 we prove Theorem 2.2. Our approach to obtaining the solution \( \tilde{q}(s) \) of (2.11) is through the Regularized Master Equations (2.25). Since these equations are obtained by a simple transformation of the Schwinger–Dyson equations (2.22), by our work in Section 4 we know that (2.25) has a unique solution \( \tilde{\mathbf{r}}(s,t) \) satisfying \( \tilde{\mathbf{r}}(s,t) > 0 \), where we write \( s = |z|. \) We then show that the pointwise limit \( \tilde{r}_i(s) = \lim_{t \downarrow 0} \tilde{r}_i(s,t) \) exists, and moreover that \( \tilde{\mathbf{q}} = \tilde{\mathbf{r}} \) and is the unique solution to (2.11). Having properly defined \( \tilde{\mathbf{q}}(s) \), our main task is to show that the distribution \( -(2\pi)^{-1} \Delta h_n(z) \) in fact defines a density on the set \( \mathcal{D} := \{ z \in \mathbb{C} : |z| \neq 0, |z| \neq \sqrt{\text{Re}(V)} \} \) and to provide an expression for this density. The general approach towards solving this problem can be found in the physics literature (see [31]). Define on \( \mathbb{C} \times (0, \infty) \) the functions

\[
U_n^Y(z,t) := -\frac{1}{2n} \log \det((Y - z)^*(Y - z) + t^2), \quad \text{and}
\]

\[
U_n(z,t) := -\frac{1}{2} \int_0^\infty \log(x^2 + t^2)\hat{\nu}_{n,z}(dx).
\]

For fixed \( t > 0 \), these functions can be seen as regularized versions of the logarithmic potentials \( U_{\mu_n}^Y(z) \) and \( U_{\mu_n}(z) \) respectively, which converge back as \( t \downarrow 0 \), in \( \mathcal{D}'(\mathbb{C}) \). Now, by Jacobi’s formula
for the derivative of a determinant,
\[ \partial_{\bar{z}} U_n^Y(z,t) = \frac{1}{2n} \text{tr} (Y - z)((Y - z)^*(Y - z) + t^2)^{-1}. \]

On the other hand, let us consider again the resolvent \( R_n(z,\eta) \) introduced above in (2.21). By setting \( \eta = it \), using the well-known formula for the inverse of a partitioned matrix [45, §0.7.3] and writing
\[ R_n(z, it) = \left( \begin{array}{cc} G_n(z, it) & F_n(z, it) \\ F_n'(z, it) & \tilde{G}_n(z, it) \end{array} \right), \]
we get that \( \partial_{\bar{z}} U_n^Y(z,t) \) coincides with \( n^{-1} \text{tr} F_n(z, it) \). Relying on the asymptotic analysis made in Section 4 on the resolvent \( R_n \), we can easily obtain an expression for \( \partial_{\bar{z}} U_n(z,t) \) by considering the asymptotic behavior of \( n^{-1} \text{tr} F_n(z, it) \). We then conclude by studying the equation
\[ \Delta U_{\mu_n} = 4 \lim_{t\downarrow 0} \partial_z \partial_{\bar{z}} U_n(z,t). \]
Section 7.1 is devoted to these questions.

In Section 7.2 we conclude the proof of Theorem 2.4. As we noted above, the key is that (2.26) is easily obtained under the double stochasticity assumption by examining the explicit solution \( \vec{r} \) to the Regularized Master Equations.

4. Sufficient conditions for \( A_2 \) to hold. This topic is covered in Section 8.

While (2.26) can be proved in a few lines under \( A_3 \) (see Proposition 2.5) or \( A_4 \) (see Proposition 2.6), establishing such a bound under the more general robust irreducibility assumption \( A_5 \) is significantly more technical. Here it is helpful to view the standard deviation profile in terms of the associated directed graph \( \Gamma(A(\sigma_0)) \) (which was defined in Section 2.1). The basic idea is that the equations (2.25) encode relationships between the size of components \( r_i(t), \tilde{r}_i(t) \) at a vertex \( i \) to the sizes of the components at neighboring vertices. Assuming \( r_{i_0}(t) \) is large at some vertex \( i_0 \), we can use the robust irreducibility assumption to propagate the property of having large \( r_i(t) \) to most of the other vertices \( i \). We can also use the equations (2.25) to show that \( \tilde{r}_i(t) \) will consequently be small for most \( i \). However, this yields a contradiction due to the crucial trace identity \[ \sum_{i=1}^n r_i(t) = \sum_{i=1}^n \tilde{r}_i(t), \]
which essentially comes from the fact that the matrix \( R \) in (2.21) satisfies
\[ \sum_{i=1}^n R_{ji} = \sum_{i=1}^n R_{n+i,n+i}. \]
See Section 8 for further details.

We remark that under the stronger broad connectivity assumption on the standard deviation profile (see Remark 2.10), Wegner-type estimates that are sufficient for the purposes of this paper were obtained by the first author by a completely different argument, following a geometric approach introduced by Tao and Vu in [65] – see [25, Theorem 4.5.1].

3. Important special cases, remarks and open questions

In Sections 3.1 and 3.2, we respectively study the case of a separable variance profile \( \sigma^2_{i,j} = d_i d_j \) and of a sampled variance profile \( \sigma^2_{i,j} = \sigma^2(i/n, j/n) \) for \( \sigma : [0,1]^2 \to (0,\infty) \) a continuous function. In Section 3.3, we provide various examples where the deterministic equivalent \( \mu_n \) exhibits different behavior at \( z = 0 \). In Section 3.4, we provide simulations associated to band variance profiles. Finally we list a series of open questions in Section 3.5.
3.1. Separable variance profile. Here we are interested in the case where the standard deviation profile takes the following form.

**A6** (Separable profile). For each \( n \geq 1 \) there are deterministic vectors \( d_n, \tilde{d}_n \in \mathbb{R}_+^n \) with components \( d_i^{(n)}, \tilde{d}_i^{(n)} \), respectively, such that

\[
A_n \circ A_n = \left( (\sigma_{ij}^{(n)})^2 \right) = (d_i^{(n)} \tilde{d}_i^{(n)}) = d_n \tilde{d}_n^T.
\]

This type of model was considered in the context of linear dynamics on structured random networks in [1].

In the sequel, we drop the dependence in \( n \) and simply write \( A, V, d, \tilde{d}, d_i, \tilde{d}_i \). As will be shown in the next theorem, the system (1.5) of 2n equations simplifies into a single equation.

**Theorem 3.1** (Separable variance profile). For each \( n \geq 1 \), let \( A_n = (\sigma_{ij}) \) be a \( n \times n \) matrix with nonnegative elements. Assume that **A1** and **A6** hold. In this case \( V_n = \frac{1}{n} d_n \tilde{d}_n^T \) and \( \rho(V) = \frac{1}{n} \langle d, \tilde{d} \rangle \).

1. For each \( s \in (0, \sqrt{\rho(V)}) \) there exists a unique positive solution \( u_n(s) \) to the equation

\[
\frac{1}{n} \sum_{i \in [n]} \frac{d_i \tilde{d}_i}{s^2 + d_i \tilde{d}_i u_n(s)} = 1.
\]

Moreover, the limit \( \lim_{s \uparrow 0} u_n(s) \) exists and is equal to one: \( u_n(0) := \lim_{s \uparrow 0} u_n(s) = 1 \). If one sets \( u_n(s) = 0 \) for \( s \geq \sqrt{\rho(V)} \), then \( s \mapsto u_n(s) \) is continuous on \( \mathbb{R}_+ \) and continuously differentiable on \( (0, \sqrt{\rho(V)}) \).

2. The function \( F_n(s) := 1 - u_n(s), s \geq 0 \) defines a rotationally invariant probability measure \( \mu_n \) by

\[
\mu_n(\{ z : 0 \leq |z| \leq s \}) = F_n(s), \quad s \geq 0.
\]

In particular, \( \mu_n(\{0\}) = 0 \).

3. On the set \( \{ z : |z| < \sqrt{\rho(V)} \} \), \( \mu_n \) admits the density

\[
\varphi_n(|z|) = \frac{1}{\pi} \left( \frac{\sum_{i \in [n]} \frac{d_i \tilde{d}_i}{(\sum_{i \in [n]} |d_i \tilde{d}_i u_n(|z|)|)^2}}{\left( \sum_{i \in [n]} (|z|)^2 + d_i \tilde{d}_i u_n(|z|) \right)^2} \right)^{-1},
\]

and the support of \( \mu_n \) is exactly \( \{ z : |z| \leq \sqrt{\rho(V)} \} \).

4. In particular, the density is bounded at \( z = 0 \) with value

\[
\varphi_n(0) = \frac{1}{n \pi \sum_{i \in [n]} d_i \tilde{d}_i}.
\]

Let \( (Y_n)_{n \geq 1} \) be as in Definition 1.2 and assume that **A0** and **A2** hold.

5. Asymptotically,

\[
\mu_n^Y \sim \mu_n \quad \text{in probability} \quad (n \to \infty).
\]

**Proof.** This theorem is essentially a specification of Theorems 2.2 and 2.3 to the case of the variance profile \( d \tilde{d}^T \). Introduce the quantities \( \alpha_n = \frac{1}{n} \langle d, q \rangle \) and \( \tilde{\alpha}_n = \frac{1}{n} \langle \tilde{d}, \tilde{q} \rangle \) which satisfy the system

\[
1 = \frac{1}{n} \sum_{i \in [n]} \frac{d_i \tilde{d}_i}{s^2 + d_i \tilde{d}_i \alpha_n \tilde{\alpha}_n} \quad \text{and} \quad \alpha_n \sum_{i \in [n]} \frac{d_i}{s^2 + d_i \alpha_n \tilde{\alpha}_n} = \tilde{\alpha}_n \sum_{i \in [n]} \frac{d_i}{s^2 + d_i \alpha_n \tilde{\alpha}_n}
\]

for \( s \in (0, \sqrt{\rho(V)}) \) and are equal to zero if \( s \geq \sqrt{\rho(V)} \). The function \( F_n \) given in (2.12) becomes

\[
F_n(s) = 1 - \alpha_n(s) \tilde{\alpha}_n(s).
\]

Set \( u_n(s) = \alpha_n(s) \tilde{\alpha}_n(s) \). Notice that \( u_n \) satisfies the first equation in
Theorem 3.1-(1). All the other properties of \( u_n \) follow from those of \( q, \tilde{q} \), except that \( u_n(0) = 1 \). To prove the later introduce \( \xi_{\min} = n^{-1} \sum_{i \in [n]} d_i \tilde{d}_i > 0 \) and \( d_{\max} = \max(d_i, \tilde{d}_i, i \in [n]) \). Then

\[
\frac{\xi_{\min}}{s^2 + d_{\max}^2 u_n(s)} \leq 1 \leq \frac{1}{u_n(s)}.
\]

We deduce that \( u_n(s) \) is bounded away from zero and upper bounded as \( s \downarrow 0 \). Taking the limit in the equation satisfied by \( u_n(s) \) as \( s \downarrow 0 \) along a converging subsequence finally yields that \( u_n(s) \xrightarrow{s \to 0} 1 \).

We do not prove items (3)-(4) since they can be proved as in Theorem 3.2-(3)-(4) below. Item (5) is straightforward.

**Remark 3.1 (Girko’s sombrero probability).** Consider the separable variance profile \( \tilde{d} \tilde{d}^T \) with the first \( k \) entries of \( d \) equal to \( a > 0 \), the last \( n - k \) equal to \( b > 0 \) and all the entries of \( \tilde{d} \) equal to 1. Denote by \( \alpha = \frac{k}{n} \), by \( \beta = \frac{n-k}{n} \) and by \( \rho = \alpha a + \beta b \) the spectral radius of \( V_n = \frac{1}{n} d \tilde{d}^T \). As a corollary of the previous theorem, we obtain

\[
\varphi_n(|z|) = \frac{1}{2 \pi \rho a b} (a + b) - \frac{|z|^2 (a - b)^2 + ab |2(\alpha a + \beta b) - (a + b)|}{\sqrt{|z|^4 (a - b)^2 + 2 |z|^2 ab |2(\alpha a + \beta b) - (a + b)| + a^2 b^2}}
\]

(3.1)

for \( |z| < \sqrt{\rho} \) and \( \varphi_n(|z|) = 0 \) elsewhere. This formula was also derived and further studied in [1, Eq. (2.63)]. In the case where \( \alpha = \beta = \frac{1}{2} \), we recover Girko’s sombrero probability distribution [35, Section 26.12]:

\[
\varphi_n(|z|) = \frac{1}{2 \pi \rho a b} (a + b) - \frac{|z|^2 (a - b)^2}{\sqrt{|z|^4 (a - b)^2 + a^2 b^2}}
\]

for \( s < \sqrt{\frac{a+b}{2}} \).

In the case \( a = b \), we recover the circular law. To compute \( \varphi_n \), we proceed as follows: Theorem 3.1 yields the equation

\[
\frac{\alpha a}{s^2 + a u_n(s)} + \frac{\beta b}{s^2 + b u_n(s)} = 1
\]

equivalent to

\[
ab u_n^2(s) + [s^2(a + b) - ab] u_n(s) + s^4 - (\alpha a + \beta b) s^2 = 0,
\]

and with positive solution for \( s < \sqrt{\rho} \)

\[
u_n(s) = -[s^2(a + b) - ab] + \sqrt{s^4(a - b)^2 + 2 s^2 ab |2(\alpha a + \beta b) - (a + b)| + a^2 b^2} \]

Now \( F_n(s) = 1 - u_n(s) \) and by Remark 2.8, the density is given by

\[
\varphi_n(|z|) = \frac{1}{2 \pi |z|} \frac{\partial F_n(s)}{\partial s} \bigg|_{s = |z|} = - \frac{1}{2 \pi |z|} \frac{\partial u_n(s)}{\partial s} \bigg|_{s = |z|}.
\]

A short computation now yields (3.1).

If the quantities \( d_i, \tilde{d}_i \) correspond to evaluations of continuous functions \( d, \tilde{d} : [0, 1] \to (0, \infty) \) at regular samples (\( \frac{j}{n} \)), then one obtains a genuine limit in the previous theorem. Notice that in this case A1 and A3 hold (and therefore A2 as well).

**Theorem 3.2 (Sampled and separable variance profile).** Let \( d, \tilde{d} : [0, 1] \to (0, \infty) \) be two continuous functions and define a variance profile \( (\sigma_{ij}^2) \) by

\[
\sigma_{ij}^2 = d \left( \frac{i}{n} \right) \tilde{d} \left( \frac{j}{n} \right).
\]

Denote by \( \rho_{\infty} = \int_0^1 d(x) \tilde{d}(x) \, dx \).
(1) For any \( s \in (0, \sqrt{\rho_{\infty}}) \) there exists a unique positive solution \( u_{\infty}(s) \) to the equation

\[
\int_0^1 \frac{d(x)\tilde{d}(x)}{s^2 + d(x)d(x)u_{\infty}(s)} \, dx = 1.
\]

If one sets \( u_{\infty}(s) = 0 \) for \( s \geq \sqrt{\rho_{\infty}} \), then \( s \mapsto u_{\infty}(s) \) is continuous on \( \mathbb{R}_+ \). Moreover, the limit \( u_{\infty}(0) := \lim_{s \downarrow 0} u_{\infty}(s) \) exists and \( u_{\infty}(0) = 1 \).

(2) The function

\[
F_\infty(s) := 1 - u_{\infty}(s), \quad s \geq 0
\]

defines a rotationally invariant probability measure \( \mu_\infty \) by

\[
\mu_\infty(\{|z| : 0 \leq |z| \leq s\}) = F_\infty(s), \quad s \geq 0, \quad \text{and} \quad \mu_\infty(\{0\}) = 0.
\]

(3) The function \( s \mapsto u_{\infty}(s) \) is continuously differentiable on \((0, \sqrt{\rho(V)})\) and \( \mu_\infty \) admits the density

\[
\varphi_\infty(|z|) = \frac{1}{\pi} \left( \int_0^1 \frac{d(x)\tilde{d}(x)}{(|z|^2 + d(x)d(x)u_{\infty}(|z|))^2} \, dx \right) \left( \int_0^1 \frac{d^2(x)d^2(x)}{(|z|^2 + d(x)d(x)u_{\infty}(|z|))^2} \, dx \right)^{-1}
\]

on the set \( \{z : |z| < \sqrt{\rho_{\infty}}\} \) and \( \varphi_\infty = 0 \) for \( |z| > \sqrt{\rho_{\infty}} \). In particular, the support of \( \mu_\infty \) is equal to \( \{z : |z| \leq \sqrt{\rho_{\infty}}\} \).

(4) This density is bounded at \( z = 0 \) with value

\[
\varphi_\infty(0) = \frac{1}{\pi} \int_0^1 \frac{dx}{d(x)d(x)}.
\]

Let \((Y_n)_{n \geq 1}\) be as in Definition 1.2 and assume that A0 holds.

(5) Asymptotically,

\[
\mu_n^{I \sim \frac{w}{n \to \infty}} \mu_\infty \quad \text{in probability}.
\]

The proof of Theorem 3.2 is postponed to Section A.1.

3.2. Sampled variance profile. Here, we are interested in the case where

\[
\sigma_{ij}^2(n) = \sigma^2 \left( \frac{i}{n}, \frac{j}{n} \right),
\]

where \( \sigma \) is a continuous nonnegative function on \([0, 1]^2\). Notice that A1 holds and denote by

\[
\sigma_{\max} = \max_{x, y \in [0, 1]} \sigma(x, y) \quad \text{and} \quad \sigma_{\min} = \min_{x, y \in [0, 1]} \sigma(x, y).
\]

For the sake of simplicity, we will restrict ourselves to the case where \( \sigma \) takes its values in \((0, \infty)\), i.e. where \( \sigma_{\min} > 0 \), which implies that A3 holds.

We shall use some results from the Krein–Rutman theory (see for instance [26]), which generalizes the spectral properties of nonnegative matrices to positive operators on Banach spaces. To the function \( \sigma^2 \) we associate the linear operator \( \mathbf{V} \), defined on the Banach space \( C([0, 1]) \) of continuous real-valued functions on \([0, 1]\) as

\[
(\mathbf{V}f)(x) := \int_0^1 \sigma^2(x, y)f(y) \, dy.
\]

By the uniform continuity of \( \sigma^2 \) on \([0, 1]^2\) and the Arzela–Ascoli theorem, it is a standard fact that this operator is compact [57, Ch. VI.5]. Let \( C^+([0, 1]) \) be the convex cone of nonnegative elements of \( C([0, 1]) \):

\[
C^+([0, 1]) = \{ f \in C([0, 1]) \, : \, f(x) \geq 0 \quad \text{for} \ x \in [0, 1] \}.
\]
Since $\sigma_{\min} > 0$, the operator $V$ is strongly positive, i.e. it sends any element of $C^+([0,1]) \setminus \{0\}$ to the interior of $C^+([0,1])$, the set of continuous and positive functions on $[0,1]$. Under these conditions, it is well-known that the spectral radius $\rho(V)$ of $V$ is non-zero, and it coincides with the so-called Krein–Rutman eigenvalue of $V$ [26, Theorem 19.2 and 19.3].

To be consistent with our notation for nonnegative finite dimensional vectors, we shall write here $f \succcurlyeq 0$ when $f \in C^+([0,1]) \setminus \{0\}$, and $f \succ 0$ when $f(x) > 0$ for all $x \in [0,1]$.

**Theorem 3.3** (Sampled variance profile). Assume that there exists a continuous function $\sigma: [0,1]^2 \to (0,\infty)$ such that

$$\sigma_{ij}^{(n)} = \sigma\left(\frac{i}{n}, \frac{j}{n}\right).$$

Let $(Y_n)_{n \geq 1}$ be a sequence of random matrices as in Definition 1.2 and assume that $A_0$ holds. Then,

1. The spectral radius $\rho(V_n)$ of the matrix $V_n = n^{-1}(\sigma_{ij}^2)$ converges to $\rho(V)$ as $n \to \infty$, where $V$ is the operator on $C([0,1])$ defined by (3.2).
2. Given $s > 0$, consider the system of equations:

$$\begin{align*}
Q_\infty(x,s) &= \frac{\int_0^1 \sigma^2(y,x)Q_\infty(y,s)\,dy}{s^2 + \int_0^1 \sigma^2(y,x)Q_\infty(y,s)\,dy \int_0^1 \sigma^2(x,y)\bar{Q}_\infty(y,s)\,dy}, \\
\bar{Q}_\infty(x,s) &= \frac{\int_0^1 \sigma^2(x,y)\bar{Q}_\infty(y,s)\,dy}{s^2 + \int_0^1 \sigma^2(y,x)Q_\infty(y,s)\,dy \int_0^1 \sigma^2(x,y)\bar{Q}_\infty(y,s)\,dy}, \\
\int_0^1 Q_\infty(y,s)\,dy &= \int_0^1 \bar{Q}_\infty(y,s)\,dy.
\end{align*}$$

(3.3)

with unknown parameters $Q_\infty(\cdot, s), \bar{Q}_\infty(\cdot, s) \in C^+([0,1])$. Then,

(a) for $s \geq \sqrt{\rho(V)}$, $Q_\infty(\cdot, s) = \bar{Q}_\infty(\cdot, s) = 0$ is the unique solution of this system.

(b) for $s \in (0, \sqrt{\rho(V)})$, the system has a unique solution $Q_\infty(\cdot, s) + \bar{Q}_\infty(\cdot, s) \succcurlyeq 0$. This solution satisfies $$Q_\infty(\cdot, s), \bar{Q}_\infty(\cdot, s) \succ 0.$$ (c) The functions $Q_\infty, \bar{Q}_\infty: [0,1] \times (0,\infty) \to [0,\infty)$ are continuous, and continuously extended to $[0,1] \times [0,\infty)$, with

$$Q_\infty(\cdot,0), \bar{Q}_\infty(\cdot,0) \succ 0.$$ (3) The function

$$F_\infty(s) := 1 - \int_{[0,1]^2} Q_\infty(x,s) \bar{Q}_\infty(y,s) \sigma^2(x,y)\,dx\,dy, \quad s \in (0,\infty)$$

converges to zero as $s \downarrow 0$. Setting $F_\infty(0) := 0$, the function $F_\infty$ is an absolutely continuous function on $[0,\infty)$ which is the CDF of a probability measure whose support is contained in $[0, \sqrt{\rho(V)}]$, and whose density is continuous on $[0, \sqrt{\rho(V)}]$.

(4) Let $\mu_\infty$ be the rotationally invariant probability measure on $\mathbb{C}$ defined by the equation $\mu_\infty(\{z: 0 \leq |z| \leq s\}) := F_\infty(s), \quad s \geq 0$.

Then, $\mu_n^Y \xrightarrow{w} \mu_\infty$ in probability.

The proof of Theorem 3.3 is an adaptation of the proofs of Lemmas 5.5 and 5.6 below to the context of Krein–Rutman’s theory for positive operators in Banach spaces, and is postponed to Appendix A.2.
3.3. Behavior of the probability distribution $\mu_n$ near zero. Recall the definitions

$$F_n(s) = 1 - \frac{1}{n} \langle q(s), V\tilde{q}(s) \rangle, \quad \mu_n \{ z \in \mathbb{C}, \ |z| \leq s \} = F_n(s), \quad \varphi_n(|z|) = \frac{1}{2\pi|z|} F_n'(|z|)$$

provided in Theorem 2.3 and in Remark 2.8.

The behavior of $F_n$ near zero is an interesting problem. By Theorem 2.3, $F_n$ admits a limit as $s \downarrow 0$. Is this limit positive (atom) or equal to zero (no atom)? Is its derivative finite at $z = 0$ (finite density) or even null (vanishing density)? or does it blow up at $z = 0$? We provide hereafter some examples that shed some light on these questions.

In Proposition 3.4, we study the density $\varphi_n$ at zero in the case where $A_3$ holds (positive variance profile). In this case, we provide an explicit formula for the density at zero. This result complements a recent result by Alt et al. [11, Lemma 4.1] where it is proved that under $A_3$, the density is positive for $|z| \leq \sqrt{\rho(V_n)}$.

In Proposition 3.5, we provide an example of a simple variance profile with some vanishing entries where $\mu_n$ admits a closed-form expression with an atom and a vanishing density at $z = 0$.

In Proposition 3.6, we provide an example of a symmetric separable sampled variance profile $\sigma_{ij} = d(i/n)d(j/n)$ where function $d$ is continuous and vanishes at zero. Depending on the function $d$, the density $\varphi_\infty(|z|)$ may blow up at $z = 0$ or not.

**Proposition 3.4** (No atom and bounded density near zero). Consider a sequence $(V_n)$ of normalized variance profiles and assume that $A_1$ and $A_3$ hold. Let $\tilde{q}(s)$ be as in Theorem 2.2, and let $\mu_n$ be as in Theorem 2.3.

1. The quantities $q(0) = \lim_{s \downarrow 0} q(s)$ and $\tilde{q}(0) = \lim_{s \downarrow 0} \tilde{q}(s)$ are well defined and satisfy

$$q_i(0)(V_n \tilde{q}(0))_i = 1 \quad \text{and} \quad \tilde{q}_i(0)(V_n^T q(0))_i = 1, \quad i \in [n].$$

In particular, probability measure $\mu_n$ has no atom at zero: $\mu_n(\{0\}) = 0$.

2. Denote by $\varphi_n$ the density of $\mu_n$, i.e. $\mu_n \{ z \in \mathbb{C} : |z| \leq s \} = \int_{|z|\leq s} \varphi_n(|z|)\ell(dz)$, then

$$\varphi_n(0) = \lim_{z \to 0} \varphi_n(|z|) \quad \text{exists and its value is given by}$$

$$\varphi_n(0) = \frac{1}{n} \sum_{i \in [n]} \frac{1}{(V_n^T q(0))_i (V_n q(0))_i} = \frac{1}{n} \sum_{i \in [n]} q_i(0)\tilde{q}_i(0).$$

In particular, there exist finite constants $\kappa, K$ independent of $n \geq 1$ such that

$$0 < \kappa \leq \varphi_n(0) \leq K.$$

The proof of Proposition 3.4 is postponed to Appendix A.3 as it relies on Lemma 5.6 and its proof.

**Proposition 3.5** (Example with an atom and vanishing density at zero). Denote by $J_m$ the $m \times m$ matrix whose elements are all equal to one. Let $k \geq 1$ be a fixed integer, assume that $n = km$ ($m \geq 1$) and consider the $n \times n$ matrix

$$A_n = \begin{pmatrix} 0 & J_m & \cdots & J_m \\ J_m & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ J_m & 0 & \cdots & 0 \end{pmatrix}.$$  \hspace{1cm} (3.4)

Associated to matrix $A_n$ is the sequence of normalized variance profiles $V_n = \frac{1}{n} A_n \circ A_n$ with spectral radius $\rho(V_n) = \frac{\sqrt{k-1}}{k}$. Denote by $\rho^* = \sqrt{\rho(V_n)} = \frac{\sqrt{k-1}}{\sqrt{k}}$. Then

1. Assumptions $A_1$ and $A_2$ hold true.
The function $F_n$ defined in Theorem 2.3 does not depend on $n$ and is given by

$$F_n(s) := F_\infty(s) = \frac{1}{k} \sqrt{(k-2)^2 + 4k^2 s^4} \quad \text{if} \quad 0 \leq s \leq \rho^*,$$

and $F_\infty(s) = 1$ if $s > \rho^*$. In particular, $F_\infty(0) = 1 - \frac{2}{k}$ and $\lim_{s \to \rho^*} F_\infty(s) = 1$.

(3) The density $\varphi_n := \varphi_\infty$ and the measure $\mu_n := \mu_\infty$ do not depend on $n$ and are given by

$$\varphi_\infty(|z|) = \frac{4k}{\pi} \frac{|z|^2}{\sqrt{(k-2)^2 + 4k^2 |z|^4}} 1_{\{|z| \leq \rho^*\}},$$

$$\mu_\infty(dz) = \left(1 - \frac{2}{k}\right) \delta_0(dz) + \frac{4k}{\pi} \frac{|z|^2}{\sqrt{(k-2)^2 + 4k^2 |z|^4}} 1_{\{|z| \leq \rho^*\}} \ell(dz).$$

In particular, $\varphi_\infty(0) = 0$.

The definition of $F_n$ readily implies that measure $\mu_n$ admits an atom at zero of weight $1 - \frac{2}{k}$ since $\mu_n(\{0\}) = F_n(0) = 1 - \frac{2}{k}$. This result can be (almost) obtained by simple linear algebra: Note that $\text{rank}(Y_n) = \text{rank}(n^{-1/2} A_n \odot X_n) \leq (m - 2)k$ for any $X_n$. Indeed, since the top-right $m \times (k-1)m$ submatrix of $Y_n$ has row-rank at most $m$, its kernel, and hence the kernel of $Y_n$, has dimension at least $m(k-2)$. Therefore, $\mu_n \nu_Y$ has an atom at zero with the weight $\frac{m(k-2)}{mk} = 1 - \frac{2}{k}$ (at least) when $n$ is a multiple of $k$.

**Remark 3.2** (Typical spacing for the random eigenvalues near zero). We heuristically evaluate the typical spacing for the random eigenvalues in a small disk centered at zero.

$$\mu_n^Y(B(0, \varepsilon)) \simeq \left(1 - \frac{2}{k}\right) + \int_{B(0, \varepsilon)} \varphi_\infty(|z|) \ell(dz)$$

If we remove the $n \left(1 - \frac{2}{k}\right) = km \left(1 - \frac{2}{k}\right) = (k-2)m$ deterministic zero eigenvalues, the typical number of random eigenvalues in $B(0, \varepsilon)$ is

$$\#\{\lambda_i \text{ random} \in B(0, \varepsilon)\} = n \times \int_{B(0, \varepsilon)} \varphi_\infty(|z|) \ell(dz) = 2\pi n \int_0^\varepsilon s \varphi_\infty(s) ds \propto n\varepsilon^4$$

Hence, if we want the number of random eigenvalues in $B(0, \varepsilon)$ to be of order $O(1)$, we need to tune $\varepsilon = n^{-1/4}$ and the typical spacing should be $n^{-1/4}$ near zero. On the other hand, the typical spacing at any point $z$ where $\varphi_\infty(|z|) > 0$ is $n^{-1/2}$. Notice that $n^{-1/4} \gg n^{-1/2}$. This is confirmed by the simulations which show some repulsion phenomenon at zero, cf. Figure 1. In particular, the optimal scale for a local law near zero should be $n^{-1/4}$.

**Proof of Proposition 3.5.** Simple computations yield $\rho(V_n) = \sqrt{k-1}/k$ and that $V_n$’s spectral measure features a Dirac mass at zero with weight $1 - \frac{2}{k}$. Assumption A1 is immediately satisfied, so is A2 as the variance profile is symmetric. Item (1) is proved. We now prove item (2) and first solve the master equations. Since the variance profile is symmetric, we have $q = \tilde{q}$ and obviously

$$q^T = (q, \ldots, q, \tilde{q}, \ldots, \tilde{q})$$

The equations satisfied by $q, \tilde{q}$ are

$$q = \frac{k(k-1)\tilde{q}}{k^2 s^2 + (k-1)^2 \tilde{q}^2} \quad \text{and} \quad \tilde{q} = \frac{kq}{k^2 s^2 + \tilde{q}^2}$$

Set $\alpha = \frac{(k-1)(k-2)}{2}$ and $\beta = \frac{k(k-1)}{2}$. We end up with the following equation for $X = q^2$:

$$s^2 X^2 + 2(k^2 s^4 + \alpha) X + k^4 s^6 - k^2 (k-1) s^2 = 0$$
Hence
\[
\Delta' = \alpha^2 + k^2(k-1)^2s^4 \quad \text{and} \quad q^2 = \frac{2k\beta s^2 - k^4s^6}{\sqrt{\Delta'} + k^2s^4 + \alpha}.
\]
Now
\[
\frac{1}{mk}(q, V_nq) = \frac{2(k-1)}{k^2} \frac{kq^2}{k^2s^2 + q^2} = \frac{2(k-1)}{k} \frac{(k-1) - k^2s^4}{\sqrt{\Delta'} + \beta}.
\]
We finally compute \(F_\infty\) for \(s \leq \rho^*\) and notice that a simplification occurs:
\[
F_\infty(s) = 1 - \frac{1}{nk}(q, V_nq) = \frac{k\sqrt{\Delta'} + k\beta - 2(k-1)^2 + 2(k-1)k^2s^4}{k(\sqrt{\Delta'} + \beta)} = \frac{2\sqrt{\Delta'}}{k(k-1)} = \frac{1}{k} \sqrt{(k-2)^2 + 4k^2s^4}.
\]
The rest of the proof is straightforward.

We denote \(u(s) \sim v(s)\) as \(s \to 0\) if \(\lim_{s \to 0} \frac{u(s)}{v(s)} = 1\).

Consider a separable variance profile \(\sigma_{ij}^2 = d(i/n)d(j/n)\) with \(d(\cdot)\) a positive function over \([0, 1]\). In Theorem 3.2, we proved that the density at zero is finite and admits the following closed-form expression:
\[
\varphi_\infty(0) = \frac{1}{\pi} \int_0^1 \frac{dx}{d^2(x)}.
\]
In order to build a distribution \(\mu_\infty\) whose density at zero blows up, we allow function \(d(\cdot)\) to vanish at \(x = 0\) and consider the following variation of Theorem 3.2.

**Proposition 3.6.** Let \(d : [0, 1] \to [0, \infty)\) be a continuous function satisfying \(d(0) = 0\) and \(d(x) > 0\) for \(x > 0\). Define a variance profile by \(\sigma_{ij}^2 = d(i/n)d(j/n)\) for \(i, j \in [n]\) and denote by \(\rho_\infty =\)
\[ \int_0^1 d^2(x) \, dx. \] Then items (1)-(3) of Theorem 3.2 hold true. Moreover,

\[ \varphi_\infty(s) \sim \frac{1}{\pi} \int_0^1 \frac{d^2(x)}{[s^2 + d^2(x)u(s)]^2} \, dx \quad \text{as} \quad s \to 0. \]

If A0 holds, then item (5) of Theorem 3.2 holds as well.

Proposition 3.6 whose proof is omitted can be proved as Theorems 3.1 and 3.2.

Applying the previous proposition to specific functions \(d(\cdot)\) yields the following examples:

**Example 3.1 (Unbounded and bounded densities near \(|z|=0\)).**

(1) Let \(d(x) = x\) then

\[ \int_0^1 \frac{x^2}{[s^2 + x^2 u(s)]^2} \, dx \sim \frac{\pi}{4s} \quad \text{hence} \quad \varphi_\infty(|z|) \sim \frac{1}{4|z|} \quad \text{as} \quad s \to 0. \]

(2) Let \(d(x) = \sqrt{x}\) then

\[ \int_0^1 \frac{x}{[s^2 + x u(s)]^2} \, dx \sim -2 \log(s) \quad \text{hence} \quad \varphi_\infty(|z|) \sim -\frac{2 \log(|z|)}{\pi} \quad \text{as} \quad s \to 0. \]

(3) Let \(d(x) = x^a\) with \(a \in (0, \frac{1}{2})\), then \(\varphi_\infty(0) = \frac{1}{\pi(1-2a)}\).

### 3.4. More examples: band matrix models.

We now provide some numerical illustrations of the results of Theorem 2.3 in the case of band matrix models. In these cases, closed-form expressions for the density seem out of reach but plots can be obtained by numerics (see also Remark 2.5).

We consider two probabilistic matrix models with complex entries (with independent Bernoulli real and imaginary parts) and sampled variance profiles associated to the following functions:

| Model A | Model B |
|---------|---------|
| \(\sigma^2(x, y) = \mathbb{1}_{\{|x-y| \leq \frac{1}{2n}\}}\) | \(\sigma^2(x, y) = (x + 2y)^2 \mathbb{1}_{\{|x-y| \leq \frac{1}{n}\}}\) |

Clearly, the function associated to Model A yields a symmetric variance profile, admissible by Proposition 2.6. Model B satisfies the broad connectivity hypothesis (see Remark 2.10), hence A5 (which is weaker than the broad connectivity assumption).

**Lemma 3.7.** Given \(\alpha \in (0, 1)\) and \(a > 0\), consider the standard deviation profile matrix \(A_n = (\sigma(i/n, j/n)_{i,j=1}^n)\) where \(\sigma^2(x, y) = (x + ay)^2 \mathbb{1}_{\{|x-y| \leq \alpha\}}\). Then, there exists a cutoff \(\sigma_0 \in (0, 1)\) such that for all \(n\) large enough, the matrix \(A_n(\sigma_0)\) satisfies the broad connectivity hypothesis (see (2.16) and (2.18)) with \(\delta = \kappa = \alpha\) for a suitable absolute constant \(c > 0\).

**Proof.** One can take the cutoff parameter \(\sigma_0\) sufficiently small that the entries \(\sigma_{ij} < \sigma_0\) within the band are confined to the top-left corner of \(A\) of dimension \(n/100\), say, at which point the argument of [24, Corollary 1.17] applies with minor modification. \(\square\)

Eigenvalue realizations for models A and B are shown on Figure 2. On Figure 3, the densities of \(\mu_n\) are shown. Plots of the functions \(F_n\) given by (2.12) are shown on Figure 4 along with their empirical counterparts.

Up to the “corner effects”, the variance profile for Model A is a scaled version of the doubly stochastic variance profile considered in Theorem 2.4. It is therefore expected that the density for Model A is “close” to the density of the circular law. This is confirmed by Figures 2a, 3a and 4a. Note in particular that \(F_n\) depicted on Figure 4a is close to a parabola, which is the radial marginal of the circular law.

Due to the form of the variance profile of Model B, a good proportion of the rows and columns of the matrix \(Y_n\) have small Euclidean norms. We can therefore expect that many of the eigenvalues of
Figure 2. Eigenvalues realizations. Setting: $n = 2000$; the circles' radii are $\sqrt{\rho(V)}$.

Figure 3. Densities of $\mu_n$.

$Y_n$ will concentrate towards zero. This phenomenon is particularly visible on the plot of the density in Figure 3b. The exact behavior of the density near zero is out of the scope of this paper.

3.5. Open questions.

Further properties of the density of deterministic equivalents. Lemma 5.6 provides an expression for the derivative of the solution $\vec{q}(s)$ to the Master Equation which could shed some light on further properties of $\mu_n$. Recently [11] showed the density is strictly positive on the closed disk with radius $\sqrt{\rho(V_n)}$ under assumptions A3. For instance, is the support of $\mu_n$ always imaginary connected? Is it possible to describe the density of $\mu_n$ near zero (as shown in the examples, a wide variety of behaviors is to be expected)? However, this expression appears difficult to analyze, and we have not pursued this.
Relaxing the robust irreducibility assumption. While control on the smallest singular value is proved under very general conditions (see Proposition 6.1), we have made the additional robust irreducibility assumption $A_5$ in order to handle the other small singular values via Wegner estimates. Would it be possible to lighten this assumption?

Almost sure convergence. As we noted in Remark 2.6, the convergence $\mu_n^Y \sim \mu_n$ in probability in Theorem 2.3 could be upgraded to almost sure convergence if we had improved lower tail estimates on the smallest singular value for nonzero scalar shifts of the matrices $Y_n$ (specifically, an improvement of the bound in Proposition 6.1 to be summable in $n$). Such an improvement may be possible by combining tools of Inverse Littlewood–Offord theory with the approach in [24].

Local law. In [23] it was shown that the circular law holds on the optimal scale of $n^{-1/2+\epsilon}$. These results were extended in [71] to $TX_n$ where $T$ is a deterministic matrix and $X_n$ is an i.i.d. random matrix. Both results rely heavily on proving an optimal local law for the empirical distribution of the singular values of scalar shifts of an i.i.d. matrix.

After the initial release of this paper, a local law was proved in [11] under $A_3$ and a stronger assumption on distribution of the matrix entries. Additionally they show that $A_3$ implies the deterministic density is bounded from above and below. It remains an open problem to prove a local law on the optimal scale when the density vanishes or is unbounded.

Extension to sparse models. As we noted in Remark 2.2, our assumptions require the number of non-zero entries to be a constant proportion of the total number of entries. This is required both to bound the smallest singular value of the shifted random matrices and to prove effective bounds on the Stieltjes transform. We expect that our results should extend to certain matrices with density $\sim n^{\varepsilon-1}$ for arbitrary fixed $\varepsilon \in (0,1)$, suitably rescaled. An interesting first case to consider is random band matrices with shrinking bandwidth. The limit of the empirical distribution of the singular values was recently computed in [46], but bounds on the smallest singular value were not considered.

Extension to allow heavy-tailed entries. In a similar direction, it would be interesting to prove an analogue of Theorem 2.3 for the case that the entries $X_{ij}$ lie in the basin of attraction of an $\alpha$-stable law for some $\alpha \in (0,2)$. In this case we expect the deterministic equivalents $\mu_n$ will not have compact support. The limiting empirical distribution of singular values for such matrices (allowed to be rectangular with bounded eccentricity) was studied by Belinschi, Dembo and Guionnet in [19].
For the case that the entries are i.i.d. the limiting empirical spectral distribution was established by Bordenave, Caputo and Chafaï in [20].

4. Asymptotics of singular values distributions

In the outline of the proof (cf. section 2.5) of the main result, we introduced the symmetrized empirical distribution of the singular values of $Y_n - z$:

$$L_{n,z} := \frac{1}{2n} \sum_{i=1}^{n} \delta_{-s_i,z} + \frac{1}{2n} \sum_{i=1}^{n} \delta_{s_i,z},$$

where the $s_i$'s are the singular values of $Y_n - z$. It is well-known that the spectrum of the Hermitian matrix model

$$Y_n^z = \begin{pmatrix} 0 & Y - z \\ Y^* - z^* & 0 \end{pmatrix}$$

is precisely the set $\{-s_i, i \in [n]\} \cup \{s_i, i \in [n]\}$, and hence $L_{n,z}$ is the empirical spectral distribution for $Y_n^z$.

For $\eta \in \mathbb{C}_+$, the resolvent of $Y_n^z$ (as a function of $\eta$) is written

$$R(z, \eta) := \left( \begin{array}{cc} \eta & Y - z \\ Y^* - z^* & -\eta \end{array} \right)^{-1} := \begin{pmatrix} G(z, \eta) & F(z, \eta) \\ \tilde{G}(z, \eta) & \tilde{F}(z, \eta) \end{pmatrix},$$

(4.1)

where, by the well-known formula for the inverse of a partitioned matrix [45, §0.7.3],

$$G(z, \eta) = \eta \left( (Y - z)(Y - z)^* - \eta^2 \right)^{-1},$$

$$\tilde{G}(z, \eta) = \eta \left( (Y - z)^*(Y - z) - \eta^2 \right)^{-1},$$

(4.2)

$$F(z, \eta) = (Y - z) \left( (Y - z)^*(Y - z) - \eta^2 \right)^{-1},$$

and

$$\tilde{F}(z, \eta) = (Y - z)^*(Y - z) - \eta^2 \left( Y - z \right)^*.$$

The main objective of this section is to provide deterministic counterparts of the normalized traces of these matrix functions.

We begin by deriving the Schwinger–Dyson equations – a system of equations approximately satisfied by the diagonal entries of the matrices in (4.2). We then show the Schwinger–Dyson equations have a unique solution corresponding to Stieltjes transforms of probability measures and analyze the properties of the solution. Finally we estimate the difference between (4.2) and the true solution of the Schwinger–Dyson equations, which in turn is used to estimate the difference between the empirical spectral measure of $Y_n^z$ and its deterministic counterpart.

4.1. Derivation of the Schwinger–Dyson equations. In this subsection we specialize to the case that the entries of $X$ are i.i.d. standard complex Gaussian variables. Later we will compare a general matrix with a Gaussian matrix, at which point we will label the Gaussian matrix and associated quantities with a superscript $N$; however, we omit the superscript in the present subsection.

For a resolvent $R$ as defined in (4.1) with complex entries, the following differentiation formulas hold true and will be needed in the sequel:

$$\frac{\partial R_{ij}}{\partial Y_{k\ell}} = -R_{ik} R_{n+\ell,j}, \quad \frac{\partial R_{ij}}{\partial Y_{ik}} = -R_{i,n+k} R_{\ell,j}, \quad 1 \leq k, \ell \leq n, \quad 1 \leq i, j \leq 2n.$$  

(4.3)

We will heavily rely on the variance estimates provided in Proposition B.2 and Corollary B.3.
Denote by \( Y = \begin{bmatrix} 0 & Y \ Y^* & 0 \end{bmatrix} \). The equation \( R^{-1} R = I_{2n} \) yields
\[
\eta R + Y R + \begin{bmatrix} -z F' & -z \tilde{G} \\
-z* G & -z* F \end{bmatrix} = I_{2n} .
\] (4.4)

Given matrices \( G \) and \( \tilde{G} \), it will be convenient to introduce the vectors
\[
g = (G_{11} \cdots G_{nn})^T \quad \text{and} \quad \tilde{g} = (\tilde{G}_{11} \cdots \tilde{G}_{nn})^T .
\]

Taking \( i \in [n] \) yields
\[
-\eta \mathbb{E} G_{ii} + \mathbb{E} (Y R)_{ii} - z \mathbb{E} F'_{ii} = 1 .
\] (4.5)

Applying the integration by part formula for complex Gaussian random variables (see for instance [54, (2.1.40)]) together with (4.3) yields
\[
\mathbb{E} (Y R)_{ii} = \sum_{\ell=1}^{n} \mathbb{E} Y_{i\ell} R_{n+\ell,i} = \sum_{\ell=1}^{n} \frac{\sigma_{i\ell}^2}{n} \mathbb{E} \left[ \frac{\partial R_{n+\ell,i}}{\partial Y_{i\ell}} \right] = -\sum_{\ell=1}^{n} \frac{\sigma_{i\ell}^2}{n} \mathbb{E} (R_{n+\ell,n+\ell} R_{i\ell} ) .
\]

Plugging this into (4.5) yields
\[
-\mathbb{E} (\eta + [V_n \tilde{g}]) G_{ii} - z \mathbb{E} F'_{ii} = 1 .
\] (4.6)

Specializing again Equation (4.4) for \( i \in [n] \) yields, with similar arguments,
\[
-\eta \mathbb{E} F_{ii} - \mathbb{E} [V_n \tilde{g}]_{ii} F_{ii} - z \mathbb{E} \tilde{G}_{ii} = 0 .
\] (4.7)

Arguing similarly with the help of the following integration by parts formula, valid for \( i > n \),
\[
\mathbb{E} (Y R)_{ij} = \sum_{\ell=1}^{n} \mathbb{E} (Y_{i\ell} R_{\ell,j} ) = \sum_{\ell=1}^{n} \frac{\sigma_{i\ell}^2}{n} \mathbb{E} \left[ \frac{\partial R_{ij}}{\partial Y_{i\ell}} \right] = -\sum_{\ell=1}^{n} \frac{\sigma_{i\ell}^2}{n} \mathbb{E} (R_{n+\ell,n+\ell} R_{ij} )
\]
yields the following equations
\[
-\mathbb{E} (\eta + [V_n \tilde{g}]) \tilde{G}_{ii} - z^* \mathbb{E} F_{ii} = 1 ,
\]
\[
-\mathbb{E} (\eta + [V_n \tilde{g}]) F_{ii} - z \mathbb{E} G_{ii} = 0 .
\] (4.8) (4.9)

Notice that equations (4.6)-(4.9) can be compactly written
\[
\mathbb{E} \left[ \begin{array}{cc} \text{diag}(V_n \tilde{g}) & -z \\
-z^* & -\text{diag}(V_n \tilde{g}) & -\eta \end{array} \right] R = I_{2n} .
\] (4.10)

Using Cauchy-Schwarz inequality and the estimates in Proposition B.2, we get
\[
\mathbb{E} [V_n \tilde{g}]_{ii} F_{ii} - \mathbb{E} [V_n \tilde{g}]_{ii} \mathbb{E} F_{ii} = \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right) .
\]

Hence
\[
- (\eta + \mathbb{E} [V_n \tilde{g}]_{ii}) \mathbb{E} F_{ii} = z \mathbb{E} \tilde{G}_{ii} + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right)
\]
by (4.7). In particular,
\[
-\mathbb{E} F_{ii} = z \frac{\mathbb{E} \tilde{G}_{ii}}{\eta + \mathbb{E} [V_n \tilde{g}]_{ii}} + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right)
\]
since \(|\eta + \mathbb{E} [V_n \tilde{g}]_{ii}|^{-1} \leq \text{Im}^{-1}(\eta)\). On the other hand, using the same decorrelation argument in equation (4.8), we obtain
\[
-\mathbb{E} (\eta + [V_n \tilde{g}]) \mathbb{E} \tilde{G}_{ii} - z^* \mathbb{E} F_{ii} = 1 + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right) .
\]
Combining these two equations, we finally get
\[ \mathbb{E}G_{ii} \left\{ -\eta + \mathbb{E}[V_n^T g]_i + \frac{|z|^2}{\eta + \mathbb{E}[V_n g]_i} \right\} = 1 + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right) . \]

Using the property (2.3) twice, one has
\[ \left| -\eta + \mathbb{E}[V_n^T g]_i + \frac{|z|^2}{\eta + \mathbb{E}[V_n g]_i} \right|^{-1} \leq \frac{1}{\text{Im}(\eta)} . \]

Hence
\[ \mathbb{E}G_{ii} = \frac{1}{-\eta + \mathbb{E}[V_n g]_i + \frac{|z|^2}{\eta + \mathbb{E}[V_n g]_i}} + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right) . \] (4.11)

Combining similarly equations (4.6) and (4.9) and decorrelating when needed with the help of Proposition B.2, we obtain the companion equation:
\[ \mathbb{E}G_{ii} = \frac{1}{-\eta + \mathbb{E}[V_n g]_i + \frac{|z|^2}{\eta + \mathbb{E}[V_n g]_i}} + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right) . \] (4.12)

We now introduce an unperturbed version of equations (4.11) and (4.12).

4.2. Schwinger–Dyson equations. For fixed \( \eta \in \mathbb{C}_+ \) and \( z \in \mathbb{C} \), consider the following system of equations in \( 2n \) unknowns \( p_1, \ldots, p_n, \bar{p}_1, \ldots, \bar{p}_n \in \mathbb{C} \), hereafter referred to as the Schwinger–Dyson equations:
\[ \begin{cases} p_i = \frac{[V_n^T p]_i + \eta}{|z|^2 - ([V_n \bar{p}]_i + \eta)([V_n^T p]_i + \eta)} ; & i \in [n] \\ \bar{p}_i = \frac{[V_n \bar{p}]_i + \eta}{|z|^2 - ([V_n \bar{p}]_i + \eta)([V_n^T p]_i + \eta)} \end{cases} \] (4.13)

where \( p, \bar{p} \) are the \( n \times 1 \) column vectors with components \( p_i, \bar{p}_i \), respectively. We introduce the notation \( \tilde{b} = \begin{pmatrix} b \\ \bar{b} \end{pmatrix} \) for any two \( n \times 1 \) vectors \( b \) and \( \bar{b} \) with complex components and the following definitions:
\[ \Upsilon(\tilde{b}, \eta) := \text{diag} \left( \frac{1}{|z|^2 - ([V_n \tilde{b}]_i + \eta)([V_n^T \tilde{b}]_i + \eta)} ; i \in [n] \right) , \] (4.14)
\[ := \text{diag}(\Upsilon_i(\tilde{b}, \eta) ; i \in [n]) , \]

and
\[ \mathcal{J}(\tilde{b}, \eta) := \begin{pmatrix} \Upsilon(\tilde{b}, \eta) & 0 \\ 0 & \Upsilon(\tilde{b}, \eta) \end{pmatrix} \tilde{b} + \eta \begin{pmatrix} \Upsilon(\tilde{b}, \eta) \mathbf{1} \\ \Upsilon(\tilde{b}, \eta) \mathbf{1} \end{pmatrix} . \] (4.15)

Then (4.13) can be compactly written as
\[ \tilde{p} = \mathcal{J}(\tilde{p}, \eta) . \] (4.16)

Both \( \Upsilon \) and \( \mathcal{J} \) depend on \( z \) as well (and to be even more precise, on \(|z|\)). We will not indicate this dependence in the sequel.

We now collect properties of solutions to (4.13).

**Proposition 4.1** (Schwinger–Dyson equations). For all fixed \( \eta \in \mathbb{C}_+ \) and \( z \in \mathbb{C} \), let \( p = (p_i) \) and \( \bar{p} = (\bar{p}_i) \) be two \( n \times 1 \) vectors which solve (4.13).

1. The system (4.13) admits a unique solution \( \tilde{p} \) satisfying \( \text{Im} \tilde{p} > 0 \).
(2) For any initial vector \( \vec{p}_0 \) with \( \text{Im} \vec{p}_0 > 0 \), the iterations

\[
\vec{p}_{k+1} = \mathcal{J}(\vec{p}_k)
\]

converge to this solution \( \vec{p} \) as \( k \to \infty \).

(3) For all \( z \in \mathbb{C} \) and \( i \in [n] \), the functions

\[
\eta \mapsto p_i(\eta) \quad \text{and} \quad \eta \mapsto \tilde{p}_i(\eta)
\]

are Stieltjes transforms of symmetric probability measures on \( \mathbb{R} \) respectively denoted by \( \mu_i \) and \( \tilde{\mu}_i \).

(4) Moreover, \( \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \tilde{p}_i \) and the common value

\[
\eta \mapsto \frac{1}{n} \sum_{i=1}^{n} p_i(\eta) = \frac{1}{n} \sum_{i=1}^{n} \tilde{p}_i(\eta)
\]

is the Stieltjes transform if a symmetric probability measure \( \tilde{\nu}_{n,z} \). We denote this Stieltjes transform by \( \eta \mapsto \tilde{g}_{\nu_{n,z}}(\eta) \).

(5) The sequences of probability measures \( (\mu_i; i \leq n; n \geq 1) \), \( (\tilde{\mu}_i; i \leq n; n \geq 1) \) and \( (\tilde{\nu}_{n,z}; n \geq 1) \) are tight. In particular,

\[
\sup_{n \geq 1} \int |x|^2 \tilde{\nu}_{n,z}(dx) < \infty.
\]

We henceforth refer to the solution \( \vec{p} = \vec{p}(\eta) \), \( \text{Im} \vec{p} > 0 \) as the solution to the Schwinger-Dyson equations.

**Proof.** The proof of part (1) relies on Earle-Hamilton’s theorem [41], which was first used (to the author’s knowledge) in random matrix theory by [43].

Let \( \mathcal{D} \) be the domain of \( \mathbb{C}^{2n} \) defined as \( \mathcal{D} := \{ \vec{p} \in \mathbb{C}^{2n} : \text{Im} \vec{p} > 0, \| \vec{p} \|_\infty < b \} \), where \( \| \cdot \|_\infty \) is the supremum norm. For a convenient choice of the constant \( b > 0 \), we will show that the holomorphic function \( \mathcal{J} \) on the domain \( \mathcal{D} \) satisfies the property that there is an \( \varepsilon > 0 \) such that the \( \varepsilon \)-neighborhood of \( \mathcal{J}(\mathcal{D}) \) lies in \( \mathcal{D} \). The Earle-Hamilton theorem then states that \( \mathcal{J} \) is a strict contraction with respect to the so-called Carathéodory-Riffen-Finsler metric, and the results of the proposition follow at once.

Write \( (\mathcal{J}(\vec{p}))_i = 1/d_i \) for \( i \in [n] \). For \( \vec{p} \in \mathcal{D} \), we have

\[
\text{Im} d_i = -\text{Im}(\text{Vec}(\vec{p})_i + \eta) - |z|^2 \frac{\text{Im}(\text{Vec}(\vec{p})_i + \eta)}{\| \text{Vec}(\vec{p})_i + \eta \|} \leq -\text{Im} \eta, \quad \text{and} \quad |d_i| \leq |\eta| + b + \frac{|z|^2}{\text{Im} \eta}.
\]

Therefore,

\[
\text{Im}(\mathcal{J}(\vec{p}))_i = \frac{-\text{Im} d_i}{|d_i|^2} \leq \frac{-\text{Im} \eta}{(|\eta| + b + |z|^2/\text{Im} \eta)^2} > 0, \quad \text{and} \quad |(\mathcal{J}(\vec{p}))_i| = \frac{1}{|d_i|} \leq \frac{1}{\text{Im} \eta}, \quad (4.17)
\]

and the case \( n+1 \leq i \leq 2n \) is handled similarly. By choosing \( b > 1/\text{Im} \eta \), we get the desired result on \( \mathcal{J}(\mathcal{D}) \).

Part (2) is a by-product of Earle-Hamilton’s theorem.

We now address part (3) and rely on the implicit function theorem for holomorphic functions (see for instance [32, Theorem 7.6, Chapter 1]). In order to prove that \( \eta \mapsto \vec{p}(\eta) \) is holomorphic, first notice that

\[
\vec{p} - \mathcal{J}(\vec{p}, \eta) = 0.
\]

If, for all \( \eta \in \mathbb{C}^+ \), the Jacobian \( J(\vec{p}, \eta) \) of \( \vec{p} \mapsto \vec{p} - \mathcal{J}(\vec{p}, \eta) \) differs from zero, then the function \( \eta \mapsto \vec{p}(\eta) \) will be holomorphic on \( \mathbb{C}^+ \).

Recall the definition of \( \Upsilon \) in (4.14). In order to express \( J(\vec{p}, \eta) \), we introduce a few more notations.
For given $n \times 1$ vectors $b$ and $\tilde{b}$, define
\[
\begin{align*}
\Delta(b) &= \text{diag}(\eta + (V^T b)_i) := \text{diag}(\Delta_i(b)) \\
\tilde{\Delta}(b) &= \text{diag}(\eta + (V^T \tilde{b})_i) := \text{diag}(\tilde{\Delta}_i(b)) .
\end{align*}
\] (4.18)

Recall that matrix $A$ is the standard deviation profile and satisfies $V = \frac{A}{\sqrt{n}} \circ \frac{A}{\sqrt{n}}$. For a given $2n \times 1$ vector $\tilde{b} = \begin{pmatrix} b \\ \tilde{b} \end{pmatrix}$, introduce the $2n \times 2n$ matrix
\[
\mathcal{A}(\tilde{b}) = \begin{pmatrix}
\frac{|z\mathbf{Y}(\tilde{b}) A^T}{\sqrt{n}} & \frac{\mathbf{Y}(\tilde{b}) \Delta(b) A}{\sqrt{n}} \\
\frac{\mathbf{Y}(\tilde{b}) \tilde{\Delta}(b) A^T}{\sqrt{n}} & \frac{|z\mathbf{Y}(\tilde{b}) A}{\sqrt{n}}
\end{pmatrix} .
\] (4.19)

Then straightforward computations yield
\[
J(\tilde{p}, \eta) = \det (I_{2n} - \mathcal{A}(\tilde{p}) \circ \mathcal{A}(\tilde{p})) .
\] (4.20)

On the other hand, straightforward but lengthy computations yield
\[
\text{Im}(\tilde{p}) = \mathcal{A}(\tilde{p}) \circ \overline{\mathcal{A}(\tilde{p})} \text{Im}(\tilde{p}) + v ,
\]
where $v > 0$ (see Section B.4 for details). By Proposition B.4,
\[
\rho \left( \mathcal{A}(\tilde{p}) \circ \overline{\mathcal{A}(\tilde{p})} \right) < 1 .
\]

Applying [45, Theorem 8.1.18], we have
\[
\rho \left( \mathcal{A}(\tilde{p}) \circ \overline{\mathcal{A}(\tilde{p})} \right) \leq \rho \left( \mathcal{A}(\tilde{p}) \circ \overline{\mathcal{A}(\tilde{p})} \right) < 1 .
\]

Hence $I_{2n} - \mathcal{A}(\tilde{p}) \circ \mathcal{A}(\tilde{p})$ is invertible and its determinant is nonzero. We focus on $p_i$, the same arguments will work for $\tilde{p}_i$. The implicit function theorem yields that $\eta \mapsto p_i(\eta)$ is an analytic map from $\mathbb{C}^+$ onto itself. It remains to prove that $\lim_{t \to \infty} itp_i(it) = -1$.

Since $\tilde{p} = \mathcal{J}(\tilde{p}, \eta)$, (4.17) implies that $|t p_i(it)| \leq 1$. Combining this estimate with (4.13), implies that $\lim_{t \to \infty} itp_i(it) = -1$.

It remains to prove that the probability measures associated to the $p_i$’s are symmetric. To this end, simply observe that, given $\eta \in \mathbb{C}_+$, if $\tilde{p} = (\tilde{p}, \tilde{p})$ is the solution with $\text{Im}(\tilde{p}) > 0$ of the Schwinger–Dyson equations (4.13), then $-\tilde{p}$ is the unique solution with $\text{Im}(-\tilde{p}) < 0$ of the analogous system obtained by replacing $\eta$ with $-\eta$. The result follows from the application of Lemma B.1. Part (3) of the theorem is established.

We now prove part (4), that is $\sum_i p_i = \sum_i \tilde{p}_i$. Getting back to the system (4.13), we have
\[
\sum_{i=1}^n p_i([V \tilde{p}]_i + \eta) = \sum_{i=1}^n \frac{([V \tilde{p}]_i + \eta)([V^T p]_i + \eta)}{([V \tilde{p}]_i + \eta)([V^T p]_i + \eta) + |z|^2} = \sum_{i=1}^n \tilde{p}_i([V^T p]_i + \eta) .
\]

But
\[
\sum_{i=1}^n p_i[V \tilde{p}]_i = \sum_{i=1}^n p_i \sigma^2_{i,\ell}\tilde{p}_i = \sum_{i=1}^n \tilde{p}_i [V^T p]_i .
\]

Since $\eta \neq 0$, we get the desired result.

We now prove part (5). Notice first that
\[
\text{Re} \left\{ t^2 (itp_i(it) + 1) \right\} = \int \frac{t^2 \lambda^2}{t^2 + \lambda^2} \mu_i(d\lambda) \quad \Rightarrow \quad \text{Re} \left\{ t^2 (itp_i(it) + 1) \right\} \xrightarrow{t \to \infty} \int \lambda^2 \mu_i(d\lambda) .
\]
Now, \( p(it) = ir(t) \) with \( r(t) = (r_i(t)) \) and \( r_i(t) > 0 \). In fact,

\[
p_i(it) = \int_{(-\infty,0)} \frac{1}{\lambda-it} \mu_i(d\lambda) + \int_{(0,\infty)} \frac{1}{\lambda-it} \mu_i(d\lambda) - \frac{1}{it} \mu_i(\{0\})
\]

\[
= \int_{(0,\infty)} \left( \frac{1}{\lambda-it} + \frac{1}{\lambda-it} \right) \mu_i(d\lambda) + \frac{1}{it} \mu_i(\{0\})
\]

\[
= 2i \int_{(0,\infty)} \frac{t}{\lambda^2 + t^2} \mu_i(d\lambda) + \frac{1}{it} \mu_i(\{0\}) = ir_i(t)
\]  

(4.21)

Similarly, \( \tilde{p}(it) = i\tilde{r}(t) \) with \( \tilde{r}(t) = (\tilde{r}_i(t)) \) and \( \tilde{r}_i(t) > 0 \). Notice for future use that

\[
\lim_{t \to \infty} tr_i(t) = 1 \quad \text{and} \quad \lim_{t \to \infty} \tilde{r}_i(t) = 1
\]

(4.22)

We now rely on the equation (4.13) satisfied by \( p_i \). We have

\[
\text{i}tp_i(it) + 1 = \text{i}t |z|^2 - ((V^T p)_i) + \text{i}t (V^T \tilde{p})_i + 1
\]

\[
= |z|^2 + t([V^T \tilde{r}]_i) + \text{i}t ([V^T r(t)]_i [V^T \tilde{r}(t)]_i)
\]

Multiplying by \( t^2 \), taking the limit as \( t \to \infty \) and taking into account (4.22), we finally get

\[
\int \lambda^2 \mu_i(d\lambda) = |z|^2 + |V1_n|_i \leq |z|^2 + \sigma_{\text{max}}^2.
\]

The same estimate holds for \( \int \lambda^2 \tilde{\mu}_i(d\lambda) \) and \( \int \lambda^2 \tilde{\nu}_{n,z}(d\lambda) \), hence the required tightness. \( \square \)

4.3. Asymptotics of the spectral measure \( \mathcal{L}_{n,z} \) and the Hermitian resolvent.

**Theorem 4.2.** Assume \( \mathbf{A0} \) and \( \mathbf{A1} \) hold, and let \( \tilde{\nu}_{n,z} \) be defined as in Proposition 4.1-(4). Then for all \( z \in \mathbb{C} \), \( \tilde{\nu}_{n,z} \) is tight, and

\[
\mathcal{L}_{n,z} \sim \tilde{\nu}_{n,z}
\]

almost surely. Moreover, for any \( \varepsilon > 0 \), \( x \mapsto \log |x| \) is \( \tilde{\nu}_{n,z} \)-integrable on the set \( \{|x| \geq \varepsilon\} \) and

\[
\int_{\{|x| \geq \varepsilon\}} \log |x| \mathcal{L}_{n,z}(dx) - \int_{\{|x| \geq \varepsilon\}} \log |x| \tilde{\nu}_{n,z}(dx) \overset{\text{a.s.}}{\xrightarrow{n \to \infty}} 0.
\]

We will sometimes refer to \( \tilde{\nu}_{n,z} \) as the deterministic equivalent of \( \mathcal{L}_{n,z} \).

The proof of Theorem 4.2 is postponed to Section 4.5. Notice that the first part (\( \mathcal{L}_{n,z} \sim \tilde{\nu}_{n,z} \)) is a variation of classical results, see for example [39]. It will be a direct consequence of the forthcoming theorem on the asymptotics of Hermitian resolvent.

In order to get some insight on the asymptotics of the spectral measure \( \mu_{n,z} \), we need more than the asymptotics of \( \mathcal{L}_{n,z} \). We rewrite hereafter the Schwinger–Dyson equations of Proposition 4.1 in a more suitable way for the forthcoming analysis. In what follows, the dependence in \( |z| \) is implicit and will be recalled if necessary.

We now introduce the deterministic equivalents to \( F \) and \( G \), defined in (4.2). Let \( \tilde{p} = (p, \tilde{p}) \) be the solution of the Schwinger–Dyson equations (4.13). Define the \( n \times n \) diagonal matrices \( P, \tilde{P} \), \( \Theta \) and \( \tilde{\Theta} \) by

\[
P := \text{diag}(p) \quad \tilde{P} := \text{diag}(\tilde{p})
\]

and

\[
\Theta := \text{diag}((V_n \tilde{p})_i, \ i \in [n]) \quad \tilde{\Theta} := \text{diag}((V_n^T p)_i, \ i \in [n])
\]
After easy massaging, the Schwinger–Dyson equations \( \mathbf{\bar{P}} = \mathcal{J}(\mathbf{\bar{P}}, \eta) \) are equivalent to:

\[
P = \left( -\Theta + \eta \right) \left| z \right|^2 (\Theta + \eta)^{-1}^{-1} \quad \text{and} \quad \bar{P} = \left( -\Theta + \eta \right) \left| z \right|^2 (\Theta + \eta)^{-1}^{-1}.
\]

Consider \( 2n \times 2n \) matrix \( \mathbf{S} \) defined as

\[
\mathbf{S} := -\left( \Theta(\left| z \right|, \eta) + \eta \frac{z}{\Theta(\left| z \right|, \eta) + \eta} \right)^{-1}.
\]

This definition is similar to equation (4.10) satisfied by the entries of the resolvent \( \mathbf{R} \). By the formula for the inverse of a partitioned matrix [45, §0.7.3], it holds that

\[
\mathbf{S} = \begin{pmatrix} P(\left| z \right|, \eta) & B(\left| z \right|, \eta) \\ B'(\left| z \right|, \eta) & \bar{P}(\left| z \right|, \eta) \end{pmatrix},
\]

where

\[
B(z, \eta) = -z \left( \Theta(\left| z \right|, \eta) + \eta \right)^{-1} \bar{P}(\left| z \right|, \eta) = -z P(\left| z \right|, \eta) \left( \Theta(\left| z \right|, \eta) + \eta \right)^{-1},
\]

and \( B'(z, \eta) \) can be made explicit in a similar fashion, but will not be used.

**Notation 4.3.** Let \( \alpha_n = \alpha_n(z, \eta) \) and \( \beta_n = \beta_n(z, \eta) \) be complex sequences such that there exist some constant \( C > 0 \) and some integers \( c_0, c_1 \) all independent from \( \eta \) and \( n \) but which may depend on \( z \) such that

\[
|\alpha_n| \leq \frac{C|\eta|^{c_1}}{\Im c_0(\eta) \wedge 1} |\beta_n|.
\]

We denote this by \( \alpha_n = \mathcal{O}_\eta(\beta_n) \). If \( \alpha_n = \alpha_n' \) and \( \beta_n = \beta_n' \) depend on some extra parameter \( i \in \mathcal{I} \), then the notation \( \mathcal{O}_\eta() \) in \( \alpha_n' = \mathcal{O}_\eta(\beta_n') \) must be understood uniform in \( i \). If \( \alpha_n \) and \( \beta_n \) are vectors or matrices, the notation \( \alpha_n = \mathcal{O}_\eta(\beta_n) \) corresponds to a uniform entrywise relation.

**Theorem 4.4.** Assume A0 and A1 hold. Then almost surely, for every \( z \in \mathbb{C} \) and \( \eta \in \mathbb{C}_+ \),

\[
\frac{1}{n} \left( \text{tr} \ G(z, \eta) \quad \text{tr} \ F(z, \eta) \right) - \frac{1}{n} \left( \text{tr} \ P(\left| z \right|, \eta) \quad \text{tr} \ B(\left| z \right|, \eta) \right) \xrightarrow{n \to \infty} 0.
\]

Moreover,

\[
\frac{1}{n} \left( \text{tr} \ \mathbb{E}G(z, \eta) \quad \text{tr} \ \mathbb{E}F(z, \eta) \right) - \frac{1}{n} \left( \text{tr} \ P(\left| z \right|, \eta) \quad \text{tr} \ B(\left| z \right|, \eta) \right) = \mathcal{O}_\eta \left( \frac{1}{\sqrt{n}} \right).
\]

The rate provided in the second part of the statement is not likely to be optimal, but it is sufficient for our purposes.

The proof of Theorem 4.4 is split into three intermediate results which are packaged in the following statements.

**Proposition 4.5.** Assume A0 holds and let \( z \in \mathbb{C} \) and \( \eta \in \mathbb{C}_+ \). Then almost surely,

\[
\frac{1}{n} \left( \text{tr} \ G(z, \eta) \quad \text{tr} \ F(z, \eta) \right) - \frac{1}{n} \left( \text{tr} \ \mathbb{E}G(z, \eta) \quad \text{tr} \ \mathbb{E}F(z, \eta) \right) \xrightarrow{n \to \infty} 0.
\]

**Proof.** This is a direct application of [22, Lemma 4.21]. \( \square \)

To manage the expectation terms \( n^{-1} \text{tr} \mathbb{E}(\cdot) \), we introduce the Gaussian counterparts of the quantities of interest. Consider a family of i.i.d. standard complex random variables \( (X_{ij}^N; 1 \leq i, j \leq n, \) where \( X_{ij}^N = (U + iU')/\sqrt{2} \), with \( U, U' \) being independent real \( \mathcal{N}(0,1) \) random variables. Notice in particular that

\[
\mathbb{E}X_{ij}^N = 0, \quad \mathbb{E}(X_{ij}^N)^2 = 0 \quad \text{and} \quad \mathbb{E}|X_{ij}^N|^2 = 1.
\]
Similarly, let \( Y_{ij}^N = \frac{x_{ij}}{\sqrt{n}} X_{ij}^N \), and let \( R_i^N, \ G_i^N, \ \tilde{G}_i^N, \ F_i^N, \) and \( F_i^N \) the matrix functions associated with the matrix \( Y_{ij}^N = (Y_{ij}^N) \) as in (4.1)-(4.2). Then we have the following proposition.

**Proposition 4.6.** Assume \( A_0 \) and \( A_1 \) hold. Let \( z \in \mathbb{C} \) and \( \eta \in \mathbb{C}_+ \). Then

\[
\frac{1}{n} \left( \text{tr} \ E G(z, \eta) \right) - \frac{1}{n} \left( \text{tr} \ E F(z, \eta) \right) = O_\eta \left( \frac{1}{n^{3/2}} \right).
\]

The proof of Proposition 4.6 relies on fairly standard arguments and is thus postponed to Appendix B.3.

**Proposition 4.7.** Assume \( A_1 \) holds, and let \( z \in \mathbb{C} \) and \( \eta \in \mathbb{C}_+ \). Then

\[
\frac{1}{n} \left( \text{tr} \ E G_i^N(z, \eta) \right) - \frac{1}{n} \left( \text{tr} \ E F_i^N(z, \eta) \right) = O_\eta \left( \frac{1}{n^{3/2}} \right).
\]

The proof of Proposition 4.7 follows hereafter in Section 4.4. It relies on an inequality on quadratic forms of independent interest (cf. Lemma B.5).

Theorem 4.4 now follows immediately from Propositions 4.5, 4.6 and 4.7.

From Theorem 4.4, we will deduce the asymptotic behavior of the empirical distribution \( L_{n,z} \) of the singular values of \( Y_n - z \) by analyzing the convergence of \( n^{-1} \text{tr} \ G(z, \eta) \). Moreover, for any \( \varepsilon > 0 \), the asymptotic behavior of \( \int_{\{|x| \geq \varepsilon\}} \log |x| \ L_{n,z}(dx) \) will be identified.

### 4.4. Proof of Proposition 4.7.

With these notations in hand and the definition of \( \Upsilon \) in (4.14) and of \( \Delta \) and \( \tilde{\Delta} \) in (4.18), we obtain

\[
\begin{align*}
\mathbb{E} G_{ii} - p_i &= \frac{1}{\eta + |V E \hat{g}|} - \frac{1}{\eta + |V \hat{p}|} + O_\eta \left( \frac{1}{n^{3/2}} \right) \\
&= Y_i (E \hat{g}) Y_i (\hat{p}) \left\{ |z|^2 [V^T (E \hat{g} - \hat{p})]_i + \Delta_i (E \hat{g}) \Delta_i (\hat{p}) [V (E \hat{g} - \hat{p})]_i \right\} + O_\eta \left( \frac{1}{n^{3/2}} \right), \quad (4.25)
\end{align*}
\]

\[
\begin{align*}
\mathbb{E} \tilde{G}_{ii} - \tilde{p}_i &= \frac{1}{\eta + |V T E \hat{g}|} - \frac{1}{\eta + |V T \hat{p}|} + O_\eta \left( \frac{1}{n^{3/2}} \right) \\
&= Y_i (E \hat{g}) Y_i (\hat{p}) \left\{ \tilde{\Delta}_i (E \hat{g}) \tilde{\Delta}_i (\hat{p}) [V^T (E \hat{g} - \hat{p})]_i + |z|^2 [V (E \hat{g} - \hat{p})]_i \right\} + O_\eta \left( \frac{1}{n^{3/2}} \right). \quad (4.26)
\end{align*}
\]

Denoting by \( \varepsilon = \mathbb{E} \hat{g} - \hat{p} \) and the definition of matrix \( A \) in (4.19), we can compactly express (4.25)-(4.26) as

\[
\varepsilon^* = A(\mathbb{E} \hat{g}) \odot A(\hat{p}) \varepsilon^* + O_\eta \left( \frac{1}{n^{3/2}} \right).
\]

The crux of the proof lies in the following proposition

**Proposition 4.8.** Let the matrix \( A \) be as in (4.19). Under the assumptions of Proposition 4.7, we have that the matrix

\[
I - A(\mathbb{E} \hat{g}) \odot A(\hat{p})
\]

is invertible and

\[
\| (I - A(\mathbb{E} \hat{g}) \odot A(\hat{p}))^{-1} \|_\infty = O_\eta (1) .
\]

The proof of Proposition 4.8 is postponed to Appendix B.4.

We are now in position to conclude. Notice that

\[
\| \mathbb{E} G_{ii} - p_i \| = \| e_i^* \varepsilon \| = \left| e_i^* (I - A(\mathbb{E} \hat{g}) \odot A(\hat{p}))^{-1} O_\eta \left( \frac{1}{n^{3/2}} \right) \right| = O_\eta \left( \frac{1}{n^{3/2}} \right),
\]
where \( e_i \) is the canonical \( 2n \times 1 \) vector \((\delta_{ij}, j \in [2n])\). A similar estimate holds for \( |\mathbb{E} \tilde{G}_{ii} - \tilde{p}_i| \).

Since these estimates are uniform in \( i \in [n] \), we obtain:

\[
\varepsilon_{\text{max}} := \max_{i \in [n]} \left( |\mathbb{E} G_{ii} - p_i|, |\mathbb{E} \tilde{G}_{ii} - \tilde{p}_i| \right) = \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right).
\]

Finally,

\[
\left| \frac{1}{n} \sum_{i \in [n]} \mathbb{E} G_{ii} - \frac{1}{n} \sum_{i \in [n]} p_i \right| \leq \frac{1}{n} \sum_{i \in [n]} |\mathbb{E} G_{ii} - p_i| \leq \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right).
\]

The same arguments apply verbatim for the term \( \frac{1}{n} \sum_i (\mathbb{E} \tilde{G}_{ii} - \tilde{p}_i) \). Consider now the term

\[
\frac{1}{n} \text{tr} \mathbb{E} F - \frac{1}{n} \text{tr} B = \frac{1}{n} \sum_{i=1}^n (\mathbb{E} F_{ii} - B_{ii})
\]

\[
= (a) \frac{1}{n} \sum_{i=1}^n \left\{ \frac{z}{-(V \mathbb{E} \bar{g}_i + \eta)} \mathbb{E} \tilde{G}_{ii} - \frac{z}{-(V \bar{p}_i + \eta)} \tilde{p}_i \right\} + \mathcal{O}_\eta \left( \frac{1}{n^{3/2}} \right),
\]

where \((a)\) follows from \((4.7)\) and \((4.24)\). One can now apply the same arguments as previously and handle similarly the term \( \frac{1}{n} \text{tr} \mathbb{E} F'(z, \eta) - \frac{1}{n} \text{tr} B'(z, \eta) \). This completes the proof of Proposition 4.7.

4.5 **Proof of Theorem 4.2.** The convergence \( \hat{L}_{n,z} \sim \tilde{\nu}_{n,z} \) is a direct consequence of Theorem 4.4. Now, it is easy to prove with the help of the law of large numbers that a.s.

\[
\limsup_n \int |x|^2 \hat{L}_{n,z}(dx) < \infty.
\]

This, together with Proposition 4.1-(5), yields

\[
\int \{|x| \geq \varepsilon\} \log |x| \hat{L}_{n,z}(dx) - \int \{|x| \geq \varepsilon\} \log |x| \tilde{\nu}_{n,z}(dx) \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad n \to \infty.
\]

The proof is complete.

5. **Proof of Theorem 2.2: Analysis of the Master Equations**

5.1 **Regularized Master Equations.** Our first step is to introduce the so-called Regularized Master Equations, which are obtained from the Schwinger–Dyson equations \((4.13)\) by taking \( \eta = it \) and substituting \( \bar{p}(it) = \text{ir}(t) \).

Given two \( n \times 1 \) vectors \( \bar{a} \) and \( \bar{a} \) with nonnegative components and fixed numbers \( s > 0 \) and \( t \geq 0 \), let \( \bar{a} = \begin{pmatrix} a \\ \bar{a} \end{pmatrix} \) and define the following quantities

\[
\Psi(\bar{a}, t) := \text{diag} \left( \frac{1}{s^2 + ((V_n \bar{a})_i + t)((V_n^t a)_i + t) ; i \in [n]} \right),
\]

\[
:= \text{diag}(\psi_i(\bar{a}, t) ; i \in [n]) ,
\]

and

\[
\mathcal{I}(\bar{a}, t) := \begin{pmatrix} \Psi(\bar{a}, t)V_n & 0 \\ 0 & \Psi(\bar{a}, t)V_n \end{pmatrix} \bar{a} + t \begin{pmatrix} \Psi(\bar{a}, t)1_n \\ \Psi(\bar{a}, t)1_n \end{pmatrix}.
\]

The proof of Proposition 2.1 amounts to showing that the vector equation admit \( \bar{\vec{r}} = \mathcal{I}(\bar{\vec{r}}, t) \) admits a unique solution \( \bar{\vec{r}} = \bar{\vec{r}}(s, t) > 0 \).
Lemma 5.1. We shall also prove that for any initial vector \( \vec{r}_0 \succcurlyeq 0 \), the iterations \( \vec{r}_{k+1} = \mathcal{I}(\vec{r}_k, t) \) converge to \( \vec{r} \) as \( k \to \infty \).

Proof of Proposition 2.1. We have proved in (4.21) that \( p(it) = \mathcal{I}(r(t)) \succcurlyeq 0 \), and similarly for \( \vec{p}(it) = \mathcal{I}(\vec{r}(t)) \succcurlyeq 0 \). The equations (2.8) satisfied by the \( r_i \)'s and the \( \vec{r}_i \)'s immediately follow from (4.13). The remaining properties follow from Proposition 4.1–(2) and (4).

5.2. Proof of Theorem 2.2. The proof makes a frequent use of some known properties of the nonnegative and irreducible matrices. For the sake of completeness, these properties are gathered in the following propositions.

Proposition 5.1 ([60, Theorems 1.1 and 5.5]). Let \( A \) and \( B \) be two square matrices such that \( 0 \preccurlyeq A \preccurlyeq B \). Then \( \rho(A) \leq \rho(B) \). Moreover, if \( B \) is irreducible, then \( \rho(A) = \rho(B) \) implies that \( A = B \).

Proposition 5.2 ([60, Theorem 1.6]). Let \( A \succcurlyeq 0 \) be a square and irreducible matrix, and let \( x \succcurlyeq 0 \) be a vector satisfying \( Ax \preccurlyeq x \). Then \( x \succcurlyeq 0 \) and \( \rho(A) \leq 1 \). Moreover, \( \rho(A) = 1 \) if and only if \( Ax = x \).

The proof of the following lemma is deferred to Section 8, see Proposition 8.2.

Lemma 5.3. Let \( V \) be a nonnegative and irreducible \( n \times n \) matrix, and let \( \vec{r}(s,t) \) be the solution of the regularized master equations (2.8). Let \( [a,b] \subset (0, \infty) \) and \( \varepsilon > 0 \), then

\[
\sup_{(s,t) \in [a,b] \times [0,\varepsilon]} \| \vec{r}(s,t) \| < \infty .
\]

In particular, \( \vec{r}(s,t) \) admits an accumulation point for fixed \( s > 0 \) as \( t \downarrow 0 \).

Next we show that any accumulation point provided by the above lemma constitutes a solution to the Master Equations (1.5):

Lemma 5.4 (Existence of solutions to the Master Equations). Let \( V \) and \( \vec{r}(s,t) \) be as in Lemma 5.3.

1. Let \( s > 0 \). If \( \vec{r}_* = (r_*, \vec{r}_*) \succcurlyeq 0 \) is an accumulation point for \( \vec{r}(s,t) \) as \( t \downarrow 0 \), then

\[
\begin{pmatrix} \vec{r}_* \\ 0 \end{pmatrix} = \begin{pmatrix} \Psi(\vec{r}_*)V^T & 0 \\ 0 & \Psi(\vec{r}_*)V \end{pmatrix} \vec{r}_* ,
\]

where we recall that \( \Psi(\vec{r}_*) = \text{diag}(\psi_i(\vec{r}_*))_{i=1}^n \) with \( \psi_i(\vec{r}_*) = (s^2 + (V\vec{r}_*)_i(V^T\vec{r}_*)_i)^{-1} \).

2. If moreover \( s^2 \in (0, \rho(V)) \), then \( \vec{r}_* \succcurlyeq 0 \).

Proof. The proof of part (1) is straightforward.

We now prove part (2) of the lemma. Let \( (t_k) \) be a positive sequence converging to zero in such a way that \( \lim_{k \to \infty} \vec{r}(s,t_k) = \vec{r}_* \). Since \( \vec{r}(s,t_k) \) satisfies (2.8), we have in particular that \( \Psi(\vec{r})(s,t_k)V\vec{r} \preccurlyeq \vec{r} \) and \( \Psi(\vec{r})(s,t_k)V^T\vec{r} \preccurlyeq \vec{r} \). From Proposition 5.2 it follows that that \( \rho(\Psi(\vec{r})(s,t_k)V) = \rho(\Psi(\vec{r})(s,t_k)V^T) < 1 \), and by the continuity of the spectral radius that \( \rho(\Psi(\vec{r}_*)V) \leq 1 \). If \( \vec{r}_* = 0 \), then \( \Psi(\vec{r}_*) = s^{-2}I \) and \( \rho(\Psi(\vec{r}_*)V) = s^{-2}\rho(V) > 1 \) since \( s^2 \in (0, \rho(V)) \), which yields a contradiction. Necessarily, \( \vec{r}_* \succcurlyeq 0 \).

Theorem 2.2–(1) and (2) are now consequences of the following lemma.

Lemma 5.5 (Uniqueness of solutions to the Master Equations). Let \( V \) be a nonnegative and irreducible \( n \times n \) matrix, and let \( \vec{q} \succcurlyeq 0 \) be a solution of the system (2.11), which exists by the previous lemma.
(1) If \( s^2 \geq \rho(V) \), then \( \bar{q} = 0 \).
(2) If \( s^2 \in (0, \rho(V)) \), then \( \bar{q} \) is unique as a solution of (2.11) satisfying \( \bar{q} \succneq 0 \). This solution satisfies \( \bar{q} \succ 0 \).

Proof. We first prove Item (1). Observe that \( \Psi(\bar{q})V^T \) is a nonnegative irreducible matrix for all \( \bar{q} \succneq 0 \). Assume that \( q \succneq 0 \). Then \( \rho(\Psi(\bar{q})V^T) = 1 \) and \( q \succ 0 \) by Proposition 5.2. From the equation \( 1^Tq = 1^T\bar{q} \) obtained from (2.11), we have \( \bar{q} \succneq 0 \). Therefore, \( \bar{q} \succ 0 \) by an argument similar to the one used for \( q \). Consequently, \((V^Tq)_i(V\bar{q})_i > 0 \) for all \( i \in [n] \), leading to the contradiction

\[
1 = \rho(\Psi(\bar{q})V^T) < \rho(s^{-2}V^T) = s^{-2}\rho(V) \leq 1
\]

where the strict inequality is due to Proposition 5.1. Hence \( \bar{q} = 0 \).

We now turn to Item (2). The argument \( \bar{q} \succneq 0 \Rightarrow \bar{q} \succ 0 \) is identical to Item (1).

The first step towards establishing uniqueness of the solution is showing that if \( \bar{q} = (q^T, \bar{q}^T)^T \) and \( \bar{q}' = (q'^T, \bar{q}'^T)^T \) are two positive solutions such that \( \bar{q} \neq \bar{q}' \), then \( \Psi(\bar{q}) \neq \Psi(\bar{q}') \). Assume the contrary. The equation \( q = \Psi(\bar{q})V^Tq \) shows that 1 is the Perron–Frobenius eigenvalue of the irreducible matrix \( \Psi(\bar{q})V^T \) (Proposition 5.2). Since its eigenspace has the dimension one, we get that \( q = \alpha q' \) for some \( \alpha > 0 \). A similar argument shows that \( \bar{q} = \tilde{\alpha} \bar{q}' \) for some \( \tilde{\alpha} > 0 \). Using the assumption \( \Psi(\bar{q}) = \Psi(\bar{q}') \) again and inspecting the expressions of these terms, we get that \( \alpha = \tilde{\alpha}^{-1} \). Moreover, the equations \( 1^Tq = 1^T\bar{q} \) and \( 1^Tq = 1^T\bar{q}' \) show that \( \alpha = \tilde{\alpha} \). This implies that \( \bar{q} = \bar{q}' \), a contradiction.

To establish the uniqueness, let us still consider the two positive solutions \( \bar{q} \neq \bar{q}' \). Write \( \Psi = \text{diag}(\psi_i) = \Psi(\bar{q}) \) and \( \Psi' = \text{diag}(\psi'_i) = \Psi(\bar{q}') \), and define the vectors

\[
\varphi = \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix} = V\bar{q}, \quad \varphi' = \begin{pmatrix} \varphi'_1 \\ \vdots \\ \varphi'_n \end{pmatrix} = V^Tq, \quad \bar{\varphi} = \begin{pmatrix} \varphi \\ \varphi' \end{pmatrix},
\]

and their similarly defined analogues \( \varphi', \bar{\varphi}' \), and \( \bar{\varphi}' \). It holds by the irreducibility of \( V \) that \( \varphi, \varphi' > 0 \). We now write

\[
\varphi_i = \frac{1}{n} \sum_{\ell=1}^n \sigma_{i,\ell}^2 \psi_\ell \varphi_\ell = \frac{1}{n} \sum_{\ell=1}^n \sigma_{i,\ell}^2 \psi_\ell^2 (s^2 \varphi_\ell + \varphi_\ell^2 \varphi_\ell)
\]

and a similar equation for \( \bar{\varphi}_i \), giving rise to the identity

\[
\bar{\varphi} = \begin{pmatrix} s^2V \psi^2 & V \psi^2 \Phi^2 \\ V^T \psi^2 \Phi \Phi^2 & s^2V^T \psi^2 \end{pmatrix} \bar{\varphi}
\]

where \( \Phi = \text{diag}(\varphi_i) \) and \( \bar{\Phi} = \text{diag}(\bar{\varphi}_i) \). Equivalently, the nonnegative matrix

\[
K_{\bar{q}} := \begin{pmatrix} s^2 \Phi^{-1}V \Phi^2 & \Phi^{-1}V \psi^2 \Phi^2 \\ \Phi^{-1}V^T \psi^2 \Phi & s^2 \Phi^{-1}V^T \psi^2 \end{pmatrix}
\]
satisfies $K_{q1} = 1$. Considering now the two solutions $\varphi$ and $\varphi'$, we can write

$$
\varepsilon_i := \left| \frac{\varphi_i - \varphi'_i}{\sqrt{\varphi_i \varphi'_i}} \right|
$$

$$
= \frac{1}{\sqrt{\varphi_i \varphi'_i}} \frac{1}{n} \sum_{\ell=1}^{n} \sigma^2_{i,\ell} (\psi_{\ell} \varphi_{\ell} - \psi'_{\ell} \varphi'_{\ell})
$$

$$
= \frac{1}{n} \sum_{\ell=1}^{n} \left( \frac{\sigma^2_{i,\ell} s^2 \psi_{\ell} \varphi_{\ell}}{\sqrt{\varphi_i \varphi'_i}} - \frac{\sigma^2_{i,\ell} \psi_{\ell} \varphi_{\ell} \sqrt{\varphi_i \varphi'_i}}{\sqrt{\varphi_i \varphi'_i}} \right)
$$

$$
\leq \frac{1}{n} \sum_{\ell=1}^{n} \left( \frac{\sigma^2_{i,\ell} s^2 \psi_{\ell} \varphi_{\ell}}{\sqrt{\varphi_i \varphi'_i}} \right)
$$

for every $i \in [n]$, and we also have a similar inequality for $\bar{\varepsilon}_i := \left| (\bar{\varphi}_i - \bar{\varphi}'_i)/\sqrt{\bar{\varphi}_i \bar{\varphi}'_i} \right|$. It results that the vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n, \bar{\varepsilon}_1, \ldots, \bar{\varepsilon}_n)^T$ satisfies the inequality $\varepsilon \ll K_{q,q} \varepsilon$, where

$$
K_{q,q} := \begin{pmatrix}
\sqrt{s^2 (\Phi \Phi')^{-1/2} V \Psi \Psi' (\Phi \Phi')^{1/2}} & (\Phi \Phi')^{-1/2} V \Psi \Psi' \Phi' (\Phi \Phi')^{1/2} \\
(\Phi \Phi')^{-1/2} V^T T \Psi \Psi' (\Phi \Phi')^{1/2} & s^2 (\Phi \Phi')^{-1/2} V^T T \Psi \Psi' (\Phi \Phi')^{1/2}
\end{pmatrix},
$$

and $\Phi' = \text{diag}(\varphi'_1, \ldots, \varphi'_n)$. By applying the Cauchy-Schwarz inequality to the scalar products $x_m 1, m = 1, \ldots, n$, where $x_m$ is the row $m$ of $K_{q,q}$, we get that $K_{q,q} 1 \ll (K_{q1}) \circ (K_{q1}) = 1$.

Now, for any $k \in \mathbb{N}$, we have

$$
\begin{pmatrix}
V & V^T \\
V^T & V^T
\end{pmatrix}^k \preceq \begin{pmatrix}
V^k & V^k \\
(V^T)^k & (V^T)^k
\end{pmatrix}.
$$

Since $\varphi, \varphi' > 0$ and $V$ is irreducible, it holds that for any $(i, j) \in [2n]^2$, there exists $k \in [n]$ such that $[K_{q,q}]_{ij} > 0$, implying that $K_{z,q}$ is irreducible. Relying on these results, we will show that there exists $i \in [n]$ such that $x_i 1 < 1$, i.e. Cauchy-Schwarz inequality is strict for this row vector. Proposition 5.2 will then show that $\rho(K_{q,q}) < 1$. By consequence, the only solution to the inequality $\varepsilon \ll K_{q,q} \varepsilon$ will be $\varepsilon = 0$, contradicting the assertion $\bar{q} \neq \bar{q}'$, and the uniqueness of the solution of (2.11) for $\bar{q} \neq 0$ will follow.

Recalling that $\Psi(\bar{q}) \neq \Psi(\bar{q}')$, there exists $\ell \in [n]$ such that $\varphi_{\ell} \bar{\varphi}_{\ell} \neq \varphi'_{\ell} \bar{\varphi}'_{\ell}$. Since $V$ is irreducible, no column of this matrix is zero. Therefore, we can choose $i \in [n]$ such that $\sigma^2_{i,\ell} > 0$. Consider the vector

$$
v_i := \left( \frac{s \sigma_{i,m} \psi_m(\bar{q}) \sqrt{\varphi_m}}{\sqrt{n} \sqrt{\varphi_i}} \right)^n_{m=1}, \left( \frac{s \sigma_{i,m} \psi_m(\bar{q}) \sqrt{\varphi_m}}{\sqrt{n} \sqrt{\varphi_i}} \right)^n_{m=1} = (v_{1,m})^{n}_{m=1}, (v_{2,m})^{n}_{m=1}
$$

and his analogue $v'_i$ (with the obvious notations). Consider also the $2 \times 2$ matrix

$$
M = \begin{pmatrix}
v_{1,\ell} & v_{2,\ell} \\
v'_{1,\ell} & v'_{2,\ell}
\end{pmatrix} = \begin{pmatrix}
\frac{\sigma_{i,\ell} \psi_i(\bar{q})}{\sqrt{n} \sqrt{\varphi_i}} & \frac{\sigma_{i,\ell} \psi_i(\bar{q})}{\sqrt{n} \sqrt{\varphi_i}} \\
\frac{\sigma_{i,\ell} \psi'_i(\bar{q})}{\sqrt{n} \sqrt{\varphi'_i}} & \frac{\sigma_{i,\ell} \psi'_i(\bar{q})}{\sqrt{n} \sqrt{\varphi'_i}}
\end{pmatrix} \begin{pmatrix}
\sqrt{\varphi_{\ell}} & \sqrt{\varphi'_{\ell}} \\
\sqrt{\varphi_{\ell}} & \sqrt{\varphi'_{\ell}}
\end{pmatrix} \begin{pmatrix} s \\ 1 \end{pmatrix} := M_1 M_2 M_3.
$$

Since $\det(M_2) = \sqrt{\varphi_{\ell} \varphi'_{\ell} \sqrt{\varphi_{\ell}} - \sqrt{\varphi'_{\ell}}} \neq 0$, the vectors $v_i$ and $v'_i$ are not colinear. Therefore, $x_i 1 = \langle v_i, v'_i \rangle < \|v_i\|^2 \|v'_i\|^2 = (K_{q1})_i (K_{q1})_i = 1$.

Lemma 5.5 is proved.

It remains to establish Theorem 2.2–(3). This is a consequence of following lemma, which also provides an expression for $\nabla \bar{q}(s)$ on $(0, \rho(V)^{1/2})$. 

\qed
Lemma 5.6. Assume that the nonnegative $n \times n$ matrix $V$ is irreducible. Then the function $s \mapsto \bar{q}(s)$ is continuous on $(0, \infty)$, and is continuously differentiable on $(0, \rho(V)^{1/2}) \cup (\rho(V)^{1/2}, \infty)$. Setting

$$M(s) = \begin{pmatrix} s^2 \Psi(\bar{q}(s))^2 V^T & -\Psi(\bar{q}(s))^2 \Phi(s)^2 V \\ -\Psi(\bar{q}(s))^2 \Phi(s)^2 V^T & s^2 \Psi(\bar{q}(s))^2 V \end{pmatrix},$$

$$\Phi(s) = \text{diag}(\varphi_i(s))_{i=1}^n, \quad \bar{\Phi}(s) = \text{diag}(\varphi_i(s))_{i=1}^n, \quad \varphi_i(s) = (V \bar{q}(s))_i, \quad \bar{\varphi}_i(s) = (V^T q(s))_i,$$

$$A(s) = \begin{pmatrix} 2n - M(s) \\ \frac{1}{1_n} - \frac{1}{1_n} \end{pmatrix} \in \mathbb{R}^{(2n+1) \times 2n}, \quad \text{and} \quad b(s) = - \begin{pmatrix} \Psi(\bar{q}(s))^2 V^T q(s) \\ \Psi(\bar{q}(s))^2 V \bar{q}(s) \end{pmatrix} \in \mathbb{R}^{2n+1},$$

the matrix $A(s)$ has a full column rank, and

$$\nabla \bar{q}(s) = 2sA(s)^{-1}b(s),$$

where $A(s)^{-1}$ is the left inverse of $A(s)$. On $(\rho(V)^{1/2}, \infty)$, it holds that $\bar{q}(s) = \nabla \bar{q}(s) = 0.$

Proof. We already know that $\bar{q}(s) = 0$ on $[\rho(V)^{1/2}, \infty)$, and we can easily check that $\nabla \bar{q}(s) = 2sA(s)^{-1}b(s)$ on $((0, \rho(V)^{1/2})$.

Let us show that $\bar{q}(s)$ is continuous on $[0, \rho(V)^{1/2})$. Fix $s \in (0, \rho(V)^{1/2})$. When $u$ belongs to a small neighborhood of $s$ in $(0, \rho(V)^{1/2})$, the function $\bar{q}(u)$ is bounded by Lemma 5.3, since $\bar{q}(u)$ is the limit as $t \downarrow 0$ of the bounded function $\bar{q}(u, t)$. Let $u_k \rightarrow_k s$ be such that $\bar{q}(u_k) \rightarrow_k \bar{q}_s$. The vector $\bar{q}_s$ is clearly a solution to (2.11). Observing that $\rho(\Psi(\bar{q}_s)V) = 1$, we get by the continuity of the spectral radius that $\rho(\Psi(\bar{q}_s)V) = 1$. If $\bar{q}_s$ were equal to zero, then we would have $\rho(\Psi(\bar{q}_s)V) = \rho(s^{-2}V) > 1$, a contradiction. Therefore $\bar{q}_s \neq 0$, and by Lemma 5.5–(2), $\bar{q}_s = \bar{q}(s)$ since $\bar{q}(s)$ is the only nonnegative and non zero solution to (2.11).

To obtain the continuity of $\bar{q}(s)$ on $(0, \infty)$, all what remains to prove is that $\bar{q}(u) \rightarrow 0$ as $u \uparrow \rho(V)^{1/2}$. Relying on Lemma 5.3, take a sequence $u_k \uparrow_k \rho(V)^{1/2}$ such that $\bar{q}(u_k) \rightarrow_k \bar{q}_s$. Then, the same argument as in the proof of Lemma 5.5–(1) shows that $\bar{q}_s = 0$.

To establish the differentiability of $\bar{q}(s)$ on $(0, \rho(V)^{1/2})$, we start by writing

$$q_i(s) = \psi_i \varphi_i = \psi_i^2 (s^2 \tilde{\varphi}_i + \varphi_i^2 \tilde{\varphi}_i) = \psi_i^2 (s^2 (V^T q)_i + \varphi_i^2 (V \bar{q})_i).$$

Doing a similar derivation for $\bar{q}_i(s)$, we get the equation $\bar{q}(s) = N(s)q(s)$, where

$$N(s) = \begin{pmatrix} s^2 \Psi(\bar{q}(s))^2 V^T & \Psi(\bar{q}(s))^2 \Phi(s)^2 V \\ \Psi(\bar{q}(s))^2 \Phi(s)^2 V^T & s^2 \Psi(\bar{q}(s))^2 V \end{pmatrix}.$$ 

As in the proof of Lemma 5.5–(2), we can show that $N(s)$ is irreducible. Thus, the Perron–Frobenius eigenvalue of $N(s)$ is equal to one, it is algebraically simple, and its associated eigenspace is generated by $\bar{q}(s)$.

Now, given two real numbers $s, s' \in (0, \rho(V)^{1/2})$ with $s \neq s'$, we have

$$q_i - q_i' = \psi_i \varphi_i - \psi_i' \varphi_i' = \psi_i \psi_i' ((\psi_i')^{-1} \varphi_i - \psi_i^{-1} \varphi_i')$$

$$= \psi_i \psi_i' ((s^2 - s'^2) \tilde{\varphi}_i + s^2 (\varphi_i - \tilde{\varphi}_i) - \varphi_i \varphi_i' (\varphi_i - \tilde{\varphi}_i))$$

where we set $q_i = q(s)$ and $q_i' = q(s')$, and we used the same notational shortcut for all the other quantities. We thus have

$$\frac{q_i - q_i'}{s^2 - s'^2} = - \left( \frac{\Psi \Psi' V^T q}{s^2 - s'^2} \right)_i + \left( \frac{s^2 \Psi \Psi' V^T q - q'}{s^2 - s'^2} \right)_i - \left( \frac{\Psi \Psi' \Phi \Phi' V (\bar{q} - \bar{q}')}{s^2 - s'^2} \right)_i.$$ 

Doing a similar derivation for $\bar{q}_i - \bar{q}_i'$, we obtain the system

$$\left( I - M(s, s') \right) \frac{\bar{q} - \bar{q}'}{s^2 - s'^2} = a(s, s')$$
where
\[
M(s, s') = \begin{pmatrix} s^2 \Psi' V^T & -\Psi \Psi' \Phi' \Phi' V \\ -\Psi \Psi' \Phi' \Phi' V^T & s^2 \Psi' V \end{pmatrix} \quad \text{and} \quad a(s, s') = -\left( \Psi' V^T \bar{q} \right).
\]
Using in addition the identity \( \sum q_i = \sum \tilde{q}_i \), we get the system
\[
A(s, s')(\bar{q} - \bar{q}')/(s^2 - s'^2) = b(s, s'),
\]
where
\[
A(s, s') = \left( I - M(s, s') \right) \in \mathbb{R}^{(2n+1) \times 2n} \quad \text{and} \quad b(s, s') = \begin{pmatrix} a(s, s') \\ 0 \end{pmatrix}.
\]
By the continuity of \( \bar{q}(s) \), \( A(s, s') \to A(s) \) and \( b(s, s') \to b(s) \) as \( s' \to s \). It is easy to see that \( (x^T, \bar{x}^T)^T \) is an eigenvector of \( M(s) \) if and only if \( (x^T, -\bar{x})^T \) is an eigenvector of \( N(s) \). Thus, the right null space \( I - M(s) \) is spanned by \( v(s) := (q(s)^T, -\bar{q}(s)^T)^T \). But we clearly have \( A(s)v(s) \neq 0 \), hence the matrix \( A(s) \) has a full column rank. Thus, for \( s' \) close enough to \( s \),
\[
\bar{q} - \bar{q}'/
\]
which shows that \( \bar{q}(s) \) is differentiable for any \( s \in (0, \rho(V)^{1/2}) \), with the gradient \( 2sA(s)^{-1}b(s) \).

The continuity of this gradient follows from the continuity of \( A(s) \) and \( b(s) \) and the fact that \( A(s) \) has a full column rank for any \( s \in (0, \rho(V)^{1/2}) \).

6. PROOF OF THEOREM 2.3-(1): TAIL ESTIMATES AND ASYMPTOTICS OF THE LOGARITHMIC POTENTIAL

The main purpose of this section is to show that the logarithmic potential \( U_{\mu_n^Y} \) is close to \( h_n(z) = -\int \log |x| \nu_{n,z}(dx) \) for large \( n \), and moreover that \( h_n \) is the logarithmic potential of a probability measure \( \mu_n^Y \). Recall that in Theorem 4.2 we have already established the almost sure convergence of the truncated potentials:
\[
\int_{\{|x| \geq \epsilon\}} \log |x| \hat{L}_{n,z}(dx) - \int_{\{|x| \geq \epsilon\}} \log |x| \nu_{n,z}(dx) \xrightarrow{\text{a.s.}} 0.
\]
Thus, we need to show that these measures uniformly integrate the singularity of \( x \mapsto \log |x| \) at 0. The proof has two main ingredients. The first is a result from [24] by the first author (stated in Proposition 6.1 below) that provides control on the smallest singular value of \( \hat{Y}_n - z \).

The second is the control of the remaining small singular values of \( \hat{Y}_n - z \) via the quantity \( E \text{Im} g_{L_{n,z}}(it) \) when \( t \) is close to zero. This leads to an estimate of the type
\[
E \text{Im} \hat{L}_{n,z}((-x, x)) \leq 2C \max(x, n^{-70}).
\]
Bounds of this form on the expected density of states for random Hermitian operators are sometimes referred to as local Wegner estimates (after the work [68]). Their application to the convergence of the empirical spectral distribution for non-Hermitian random matrices goes back to Bai’s proof of the circular law [17]; our presentation of the argument is closer to the one in [38]. Since we have obtained a quantitative comparison between \( E g_{L_{n,z}}(it) \) and \( \text{Im} g_{\nu_{n,z}}(it) \) in Theorem 4.4, the problem is reduced to obtaining bounds for the deterministic quantity
\[
\text{Im} g_{\nu_{n,z}}(it) = \frac{1}{n} \sum_{i=1}^{n} r_i(|z|, t)
\]
that are uniform in \( n \) and \( t \), where \( \{r_i\}_{i=1}^{n} \) were defined in Proposition 2.1. This is provided by Assumption A2.

To obtain the deterministic equivalents \( \mu_n \) for the ESDs \( \mu_n^Y \) we rely on a meta-model argument, which has been used before in [27, 51]. The idea is that for fixed \( n \) we can define a sequence \( \{Y_n^{(m)}\}_{m \geq 1} \) of \( nm \times nm \) random matrices as in Definition 1.2, where the standard deviation profiles
$A_n^{(m)}$ are obtained by replacing each entry $\sigma_{ij}$ of $A_n$ by an $m \times m$ block with entries all equal to $\sigma_{ij}$. We can then show that the logarithmic potentials of the associated ESDs converge to $h_n$ as $m \to \infty$, which will allow us to deduce that $h_n$ is itself the logarithmic potential of a probability measure. This argument is described in more detail in Section 6.2 below.

6.1. Control on small singular values. The following result, obtained by one of the authors in [24], gives an estimate on the lower tail of the smallest singular value $s_{n,z}$ of $Y_n - z$.

**Proposition 6.1** ([24], Theorem 1.19 and Corollary 1.22). Assume $A_0$ and $A_1$ hold, and fix $z \in \mathbb{C} \setminus \{0\}$. There exist constants $C(|z|, M_0, \sigma_{\text{max}}), \alpha(\varepsilon), \beta(|z|, \varepsilon, M_0\sigma_{\text{max}}) > 0$ such that for all $n \geq 1$,

$$
\mathbb{P}\left(s_{n,z} \leq n^{-\beta}\right) \leq C n^{-\alpha}.
$$

(6.1)

**Remark 6.1.** Similar bounds have been obtained under stronger assumptions on the standard deviation profile. For instance, (6.1) follows from [22, Lemma A.1] if we additionally assume $A_3$ (and in fact this result does not require $A_0$). Further assuming that $A_n$ is composed of a bounded number of blocks of equal size with constant entries, [5, Corollary 5.2] gives (6.1) with $\alpha > 0$ as large as we please (and $\beta = \beta(\alpha)$). An easy argument also gives (6.1) for arbitrary fixed $\alpha > 0$ and $\beta(\alpha)$ under $A_3$ and replacing $A_0$ with a bounded density assumption — see [22, Section 4.4]. For the case that the entries $X_{ij}$ are real Gaussian variables and $A_n(\sigma_0)$ is $(\delta, \kappa)$-broadly connected for some fixed $\sigma_0, \delta, \kappa \in (0, 1)$ (see Definition 2.18), (6.1) holds with arbitrary $\alpha > 0$ and $\beta = \alpha + 1$ by [59, Theorem 2.3].

We now consider the other small singular values of $Y_n - z$. The key is the uniform control on solutions to the Regularized Master Equations (2.8) provided by Assumption $A_2$:

$$
\sup_{n} \sup_{t \in (0, 1]} \frac{1}{n} \sum_{i=1}^{n} r_{i}(|z|, t) \leq C.
$$

Combined with Theorem 4.4 the above estimate allows us to establish **local Wegner estimates** for the empirical singular value distributions $L_{n,z}$ and associated deterministic equivalents $\check{v}_{n,z}$:

**Corollary 6.2** (Wegner estimates). Let $A_0$, $A_1$ and $A_2$ hold. Then, for all $z \in \mathbb{C} \setminus \{0\}$ there exist constants $C, \gamma_0 > 0$ such that for all $x > 0$,

$$
\check{v}_{n,z}((x, -x)) \leq C x
$$

(6.2)

and

$$
\mathbb{E}\check{L}_{n,z}((x, -x)) \leq C (x \vee n^{-\gamma_0}).
$$

(6.3)

**Proof.** We rely on the following elementary estimate for the Stieltjes transform of a probability measure $\mu$ (see for instance [38, Lemma 15]):

$$
\text{Im}g_{\mu}(it) = t \int \frac{\mu(d\lambda)}{\lambda^2 + t^2} \geq t \int_{-t}^{t} \frac{\mu(d\lambda)}{\lambda^2 + t^2} \geq \frac{1}{2t} \mu((-t, t)).
$$

(6.4)

Recall that

$$
\text{Im}g_{\check{v}_{n,z}}(it) = \frac{1}{n} \sum_{i \in [n]} r_{i}(|z|, t).
$$

The first Wegner estimate (6.2) is a straightforward consequence of Assumption $A_2$ and the estimate (6.4).

We now establish the second Wegner estimate (6.3) and first prove that there exists $\gamma_0 > 0$ such that

$$
\sup_{t \geq n^{-\gamma_0}} \mathbb{E}\text{Im}g_{L_{n,z}}(it) \leq C
$$

(6.5)
for all $n \geq 1$. For $t \geq 1$, $\mathbb{E} \Im g_{L_{n,z}}(it) \leq 1$ by the mere definition of a Stieltjes transform. Assume $t < 1$ and recall that $\mathbb{E} \Im g_{L_{n,z}}(it) = n^{-1} \text{tr} \Im \mathbb{E} G(z,it)$. By Theorem 4.4, there exist constants $c_0, C > 0$ such that

$$\left| \frac{1}{n} \text{tr} \Im \mathbb{E} G(z,it) - \frac{1}{n} \sum_{i=1}^{n} r_i(|z|,t) \right| \leq \frac{C}{\sqrt{nt^{c_0}}}.$$  

By A2, we therefore get that

$$\mathbb{E} \Im g_{L_{n,z}}(it) \leq C(t^{-c_0}n^{-1/2} + 1).$$  

By letting now $t \geq n^{-\gamma_0}$ with $\gamma_0 = 1/(2c_0)$, we obtain (6.5). Combining this result with (6.4), we get

$$\mathbb{E} \bar{L}_{n,z}((-x,x)) \leq \mathbb{E} \bar{L}_{n,z}((-x \lor n^{-\gamma_0}, x \lor n^{-\gamma_0})) \leq 2C(x \lor n^{-\gamma_0})$$  

which is the desired result.

6.2. Comparison of logarithmic potentials via a meta-model. We now turn to the task of finding the measures $\mu_n$ from Theorem 2.3 which serve as a sequence of deterministic equivalents for the ESDs $\mu_n^Y$. A first idea is to try to show that for every $\psi \in \mathbb{C}_c^\infty(\mathbb{C})$,

$$\int \psi(z) \mu_n^Y(dz) = -\frac{1}{2\pi} \int \Delta \psi(z) U_n^Y(z) dz = \frac{1}{2\pi} \int \Delta \psi(z) \left( \int_{\mathbb{R}} \log|x| \bar{L}_{n,z}(dx) \right) dz$$

is “close” to $-(2\pi)^{-1} \int \Delta \psi(z) h_n(z) dz$. However, there is a difficulty in directly applying this approach, related to the fact that $\bar{L}_{n,z}$ does not converge in general with no further assumption on the variance profile matrices $V_n$.

To circumvent this difficulty, we rely on a meta-model argument, which has been used in [27, 51], and which we now describe. Let $n$ be fixed, consider the standard deviation profile $A_n = (\sigma_{ij})$ and the normalized variance profile $V_n = \left( \frac{1}{n} \sigma_{ij}^2 \right)$. Recall the associated Schwinger–Dyson equations as provided in Proposition 4.1 and the solution $\tilde{\rho} = \left( \frac{\rho}{\tilde{\rho}} \right)$, of dimension $2n \times 1$. Define the meta-model in the following way: for an integer $m \geq 1$, consider the $nm \times nm$ standard deviation profile matrix defined as

$$A_n^{(m)} = \begin{pmatrix} A_n & \cdots & A_n \\ \vdots & \ddots & \vdots \\ A_n & \cdots & A_n \end{pmatrix} = (1_m 1_m^T) \otimes A_n,$$

associated to the normalized variance profile $V_n^{(m)} = (1_m 1_m^T) \otimes m^{-1} V_n$, and the random matrix

$$Y_n^{(m)} = \left( \frac{A_n^{(m)}}{\sqrt{mn}} X^{(nm)} \right)_{i,j \in [nm]}.$$  

(6.6)

Denote by $\tilde{L}_{n,z}^{(m)}$ the symmetrized empirical distribution of the singular values of $Y_n^{(m)} - z I_{nn}$. Due to the specific form of $V_n^{(m)}$, it is straightforward to check that the solutions of the canonical equations associated to this model are provided by

$$\tilde{\rho}_m = \left( \frac{p_m}{\tilde{p}_m} \right) \quad \text{where} \quad p_m = \left( \begin{array}{c} p \\ \vdots \\ p \end{array} \right) \quad \text{and} \quad \tilde{p}_m = \left( \begin{array}{c} \tilde{p} \\ \vdots \\ \tilde{p} \end{array} \right),$$

where $p_m$ and $\tilde{p}_m$ are $nm \times 1$ vectors. As an important consequence, we have:

$$g_{\tilde{L}_{n,z}^{(m)}}(\eta) = \frac{1}{mn} \sum_{i=1}^{mn} [\tilde{p}_m]_i = \frac{1}{n} \sum_{i=1}^{n} [\tilde{p}]_i = g_{L_{n,z}}(\eta).$$
Hence the Stieltjes transform $\hat{g}_\nu(n)\nu$ of $\nu(n)\nu$ does not depend on $m$ and is equal to $\nu(n)\nu$. Finally, if $A_0$ and $A_1$ are satisfied for $Y_n$, they are also satisfied for $Y^{(m)}_n$. In particular, $L^{(m)}_{n,z} \sim \nu(n)\nu$ admits a genuine limit as $m \to \infty$:   \[
abla_{m \to \infty} \nu(n)\nu \quad \text{a.s.} \tag{6.7}\]

since $\nu^{(m)}_{n,z} = \nu(n)\nu$.

We now state our proposition giving the existence of the measures $\mu_n$, the proof of which will occupy the bulk of the remainder of this section.

**Proposition 6.3.** Let $A_0$, $A_1$ and $A_2$ hold. Then the following hold.

1. For all $n \geq 1$ and $z \in \mathbb{C} \setminus \{0\}$, the function
   \[h_n(z) = -\int_{\mathbb{R}} \log|z| \nu(n,z)(dx)\]
   is well-defined and for every compact set $K \subset \mathbb{C}$,
   \[\sup_n \int_K |h_n(z)|^2 \, dz < \infty.\]
   Moreover, $h_n(z)$ coincides with the logarithmic potential $U_{\mu_n}(z)$ of a probability measure $\mu_n$ on $\mathbb{C}$.

2. For $\mu_n$ as defined in part (1), there exists a constant $C > 0$, independent of $n$, such that for all $M > 0$,
   \[\mu_n(\{z \in \mathbb{C}; |z| > M\}) \leq \frac{C}{M^2}.\]

We will make use of the following technical lemmas, whose proofs are deferred to Appendix C.

**Lemma 6.4.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space, $\zeta$ a finite positive measure on $\mathbb{C}$ and $f_n : \Omega \times \mathbb{C} \to \mathbb{R}$ measurable functions satisfying
   \[\sup_n \int_{\mathbb{C}} |f_n(\omega, z)|^{1+\alpha} \mathbb{P} \otimes \zeta(d\omega \times dz) \leq C\]
   for some constants $\alpha, C > 0$.

   Let $g : \mathbb{C} \to \mathbb{R}$ be a measurable function such that for $\zeta$-almost all $z \in \mathbb{C}$,
   \[f_n(\omega, z) \xrightarrow{\mathbb{P}} g(z).\]

   Then $\int_{\mathbb{C}} |g(z)|^{1+\alpha} \zeta(dz) \leq C$, and
   \[\int_{\mathbb{C}} f_n(\omega, z) \zeta(dz) \xrightarrow{\mathbb{P}} \int g(z) \zeta(dz).\]

**Lemma 6.5.** Let $(\zeta_n)$ be a sequence of random probability measures on $\mathbb{C}$. Assume that a.s. $(\zeta_n)$ is tight and that there exists a locally integrable function $h : \mathbb{C} \to \mathbb{R}$ such that for all $\psi \in C^\infty_c(\mathbb{C})$,
   \[\int \psi(z) \zeta_n(dz) \xrightarrow{\mathbb{P}} \frac{1}{2\pi} \int \Delta \psi(z) h(z) \, dz.\]

   Then there exists a non-random probability measure $\zeta$ on $\mathbb{C}$ with logarithmic potential $h$, i.e.
   \[h(z) = -\int \log|z - u| \zeta(du)\]
   for almost all $z \in \mathbb{C}$ such that
   \[\zeta_n \xrightarrow{\mathbb{w}} \zeta\]
in probability.
**Proof of Proposition 6.3.** We prove the first point of the proposition.

With $Y^{(m)}_n$ as in (6.6), denote by $\mu_{n,m}$ the spectral measure of $Y^{(m)}_n$, and recall that

$$U_{\mu_{n,m}}(z) = -\int_\mathbb{R} \log |x| \hat{L}_{n,z}^{(m)}(dx).$$

The proof consists of the following three steps:

1. To show that for every $z \in \mathbb{C} \setminus \{0\}$, $x \mapsto \log |x|$ is $\nu_{n,z}$-integrable and

$$U_{\mu_{n,m}}(z) \xrightarrow{P}{m \to \infty} h_n(z). \quad (6.11)$$

2. To show that the function $h_n(z)$ is measurable.

3. To show that for any compact set $\mathcal{K} \subset \mathbb{C}$,

$$\sup_m \mathbb{E} \int_{\mathcal{K}} |U_{\mu_{n,m}}(z)|^2 \, dz \leq C \quad (6.12)$$

for some constant $C > 0$ independent of $n$.

The three previous steps being proved, the assumptions of Lemma 6.4 are fulfilled and the lemma yields

$$\int_{\mathcal{K}} |h_n(z)|^2 \, dz \leq C$$

where $C$ does not depend upon $n$. Moreover,

$$\int \psi(z) \mu_{n,m}(dz) = -\frac{1}{2\pi} \int \Delta \psi(z) U_{\mu_{n,m}}(z) \, dz \xrightarrow{P}{m \to \infty} -\frac{1}{2\pi} \int \Delta \psi(z) h_n(z) \, dz$$

for every $\psi \in \mathcal{C}_c^\infty(\mathbb{C})$.

It remains to apply Lemma 6.5 to conclude that $h_n$ is the logarithmic potential of a probability distribution $\mu_n$ on $\mathbb{C}$, i.e.

$$h_n(z) = -\int \log |z - \lambda| \mu_n(d\lambda),$$

and point (1) of Proposition 6.3 will be proved.

Let us address Step 1 and prove (6.11). We first split the integrals of interest:

$$U_{\mu_{n,m}}(z) = -\int_{\{|x|<\varepsilon\}} \log |x| \hat{L}_{n,z}^{(m)}(dx) - \int_{\{|x|\geq\varepsilon\}} \log |x| \hat{L}_{n,z}^{(m)}(dx),$$

$$h_n(z) = -\int_{\{|x|<\varepsilon\}} \log |x| \tilde{\nu}_{n,z}(dx) - \int_{\{|x|\geq\varepsilon\}} \log |x| \tilde{\nu}_{n,z}(dx).$$

Taking into account the convergence (6.7) and applying Theorem 4.2 to the sequence $(\hat{L}_{n,z}^{(m)})_m$, we deduce that $x \mapsto \log |x|$ is $\tilde{\nu}_{n,z}$-integrable near infinity, and that

$$\int_{\{|x|\geq\varepsilon\}} \log |x| \hat{L}_{n,z}^{(m)}(dx) \xrightarrow{a.s.}{m \to \infty} \int_{\{|x|\geq\varepsilon\}} \log |x| \tilde{\nu}_{n,z}(dx). \quad (6.13)$$
We now handle the remaining integrals.

\[
\left| \int_{\{|x|<\varepsilon\}} \log |x| \tilde{\nu}_{n,z}(dx) \right| \leq \int_{\{|x|<\varepsilon\}} |\log |x|| \tilde{\nu}_{n,z}(dx)
\]

\[
= \int_{\mathbb{R}} \nu_{n,z} \{|x|<\varepsilon, \log |x| \leq -y\} dy
\]

\[
\leq C \int_{0}^{\infty} \left( \exp(-y) \wedge \varepsilon \right) dy
\]

\[
= C\varepsilon(1 - \log \varepsilon),
\]

where \((a)\) follows from Wegner’s estimate \((6.2)\) in Corollary 6.2.

Let \((s_{i,z})_{i\in[mn]}\) be the singular values of \(Y_{n}^{(m)} - z\) ordered as \(s_{1,z} \geq \cdots \geq s_{mn,z}\). For \(z \in \mathbb{C} \setminus \{0\}\) and \(\beta > 0\) the exponent as in Proposition 6.1, we introduce the event

\[
G_{m} := \left\{ s_{i,z} \geq (mn)^{-\beta}, i \in [mn] \right\}.
\]

For all \(\tau > 0\),

\[
P \left\{ \left| \int_{\{|x|<\varepsilon\}} \log |x| L_{n,z}^{(m)}(dx) \right| > \tau \right\}
\]

\[
\leq P \left\{ \mathbb{1}_{G_{m}} \int_{\{|x|<\varepsilon\}} \log |x| L_{n,z}^{(m)}(dx) > \frac{\tau}{2} \right\} + P \left\{ \mathbb{1}_{G_{m}^{c}} \int_{\{|x|<\varepsilon\}} \log |x| L_{n,z}^{(m)}(dx) > \frac{\tau}{2} \right\}.
\]

Noticing that

\[
\left\{ \mathbb{1}_{G_{m}^{c}} \int_{\{|x|<\varepsilon\}} \log |x| L_{n,z}^{(m)}(dx) > \frac{\tau}{2} \right\} \subset \left\{ s_{mn,z} \leq (mn)^{-\beta} \right\}
\]

for \(m\) large enough, Proposition 6.1 yields that

\[
P \left\{ \mathbb{1}_{G_{m}} \int_{\{|x|<\varepsilon\}} \log |x| L_{n,z}^{(m)}(dx) > \frac{\tau}{2} \right\} \leq \frac{C}{(mn)^{\alpha}}.
\]

Recall the constant \(\gamma_{0}\) in Corollary 6.2. Choose \(m\) large enough and \(\gamma \leq \gamma_{0}\) small enough so that \((mn)^{-\beta} \leq (mn)^{-\gamma} \leq \varepsilon \leq 1\). We now estimate

\[
E \mathbb{1}_{G_{m}} \int_{\{|x|<\varepsilon\}} |\log |x|| \mathcal{L}_{n,z}^{(m)}(dx)
\]

\[
= E \left( \frac{1}{mn} \sum_{i\in[mn]} \log(s_{i,z}) \mathbb{1}_{s_{i,z} \in[(mn)^{-\beta},\varepsilon]} \right)
\]

\[
= \int_{\{(mn)^{-\beta} \leq |x| \leq (mn)^{-\gamma}\}} \log |x| E \mathcal{L}_{n,z}^{(m)}(dx) + \int_{\{(mn)^{-\gamma} < |x| < \varepsilon\}} \log |x| E \mathcal{L}_{n,z}^{(m)}(dx)
\]

\[
:= I_{1} + I_{2}.
\]

By the Wegner estimate \((6.3)\), we obtain

\[
I_{1} \leq \beta \log(mn) E \mathcal{L}_{n,z}^{(m)}([-(mn)^{-\gamma}, (mn)^{-\gamma}]) \leq C \beta (mn)^{-\gamma} \log(mn).
\]
On the other hand, another application of the same Wegner estimate yields

\[
I_2 = \int_0^\infty \mathbb{E}[\tilde{L}_{n,z}^{(m)} \left( \{ x : |\log |x| \mathbb{I}_{[(mn)^{-\gamma}]}(|x|) \geq y \} \right) \, dy \\
= \int_0^\infty \mathbb{E}[L_{n,z}^{(m)} \left( \{ -e^{-y} \wedge \varepsilon, -(mn)^{-\gamma} \} \cup \{(mn)^{-\gamma}, e^{-y} \wedge \varepsilon \} \right) \, dy \\
\leq C \int_0^\infty (\exp(-y) \wedge \varepsilon) \, dy \\
= C\varepsilon(1 - \log \varepsilon).
\]

Therefore, by Markov’s inequality, we finally obtain

\[
P\left( \left| \int_{\{|x|<\varepsilon\}} \log |x| \tilde{L}_{n,z}^{(m)} (dx) \right| \geq \tau \right) \leq \frac{1}{\tau} 2C[(mn)^{-\gamma} \log(mn) + \varepsilon(1 - \log \varepsilon)] + C(mn)^{-\alpha}.
\]

Thus, for all \(\tau, \tau' > 0\), we can choose \(\varepsilon > 0\) small enough so that

\[
P\left( \left| \int_{\{|x|<\varepsilon\}} \log |x| \tilde{L}_{n,z}^{(m)} (dx) \right| > \tau \right) < \tau'
\]

for \(m\) large enough. Gathering this result with (6.13) and (6.14) yields (6.11), and Step 1 is proved.

We now address Step 2 and study the measurability of \(h_n(z)\). Define on \(\mathbb{C} \times (0, \infty)\) the functions

\[
U_{n,m}^Y(z, t) := -\frac{1}{2nm} \log(\det((Y_n^{(m)} - z)^{\ast}(Y_n^{(m)} - z) + t^2)) \\
= -\frac{1}{2} \int_{\mathbb{R}} \log(x^2 + t^2) L_{n,z}^{(m)}(dx), \\
U_n(z, t) := -\frac{1}{2} \int_{\mathbb{R}} \log(x^2 + t^2) \nu_{n,z}(dx).
\]

(6.15)

Given \(z, z' \in \mathbb{C}\), Hoffman-Wielandt’s theorem applied to \(Y_n^{(m)} - z\) and \(Y_n^{(m)} - z'\) yields

\[
\max_{i \in [mn]} |s_{i,z} - s_{i,z'}| \leq |z - z'|.
\]

Thus

\[
|U_{n,m}^Y(z, t) - U_{n,m}^Y(z', t)| = \frac{1}{2nm} \left| \sum_{i \in [mn]} \log(1 + \frac{s_{i,z}^2}{t^2}) - \log(1 + \frac{s_{i,z'}^2}{t^2}) \right| \\
\leq \frac{1}{2t^2} \max_{i \in [mn]} |s_{i,z} - s_{i,z'}| \\
\leq \frac{|z - z'|}{2t^2}
\]

and it follows that for any fixed \(t > 0\) the family \(\{ z \mapsto U_{n,m}^Y(z, t) \}_{m \geq 1}\) is uniformly equicontinuous. Since from Theorem 4.2 we have

\[
U_{n,m}^Y(z, t) \xrightarrow{a.s. \, m \to \infty} U_n(z, t)
\]

it follows that \(z \mapsto U_n(z, t)\) is continuous for any fixed \(t > 0\). Finally, since \(x \mapsto \log |x|\) is \(\nu_{n,z}\)-integrable near zero for any \(z \neq 0\) by (6.14),

\[
U_n(z, t) \xrightarrow{t \to 0} h_n(z).
\]

The measurability of \(h_n\) follows and Step 2 is proved.
We now address Step 3 and prove (6.12). Observe that on any compact set \( K \subset \mathbb{C} \), there exists a constant \( C_K \) such that
\[
\int_K (|\lambda - z|^2)^2 \, dz \leq C_K (1 + |\lambda|^2)
\]
for all \( \lambda \in \mathbb{C} \). Denote by \( (\lambda_i; i \in [nm]) \) the eigenvalues of \( Y_n^{(m)} \). We have
\[
\mathbb{E} \int_K |U_{\mu_{n,m}}(z)|^2 \, dz \leq \mathbb{E} \left( \frac{1}{nm} \sum_{i \in [mn]} \int_K (|\lambda_i^{(m)} - z|^2)^2 \, dz \right) \leq C_K \left( 1 + \mathbb{E} \int |\lambda|^2 \mu_{n,m}(d\lambda) \right).
\]
By the Weyl comparison inequality for eigenvalues and singular values (cf. e.g. [44, Theorem 3.3.13]),
\[
\int |\lambda|^2 \mu_{n,m}(d\lambda) = \frac{1}{nm} \sum_{i=1}^{nm} |\lambda_i|^2 \leq \frac{1}{nm} \sum_{i=1}^{nm} s_{i,0}^2 = \frac{1}{nm} \operatorname{tr} \left( Y_n^{(m)}(Y_n^{(m)})^* \right) \leq \frac{\sigma_{\max}^2}{(nm)^2} \sum_{i,j=1}^{nm} |X_{ij}^{(nm)}|^2.
\]
Taking the expectation of the previous inequality finally yields
\[
\mathbb{E} \int_K |U_{\mu_{n,m}}(z)|^2 \, dz \leq C.
\]
Step 3 is proved.

We now prove point (2) of Proposition 6.3. Given \( M > 0 \), we get from Lemma 6.4 that
\[
\limsup_m \mu_{n,m}^Y(\{z \in \mathbb{C}; |z| > M\}) \leq \limsup_m \frac{1}{M^2} \int_{|z| > M} |\lambda|^2 \mu_{n,m}(d\lambda) \leq \frac{C}{M^2} \quad \text{a.s.}
\]
where \( C > 0 \) is independent of \( n \). Let \( \psi \) be a nonnegative \( C_c^\infty(\mathbb{C}) \) function equal to one for \(|z| < M\) and to zero if \(|z| > M + 1\). As a byproduct of Lemma 6.5,
\[
\mu_{n,m}^Y \xrightarrow{w} \mu_n \quad m \to \infty
\]
amost surely. Consequently, on a set of probability one,
\[
\mu_n(\{z \in \mathbb{C}; |z| \leq M + 1\}) \geq \int \psi(z) \mu_n(dz) = \lim_m \int \psi(z) \mu_{n,m}(dz) \geq 1 - \frac{C}{M^2}.
\]
Proposition 6.3 is proved.

\[ \square \]

6.3. Conclusion of the proof of Theorem 2.3-(i). We can now complete the proof of Theorem 2.3-(i) and prove that \( \mu_n^Y \sim \mu_n \) in probability, with \( \mu_n \) defined in Proposition 6.3.

By Proposition 6.3, the sequence \( (\mu_n) \) is tight. It remains to prove that for all \( \varphi \in C_c(\mathbb{C}) \),
\[
\int \varphi \mu_n^Y - \int \varphi \mu_n \to 0 \quad \text{in probability.}
\]
By the density of \( C_c^\infty(\mathbb{C}) \) in \( C_c(\mathbb{C}) \), it is enough to show that
\[
\int \psi(z) \mu_n^Y(dz) - \int \psi(z) \mu_n(dz) = -\frac{1}{2\pi} \int \Delta \psi(z)(U_{\mu_n^Y}(z) - U_{\mu_n}(z)) \, dz \xrightarrow{n \to \infty} 0
\]
for all \( \psi \in C_c^\infty(\mathbb{C}) \). By mimicking the proof of Proposition 6.3, where \( Y_n^{(m)} \) and \( m \) are replaced with \( Y_n \) and \( n \) respectively, we straightforwardly obtain that \( U_{\mu_n^Y}(z) - U_{\mu_n}(z) \to 0 \) in probability for every \( z \in \mathbb{C} \setminus \{0\} \). This proof also shows that \( \sup_n \mathbb{E} \int_K |U_{\mu_n^Y}(z)|^2 \, dz < \infty \) for all compact sets \( K \subset \mathbb{C} \). We also know by Proposition 6.3 that \( \sup_n \mathbb{E} \int_K |U_{\mu_n}(z)|^2 \, dz < \infty \). The result now follows from Lemma 6.4.
7. Conclusion of proofs of Theorems 2.3 and 2.4

7.1. Proof of Theorem 2.3: Identification of \( \mu_n \). We established in the previous section that \( \mu_n \sim \mu \) in probability. To finish the proof of Theorem 2.3, it remains to show that \( \mu_n \) is rotationally invariant, and that its radial cumulative distribution function

\[
\mu_n \{ z \in \mathbb{C}; |z| \leq r \}
\]

coincides with the function \( F_n \) specified in the statement of the theorem. These facts, along with the properties of \( F_n \), are established in Lemma 7.2 below.

For the remainder of this section, we set

\[
b_n(z, t) := -\frac{z}{2n} \text{tr} \Psi(\|z\|, t) \quad \text{and} \quad b_n(z) = -\frac{z}{2n} \text{tr} \Psi(\|z\|),
\]

(7.1)

where \( \Psi(\cdot, t) \) and \( \bar{r}(\cdot, t) \) are respectively defined in (5.1) and Proposition 2.1, while \( \Psi(\cdot) \) and \( \bar{q}(\cdot) \) are respectively defined in (2.10) and Theorem 2.2.

Lemma 7.1. Under the same assumptions as in Theorem 2.3, the function \( z \mapsto b_n(z) \) is locally integrable on \( \mathbb{C} \), and

\[
\partial z U_{\mu_n}(z) = b_n(z)
\]

in \( \mathcal{D}'(\mathbb{C}) \).

Proof. Recall the definition of \( \mathcal{U}_n(z, t) \) in (6.15). Recall also that by Proposition 6.3-(i), the probability measure \( \mu_n \) is such that:

\[
U_{\mu_n}(z) = -\int_{\mathbb{R}} \log|x| \tilde{\nu}_{n, z}(dx).
\]

We first prove that

\[
U_n(z, t) \xrightarrow{\mathcal{D}'(\mathbb{C})} U_{\mu_n}(z).
\]

(7.2)

Recall from the proof of Proposition 6.3 that \( z \mapsto \mathcal{U}_n(z, t) \) is continuous for any fixed \( t > 0 \). It is moreover clear from the expressions of \( \mathcal{U}_n(z, t) \) and \( U_{\mu_n}(z) \) that \( \mathcal{U}_n(z, t) \uparrow U_{\mu_n}(z) \) as \( t \downarrow 0 \). Therefore, \( \mathcal{U}_n(t, z) \) is bounded from below on compact subsets of \( \mathbb{C} \), uniformly for \( t \in [0, t_0] \) for any fixed \( t_0 > 0 \). Recall that \( U_{\mu_n}(z) \), being a logarithmic potential, is locally integrable (as can be seen by Fubini’s theorem). Thus, on any compact subset of \( \mathbb{C} \),

\[
0 \leq U_{\mu_n}(z) - \mathcal{U}_n(t, z) \leq |U_{\mu_n}(z)| + \text{constant}
\]

and (7.2) immediately follows from the dominated convergence theorem. By a property of the convergence in \( \mathcal{D}'(\mathbb{C}) \), this implies the convergence of the distributional derivative

\[
\partial z \mathcal{U}_n(z, t) \xrightarrow{\mathcal{D}'(\mathbb{C})} \partial z U_{\mu_n}(z).
\]

(7.3)

We now prove that for all \( t > 0 \),

\[
\partial z \mathcal{U}_n(z, t) = b_n(z, t)
\]

(7.4)

in \( \mathcal{D}'(\mathbb{C}) \). We shall rely on a meta-model argument. Recall the meta-model \( Y_{n}^{(m)} \) introduced in (6.6), its limiting property (6.7), and the definition of \( \mathcal{U}_{\mu_n}^{Y_n}(z, t) \) in (6.15).
Fix \( t > 0 \). By Theorem 4.2, \( \mathcal{U}_{n,m}^Y(z,t) \to \mathcal{U}_n(z,t) \) almost surely as \( m \to \infty \) for all \( z \in \mathbb{C} \). Furthermore, recalling the notation \((s_{i,z})_{i\in[mn]}\) for the singular values of \( Y_n^{(m)} - z \), we have

\[
|\mathcal{U}_{n,m}^Y(z,t)|^2 = \left| \log(t) + \frac{1}{2mn} \sum \log \left( 1 + \frac{s_{i,z}^2}{t^2} \right) \right|^2 \\
\leq 2 \left| \log(t) \right|^2 + 2 \left| \frac{1}{2mn} \sum \log \left( 1 + \frac{s_{i,z}^2}{t^2} \right) \right|^2 \\
\leq 2 \left| \log(t) \right|^2 + \frac{1}{2mn} \sum \left| \log \left( 1 + \frac{s_{i,z}^2}{t^2} \right) \right|^2 \leq 2 \left| \log(t) \right|^2 + \frac{1}{t^2mn} \sum s_{i,z}^2
\]

where \((a)\) follows from the elementary inequality \( 2^{-1} \log^2(1 + x) \leq x \), valid for \( x \geq 0 \). In particular, this implies that

\[
\mathbb{E} \int_K |\mathcal{U}_{n,m}^Y(z,t)|^2 \, dz \leq (\log t)^2 + \frac{1}{t^2} \mathbb{E} \int_K \frac{\text{tr} (Y_n^{(m)} - z)^*(Y_n^{(m)} - z) mn}{\sum n,m} \, dz \leq C
\]
on every compact set \( K \subset \mathbb{C} \). By Lemma 6.4, we get that \( \mathcal{U}_n(\cdot, t) \) is locally integrable on \( \mathbb{C} \), and that

\[
\int \partial_\psi(z) \mathcal{U}_{n,m}^Y(z,t) \, dz \xrightarrow{m \to \infty} \int \partial_\psi(z) \mathcal{U}_n(z,t) \, dz \quad (7.5)
\]

for all \( \psi \in C^\infty_c(\mathbb{C}) \). An integration by parts along with Jacobi’s formula shows that for all \( \omega \in \Omega \), the distributional derivative \( \partial_\omega \mathcal{U}_{n,m}^Y(z,t) \) coincides with the pointwise derivative, which is given by

\[
\partial_\omega \mathcal{U}_{n,m}^Y(z,t) = \frac{1}{2nm} \text{tr} (Y_n^{(m)} - z)((Y_n^{(m)} - z)^*(Y_n^{(m)} - z) + t^2)^{-1}.
\]

On the other hand, we know from Theorem 4.4 that \( \partial_\omega \mathcal{U}_{n,m}^Y(z,t) \to b_n(z,t) \) almost surely as \( m \to \infty \), for all \( z \in \mathbb{C} \). Moreover, from a singular value decomposition of \( Y_n^{(m)} - z \) we easily see that \( |\partial_\omega \mathcal{U}_{n,m}^Y(z,t)| \leq (4t)^{-1} \). Consequently, we get by Lemma 6.4 again that

\[
\int \partial_\psi(z) \mathcal{U}_{n,m}^Y(t,z) \, dz = - \int \psi(z) \partial_\omega \mathcal{U}_{n,m}^Y(t,z) \, dz \xrightarrow{m \to \infty} - \int \psi(z) b_n(z,t) \, dz.
\]

Comparing with \((7.5)\), we obtain that \( \partial_\omega \mathcal{U}_n(z,t) = b_n(z,t) \) in \( D'(\mathbb{C}) \).

We now consider the limit in \( t \downarrow 0 \) in \((7.4)\). Since \( |b_n(z,t)| \leq |2z|^{-1} \), the dominated convergence theorem yields

\[
b_n(z,t) \xrightarrow{D'(\mathbb{C})} b_n(z) \quad (t \downarrow 0).
\]

Combining this convergence together with \((7.3)\) and \((7.4)\), we obtain the desired result. \( \square \)

In order to characterize the probability measure \( \mu_n \), we use the equation \( \mu_n = -(2\pi)^{-1} \Delta U_{\mu_n} \) and rely on the smoothness properties of \( \Delta U_{\mu_n} \) that can be deduced from Lemma 5.6. We recall that \( \bar{q}(s) \) is defined in the statement of Theorem 2.2.

**Lemma 7.2.** The probability measure \( \mu_n \) is rotationally invariant. On \((0, \infty)\), the distribution function \( F_n(s) := \mu_n(\{z: |z| \leq s\}) \) satisfies

\[
F_n(s) = 1 - \frac{1}{n} \langle q(s), V \bar{q}(s) \rangle.
\]

The support of \( \mu_n \) is contained in \( \{z: |z| \leq \sqrt{\rho(V)}\} \). Finally, \( F_n \) is absolutely continuous on \((0, \infty)\), and has a continuous density on \((0, \sqrt{\rho(V)})\).
Before entering the proof, we note that the rotational invariance of \( \mu_n \) can be “guessed” from the form of the Schwinger–Dyson equations of Proposition 4.1. Indeed, from this one sees that the Stieltjes transform \( g_{\nu_{n,z}}(\eta) = n^{-1}\text{tr} \ P(|z|, \eta) \) of \( \nu_{n,z} \) depends on \( z \) only through its absolute value, and this is therefore also the case for \( U_{\mu_n}(z) \). It is easy to check that this yields the rotational invariance of \( \mu_n \).

**Proof.** First, we show that \( \mu_n(C) = 0 \), where \( C \) is the circle with center zero and radius \( \sqrt{\rho(V)} \).

Consider a smooth function \( \phi : \mathbb{R} \to [0, 1] \) with support in \([-1, 1]\) and value \( \phi(0) = 1 \), and the function
\[
g_\varepsilon(z) = \phi \left( \frac{|z| - \sqrt{\rho(V)}}{\varepsilon} \right), \quad z \in \mathbb{C}
\]
with support in the annulus \( \{ z : \sqrt{\rho(V)} - \varepsilon \leq |z| \leq \sqrt{\rho(V)} + \varepsilon \} \). We have
\[
\int g_\varepsilon(z) \mu_n(dz) = -\frac{1}{2\pi} \int g_\varepsilon(z) \Delta U_{\mu_n}(z) \ell(dz) = \frac{2}{\pi} \int \partial_z g_\varepsilon(z) b_n(z) \ell(dz)
\]
where \( b_n \) is defined in (7.1). Notice that
\[
\lim_{\varepsilon \downarrow 0} \int g_\varepsilon(z) \mu_n(dz) = \mu_n(C).
\]
By replacing \( |z| = \sqrt{x^2 + y^2} \) and computing \( \partial_z = \frac{1}{2} (\partial_x - i \partial_y) \), we get
\[
\partial_z g_\varepsilon(z) = \frac{z}{2\varepsilon |z|} \phi' \left( \frac{|z| - \sqrt{\rho(V)}}{\varepsilon} \right).
\]
Hence, replacing \( b_n \) by its expression in (7.1), we obtain
\[
\frac{2}{\pi} \int \partial_z g_\varepsilon(z) b_n(z) \ell(dz) = -\frac{1}{2\varepsilon n \pi} \int |z| \phi' \left( \frac{|z| - \sqrt{\rho(V)}}{\varepsilon} \right) \text{tr} \Psi(\tilde{\Phi}(|z|)) \ell(dz)
\]
\[
\overset{(a)}{=} -\frac{1}{2\varepsilon n \pi} \int_{\theta=0}^{2\pi} \sqrt{\rho(V) + \varepsilon} \phi' \left( \frac{\rho - \sqrt{\rho(V)}}{\varepsilon} \right) \text{tr} \Psi(\tilde{\Phi}(\rho)) \rho^2 d\rho d\theta
\]
\[
\overset{(b)}{=} -\frac{1}{\varepsilon n} \int_{-1}^{1} \left( \sqrt{\rho(V) + \varepsilon u} \right)^2 \phi'(u) \text{tr} \Psi(\tilde{\Phi}(\sqrt{\rho(V) + \varepsilon u})) \varepsilon du
\]
where \( (a) \) follows from a change of variables in polar coordinates and \( (b) \), from the change of variable \( u = \frac{\rho - \sqrt{\rho(V)}}{\varepsilon} \). Since \( n^{-1}\text{tr} \Psi(\tilde{\Phi}(|z|)) \leq |z|^{-2} \), the dominated convergence theorem yields
\[
-\frac{1}{n} \int_{-1}^{1} \left( \sqrt{\rho(V) + \varepsilon u} \right)^2 \phi'(u) \text{tr} \Psi(\tilde{\Phi}(\sqrt{\rho(V) + \varepsilon u})) du
\]
\[
\overset{\varepsilon \downarrow 0}{\longrightarrow} -\frac{\rho(V)}{n} \text{tr} \Psi(\tilde{\Phi}(\sqrt{\rho(V)})) \int_{-1}^{1} \phi'(u) du
\]
\[
= -\frac{\rho(V)}{n} \text{tr} \Psi(\tilde{\Phi}(\sqrt{\rho(V)})) [\phi(1) - \phi(-1)] = 0.
\]
Equating with (7.6), we finally conclude \( \mu_n(C) = 0 \).

By Theorem 2.2–(3), the mapping \( z \mapsto \tilde{\Phi}(|z|) \) is continuously differentiable on the open set \( \mathcal{D} := \{ z \in \mathbb{C} : |z| \neq 0, |z| \neq \rho(V)^{1/2} \} \). Therefore, \( b_n(z) \) is continuously differentiable on this set, and for any \( \gamma \in C_\infty^\infty(\mathcal{D}) \), we get
\[
\int_{\mathbb{C}} g(z) \mu_n(dz) = -\frac{1}{2\pi} \int_{\mathbb{C}} g(z) \Delta U_{\mu_n}(z) dz = -\frac{2}{\pi} \int_{\mathbb{C}} g(z) \partial_z b_n(z) dz = \int_{\mathbb{C}} g(z) f_n(z) dz
\]
where the density \( f_n(z) \) is given by
\[
 f_n(z) := -\frac{2}{\pi} \partial_z b_n(z) = \frac{1}{n\pi} \partial_z \left( z \text{tr} \Psi(\tilde{q}(|z|)) \right) = \frac{1}{n\pi} \left\{ \text{tr} \Psi(\tilde{q}(|z|)) + |z|^2 \partial_z \text{tr} \Psi(\tilde{q}(|z|)) \right\}
\]
\[
= \frac{1}{n\pi} \partial_z \left\{ |z|^2 \text{tr} \Psi(\tilde{q}(|z|)) \right\} = \frac{-1}{n\pi} \sum_{i=1}^{n} \partial_z \left( \frac{\varphi_i \varphi_i^t}{|z|^2 + \varphi_i \varphi_i^t} \right)
\]
\[
= \frac{-1}{n\pi} \partial_z (q(|z|), V \tilde{q}(|z|)).
\]

Since \( f_n(z) \) depends only on \( |z| \), this density is rotationally invariant. From Theorem 2.2–(1), \( f_n(z) = 0 \) for \( |z| > \sqrt{\rho(V)} \). Thus, the support of \( \mu_n \) is contained in \( B(0, \sqrt{\rho(V)}) \). Moreover, \( F_n(\sqrt{\rho(V)}) = 1 = \lim_{v \uparrow \sqrt{\rho(V)}} F_n(v) \), since \( \mu_n(C) = 0 \). Given \( 0 < s < v < \sqrt{\rho(V)} \), we have
\[
F_n(v) - F_n(s) = \int_{B(0,v) \setminus B(0,s)} f_n(z) \, dz = \int_{0}^{2\pi} d\theta \int_{s}^{v} \frac{-1}{2\pi r n} \partial_r(q(r), V \tilde{q}(r)) \, r \, dr
\]
\[
= \frac{1}{n} \langle q(s), V \tilde{q}(s) \rangle - \frac{1}{n} \langle q(v), V \tilde{q}(v) \rangle.
\]

By taking \( v \uparrow \sqrt{\rho(V)} \), \( \langle q(v), V \tilde{q}(v) \rangle \rightarrow 0 \), and we get the expression (2.12). Finally, the continuity of the density of \( F_n \) on \( (0, \sqrt{\rho(V)}) \) follows from Theorem 2.2–(3). \( \square \)

7.2. Proof of Theorem 2.4. In Example 2.2, it has been proved that Theorem 2.4 holds under the additional assumption that the matrices \( A_n \) are irreducible. Now for the general case, by conjugating \( Y_n \) by a permutation matrix we may assume \( A_n \) takes the form
\[
 A_n = \begin{pmatrix}
 A_n^{(1)} & 0 & 0 & \cdots & 0 \\
 0 & A_n^{(2)} & 0 & \cdots & 0 \\
 0 & 0 & A_n^{(3)} & \cdots & 0 \\
 \vdots & \ddots & \ddots & \ddots & \vdots \\
 0 & 0 & 0 & \cdots & A_n^{(m)}
\end{pmatrix}
\]
(7.7)

where \( A_n^{(1)}, \ldots, A_n^{(m)} \) are square irreducible matrices of respective dimension \( n_1 \geq \cdots \geq n_m \). Indeed, for \( A_n \) a general nonnegative matrix we can achieve this with the upper triangular blocks not necessarily zero, but these are forced to be zero by the stochasticity condition. Also, by A1 and the row-sum constraint applied to the last row of \( A_n = (\sigma_{ij}) \), \( 1 = \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij}^2 \leq \frac{m}{n} \sigma_{\text{max}}^2 \) so in fact we have
\[
n_1, \ldots, n_m \geq n/\sigma_{\text{max}}^2.
\]
(7.8)

Denote the corresponding submatrices of \( X_n \) by \( X_n^{(k)} \) and set
\[
 Y_n^{(k)} = \frac{1}{\sqrt{n}} A_n^{(k)} \odot X_n^{(k)} = \frac{1}{\sqrt{n_k}} B_n^{(k)} \odot X_n^{(k)}
\]
(7.9)

where we set \( B_n^{(k)} = (n_k/n)^{1/2} A_n^{(k)} \). For each \( k \) we have:

1. \( A_n^{(k)} \) is irreducible,
2. \( \frac{1}{n_k} B_n^{(k)} \odot B_n^{(k)} \) is doubly stochastic, and
3. \( n_k \to \infty \) as \( n \to \infty \) (by (7.8)).

Thus, for each \( k \) the ESD \( \mu_n^{(k)} \) of \( Y_n^{(k)} \) converges weakly in probability to \( \mu_{\text{circ}} \). Since the \( \mu_n^Y \) is the weighted sum:
\[
\mu_n^Y = \frac{(n_1/n)}{\mu_n^{(1)}} + \cdots + \frac{(n_m/n)}{\mu_n^{(m)}},
\]
we get that \( \mu_n^Y \) converges weakly in probability to \( \mu_{\text{circ}} \). This concludes the proof of Theorem 2.4.
8. BOUNDEDNESS OF SOLUTIONS TO THE REGULARIZED MASTER EQUATIONS

In this section we are concerned with establishing bounds on the solution \( \vec{\tau}(s,t) \succ 0 \) to the Regularized Master Equations (2.8) that are uniform in the regularization parameter \( t > 0 \).

Here we will view the standard deviation profile \( A \) and all parameters as fixed. Hence, we fix \( n \geq 1 \) and consider an arbitrary nonnegative \( n \times n \) matrix \( A = (\sigma_{ij}) \). Putting \( V = \frac{1}{n} A \odot A \) and fixing \( s,t > 0 \), we let \( \vec{\tau} = \vec{\tau}(s,t) \) denote the unique solution to the Regularized Master Equations satisfying \( \vec{\tau} \succ 0 \), as is provided by Proposition 2.1.

8.1. Some preparation and proofs of Propositions 2.5 and 2.6. We begin by recording some key estimates and identities that will be used repeatedly in the sequel. We may write the Regularized Master Equations (2.8) in component form as

\[
\frac{1}{r_i} = \varphi_i + \frac{s^2}{\bar{\varphi}_i}, \quad \frac{1}{\bar{r}_i} = \bar{\varphi}_i + \frac{s^2}{\varphi_i}
\]

where

\[
\bar{\varphi}_i := t + (V^T \tau)_i, \quad \varphi_i := t + (V \bar{\tau})_i
\]  

(from Proposition 2.1 we have \( r_i, \bar{r}_i > 0 \) for all \( i \in [n] \), so we are free to take reciprocals). Additionally from Proposition 2.1 we have trace identity:

\[
\frac{1}{n} \sum_{j=1}^{n} r_j = \frac{1}{n} \sum_{j=1}^{n} \bar{r}_j.
\]

From (8.1) it is immediate that

\[
\frac{1}{2} \leq \min \left( \frac{1}{\varphi_i / s^2, 1/\bar{\varphi}_i} \right), \quad \frac{1}{2} \leq \min \left( \frac{1}{\varphi_i / s^2, 1/\bar{\varphi}_i} \right) \leq 1
\]

for all \( i \in [n] \). We can similarly bound the product

\[
r_i \bar{r}_i = \frac{\varphi_i \bar{\varphi}_i}{(s^2 + \varphi_i \bar{\varphi}_i)^2} \leq \frac{1}{s^2} \min \left( \frac{\varphi_i \bar{\varphi}_i}{s^2, \varphi_i \bar{\varphi}_i} \right) \leq 1/s^2.
\]

Hence, if for some \( i \in [n] \) one of \( r_i, \bar{r}_i \) is large, the other is small. Finally, we note the trivial upper bounds

\[
r_i, \bar{r}_i \leq 1/t \quad \forall i \in [n].
\]

We now prove Proposition 2.5.

**Proof of Proposition 2.5.** Assume towards a contradiction that \( \frac{1}{n} \sum_{j=1}^{n} r_j > 1/\sigma_{\min} \). Then there exists \( i \in [n] \) such that \( r_i > 1/\sigma_{\min} \). From (8.4) it follows that

\[
1/\sigma_{\min} < 1/\varphi_i = 1/(t + (V \bar{\tau})_i)
\]

and so

\[
\sigma_{\min} > (V \bar{\tau})_i = \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij} \bar{r}_j \geq \frac{\sigma_{\min}^2}{n} \sum_{j=1}^{n} \bar{r}_j.
\]

Rearranging, we find \( \frac{1}{n} \sum_{j=1}^{n} \bar{r}_j < 1/\sigma_{\min} \). From (8.3) it follows that \( \frac{1}{n} \sum_{j=1}^{n} r_j < 1/\sigma_{\min} \), a contradiction. \( \square \)

We now prove Proposition 2.6.
Proof of Proposition 2.6. If \( V = \frac{1}{n} A \odot A \) is symmetric, then the \( 2n \) regularized master equations merge into \( n \) equations since \( \mathbf{r} = \mathbf{\tilde{r}} \). In fact since \( V^T = V \) then if \( \mathbf{\tilde{r}} = (\mathbf{r}^T \mathbf{r}^T)^T \) is a solution of the Regularized Master Equations, so is \( \mathbf{\tilde{r}} = (\mathbf{r}^T \mathbf{r}^T)^T \). By uniqueness, \( \mathbf{r} = \mathbf{\tilde{r}} \). Hence the Regularized Master Equations write

\[
\mathbf{r}_i = \frac{(V\mathbf{r})_i + t}{s^2 + (V\mathbf{r})^2_i}, \quad i \in [n].
\]

An elementary analysis of the function \( f(x) = \frac{x}{s^2 + x^2} \) yields \( \sup_{x \in [0,\infty)} f(x) \leq (2s)^{-1} \). Hence

\[
\frac{1}{n} \sum_{i \in [n]} r_i = \frac{1}{n} \sum_{i \in [n]} f((V\mathbf{r})_i + t) \leq \frac{1}{2s} \quad \text{for all } t > 0.
\]

\[ \square \]

Our main objective now is to establish the following, which immediately yields Theorem 2.8.

Proposition 8.1. Assume \( \sigma_{ij} \leq \sigma_{\text{max}} \) for all \( i, j \in [n] \) and some \( \sigma_{\text{max}} < \infty \), and that \( A(\sigma_0) \) is \((\delta, \kappa)\)-robustly irreducible for some \( \sigma_0, \delta, \kappa \in (0, 1) \) (see Definition 2.7). For every fixed \( z \in \mathbb{C} \setminus \{0\} \) there exists a constant \( K = K(z, \sigma_0, \sigma_{\text{max}}, \delta, \kappa) < \infty \) such that

\[
\frac{1}{n} \sum_{i=1}^n r_i \leq K.
\]

In the proof of Proposition 2.5 we were able to pass from a lower bound on a single component \( r_i \) to an upper bound on the average \( \frac{1}{n} \sum_{j=1}^n \mathbf{\tilde{r}}_j \) in one line. This will not be possible when we allow some of the variances \( \sigma_{ij}^2 \) to be zero. The basic outline of our arguments is as follows:

1. Assume towards a contradiction that \( \frac{1}{n} \sum_{i=1}^n r_i \) is large. By the pigeonhole principle there exists \( i_0 \in [n] \) such that \( r_{i_0} \) is large.
2. Use the estimates (8.4)–(8.6) together with assumptions on the connectivity properties of the associated directed graph to iteratively “grow” the set of indices \( i \) for which we know \( r_i \) is large.
3. Once we have shown \( r_i \) is large for (almost) all \( i \in [n] \), by (8.5) it follows that \( \mathbf{\tilde{r}}_i \) is small for (almost) all \( i \in [n] \). We then apply the trace constraint (8.3) to derive a contradiction.

We emphasize that the key idea for our proofs to bound \( \frac{1}{n} \sum_i r_i \) will be to play (8.5) against the trace constraint (8.3).

Recall the graph theoretic notation from Section 2.1. In this section we abbreviate

\[
\mathcal{N}_+(i) := \mathcal{N}_{A(\sigma_0)}(i), \quad \mathcal{N}_-(i) := \mathcal{N}_{A(\sigma_0)^T}(i)
\]

for the in- and out-neighborhoods of a vertex \( i \) in the graph \( \Gamma = \Gamma(A(\sigma_0)) \), and similarly define \( \mathcal{N}_-^{(\delta)}(i) \). For parameters \( \alpha, \beta > 0 \) we define the sets

\[
S_\alpha = \{ i \in [n] : r_i \geq \alpha \| \mathbf{\tilde{\varphi}} \|_\infty \}, \quad T_\beta = \{ i \in [n] : \tilde{r}_i < \beta / \| \mathbf{\tilde{\varphi}} \|_\infty \}
\]

where here and in the sequel we write \( \| \mathbf{\tilde{\varphi}} \|_\infty = \max_{i \in [n]} \tilde{\varphi}_i \) and similarly \( \| \varphi \|_\infty = \max_{i \in [n]} \varphi_i \). (Note that by (8.2) we have \( \| \mathbf{\tilde{\varphi}} \|_\infty \geq t > 0 \).)

8.2. Qualitative boundedness. In this subsection we establish Proposition 2.9 and Lemma 5.3, which give an \( n \)-dependent bound on the components of \( \mathbf{r}, \mathbf{\tilde{r}} \) assuming only that the standard deviation profile is irreducible. This was used in Section 5 with a compactness argument to establish existence of solutions to the Master Equations. While Lemma 5.3 follows from Proposition 8.1 under the robust irreducibility assumption \( \text{A5} \) (which we assume for our main result), we prove this lemma separately for two reasons:
There are positive constants $\beta$ and $K_0$ such that the following holds. Let $s \geq s_0$, and suppose $A$ is irreducible with $\sigma_{ij} \in \{0\} \cup [\sigma_0, \sigma_{\max}]$ for all $1 \leq i, j \leq n$. Then $\frac{1}{n} \sum_{i=1}^{n} r_i \leq K_0$.

Proposition 8.2 and Lemma 5.3 are an immediate consequence of the following:

**Proposition 8.2.** For any fixed $0 < s_0 \leq \sigma_{\max}$ and $s_0 > 0$ there is a constant $K_0(s, s_0, \sigma_0, \sigma_{\max})$ such that the following holds. Let $s \geq s_0$, and suppose $A$ is irreducible with $\sigma_{ij} \in \{0\} \cup [\sigma_0, \sigma_{\max}]$ for all $1 \leq i, j \leq n$. Then $\frac{1}{n} \sum_{i=1}^{n} r_i \leq K_0$.

We now begin the proof of Proposition 8.2. First we dispose of the case that $s > \sigma_{\max}$:

**Claim 8.3.** Suppose $s > \sigma_{\max}$. Then

$$r_i, \bar{r}_i \leq \min \left( \frac{t}{s^2 - \sigma_{\max}^2}, \frac{1}{t} \right) \leq (s^2 - \sigma_{\max}^2)^{-1/2}. \quad (8.9)$$

**Proof.** Suppose $r_i = \max_{i \in [n]} r_i$. From (8.4) we have

$$r_i \leq \frac{1}{s^2} \bar{\varphi}_i = \frac{1}{s^2} (t + (V^T r)_i) \leq \frac{1}{s^2} (t + \sigma_{\max}^2 r_i) .$$

Rearranging we obtain $r_i \leq t/(s^2 - \sigma_{\max}^2)$, which combines with (8.6) to give the desired uniform bound for $r_i, i \in [n]$. The same bound is obtained for $\bar{r}_i$ by similar lines. \qed

Without loss of generality we take $\sigma_{\max} = 1$. By Claim 8.3 we may assume $0 < s_0 \leq s \leq 1$. We may also assume $t \leq 1$. Indeed, otherwise it follows from (8.6) that $\frac{1}{n} \sum_{i=1}^{n} r_i < 1$ and we are done. Let $K > 0$ to be chosen later depending on $n, \sigma_0$ and $s_0$, but independent of $t$, and assume

$$\frac{1}{n} \sum_{i=1}^{n} r_i \geq K. \quad (8.10)$$

We will derive a contradiction for $K$ sufficiently large.

In the following lemma we use the irreducibility of $A_n$ to show that if $T_\beta$ is non-empty for some $\beta$ sufficiently small, then $T_{\beta'} = [n]$ for a somewhat larger value of $\beta'$. This will allow us to assume a uniform lower bound on the components $\bar{r}_i$.

**Lemma 8.4.** There are positive constants $C_0(\sigma_0, n), \beta_0(\sigma_0, n)$ such that for all $\beta \leq \beta_0$, if $T_{\beta}$ is non-empty then $T_{C_0 \beta} = [n]$.

**Proof.** Let $\beta > 0$ to be taken sufficiently small depending on $\sigma_0, n$, and suppose $T_{\beta}$ is non-empty. Then there exists $i \in [n]$ such that $\bar{r}_i < \beta/\|\bar{\varphi}\|_\infty$. From (8.4) it follows that

$$\frac{1}{2} \min(\varphi_i/s^2, 1/\bar{\varphi}_i) < \beta/\|\bar{\varphi}\|_\infty .$$

Assuming $\beta \leq 1/2$ it follows that

$$2s^2 \beta/\|\bar{\varphi}\|_\infty > \varphi_i \geq (V \bar{r})_i \geq \sigma_0^2 \frac{1}{n} \sum_{j \in N_+(i)} \bar{r}_j$$

and hence

$$\bar{r}_j < \frac{2s^2 n}{\sigma_0^2} \frac{\beta}{\|\bar{\varphi}\|_\infty} \quad \forall j \in N_+(i). \quad (8.11)$$

Again from (8.4), if we further assume $\beta \leq \sigma_0^2/4s^2n$ then it follows that

$$\frac{4s^4 n \beta}{\sigma_0^2 \|\bar{\varphi}\|_\infty} > \varphi_j \quad \forall j \in N_+(i). \quad (8.12)$$
Now let $k \in [n]$ be arbitrary. By the irreducibility of $A_n$ there exists a directed path in the associated digraph $\Gamma_n$ from vertex $k$ to vertex $i$ of length at most $n$. Applying the above lines iteratively along each edge of the path we find

$$\bar{r}_k \leq \left( \frac{2s^2 n}{\sigma_0^2} \right)^n \frac{\beta}{\|\varphi\|_\infty} \quad (8.13)$$

if we take $\beta \leq \frac{1}{2} \left( \frac{\sigma_0^2}{2s^2 n} \right)^{n-1}$. Since $k$ was arbitrary, the result follows by setting $C_0 = (2n/\sigma_0^2)^n$ and $\beta_0 = \frac{1}{2} (\sigma_0^2/2n)^{n-1}$ (here we have used our assumption $s \leq 1$).

If $T_{C_0 \beta} = [n]$ for some $\beta \leq \beta_0$ then by the trace identity (8.3),

$$\frac{1}{n} \sum_{i=1}^n r_i = \frac{1}{n} \sum_{i=1}^n \bar{r}_i \leq C_0 \beta / \|\varphi\|_\infty. \quad (8.14)$$

On the other hand, from (8.4) we have

$$r_j \leq \bar{r}_j / s^2 \leq \|\varphi\|_\infty / s^2 \quad (8.15)$$

for all $j \in [n]$. In particular,

$$\frac{1}{n} \sum_{j=1}^n r_j \leq \|\varphi\|_\infty / s^2 \quad (8.16)$$

Hence,

$$\frac{1}{n} \sum_{i=1}^n r_i \leq \min \left( \frac{C_0 \beta / \|\varphi\|_\infty}{s^2}, \frac{\|\varphi\|_\infty}{s^2} \right) \leq (C_0 \beta / s^2)^{1/2} \leq (C_0 \beta_0 / s_0^2)^{1/2} \quad (8.17)$$

which contradicts (8.10) if $K$ is sufficiently large. Hence we may assume $T_{\beta_0}$ is empty for $\beta_0(\sigma_0, n)$ as in Lemma 8.4. Thus,

$$\bar{r}_i \geq \beta_0 / \|\varphi\|_\infty \quad \forall i \in [n]. \quad (8.18)$$

Now we find a value of $\alpha$ for which $S_\alpha$ is already of linear size:

**Lemma 8.5.** Assume $K \geq 2 / s^2$. Then $|S_{1/4}| \geq (\sigma^2 / 4)n$.

**Proof.** From our assumption and (8.16),

$$2 / s^2 \leq K \leq \frac{1}{n} \sum_{j=1}^n r_j \leq \|\varphi\|_\infty / s^2 \quad (8.19)$$

so $\|\varphi\|_\infty \geq 2$. Let $i \in [n]$ such that $\bar{r}_i = \|\varphi\|_\infty$. We have

$$\|\varphi\|_\infty = \bar{r}_i = t + \frac{1}{n} \sum_{j=1}^n \sigma_j r_j \leq t + \frac{1}{n} \sum_{j=1}^n r_j.$$ 

Since $\|\varphi\|_\infty \geq 2$ and $t \leq 1$, $\frac{1}{n} \sum_{j=1}^n r_j \geq \frac{1}{2} \|\varphi\|_\infty$. Now again by (8.15),

$$\|\varphi\|_\infty n / 2 \leq \sum_{j \in S_{1/4}} r_j + \sum_{j \in S_{1/4}'} r_j \leq (\|\varphi\|_\infty / s^2) |S_{1/4}| + \|\varphi\|_\infty n / 4$$

and the result follows by rearranging. \(\square\)

Next we seek to show that we can enlarge $S_\alpha$ by lowering $\alpha$. By irreducibility we can find a vertex $i^* \in S_\alpha$ that is connected to $S_\infty^c$. We can use this to show that the average of the components $r_k$ over $S_\alpha^c$ is bounded below by $c r_{i^*}$ for some small $c > 0$ depending on $\alpha, n, s, \sigma_0$. From the pigeonhole principle we obtain $k \in S_\alpha^c$ with $r_k \geq c r_{i^*} \geq c \alpha \|\varphi\|_\infty$. Taking $\alpha' = c \alpha$, we will then have shown $|S_{\alpha'}| \geq |S_\alpha| + 1$. 
We begin by relating the values of $r, \tilde{r}$ on a fixed set of vertices $S \subset [n]$ to the values taken on $S^c$. For an $n \times n$ matrix $M$ and $S, T \subset [n]$ nonempty we write $M_{S \times T}$ for the $|S| \times |T|$ submatrix of $M$ with entries indexed by $S \times T$. The following lemma will also be used in the proof of Proposition 8.1.

**Lemma 8.6.** Fix a nonempty set $S \subset [n]$, and recall the diagonal matrix $\Psi$ from (5.1). The $|S| \times |S|$ matrix $\Psi_{S \times S}^{-1} - V_{S \times S}^T$ is invertible, and we denote its inverse

$$W^S = (\Psi_{S \times S}^{-1} - V_{S \times S}^T)^{-1}. \tag{8.20}$$

In terms of $W^S$ the restrictions of $r, \tilde{r}$ to $S$ and $S^c$ satisfy

$$r_S = W^S(t + V_{S \times S}^T r_{S^c}), \quad \tilde{r}_S = (W^S)^T(t + V_{S \times S^c} \tilde{r}_{S^c}). \tag{8.21}$$

Furthermore, the entries of $W^S$ satisfy the following bounds. For all $i, j \in S$,

$$W^S_{ij} \geq 0. \tag{8.22}$$

For all $j \in S$,

$$\sum_{i \in S} W^S_{ij} \leq \tilde{r}_j/t \tag{8.23}$$

and if (8.18) holds for some $\beta_0 > 0$,

$$\sum_{i \in S} W^S_{ij} |\mathcal{N}_+(i) \cap S^c| \leq \left(\frac{n \|\tilde{\varphi}\|_\infty}{\beta_0 \sigma_0^2}\right) \tilde{r}_j. \tag{8.24}$$

**Proof.** Arguing as in the proof of Lemma 5.4(2) (using Proposition 5.2) we find the spectral radius of $\Psi_{S \times S}V_{S \times S}^T$ is strictly less than 1. Hence, $(I - \Psi_{S \times S}V_{S \times S}^T)^{-1}$ has a convergent Neumann series, and it follows that

$$W^S = \Psi_{S \times S}(I - \Psi_{S \times S}V_{S \times S}^T)^{-1} = \Psi_{S \times S} \sum_{k=0}^\infty (\Psi_{S \times S}V_{S \times S}^T)^k.$$

is well-defined. Furthermore, since all of the matrices in the above series have non-negative entries, (8.22) follows.

(8.21) is quickly obtained by rearranging the equations (2.8).

Now for (8.23) and (8.24), let $j \in S$ be arbitrary. From the second equation in (8.21),

$$\tilde{r}_j = \sum_{i \in S} W^S_{ij} \left(t + \frac{1}{n} \sum_{k \in S^c} \sigma_{ik}^2 \tilde{r}_k \right).$$

In particular,

$$\tilde{r}_j \geq t \sum_{i \in S} W^S_{ij},$$

giving (8.23), and

$$\tilde{r}_j \geq \frac{1}{n} \sum_{i \in S} W^S_{ij} \sum_{k \in S^c} \sigma_{ik}^2 \tilde{r}_k \geq \frac{\beta_0 \sigma_0^2}{n \|\tilde{\varphi}\|_\infty} \sum_{i \in S} W^S_{ij} |\mathcal{N}_+(i) \cap S^c|$$

which rearranges to give (8.24). \qed

We can use Lemma 8.6 and the irreducibility of $A$ to establish the following:

**Lemma 8.7** (Incrementing $\alpha$). Let $\alpha > 0$ such that $1 \leq |S_\alpha| \leq n - 1$. If $K$ is sufficiently large depending on $s_0, \sigma_0, n$ and $\alpha$ then there exists $\alpha' = \alpha'(\alpha, s_0, \sigma_0, n) \in (0, \alpha)$ such that $|S_{\alpha'}| \geq |S_\alpha| + 1$. 
Let us conclude the proof of Proposition 8.2 on the above lemma. Putting $\alpha_0 = 1/4$, by Lemma 8.5 we have $S_{\alpha_0} \neq \emptyset$. Taking $K$ sufficiently large depending on $s_0, \sigma_0$ and $n$ we can iterate Lemma 8.7 at most $n$ times to find $\alpha = \alpha(s_0, \sigma_0, n) > 0$ such that $S_\alpha = [n]$. Then by (8.3) and (8.4),

$$\alpha\|\tilde{\varphi}\|_\infty \leq \frac{1}{n} \sum_{j=1}^n r_j = \frac{1}{n} \sum_{j=1}^n \tilde{r}_j \leq \frac{1}{s^2 \alpha\|\varphi\|_\infty}$$

(8.25)

so we have $\|\varphi\|_\infty \leq 1/s \alpha$. Then again by (8.4) we have

$$\frac{1}{n} \sum_{j=1}^n r_j \leq \|\varphi\|_\infty \leq \frac{1}{s^2 \alpha} \leq \frac{1}{s_0^3 \alpha}$$

(8.26)

and we are done.

**Proof of Lemma 8.7.** We write $W = W^{S_\alpha}$. From the first equation in (8.21), for any $i \in S_\alpha$ we have

$$\alpha\|\varphi\|_\infty \leq r_i = \sum_{j \in S_\alpha} W_{ij} \left( t + \sum_{k \in S_\alpha} \sigma_{kj}^2 r_k \right)$$

(8.27)

Suppose first that

$$\sum_{j \in S_\alpha} W_{ij} > \frac{\alpha\|\varphi\|_\infty}{2t}$$

(8.28)

for some $i \in S_\alpha$. Then from (8.23),

$$\frac{\alpha\|\varphi\|_\infty}{2t} < \sum_{j \in S_\alpha} W_{ij} \leq \frac{1}{t} \sum_{j \in S_\alpha} \tilde{r}_j \leq \frac{1}{s^2 t} \sum_{j \in S_\alpha} \frac{1}{r_j} \leq \frac{|S_\alpha|}{t s^2 \alpha\|\varphi\|_\infty}$$

where in the third inequality we applied (8.4). Rearranging we have $\|\varphi\|_\infty \leq \sqrt{2|S|}/s \alpha \leq \sqrt{2n}/s_0 \alpha$ in this case. On the other hand, from (8.16) we have $K \leq \|\varphi\|_\infty^2/s_0^3$, and we obtain a contradiction if $K$ is sufficiently large depending on $s_0, \sigma_0, n$ and $\alpha$.

Suppose now that (8.28) does not hold for any $i \in S_\alpha$. Then rearranging (8.27) we have

$$\frac{\alpha\|\varphi\|_\infty}{2} \leq \sum_{j \in S_\alpha} \sum_{k \in S_\alpha} \sigma_{kj}^2 r_k W_{ij}$$

(8.29)

for any $i \in S_\alpha$. From the assumption that $A_{S_\alpha}$ is irreducible there exists $(i^*, j^*) \in S_\alpha \times S_\alpha^c$ such that $\sigma_{i^*j^*} \geq \sigma_0, \text{i.e. } |\mathcal{V}_+(i^*) \cap S_\alpha^c| \geq 1$. From (8.24) it follows that

$$W_{i^*j^*} \leq \left( \frac{n\|\varphi\|_\infty}{\beta_0 \sigma_0^2} \right) \tilde{r}_j$$

(8.30)

for all $j \in S_\alpha$. Inserting this bound in (8.29) we have

$$\frac{\alpha\|\varphi\|_\infty}{2} \leq \frac{n\|\varphi\|_\infty}{\beta_0 \sigma_0^2} \sum_{j \in S_\alpha} \sum_{k \in S_\alpha^c} \sigma_{kj}^2 r_k \tilde{r}_j \leq \frac{n|S_\alpha|}{\beta_0 \sigma_0^2 s^2 \alpha} \sum_{k \in S_\alpha^c} r_k$$

where in the second inequality we applied the bounds $\sigma_{ij} \leq 1$ for all $i, j \in [n]$ and $\tilde{r}_j \leq (s^2 \alpha\|\varphi\|_\infty)^{-1}$ for all $j \in S_\alpha$ (by (8.5)). Rearranging we have

$$\sum_{k \in S_\alpha^c} r_k \geq \frac{\alpha^2 s^2 \beta_0 \sigma_0^2}{2n|S_\alpha|} \|\varphi\|_\infty.$$

(8.31)

By the pigeonhole principle there exists $k \in S_\alpha^c$ such that

$$r_k \geq \frac{\alpha^2 s^2 \beta_0 \sigma_0^2}{2n|S_\alpha| |S_\alpha^c|} \|\varphi\|_\infty \geq \frac{\alpha^2 s^2 \beta_0 \sigma_0^2}{2n^3} \|\varphi\|_\infty.$$

(8.32)
Setting \( \alpha' = \alpha^2 s_0^2 \beta_0 \sigma_0^2 / 2n^3 \) we have \( k \in S_{\alpha'} \). Also, since \( \alpha' < \alpha \) we have \( S_{\alpha'} \subset S_{\alpha} \). Hence, \( |S_{\alpha'}| \geq |S_{\alpha}| + 1 \) as desired. \( \square \)

8.3. Quantitative boundedness. Now we prove Proposition 8.1. By rescaling the variance profile \( V \) we may take \( \sigma_{\max} = 1 \). By Claim 8.3 we may assume \( s \in (0, 1] \). As in the proof of Proposition 8.2 we may assume \( t \leq 1 \). We may also assume \( n \) is sufficiently large depending on \( s, \sigma_0, \delta \) and \( \kappa \).

In the remainder of the section we make use of asymptotic notation \( O(), \lesssim, \gtrsim \), allowing implied constants to depend on the parameters \( s, \sigma_0, \delta \) and \( \kappa \) (but not on \( n \) and \( t \)).

As in the proof of Proposition 8.2 we assume

\[
\frac{1}{n} \sum_{i=1}^{n} r_i \geq K
\] (8.33)

for some \( K > 0 \) and aim to derive a contradiction for \( K \) sufficiently large depending on \( s, \sigma_0, \delta \) and \( \kappa \). The argument follows the same general outline as the proof in the previous subsection. We will reuse Lemmas 8.5 and 8.6 as stated, but we will need versions of Lemmas 8.4 and 8.7 with constants independent of \( n \).

8.3.1. Lower bounding \( \tilde{r}_i \). The following is an analogue of Lemma 8.4.

Lemma 8.8. There are positive constants \( C_0(s, \sigma_0, \delta, \kappa) \), \( \beta_0(s, \sigma_0, \delta, \kappa) \) such that for all \( \beta \leq \beta_0 \), if \( T_\beta \) is non-empty then \( T_{C_0 \beta} = [n] \).

Proof. Let \( \beta > 0 \) to be taken sufficiently small and assume \( T_\beta \) is non-empty. Fix an element \( i_0 \in T_\beta \). We will grow the set \( T_\beta \) in stages by enlarging \( \beta \) by appropriate constant factors. We do this by iterative application of the following:

Claim 8.9. Let \( \beta, \varepsilon_0 \in (0, 1/2] \), and assume \( 0 < |T_\beta| \leq (1 - \varepsilon_0)n \). There exists \( C = C(\sigma_0, \delta, \varepsilon_0) > 0 \) such that if \( n \) is sufficiently large depending on \( \kappa \) and \( \varepsilon_0 \) then \( |T_{C \beta} \setminus T_\beta| \geq (\delta \varepsilon_0 / 2)n \).

Proof. By the assumption that \( A(\sigma_0) \) is \( (\delta, \kappa) \)-robustly irreducible we have

\[
|\mathcal{N}(\delta)(T_\beta)| \cap T_\beta| \geq \min(\kappa |T_\beta|, |T_\beta|) \geq \min(\kappa \varepsilon_0 n, 1) \geq 1
\]

if \( n \) is sufficiently large. Fix an element \( i \in \mathcal{N}(\delta)(T_\beta) \cap T_\beta \). By definition we have

\[
|\mathcal{N}(i) \cap T_\beta| \geq \delta |T_\beta| \geq \delta \varepsilon_0 n.
\]

Next, we claim that for any \( C > 0 \) we have

\[
|\mathcal{N}(i) \cap T_{C \beta}| \leq \frac{2s^2}{C \sigma_0^2} n. \quad (8.34)
\]

Indeed, since \( i \in T_\beta \), by (8.8) and (8.4) we have

\[
\frac{\beta}{\| \hat{\varphi} \|_\infty} > \tilde{r}_i \geq \frac{1}{2} \min \left( \frac{\varphi_i}{s^2}, \frac{1}{\varphi_i} \right).
\]

Since \( \beta \leq 1/2 \) it follows that the minimum is attained by the first argument. Thus

\[
\frac{\beta}{\| \hat{\varphi} \|_\infty} > \frac{\varphi_i}{2s^2} > \frac{1}{2s^2} \frac{1}{n} \sum_{j=1}^{n} \sigma_{ij} \tilde{r}_j \geq \frac{\sigma_0^2}{2s^2} \frac{1}{n} \sum_{j \in \mathcal{N}(i)} \tilde{r}_j.
\]
From Markov’s inequality it follows that for any $C > 0$, $\bar{r}_j < C\beta/\|\varphi\|_\infty$ for all but at most $(2s^2/C\sigma_0^2)n$ values of $j \in \mathcal{N}_+(i)$, which gives (8.34). Combining these estimates and taking $C = 8s^2/\sigma_0^2\delta \varepsilon_0$ we have

$$|T_{C\beta} \setminus T_{\beta}| \geq |\mathcal{N}_+(i) \cap T_{\beta}| - |\mathcal{N}_+(i) \cap T_{C\beta}| \geq \left(\delta \varepsilon_0 - \frac{2s^2}{C\sigma_0^2}\right)n \geq (\delta \varepsilon_0/2)n$$

as desired. \hfill \Box

Applying the above claim iteratively with $\varepsilon_0 = s^2/8$ we obtain $C'(\sigma_0, \delta, s) < \infty$ such that if $\beta$ is sufficiently small depending on $\sigma_0, \delta, s$ and $n$ is sufficiently large depending on $\kappa, s$, then

$$|T_{C'\beta}| \geq (1 - s^2/8)n.$$  \hfill (8.35)

Now let $C_0 > 0$ to be chosen later, and towards a contradiction suppose $T_{C_0\beta} \neq [n]$. Then there exists $i \in [n]$ such that $\bar{r}_i \geq C_0\beta/\|\varphi\|_\infty$. From (8.4) we have the upper bound $\bar{r}_i \leq \|\varphi\|_\infty/s^2$, so we conclude $\|\varphi\|_\infty \geq C_0s^2\beta/\|\varphi\|_\infty$. Now from our assumption $K \leq \frac{1}{4}\sum_{j=1}^n r_j = \sum_{j=1}^n \bar{r}_j$, if $K$ is sufficiently large depending on $s$ then the same argument as in the proof of Lemma 8.5 shows that $\bar{r}_j \geq \|\varphi\|_\infty/4$ for at least $(s^2/4)n$ values of $j \in [n]$. Thus,

$$\bar{r}_j \geq \frac{C_0s^2\beta}{4\|\varphi\|_\infty}$$  \hfill (8.36)

for at least $(s^2/4)n$ values of $j \in [n]$, i.e. $|T_{C_0s^2\beta/4}| < (1 - s^2/4)n$. Taking $C_0 = 4C'/s^2$ we contradict (8.35), and we conclude $T_{C_0\beta} = [n]$. \hfill \Box

Now by the same lines as in (8.14)–(8.18) we conclude

$$\bar{r}_i \geq \beta_0/\|\varphi\|_\infty \quad \forall i \in [n]$$  \hfill (8.37)

for some $\beta_0(s, \sigma_0, \delta, \kappa) > 0$. Note that we are now free to use the estimates in Lemma 8.6 with this value of $\beta_0$.

8.3.2. Upper bounding $r_i$. Here our task is essentially to modify the proof of Lemma 8.7 to show we can take $\alpha'$ sufficiently small and independent of $n$ such that $|S_{\alpha'} \setminus S_\alpha| \gtrsim n$, rather than merely nonempty. We can then conclude the proof by iterating this fact a bounded number of times.

Let us summarize the key new ideas. In the proof of Lemma 8.7 we used the irreducibility assumption to find an element $i^* \in S_\alpha$ such that the average of the components $r_k$ over $S_c^\alpha$ was of order $\sum_{i^*} r_i^* \geq \alpha \|\varphi\|_\infty$ (see (8.31)). In a similar spirit, Lemma 8.10 below controls the average of $r_k$ over $k \in S_c^\alpha$ from below by the average of $r_i$ over $i \in U_0$, for a set $U_0 \subset S_\alpha$ that is densely connected to $S_c^\alpha$. By averaging over a large set $U_0$ we are able to use the full strength of the bounds in Lemma 8.6 and avoid any dependence of the constants on $n$.

Proceeding naïvely, one can then use Lemma 8.10 to deduce

$$|S_{c\alpha^2} \setminus S_\alpha| \geq c_0\alpha |S_c^\alpha|$$

for a sufficiently small constant $c_0 = c_0(s, \sigma_0, \delta, \kappa) > 0$. However, when iterating this bound over a sequence of values $\alpha_{k+1} = c_0\alpha_k^2$, the sets $S_{\alpha_k}$ grow by an exponentially decreasing proportion of $n$, so this is not enough to find a value of $\alpha$ for which $|S_\alpha|$ is close to $n$.

Instead, in Lemma 8.11 we are able to grow $S_\alpha$ by a constant factor using a nested iteration argument, which we now describe. We would like to find some value of $\alpha' \in (0, \alpha)$ for which

$$|S_{\alpha'} \setminus S_\alpha| \geq c|S_{\alpha'}|$$  \hfill (8.38)

where $c > 0$ is small constant. Suppose that (8.38) fails. By the expansion assumption, we know that $S_{\alpha'}$ contains a fairly large set $U = \mathcal{N}_-^{(\delta)}(S_{\alpha'}^c) \cap S_{\alpha'}$ (of size at least min$(|S_{\alpha'}|, \kappa|S_{\alpha'}^c|)$) that is


densely connected to $S^c_\alpha$. In particular, if $c$ is sufficiently small depending on $\kappa$, then $U$ must have large overlap with $S_\alpha$. Denoting the overlap by $U_0$, Lemma 8.10 can now be applied to deduce

$$|S_{c_0\alpha'c'0}| \geq c_0\alpha|S^c_\alpha|$$

for some $c_0 = c_0(s, \sigma_0, \delta, \kappa) > 0$ sufficiently small. The key is that the constant of proportionality on the right hand side is independent of $\alpha'$. Hence, for fixed $\alpha$, as long as (8.38) fails we can iteratively lower $\alpha'$ to increase $|S_\alpha' \setminus S_\alpha|$ by an amount $\gtrsim \alpha|S^c_\alpha|$, until eventually (8.38) holds. This whole procedure can then be iterated a bounded number of times to obtain $\alpha''$ such $|S_{\alpha''c'}|$ is close to $n$.

Having motivated the key ideas, we turn now to the proofs.

**Lemma 8.10.** Let $\alpha \in (0, 1)$ and suppose that $0 < |S_\alpha| \leq (1 - \delta/2)n$. If $K$ is sufficiently large depending on $\alpha, s, \sigma_0, \delta, \kappa$, then for any $U_0 \subset \mathcal{N}^{(\delta)}(S^c_\alpha) \cap S_\alpha$ with $|U_0| \geq \frac{1}{10}|\mathcal{N}^{(\delta)}(S^c_\alpha) \cap S_\alpha|$ we have

$$\frac{1}{|S_\alpha^c|} \sum_{k \in S_\alpha^c} r_k \gtrsim \frac{\alpha}{|S_\alpha^c|} \sum_{i \in U_0} r_i.$$  \hspace{1cm} (8.39)

**Proof.** First we prove the comparison

$$\sum_{i \in U_0} r_i \geq 2 \sum_{j \in S_\alpha} \tilde{r}_j$$  \hspace{1cm} (8.40)

assuming $K$ is sufficiently large depending on $\alpha, s, \sigma_0, \delta, \kappa$. Indeed, if (8.40) does not hold, then by the fact that $U_0 \subset S_\alpha$ and (8.5),

$$\alpha ||\tilde{\varphi}||_\infty |U_0| \leq \sum_{i \in U_0} r_i < 2 \sum_{j \in S_\alpha} \tilde{r}_j \leq \frac{2|S_\alpha|}{s^2\alpha ||\tilde{\varphi}||_\infty}.$$

Rearranging we have

$$||\tilde{\varphi}||_\infty \leq \frac{1}{\alpha} \left( \frac{2|S_\alpha|}{s^2|U_0|} \right)^{1/2} \leq \frac{1}{\alpha} \left( \frac{|S_\alpha|}{\min(|S_\alpha|, |S^c_\alpha|)} \right)^{1/2} \lesssim 1/\alpha,$$

where in the second bound we applied the robust irreducibility assumption and our assumed bounds on $U_0$ and $S_\alpha$, and in the bound we used that both $S_\alpha$ and its complement are of linear size in $n$. From our assumption (8.33), (8.4) and the above it follows that

$$K \leq \frac{1}{n} \sum_{i=1}^n r_i \lesssim ||\tilde{\varphi}||_\infty/s^2 \lesssim 1/\alpha.$$  \hspace{1cm} (8.41)

Taking $K$ sufficiently large depending on $\alpha, s, \sigma_0, \delta$ and $\kappa$, we may assume (8.40) holds.

From (8.21) and Lemma 8.6 we have

$$\sum_{i \in U_0} r_i = t \left( \sum_{j \in S_\alpha} \sum_{i \in U_0} W_{ij}^{S_\alpha} \right) + \frac{1}{n} \sum_{j \in S_\alpha} \sum_{k \in S^c_\alpha} \left( \sum_{i \in U_0} W_{ij}^{S_\alpha} \right) \sigma_{kji} r_k \geq \sum_{j \in S_\alpha} \tilde{r}_j + \frac{1}{n} \sum_{j \in S_\alpha} \sum_{k \in S^c_\alpha} \left( \sum_{i \in U_0} W_{ij}^{S_\alpha} \right) \sigma_{kji} r_k$$

where in the second line we applied (8.22) and (8.23). Applying (8.40) and rearranging yields

$$\sum_{i \in U_0} r_i \leq \frac{2}{n} \sum_{j \in S_\alpha} \sum_{k \in S^c_\alpha} \left( \sum_{i \in U_0} W_{ij}^{S_\alpha} \right) \sigma_{kji} r_k.$$  \hspace{1cm} (8.42)

Now since $U_0 \subset \mathcal{N}^{(\delta)}(S^c_\alpha)$, for any $i \in U_0$ we have $|\mathcal{N}_+(i) \cap S^c_\alpha| \geq \delta |S^c_\alpha|$. Together with (8.22) and (8.24) this implies
By (8.47) and Lemma 8.10, by the robust irreducibility assumption and the fact that (8.44) fails,

\[ \delta |S^c_{\alpha}| \sum_{i \in U_0} W_{ij}^{S_{\alpha}} \leq \sum_{i \in S_{\alpha}} W_{ij}^{S_{\alpha}} |\mathcal{N}_+ (i) \cap S^c_{\alpha}| \leq \left( \frac{n \| \overline{\varphi} \|_\infty}{\beta_0 \sigma_0^2} \right) \tilde{r}_j. \]

Rearranging we obtain a bound on \( \sum_{i \in U_0} W_{ij}^{S_{\alpha}} \), which we substitute in (8.42) to obtain

\[ \sum_{i \in U_0} r_i \leq \frac{2 \| \overline{\varphi} \|_\infty}{\beta_0 \sigma_0^2 \delta |S^c_{\alpha}|} \sum_{j \in S_{\alpha}} \sum_{k \in S_{\alpha}} \sigma_{kj} \tilde{r}_j r_k \leq \frac{2 |S_{\alpha}|}{s^2 \alpha \beta_0 \sigma_0^2 \delta |S^c_{\alpha}|} \sum_{k \in S_{\alpha}} r_k \]

where in the second inequality we applied (8.5) to bound \( \tilde{r}_j \leq 1/s^2 \alpha \| \overline{\varphi} \|_\infty \) for all \( j \in S_{\alpha} \). The result now follows by rearranging.

\[ \square \]

**Lemma 8.11.** For any \( \alpha \in (0, 1) \) there exists \( \alpha' = \alpha'(\alpha, s, \sigma_0, \delta, \kappa) > 0 \) such that either

\[ |S_{\alpha'}| \geq (1 - \delta/2) n \]  

(8.43)

or

\[ |S_{\alpha'} \setminus S_{\alpha}| \geq \frac{1}{2} \min(|S_{\alpha'}|, \kappa |S^c_{\alpha'}|) \]  

(8.44)

(or both).

**Proof.** For \( \alpha' \in (0, \alpha) \) denote by \( P(\alpha') \) the statement that at least one of (8.43) and (8.44) holds. We will show that while \( P(\alpha') \) fails, we can lower \( \alpha' \) by a controlled amount to increase the size of \( S_{\alpha'} \setminus S_{\alpha} \) by a little bit. We can then iterate this until \( P(\alpha') \) holds.

Let \( \alpha' \in (0, \alpha) \) be arbitrary and assume \( P(\alpha') \) fails. We claim there is a constant \( c_0(s, \sigma_0, \delta, \kappa) > 0 \) such that

\[ \frac{1}{|S^c_{\alpha'}|} \sum_{k \in S^c_{\alpha'}} r_k \geq c_0 \alpha' \| \overline{\varphi} \|_\infty. \]  

(8.45)

Put

\[ U_0 = \mathcal{N}^{(\delta)}(S^c_{\alpha'}) \cap S_{\alpha}. \]  

(8.46)

By the robust irreducibility assumption and the fact that (8.44) fails,

\[ |U_0| \geq |\mathcal{N}^{(\delta)}(S^c_{\alpha'}) \cap S_{\alpha'}| - |S_{\alpha'} \setminus S_{\alpha}| \geq \frac{1}{2} |\mathcal{N}^{(\delta)}(S^c_{\alpha'}) \cap S_{\alpha'}| \]

\[ \geq \frac{1}{2} \min(|S_{\alpha'}|, \kappa |S^c_{\alpha'}|). \]  

(8.47)

By (8.47) and Lemma 8.10,

\[ \frac{1}{|S^c_{\alpha'}|} \sum_{k \in S^c_{\alpha'}} r_k \geq \frac{\alpha'}{|S_{\alpha'}|} \sum_{i \in U_0} r_i \geq \alpha' \| \overline{\varphi} \|_\infty \frac{|U_0|}{|S_{\alpha'}|} \geq \alpha' \| \overline{\varphi} \|_\infty \]  

where in the last inequality we applied (8.48) and the fact that (8.43) fails. This gives (8.45) as desired.

Now denoting

\[ U' = \left\{ k \in S^c_{\alpha'} : r_k \geq \frac{1}{2} c_0 \alpha' \| \overline{\varphi} \|_\infty \right\} \]

we have

\[ \sum_{k \in S^c_{\alpha'}} r_k \leq \sum_{k \in U'} r_k + \sum_{k \in S^c_{\alpha'} \setminus U'} r_k \leq \alpha' \| \overline{\varphi} \|_\infty |U'| + \frac{1}{2} c_0 \alpha' \| \overline{\varphi} \|_\infty |S^c_{\alpha'}| \]

where we used that by definition, \( r_k \leq \alpha' \| \overline{\varphi} \|_\infty \) for all \( k \in S^c_{\alpha'} \). Combining with (8.45) and rearranging gives

\[ |S_{\alpha' \alpha'}/2 \setminus S_{\alpha'}| \geq |U'| \geq \frac{1}{2} c_0 \alpha |S^c_{\alpha'}|. \]  

(8.49)
Since (8.49) holds as long as $P(\alpha')$ fails, we can repeatedly lower $\alpha'$ by a factor $c_0\alpha/2$ to obtain $\alpha' = \alpha'(\alpha, s, \sigma_0, \delta, \epsilon)$ such that $P(\alpha')$ holds. More explicitly, for each $k \geq 0$ put $\alpha_k = (c_0\alpha/2)^k\alpha$ and abbreviate $S_k := S_{\alpha_k}$. Then for all $k \geq 1$ such that $P(\alpha_k)$ fails we have

$$|S_{k+1} \setminus S_k| \geq \frac{1}{2} c_0 \alpha |S_k^c|$$

so

$$|S_{k+1} \setminus U_0| = |S_{k+1} \setminus S_k| + \cdots + |S_1 \setminus U_0| \\
\geq \frac{1}{2} c_0 \alpha (|S_k^c| + \cdots + |S_0^c|) \\
\geq (k + 1) \frac{1}{2} c_0 \alpha |S_{k+1}^c|.$$ 

Thus, we must have that $P(\alpha_k)$ holds for some $k \leq 2\kappa/c_0\alpha$. (This gives $\alpha'$ of size $O(1/\alpha)^{O(1/\alpha)}$.)

□

Now we conclude the proof of Proposition 8.1. From Lemma 8.5 we have $|S_{1/4}| \geq (s^2/4)n$. Applying Lemma 8.11 $O(1)$ times we obtain $\alpha'' \gtrsim 1$ such that $|S_{\alpha''}| \geq (1 - \delta/2)n$.

Now from (8.5) we have

$$\tilde{r}_j \leq \frac{1}{s^2\alpha'' \|\tilde{\varphi}\|_{\infty}}$$

for all $j \in S_{\alpha''}$. On the other hand, for any $j \in S_{\alpha''}^c$, we have

$$1/\tilde{r}_j \geq \tilde{\varphi}_j \geq (V^T r)_j \geq \frac{1}{n} \sum_{i \in S_{\alpha''}} \sigma_{ij}^2 r_i \geq \frac{1}{n} \sigma_0^2 \alpha'' \|\tilde{\varphi}\|_{\infty} |N_-(j) \cap S_{\alpha''}|.$$

From (8.50) and the robust irreducibility assumption (specifically the condition (2.16)),

$$|N_-(j) \cap S_{\alpha''}| \geq \delta n - |S_{\alpha''}^c| \geq \delta n/2.$$

Combining the previous two displays we obtain

$$\tilde{r}_j \leq \frac{2}{\delta \sigma_0^2 \alpha'' \|\tilde{\varphi}\|_{\infty}}$$

for all $j \in S_{\alpha''}^c$. Together with (8.51) we have

$$\tilde{r}_j \lesssim \frac{1}{\alpha'' \|\tilde{\varphi}\|_{\infty}}$$

for all $j \in [n]$. Applying (8.3),

$$\alpha'' \|\tilde{\varphi}\|_{\infty} n/2 \leq \alpha'' \|\tilde{\varphi}\|_{\infty} |S_{\alpha''}| \leq \sum_{j=1}^n r_j = \sum_{j=1}^n \tilde{r}_j \lesssim \frac{n}{\alpha'' \|\tilde{\varphi}\|_{\infty}}$$

and rearranging gives $\|\tilde{\varphi}\|_{\infty} \lesssim 1/\alpha'' \lesssim 1$. Finally, since

$$K \leq \frac{1}{n} \sum_{j=1}^n r_j \leq \|\tilde{\varphi}\|_{\infty}/s^2$$

by (8.4), we obtain a contradiction if $K$ is sufficiently large depending on $s, \sigma_0, \delta$ and $\kappa$. It follows that (8.33) fails for sufficiently large $K$, which concludes the proof of Proposition 8.1.

Remark 8.1. We note that in the above proof we only applied the expansion bound (2.17) to sets of size at least $\delta n/10$. 

---

**NON-HERMITIAN RANDOM MATRICES WITH A VARIANCE PROFILE**

---

**Remark 8.1.**
A.1. Proof of Theorem 3.2. Denote by $d_i = d(i/n), \tilde{d}_j = d(j/n)$. The associated vectors $d, \tilde{d}$ meet the conditions of Assumption A6. In order to study the existence of $s \mapsto u_\infty(s)$ and its properties, we introduce the solution $u_n \in [0,1]$ defined in Theorem 3.1-(1). Notice that $\rho(V) = \frac{1}{n} \tilde{d}^T d \to \rho_\infty$ as $n \to \infty$.

We prove parts (1) and (2). We establish the existence of $s \mapsto u_\infty(s)$ by relying on Arzela–Ascoli’s theorem.

Denote by

$$\delta_{\min} = \liminf_{n \geq 1} \frac{1}{n} \sum_{i \in [n]} d_i^2 \tilde{d}_i^2 > 0$$

and $d_{\max} = \sup_{x \in [0,1]} d(x) \vee \tilde{d}(x)$. Let $s, t > 0$ be such that $s, t < \sqrt{\rho_\infty}$. For $n$ large enough, $s, t < \sqrt{\rho_\infty}$ and

$$(t^2 - s^2) \frac{1}{n} \sum_{i \in [n]} d_i \tilde{d}_i$$

by simply subtracting equation in Theorem 3.1-(i) evaluated at $s$ to itself evaluated at $t$. Now

$$\frac{1}{n} \sum_{i \in [n]} \frac{d_i \tilde{d}_i}{(s^2 + d_i \tilde{d}_i u_n(s))(t^2 + d_i \tilde{d}_i u_n(t))} \leq \frac{d_{\max}^2}{s^2 t^2},$$

$$\frac{1}{n} \sum_{i \in [n]} \frac{d_i^2 \tilde{d}_i^2}{(s^2 + d_i \tilde{d}_i u_n(s))(t^2 + d_i \tilde{d}_i u_n(t))} \geq \frac{\delta_{\min}}{(\rho(V) + d_{\max}^2)^2}.$$

Plugging these two estimates into the previous equation yields

$$|u_n(t) - u_n(s)| \leq K \times \frac{t + s}{t^2 s^2} \times |t - s|,$$

where $K$ depends on $d_{\min}, \tilde{d}_{\min}, d_{\max}$ and $\tilde{d}_{\max}$. Notice in particular that $u_n$ being Lipschitz in any interval $[a, b] \subset (0, \sqrt{\rho_\infty})$ is an equicontinuous family. By Arzela–Ascoli’s theorem, the sequence $(u_n)$ is relatively compact for the supremum norm on any interval $[a, b] \subset (0, \sqrt{\rho_\infty})$. Let $u$ be an accumulation point for $s \in [a, b] \subset (0, \sqrt{\rho_\infty})$, then by continuity

$$\int_0^1 \frac{d(x) \tilde{d}(x)}{s^2 + d(x) \tilde{d}(x) u(s)} dx = 1 \quad \text{for} \quad s \in [a, b],$$

hence the existence. If $u$ and $\tilde{u}$ are two accumulation points of $(u_n)$ on $[a, b]$, then

$$(\tilde{u}(s) - u(s)) \int_0^1 \frac{d(x) \tilde{d}(x)}{(s^2 + d(x) \tilde{d}(x) u(s))(s^2 + d(x) \tilde{d}(x) \tilde{u}(s))} dx = 0.$$

By relying on the same estimates as in the discrete case, one proves that the integral on the l.h.s. is positive and hence $u = \tilde{u} := u_\infty$. The uniqueness and the continuity of a solution to (A.2) is established for $s \in (0, \sqrt{\rho_\infty})$. Using similar arguments, one can prove that $u_\infty(s) > 0$ for $s \in (0, \sqrt{\rho_\infty})$, that $s \mapsto u_\infty(s)$ satisfies the Cauchy criterion for functions as $s \downarrow 0$ and $s \uparrow \sqrt{\rho_\infty}$. In particular, $u$ admits a limit as $s \downarrow 0$ and $s \uparrow \sqrt{\rho_\infty}$ and it is not difficult to prove that

$$\lim_{s \downarrow 0} u_\infty(s) = 1 \quad \text{and} \quad \lim_{s \uparrow \sqrt{\rho_\infty}} u_\infty(s) = 0.$$
We now prove (3) and establish that \( s \mapsto u_\infty(s) \) is differentiable on \((0, \sqrt{\rho_\infty})\). By considering the continuous counterpart of equation (A.1), we obtain

\[
\frac{u_\infty(t) - u_\infty(s)}{t - s} = -(t + s) \left( \frac{\int_0^1 \frac{d(x)d(x)}{(s^2 + d(x)d(x)u_\infty(s))(t^2 + d(x)d(x)u_\infty(t))} dx}{\int_0^1 \frac{d^2(x)d^2(x)}{(s^2 + d(x)d(x)u_\infty(s))(t^2 + d(x)d(x)u_\infty(t))} dx} \right).
\]

The r.h.s. of the equation admits a limit as \( t \to s \), hence the existence and expression of \( u_\infty \)'s derivative:

\[
u'_\infty(s) = -2s \left( \int_0^1 \frac{\frac{d(x)d(x)}{(s^2 + d(x)d(x)u_\infty(s))}}{(s^2 + d(x)d(x)u_\infty(s))^2} dx \right) \left( \int_0^1 \frac{d^2(x)d^2(x)}{(s^2 + d(x)d(x)u_\infty(s)^2)} dx \right)^{-1}
\]

for \( s \in (0, \sqrt{\rho_\infty}) \). This limit is continuous in \( s \). The density follows from Remark 2.8:

\[
\varphi_\infty(|z|) = -\frac{1}{2\pi|z|} u'_\infty(|z|).
\]

Item (4) follows from the fact that \( u_\infty(0) = 1 \) and by a continuity argument.

We now establish (5). Notice first that \( F_\infty(s) = 1 - u_\infty(s) \) is the cumulative distribution function of a rotationally invariant probability measure on \( \mathbb{C} \). Since \( u_n \to u_\infty \) for \( s \geq 0 \) (some care is required to prove the convergence for \( s = \sqrt{\rho_\infty} \) but we leave the details to the reader), one has \( \mu_n \xrightarrow{w} \mu_\infty \).

Combining this convergence with Theorem 3.1-(3) yields the desired convergence. The proof of Theorem 3.2 is complete.

A.2. Proof of Theorem 3.3. Extending the maximum norm notation from vectors to functions, we also denote by \( \|f\|_\infty = \sup_{x \in [0,1]} |f(x)| \) the norm on the Banach space \( C([0,1]) \). Given a positive integer \( n \), the linear operator \( V_n \) defined on \( C([0,1]) \) as

\[
V_n f(x) := \frac{1}{n} \sum_{j=1}^{n} \sigma^2(x, j/n) f(j/n)
\]

is a finite rank operator whose eigenvalues coincide with those of the matrix \( V_n \). It is easy to check that \( V_n f \to V f \) in \( C([0,1]) \) for all \( f \in C([0,1]) \), in other words, \( V_n \) converges strongly to \( V \) in \( C([0,1]) \), denoted by

\[
V_n \xrightarrow{\text{str}} V_{n \to \infty} V
\]

in the sequel. However, \( V_n \) does not converge to \( V \) in norm, in which case the convergence of \( \rho(V_n) \) to \( \rho(V) \) would have been immediate. Nonetheless, the family of operators \( \{V_n\} \) satisfies the property that the set \( \{V_n f : n \geq 1, \|f\|_\infty \leq 1\} \) has a compact closure, being a set of equicontinuous and bounded functions thanks to the uniform continuity of \( \sigma^2 \) on \([0,1]^2 \). Following [13], such a family is named collectively compact.

We recall the following important properties, cf. [13]. If a sequence \( (T_n) \) of collectively compact operators on a Banach space converges strongly to a bounded operator \( T \), then:

i) The spectrum of \( T_n \) is eventually contained in any neighborhood of the spectrum of \( T \). Furthermore, \( \lambda \) belongs to the spectrum of \( T \) if and only if there exist \( \lambda_n \) in the spectrum of \( T_n \) such that \( \lambda_n \to \lambda \);

ii) \( (\lambda - T_n)^{-1} \xrightarrow{n \to \infty} (\lambda - T)^{-1} \) for any \( \lambda \) in the resolvent set of \( T \).

The statement (1) of the theorem follows from i).
We now provide the main steps of the proof of the statement (2). Given \( n \geq 1 \) and \( s > 0 \), let \((q^n(s))^T \in \mathbb{R}^{2n}\) be the solution of the system (2.11) that is specified by Theorem 2.2. Denote by \( q^n(s) = (q^n_1(s), \ldots, q^n_n(s)) \) and \( \tilde{q}^n = (\tilde{q}^n_1, \ldots, \tilde{q}^n_n) \) and introduce the quantities

\[
\Phi_n(x, s) := \frac{1}{n} \sum_{i=1}^{n} \sigma^2 \left( x, \frac{i}{n} \right) q^n_i(s) \quad \text{and} \quad \tilde{\Phi}_n(x, s) := \frac{1}{n} \sum_{i=1}^{n} \sigma^2 \left( \frac{i}{n}, x \right) q^n_i(s). \tag{A.3}
\]

By Proposition 2.5 (recall that A3 holds), we know that the average \( (q^n(s))_n = \frac{1}{n} \sum_{i=1}^{n} q^n_i(s) \) satisfies \( (q^n(s))_n \leq \sigma_{\min}^{-1} \). Therefore, we get from (1.5) that

\[
\|q^n(s)\|_{\infty} \leq \frac{\sigma_{\max}^2(q^n(s))_n}{s^2} \leq \frac{\sigma_{\max}^2}{\sigma_{\min}s^2}. \tag{A.4}
\]

Consequently the family \( \{\Phi_n(\cdot, s)\}_{n \geq 1} \) is an equicontinuous and bounded subset of \( C([0, 1]) \). Similarly, an identical conclusion holds for the family \( \{\Phi_n(\cdot, s)\}_{n \geq 1} \). By Arzela–Ascoli’s theorem, there exists a subsequence (still denoted by \( (n) \), with a small abuse of notation) along which \( \Phi_n(\cdot, s) \) and \( \Phi_n(\cdot, s) \) respectively converge to given functions \( \Phi_{\infty}(\cdot, s) \) and \( \Phi_{\infty}(\cdot, s) \) in \( C([0, 1]) \). Denote

\[
\Psi_n(x, s) = \frac{1}{s^2 + \Phi_n(x, s) \Phi_n(x, s)} \quad \text{and} \quad \Psi_{\infty}(x, s) = \frac{1}{s^2 + \Phi_{\infty}(x, s) \Phi_{\infty}(x, s)}.
\]

and introduce the auxiliary quantities

\[
Q_n(x, s) := \Psi_n(x, s) \Phi_n(x, s) \quad \text{and} \quad \tilde{Q}_n(x, s) := \Phi_n(x, s) \Phi_n(x, s).
\]

Then there exist \( Q_{\infty}(x, s) \) and \( \tilde{Q}_{\infty}(x, s) \) such that \( Q_n(\cdot, s) \to Q_{\infty}(\cdot, s) \) and \( \tilde{Q}_n(\cdot, s) \to \tilde{Q}_{\infty}(\cdot, s) \) in \( C([0, 1]) \). These limits satisfy

\[
Q_{\infty}(x, s) = \frac{\Phi_{\infty}(x, s)}{s^2 + \Phi_{\infty}(x, s) \Phi_{\infty}(x, s)} \quad \text{and} \quad \tilde{Q}_{\infty}(x, s) = \frac{\Phi_{\infty}(x, s)}{s^2 + \Phi_{\infty}(x, s) \Phi_{\infty}(x, s)}.
\]

Moreover, the mere definition of \( q^n \) and \( \tilde{q}^n \) as solutions of (2.11) yields that

\[
\begin{cases}
Q_n \left( \frac{i}{n}, s \right) = q^n_i(s) & 1 \leq i \leq n \\
\tilde{Q}_n \left( \frac{i}{n}, s \right) = \tilde{q}^n_i(s) & 1 \leq i \leq n.
\end{cases} \tag{A.5}
\]

Combining (A.3), (A.5) and the convergence of \( Q_n \) and \( \tilde{Q}_n \), we finally obtain the useful representation

\[
\Phi_{\infty}(x, s) = \int_0^1 \sigma^2(x, y) \tilde{Q}_{\infty}(y, s) dy \quad \text{and} \quad \tilde{\Phi}_{\infty}(x, s) = \int_0^1 \sigma^2(y, x) Q_{\infty}(y, s) dy. \tag{A.6}
\]

which yields that \( Q_{\infty} \) and \( \tilde{Q}_{\infty} \) satisfy the system (3.3).

To establish the first part of the statement (2), we show that these limits are zero if \( s^2 \geq \rho(V) \) and positive if \( s^2 < \rho(V) \), then we show that they are unique. It is known that \( \rho(V) \) is a simple eigenvalue, it has a positive eigenvector, and there is no other eigenvalue with a positive eigenvector. If \( T \) is a bounded operator on \( C([0, 1]) \) such that \( Tf - \nabla f > 0 \) for \( f \geq 0 \), then \( \rho(T) > \rho(V) \) [26, Theorem 19.2 and 19.3].

We first establish (2)-(a). Fix \( s^2 \geq \rho(V) \), and assume that \( Q_{\infty}(\cdot, s) \not\equiv 0 \). Since \( Q_{\infty}(\cdot, s) = \Psi_{\infty} V Q_{\infty}(\cdot, s) \), where \( \Psi_{\infty}(\cdot, s) \) is the limit of \( \Psi_n(\cdot, s) \) along the subsequence \( (n) \), it holds that \( Q_{\infty}(\cdot, s) > 0 \), and by the properties of the Krein–Rutman eigenvalue, that \( \rho(\Psi_{\infty} V) = 1 \). From the identity \( \int Q_{\infty}(x, s) dx = \int \tilde{Q}_{\infty}(x, s) dx \), we get that \( \tilde{Q}_{\infty}(\cdot, s) \not\equiv 0 \), hence \( \tilde{Q}_{\infty}(\cdot, s) > 0 \) by the same
argument. By consequence, $s^{-2} \mathbf{V}f - \Psi_{\infty} \mathbf{V}f > 0$ for all $f \not\equiv 0$. This leads to the contradiction $1 \geq \rho(s^{-2} \mathbf{V}) > \rho(\Psi_{\infty} \mathbf{V}) = 1$. Thus, $Q_{\infty}(\cdot, s) = \hat{Q}_{\infty}(\cdot, s) = 0$.

We now establish (2)-(b). Let $s^2 < \rho(\mathbf{V})$. By an argument based on collective compactness, it holds that

$$
\rho(\Psi_n \mathbf{V}_n) \xrightarrow{n \to \infty} \rho(\Psi_{\infty} \mathbf{V})
$$

and moreover, that $\rho(\Psi_n \mathbf{V}_n) = 1$ (see e.g. the proof of Lemma 5.5). Thus, $Q_{\infty}(\cdot, s) \not\equiv 0$ and $\hat{Q}_{\infty}(\cdot, s) \not\equiv 0$, otherwise $\rho(\Psi_{\infty} \mathbf{V}) = \rho(s^{-2} \mathbf{V}) > 1$. Since $Q_{\infty}(\cdot, s) = \Psi_{\infty} \mathbf{V}Q_{\infty}(\cdot, s)$, we get that $Q_{\infty}(\cdot, s) > 0$ and similarly, that $\hat{Q}_{\infty}(\cdot, s) > 0$.

It remains to show that the accumulation point $(Q_{\infty}, \hat{Q}_{\infty})$ is unique. The proof of this fact is similar to its finite dimensional analogue in the proof of Lemma 5.5. In particular, the properties of the Perron–Frobenius eigenvalue and its eigenspace are replaced with their Krein–Rutman counterparts, and the matrices $K\tilde{q}$ and $K\tilde{q}\tilde{q}$ in that proof are replaced with continuous and strongly positive integral operators. Note that the end of the proof is simpler in our context, thanks to the strong positivity assumption instead of the irreducibility assumption. We leave the details to the reader.

We now address (2)-(c) and first prove the continuity of $Q_{\infty}$ and $\hat{Q}_{\infty}$ on $[0, 1] \times (0, \infty)$. This is equivalent to proving the continuity of $\Phi_{\infty}$ and $\hat{\Phi}_{\infty}$ on this set. Let $(x_k, s_k) \to_k (x, s) \in [0, 1] \times (0, \infty)$. The bound

$$
0 \leq \tilde{Q}_{\infty}(y, s) \leq \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} s^2
$$

follows from (A.5) and the convergence of $\tilde{Q}_n$ to $\tilde{Q}_{\infty}$. As a consequence of (A.6), the family $\{\Phi_{\infty}(\cdot, s_k)\}_k$ is equicontinuous for $k$ large. By Arzela–Ascoli’s theorem and the uniqueness of the solution of the system, we get that $\Phi_{\infty}(\cdot, s_k) \to_k \Phi_{\infty}(\cdot, s)$ in $C([0, 1])$. Therefore, writing $|\Phi_{\infty}(x_k, s_k) - \Phi_{\infty}(x, s)| \leq ||\Phi_{\infty}(\cdot, s_k) - \Phi_{\infty}(\cdot, s)||_{\infty} + |\Phi_{\infty}(x_k, s) - \Phi_{\infty}(x, s)|$ and using the continuity of $\Phi_{\infty}(\cdot, s)$, we get that $\Phi_{\infty}(x_k, s_k) \to_k \Phi_{\infty}(x, s)$.

The main steps of the proof for extending the continuity of $Q_{\infty}$ and $\hat{Q}_{\infty}$ from $[0, 1] \times (0, \infty)$ to $[0, 1] \times [0, \infty)$ are the following. Following the proof of Proposition 3.4, we can establish that

$$
\liminf_{s \downarrow 0} \int_0^1 Q_{\infty}(x, s) \, dx > 0.
$$

The details are omitted. Since

$$
\frac{1}{Q_{\infty}(x, s)} = \frac{s^2}{\Phi_{\infty}(x, s)} + \tilde{\Phi}_{\infty}(x, s) > \sigma_{\text{min}} \int_0^1 Q_{\infty}(y, s) \, dy,
$$

we obtain that $||\tilde{Q}_{\infty}(\cdot, s)||_{\infty}$ is bounded when $s \in (0, \varepsilon)$ for some $\varepsilon > 0$. Thus, $\{\Phi_{\infty}(\cdot, s)\}_{s \in (0, \varepsilon)}$ is equicontinuous by (A.6), and it remains to prove that the accumulation point $\Phi_{\infty}(\cdot, 0)$ is unique.

This can be done by working on the system (3.3) for $s = 0$, along the lines of the proof of Lemma 5.5 and Proposition 3.4. Details are omitted.

Turning to Statement (3), the assertion $F(s) \to 0$ as $s \downarrow 0$ can be deduced from the proof of Proposition 3.4 and a passage to the limit, noting that the bounds in that proof are independent from $n$.

Consider the Banach space $\mathcal{B} = C([0, 1]; \mathbb{R}^2)$ of continuous functions

$$
\hat{f} = (f, \hat{f})^T : [0, 1] \to \mathbb{R}^2
$$

endowed with the norm $\|\hat{f}\|_{\mathcal{B}} = \sup_{t \in [0, 1]} (|f(x)| + |\hat{f}(x)|)$. In the remainder of the proof, we may use the notation shortcut $\Psi_{\infty}^s$ instead of $\Psi_{\infty}(\cdot, s)$ and corresponding shortcuts for quantities $\Phi_{\infty}(\cdot, s)$, $\hat{\Phi}_{\infty}(\cdot, s)$, $Q_{\infty}(\cdot, s)$ and $\hat{Q}_{\infty}(\cdot, s)$.
Given $s, s' \in (0, \sqrt{\rho(V)})$ with $s \neq s'$, consider the function

$$\Delta \bar{Q}_{\infty}^{s, s'} := \frac{(Q_{\infty}^s - Q_{\infty}^{s'}, \bar{Q}_{\infty}^s - \bar{Q}_{\infty}^{s'})^T}{s^2 - s'^2} \in \mathcal{B}.$$  

Let $V^T$ be the linear operator associated to the kernel $(x, y) \mapsto \sigma^2(y, x)$, and defined as

$$V^T f(x) := \int_0^1 \sigma^2(y, x) f(y) \, dy.$$  

Then, mimicking the proof of Lemma 5.6, it is easy to prove that $\Delta \bar{Q}_{\infty}^{s, s'}$ satisfies the equation

$$\Delta \bar{Q}_{\infty}^{s, s'} = M_{\infty}^{s, s'} \Delta \bar{Q}_{\infty}^{s, s'} + a_{\infty}^{s, s'},$$

where $M_{\infty}^{s, s'}$ is the operator acting on $\mathcal{B}$ and defined in a matrix form as

$$M_{\infty}^{s, s'} = \begin{pmatrix} s^2 \Psi_{\infty}^s \Psi_{\infty}^{s'} V^T & -\Psi_{\infty}^s \bar{\Psi}_{\infty}^{s'} \bar{V}^T \\ -\Psi_{\infty}^s \bar{\Psi}_{\infty}^{s'} \bar{V}^T & s^2 \Psi_{\infty}^{s'} \bar{V}^T \end{pmatrix} ,$$

and $a_{\infty}^{s, s'}$ is a function $\mathcal{B}$ defined as

$$a_{\infty}^{s, s'} = -\begin{pmatrix} \Psi_{\infty}^s \Psi_{\infty}^{s'} V^T Q_{\infty}^s \\ \Psi_{\infty}^s \Psi_{\infty}^{s'} V Q_{\infty}^s \end{pmatrix}.$$  

To proceed, we rely on a regularized version of this equation. Denoting by $1$ the constant function $1(x) = 1$ in $C([0, 1])$, and letting $v = (1, -1)^T \in \mathcal{B}$, the kernel operator $vv^T$ on $\mathcal{B}$ is defined by the matrix

$$(vv^T)(x, y) = \begin{pmatrix} 1(x)1(y) & -1(x)1(y) \\ -1(x)1(y) & 1(x)1(y) \end{pmatrix}.$$  

By the constraint $\int Q_{\infty}^s = \int \bar{Q}_{\infty}^s$, it holds that $(vv^T)\Delta \bar{Q}_{\infty}^{s, s'} = 0$. Thus, $\Delta \bar{Q}_{\infty}^{s, s'}$ satisfies the identity

$$((I - (M_{\infty}^{s, s'})^T)(I - M_{\infty}^{s, s'}) + vv^T)\Delta \bar{Q}_{\infty}^{s, s'} = (I - (M_{\infty}^{s, s'})^T)a_{\infty}^{s, s'}.$$  

(A.7)

We rewrite the left hand member of this identity as $(I - G_{\infty}^{s, s'})\Delta \bar{Q}_{\infty}^{s, s'}$ where

$$G_{\infty}^{s, s'} = M_{\infty}^{s, s'} + (M_{\infty}^{s, s'})^T - (M_{\infty}^{s, s'})^T M_{\infty}^{s, s'} - vv^T,$$

and we study the behavior of $M_{\infty}^{s, s'}$ and $G_{\infty}^{s, s'}$ as $s' \to s$.

Let $s \in (0, \sqrt{\rho(V)})$ and $s'$ belong to a small compact neighborhood $\mathcal{K}$ of $s$. Then the first component of $M_{\infty}^{s, s'} f(x)$ has the form

$$\int (\Theta_{11}(x, y, s') f(y) + \Theta_{12}(x, y, s') \bar{f}(y)) \, dy,$$

where $\Theta_{11}$ and $\Theta_{12}$ are continuous on the compact set $[0, 1]^2 \times \mathcal{K}$ by the previous results. A similar argument holds for the other component of $M_{\infty}^{s, s'} f(x)$. By the uniform continuity of these functions on this set, we get that the family $\{M_{\infty}^{s, s'} f : s' \in \mathcal{K}, \|f\|_{\mathcal{B}} \leq 1\}$ is equicontinuous, and by the Arzela–Ascoli theorem, the family $\{M_{\infty}^{s, s'} : s' \in \mathcal{K}\}$ is collectively compact. Moreover,

$$M_{\infty}^{s, s'} \xrightarrow{str} M_{\infty}^s := \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} N_{\infty}^s \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} ,$$

where

$$N_{\infty}^s := \begin{pmatrix} s^2 \Psi_{\infty}^s (\cdot, s) V^T & \Psi_{\infty}^s (\cdot, s) \bar{\Phi}_{\infty}^s (\cdot, s) V^T \\ \Psi_{\infty}^s (\cdot, s) \Phi_{\infty}^s (\cdot, s) V^T & s^2 \Psi_{\infty}^s (\cdot, s) V^T \end{pmatrix}.$$  

By a similar argument, $\{G_{\infty}^{s, s'} : s' \in \mathcal{K}\}$ is collectively compact, and $G_{\infty}^{s, s'} \xrightarrow{str} G_{\infty}^s$, where

$$G_{\infty}^s := M_{\infty}^s + (M_{\infty}^s)^T - (M_{\infty}^s)^T M_{\infty}^s - vv^T.$$
We now claim that 1 belongs to the resolvent set of the compact operator $G_{s\infty}^s$.

Repeating an argument of the proof of Lemma 5.6, we can prove that the Krein–Rutman eigenvalue of the strongly positive operator $N_{s\infty}^s$ is equal to one, and its eigenspace is generated by the vector $\tilde{Q}_{s\infty}^s := (Q_{s\infty}^s, \tilde{Q}_{s\infty}^s)^T$. From the expression of $M_{s\infty}^s$, we then obtain that the spectrum of this compact operator contains the simple eigenvalue 1, and its eigenspace is generated by the vector $(Q_{s\infty}^s, -\tilde{Q}_{s\infty}^s)$.

We now proceed by contradiction. If 1 were an eigenvalue of $G_{s\infty}^s$, there would exist a non zero vector $\tilde{f} \in B$ such that $(I - G_{s\infty}^s)\tilde{f} = 0$, or, equivalently,

$$(I - (M_{s\infty}^s)\tilde{f})^T(I - M_{s\infty}^s)\tilde{f} + vv^T \tilde{f} = 0.$$ 

Left-multiplying the left hand side of this expression by $\tilde{f}^T$ and integrating on $[0, 1]$, we get that $(I - M_{s\infty}^s)\tilde{f} = 0$ and $\int f = \int \tilde{f}$, which contradicts the fact the $\tilde{f}$ is collinear with $(Q_{s\infty}^s, -\tilde{Q}_{s\infty}^s)$.

Returning to (A.7) and observing that $\{M_{s\infty}^s s' : s' \in \mathcal{K}\}$ is bounded, we get from the convergence $(M_{s\infty}^s)^T \xrightarrow{str} (M_{s\infty}^s)^T$ that

$$(I - (M_{s\infty}^s)^T)\alpha_{s\infty}^{s,s'}_s \xrightarrow{s' \to s} (I - (M_{s\infty}^s)^T)\alpha_{s\infty}^s,$$

where

$$\alpha_{s\infty}^s(\cdot) := -\left(\Psi_{s\infty}(\cdot, s)^2 (V^T Q_{s\infty}(\cdot, s) - V^T \tilde{Q}_{s\infty}(\cdot, s))^\top\right).$$

From the aforementioned results on the collectively compact operators, it holds that there is a neighborhood of 1 where $G_{s\infty}^{s,s'}$ has no eigenvalue for all $s'$ close enough to $s$ (recall that 0 is the only possible accumulation point of the spectrum of $G_{s\infty}^s$). Moreover,

$$(I - G_{s\infty}^{s,s'})^{-1} \xrightarrow{str} (I - G_{s\infty}^s)^{-1}.$$

In particular, for $s'$ close enough to $s$, the family $\{(I - G_{s\infty}^{s,s'})^{-1}\}$ is bounded by the Banach-Steinhaus theorem. Thus,

$$\Delta Q_{s\infty}^{s,s'} \xrightarrow{s' \to s} (I - (M_{s\infty}^s)^T(I - M_{s\infty}^s) + vv^T)^{-1}(I - (M_{s\infty}^s)^T)\alpha_{s\infty}^s = (\partial_{s\infty} Q_{s\infty}^s, \partial_{s\infty} \tilde{Q}_{s\infty}^s)^T.$$

Using this result, we straightforwardly obtain from the expression of $F_{\infty}$ that this function is differentiable on $(0, \sqrt{p(V)})$. The continuity of the derivative as well as the existence of a right limit as $s \downarrow 0$ and a left limit as $s \uparrow \sqrt{p(V)}$ can be shown by similar arguments involving the behaviors of the operators $M_{s\infty}^s$ and $G_{s\infty}^s$ as $s$ varies. The details are skipped.

Since $\mu_n^Y \sim \mu_n$ in probability and since we have the straightforward convergence $\mu_n \xrightarrow{w} \mu_{\infty}$, the statement (4) of the theorem follows.

**A.3. Proof of Proposition 3.4.** We first prove item (1).

Let $\sigma_{\max}, \sigma_{\min}$ be given by A1 and A3 and denote by $\langle q \rangle_n = \frac{1}{n} \langle q, 1 \rangle$ and $\langle \tilde{q} \rangle_n = \frac{1}{n} \langle \tilde{q}, 1 \rangle$. Notice that by Theorem 2.2, $\langle q \rangle_n = \langle \tilde{q} \rangle_n$.

Let $s_0 = \sigma_{\min}/2$. We first establish useful uniform upper and lower bounds:

$$0 < \kappa \leq \langle q(s) \rangle_n \leq K < \infty,$$  \hspace{1cm} (A.8)

where $\kappa, K$ are generic constants, independent of $s \in (0, s_0]$ and $n \geq 1$. 
In fact, \((V\tilde{q})_i(V^Tq)_i \leq \sigma^4_{\max}(q^2_n)\) and \((V^Tq)_i \geq \sigma^2_{\min}(q)_n\). The definition of \(q_i\) yields

\[
q_i = \frac{(V^Tq)_i}{s^2 + (Vq)_i(V^Tq)_i} \geq \frac{\sigma^2_{\min}(q)_n}{s^2 + \sigma^4_{\max}(q^2_n)} \Rightarrow \langle q \rangle_n \geq \frac{\sigma^2_{\min}(q)_n}{s^2 + \sigma^4_{\max}(q^2_n)} .
\]

By Theorem 2.2 again, \(q \succ 0\) for \(s > 0\), so

\[
1 \geq \frac{\sigma^2_{\min}(q)_n}{s^2 + \sigma^4_{\max}(q^2_n)} \Rightarrow s^2 + \sigma^4_{\min}(q^2_n) \geq \sigma^2_{\min} .
\]

which immediately implies the lower bound in (A.8) for \(s \leq s_0\). Similarly,

\[
1 \leq \frac{\sigma^2_{\max}(q^2_n)}{s^2 + \sigma^4_{\min}(q^2_n)} \Rightarrow s^2 + \sigma^4_{\min}(q^2_n) \leq \sigma^2_{\max} .
\]

and (A.8) is proved. Note that combining (A.8) with A1 and A3, we easily obtain the following bounds

\[
0 < \kappa \leq q_i(s), \tilde{q}_i(s), (V^Tq(s))_i, (V\tilde{q}(s))_i \leq K < \infty ,
\]

where \(\kappa, K\) are again generic constants, uniform in \(s \in (0,s_0], i \in [n], n \geq 1\).

We now study the convergence of \(q(s)\) and \(\tilde{q}(s)\) as \(s \downarrow 0\). According to Sinkhorn’s theorem [62], matrix \(V > 0\) admits a unique pair of vectors \(f = (f_i) > 0, \tilde{f} = (\tilde{f}_i) > 0\) such that

\[
f_i(V\tilde{f})_i = 1, \quad \tilde{f}_i(V^Tf)_i = 1, \quad i \in [n] \quad \text{and} \quad \langle f \rangle_n = \langle \tilde{f} \rangle_n .
\]

The definition (1.5) of \(q_i(s)\) yields

\[
q_i(s)(V\tilde{q}(s))_i + s^2 \frac{q_i(s)}{(V^Tq(s))_i} = 1 .
\]

Consider a generic sequence \(s_m \downarrow 0\), and denote by \(q = (q_i), \tilde{q} = (\tilde{q}_i)\) the limits of a converging subsequence of \((q(s_m), \tilde{q}(s_m))\). Then \(q, \tilde{q} \succ 0\) and \(\langle q \rangle_n = \langle \tilde{q} \rangle_n\). Combining (A.10) and (A.11), we obtain \(q = f\) and \(\tilde{q} = \tilde{f}\). Since this argument holds for every converging subsequence, this proves that

\[
q(0) := \lim_{s \downarrow 0} q(s) \quad \text{and} \quad \tilde{q}(0) := \lim_{s \downarrow 0} \tilde{q}(s) \quad \text{exist and satisfy} \quad q_i(0)(V\tilde{q}(0))_i = 1 \quad \text{and} \quad \tilde{q}_i(0)(V^Tq(0))_i = 1 .
\]

Combining this property with the definition of \(\mu_n\), we obtain

\[
\mu_n(\{0\}) = 1 - \lim_{s \downarrow 0} \frac{1}{n} \langle q(s), V\tilde{q}(s) \rangle = 1 - \frac{1}{n} \sum_{i \in [n]} q_i(0)(V\tilde{q}(0))_i = 0 .
\]

Part (1) of the proposition is proved.

We then prove item (2). We first prove that \(s \mapsto \varphi_n(s)\) admits a limit as \(s \downarrow 0\). By Lemma 5.6, \(\nabla \tilde{q}_i(s) = 2sA(s)^{-L}b(s)\), where \(A(s)^{-L}\) and \(b(s)\) are defined there. Careful inspection of the proof of Lemma 5.6 together with Proposition 3.4-(1) shows that \(A(s) \to A(0)\) which still has full column rank, that \(A(s)^{-L} \to A(0)^{-L}\), and that \(b(s) \to b(0)\) as \(s \downarrow 0\).

Denote by \([A(s)^{-L}b(s)]_{1:n}\) the \(n \times 1\) vector made of the first \(n\) components of \(A(s)^{-L}b(s)\), and by \([A(s)^{-L}b(s)]_{n+1:2n}\) the vector made of the last \(n\) components of \(A(s)^{-L}b(s)\). By differentiating \(F_n(s)\) and relying on Remark 2.8, we obtain

\[
\varphi_n(s) = -\frac{1}{n\pi} \langle [A(s)^{-L}b(s)]_{1:n}, V\tilde{q}(s) \rangle - \frac{1}{n\pi} \langle q(s), V[A(s)^{-L}b(s)]_{n+1:2n} \rangle .
\]

Hence, function \(\varphi_n\) admits a finite limit as \(s \downarrow 0\) denoted by \(\varphi_n(0)\).
A direct computation of $\varphi_n(0)$ based on the previous formula appears difficult, so we proceed differently. Since $\varphi_n$ is continuous as $s \downarrow 0$, we have

$$
\lim_{s \downarrow 0} \frac{1}{\pi s^2} \mu_n \{ |z| \leq s \} = \lim_{s \downarrow 0} \frac{1}{\pi s^2} \int_{\{|z| \leq s\}} \varphi_n(|z|) \ell(dz) = \varphi_n(0) .
$$

(A.12)

On the other hand,

$$
\lim_{s \downarrow 0} \frac{1}{\pi s^2} \mu_n \{ |z| \leq s \} = \lim_{s \downarrow 0} \frac{1}{\pi s^2} \left( 1 - \frac{1}{n} \langle q, V\bar{q} \rangle \right) = \lim_{s \downarrow 0} \frac{1}{\pi n} \sum_{i \in [n]} \frac{1}{s^2 + (V\bar{q})_i (V^T\bar{q})_i} = \frac{1}{\pi n} \sum_{i \in [n]} (V\bar{q}(0))_i (V^T\bar{q}(0))_i ,
$$

(A.13)

where the last limit follows from Proposition 3.4-(1) and is finite by the bounds $\langle A.9 \rangle$. It remains to identify (A.12) and (A.13) to conclude. The uniform bounds over $\varphi_n(0)$ follow from (A.9).

This concludes the proof of Proposition 3.4.

**Appendix B. Remaining proofs for Section 4**

**B.1. Stieltjes transform of a symmetric probability measure.** We note that a symmetric probability distribution $\tilde{\nu}$ on $\mathbb{R}$ satisfies $\tilde{\nu}(A) = \tilde{\nu}(-A)$ for each Borel set $A \subset \mathbb{R}$.

**Lemma B.1.** A probability measure $\tilde{\nu}$ is symmetric if and only if its Stieltjes transform $g_{\tilde{\nu}}$, seen as an analytic function on $\mathbb{C} \setminus \mathbb{R}$, satisfies $g_{\tilde{\nu}}(-\eta) = -g_{\tilde{\nu}}(\eta)$.

**Proof.** The necessity is obvious from the definition of the Stieltjes transform and from the fact that $\tilde{\nu}(d\lambda) = \tilde{\nu}(-d\lambda)$. To prove the sufficiency, we use the Perron inversion formula, that says that for any function $\varphi \in C_c(\mathbb{R})$,

$$
\int_{\mathbb{R}} \varphi(x) \tilde{\nu}(dx) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_{\mathbb{R}} \varphi(x) \text{ Im } g_{\tilde{\nu}}(x + i\varepsilon) \, dx .
$$

By a simple variable change at the right hand side, and by using the equalities $g_{\tilde{\nu}}(-\eta) = -g_{\tilde{\nu}}(\eta)$ and $g_{\tilde{\nu}}(\eta) = g_{\tilde{\nu}}(\eta)$, we obtain that $\int \varphi(x) \tilde{\nu}(dx) = \int \varphi(-x) \tilde{\nu}(dx)$, as desired. \(\square\)

**B.2. Variance estimates.** In this section we collect without proofs a number of standard variance estimates. Let $(Y_n)$ be a sequence of matrices as in Definition 1.2. In the sequel we drop the subscript $n$. Denote by $(\vec{e}_i)$ the standard vector basis. We introduce the following notations:

$$
Y = (\vec{y}_1, \cdots, \vec{y}_n) \quad \text{and} \quad Q(\eta^2) = \left[ \sum_{i=2}^n (\vec{y}_i - z\vec{e}_i)(\vec{y}_i - z\vec{e}_i)^* - \eta^2 \right]^{-1} .
$$

Recall the definition of matrices $R$ and $G$ in (4.1).

**Proposition B.2.** Let $A_0$ and $A_1$ hold. Let $\Delta$ be a $n \times n$ deterministic diagonal matrix, then the following estimates hold:

$$
\text{var}(R_{ij}) = O_\eta \left( n^{-1} \right) \quad \text{for} \quad 1 \leq i, j \leq 2n , \quad (B.1)
$$

$$
\text{var} \left( \frac{1}{n} \text{ tr } \Delta G \right) = O_\eta \left( \|\Delta\|^2 n^{-2} \right) , \quad (B.2)
$$

$$
\text{var} \left[ (\vec{y}_i - z\vec{e}_1)^* \eta Q(\alpha)(\vec{y}_i - z\vec{e}_1) \right] = O_\eta \left( n^{-1} \right) \quad \text{for} \quad \alpha = 1, 2 . \quad (B.3)
$$

Similar estimates hold true if $G$ is replaced by $\bar{G}$, if one considers the columns of $Y^*$ instead of those of $Y$, etc.
These estimates can be obtained as in the proof of [51, Proposition 6.3], see also the references therein.

As a direct corollary of the previous proposition, we have:

**Corollary B.3.** Let A0 and A1 hold.

\[
\operatorname{var}\left[\left(\eta + \frac{1}{n} \operatorname{tr} \Delta G\right)^{-1}\right] = O_{\eta}\left(\|\Delta\|^2 n^{-2}\right), \tag{B.4}
\]

\[
\operatorname{var}\left[(\eta + (\bar{y}_1 - z\bar{e}_1)^*)[\eta Q](\bar{y}_1 - z\bar{e}_1)^{-1}\right] = O_{\eta}(n^{-1}). \tag{B.5}
\]

**Proof.** Let us establish (B.4). Notice first that \(|\eta + \frac{1}{n} \operatorname{tr} \Delta G|^{-1} \leq \operatorname{Im}^{-1}(\eta)|\) by (2.3).

\[
\operatorname{var}\left(\eta + \frac{1}{n} \operatorname{tr} \Delta G\right) = \mathbb{E}\left|\eta + \frac{1}{n} \operatorname{tr} \Delta G\right|^2 - \mathbb{E}\left(\eta + \frac{1}{n} \operatorname{tr} \Delta G\right)^2, \tag{B.6}
\]

\[
\leq \mathbb{E}\left|\eta + \frac{1}{n} \operatorname{tr} \Delta G\right|^2 - \mathbb{E}\left(\eta + \frac{1}{n} \operatorname{tr} \Delta G\right)^2, \tag{B.7}
\]

\[
\leq \frac{1}{\operatorname{Im}^4(\eta)} \operatorname{var}\left(\frac{1}{n} \operatorname{tr} \Delta G\right) = O_{\eta}\left(\frac{\|\Delta\|^2}{n^2}\right),
\]

where (a) follows from the fact that \(\operatorname{var}(X) = \inf_a \mathbb{E}|X - a|^2\). Estimate (B.5) can be established similarly. \(\square\)

**B.3. Proof of Proposition 4.6.** We will establish the following convergences

\[
\frac{1}{n} \operatorname{tr} \mathbb{E}G(\eta, z) - \frac{1}{n} \operatorname{tr} \mathbb{E}G^N(\eta, z) = O_{\eta}\left(\frac{1}{\sqrt{n}}\right), \tag{B.6}
\]

\[
\frac{1}{n} \operatorname{tr} \mathbb{E}F(\eta, z) - \frac{1}{n} \operatorname{tr} \mathbb{E}F^N(\eta, z) = O_{\eta}\left(\frac{1}{\sqrt{n}}\right). \tag{B.7}
\]

The analogues for \(\tilde{G} - G^N\) and \(F' - F'^N\) follow the same lines and their proof is omitted.

**Proof of (B.6).** Recall that \(G(\eta, z) = \eta [\eta (Y - z)(Y - z)^* - \eta^2]^{-1}\). Consider the column vectors of \(Y^N: Y^N = (\bar{y}_1^N, \ldots, \bar{y}_n^N)\) and the following family of interpolating matrices:

\[
Y_i = (\bar{y}_1^N, \ldots, \bar{y}_i^N, \bar{y}_{i+1}, \ldots, \bar{y}_n),
\]

with the convention that \(Y_0 = Y\) and \(Y_n = Y^N\). We can write

\[
\frac{1}{n} \mathbb{E}\operatorname{tr}(G - G^N) = \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}\operatorname{tr}\left\{\eta [(Y_i - z)(Y_i - z)^* - \eta^2]^{-1} - \eta [(Y_{i+1} - z)(Y_{i+1} - z)^* - \eta^2]^{-1}\right\},
\]

\[
= \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}T_i. \tag{B.8}
\]

We now handle the first term of the r.h.s., the other ones being handled similarly.

\[
T_1 = \eta \operatorname{tr}\left[(Y_0 - z)(Y_0 - z)^* - \eta^2\right]^{-1} - \eta \operatorname{tr}\left[(Y_1 - z)(Y_1 - z)^* - \eta^2\right]^{-1}.
\]

By the Sherman–Morrison formula, we have

\[
T_1 = -\eta \frac{(\bar{y}_1 - z\bar{e}_1)^* Q^2(\bar{y}_1 - z\bar{e}_1)}{1 + (\bar{y}_1 - z\bar{e}_1)^* Q(\bar{y}_1 - z\bar{e}_1)} + \frac{\eta (\bar{y}_1 - z\bar{e}_1)^* Q^2(\bar{y}_1 - z\bar{e}_1)}{1 + (\bar{y}_1 - z\bar{e}_1)^* Q(\bar{y}_1 - z\bar{e}_1)},
\]

\[
= -\frac{(\bar{y}_1 - z\bar{e}_1)^* [\eta Q]^2(\bar{y}_1 - z\bar{e}_1)}{\eta + (\bar{y}_1 - z\bar{e}_1)^* [\eta Q](\bar{y}_1 - z\bar{e}_1)} + \frac{(\bar{y}_1 - z\bar{e}_1)^* [\eta Q]^2(\bar{y}_1 - z\bar{e}_1)}{\eta + (\bar{y}_1 - z\bar{e}_1)^* [\eta Q](\bar{y}_1 - z\bar{e}_1)}.
\]
Denote by
\[ A = (\vec{y}_1 - z\vec{e}_1)^*|\eta Q|^2(\vec{y}_1 - z\vec{e}_1) \quad \text{and} \quad B = \eta + (\vec{y}_1 - z\vec{e}_1)^*|\eta Q|(\vec{y}_1 - z\vec{e}_1), \]
and by \( A^N \) and \( B^N \) the analogues of \( A \) and \( B \) respectively, where \( \vec{y}_1 \) is replaced with \( \vec{y}_1^N \). Then
\[ \mathbb{E}T_1 = -\mathbb{E}AB^{-1} + \mathbb{E}A^NB^{-1}. \]
Proposition B.2 yields
\[ \left| \mathbb{E}\left(A^NB^{-1}\right) - \mathbb{E}A^N\mathbb{E}B^{-1} \right| \leq \sqrt{\text{var}(A^N)}\sqrt{\text{var}(B^{-1})} = O_\eta\left(\frac{1}{n}\right). \]
Hence
\[ \mathbb{E}T_1 = -\mathbb{E}\left(\frac{A - A^N + A^N}{B - B^N}\right), \]
\[ = -\mathbb{E}\left(\frac{1}{B}(A - A^N) - \mathbb{E}A^N\left(\frac{1}{B} - \frac{1}{B^N}\right)\right), \]
\[ = -\mathbb{E}\left(\frac{1}{B}\right)\mathbb{E}(A - A^N) - \mathbb{E}A^N\mathbb{E}\left(\frac{1}{B} - \frac{1}{B^N}\right) + O_\eta\left(\frac{1}{n}\right). \]
Now, \( \mathbb{E}B \) and \( \mathbb{E}A^N \) are of order \( O_\eta(1) \), \( \mathbb{E}(A - A^N) = 0 \) and \( \mathbb{E}B = \mathbb{E}B^N \). Hence
\[ \mathbb{E}\left|\frac{1}{B} - \frac{1}{B^N}\right| = \mathbb{E}\left|\frac{B - \mathbb{E}B + \mathbb{E}B^N - B^N}{BB^N}\right| \]
\[ \leq \frac{1}{\text{Im}^2(\eta)}\left(\sqrt{\text{var}(B)} + \sqrt{\text{var}(B^N)}\right) = O_\eta\left(n^{-1/2}\right). \]
As a consequence, we obtain the following estimate \( \mathbb{E}T_1 = O_\eta\left(n^{-1/2}\right) \). Plugging this into (B.8) and proceeding similarly for the other terms \( T_i \), we finally get
\[ \frac{1}{n}\mathbb{E}\text{Tr} \left(G - \tilde{G}^N\right) = O_\eta\left(\frac{1}{\sqrt{n}}\right), \]
and (B.6) is proved.

Proof of (B.7). Recall that
\[ F(\eta, z) = (Y - z)\left[(Y - z)^*(Y - z) - \eta^2\right]^{-1}, \]
\[ F^N(\eta, z) = (Y^N - z)\left[(Y^N - z)^*(Y^N - z) - \eta^2\right]^{-1}. \]
Notice first that
\[ \frac{1}{n}\mathbb{E}\text{Tr} F(\eta, z) - \frac{1}{n}\mathbb{E}\text{Tr} F^N(\eta, z) \]
\[ = \frac{1}{n}\mathbb{E}\text{Tr} Y \left[(Y - z)^*(Y - z) - \eta^2\right]^{-1} - \frac{1}{n}\mathbb{E}\text{Tr} Y^N \left[(Y^N - z)^*(Y^N - z) - \eta^2\right]^{-1} \]
\[ = \frac{1}{\eta^2 n}\mathbb{E}\text{Tr} (\tilde{G} - \tilde{G}^N). \]
The last part of the r.h.s. can be handled as (B.6) - we omit it. We focus now of the first part of the r.h.s. and work on the columns of \( Y^* \). We introduce the notations \( Y^* = [\vec{\xi}_1, \cdots, \vec{\xi}_n] \) and its
Gaussian counterparts \((Y^N)^* = [\tilde{\xi}_1^N, \ldots, \tilde{\xi}_n^N]\). Write
\[
\frac{1}{n} \text{Etr } Y \left[ (Y - z)^*(Y - z) - \eta^2 \right]^{-1} - \frac{1}{n} \text{Etr } Y^N \left[ (Y^N - z)^*(Y^N - z) - \eta^2 \right]^{-1} = \frac{1}{n} \sum_{i=0}^{n-1} \left\{ \text{Etr } Y_i \left[ (Y_i - z)^*(Y_i - z) - \eta^2 \right]^{-1} - \text{Etr } Y_{i+1} \left[ (Y_{i+1} - z)^*(Y_{i+1} - z) - \eta^2 \right]^{-1} \right\}
\]
where
\[Y_i^* = \left( \tilde{\xi}_1^N, \ldots, \tilde{\xi}_i^N, \tilde{\xi}_{i+1}^N, \ldots, \tilde{\xi}_n^N \right), \quad Y_0^* = Y^*, \quad Y_n^* = (Y^N)^*.\]
We only handle the first term of the sum in \((B.9)\) and will prove that
\[
\Delta_1 := \text{Etr } Y \left[ (Y - z)^*(Y - z) - \eta^2 \right]^{-1} - \text{Etr } Y_1 \left[ (Y_1 - z)^*(Y_1 - z) - \eta^2 \right]^{-1},
\]
\[
= \mathcal{O}_g \left( \frac{1}{\sqrt{n}} \right). \tag{B.10}
\]
We rely again on Sherman–Morrison’s formula:
\[
\left[ (Y - z)^*(Y - z) - \eta^2 \right]^{-1} = Q - \frac{Q(\tilde{\xi}_1 - z\tilde{e}_1)(\tilde{\xi}_1 - z\tilde{e}_1)^*Q}{1 + (\tilde{\xi}_1 - z\tilde{e}_1)^*Q(\tilde{\xi}_1 - z\tilde{e}_1)}
\]
where
\[
Q = \left[ \sum_{i=2}^{n} (\tilde{\xi}_i - z\tilde{e}_1)(\tilde{\xi}_i - z\tilde{e}_1)^* - \eta^2 \right]^{-1}.
\]
Notice that \(\text{Etr } (YQ - Y_1 Q) = 0\) hence
\[
\Delta_1 = -\text{Etr } Y Q(\tilde{\xi}_1 - z\tilde{e}_1)(\tilde{\xi}_1 - z\tilde{e}_1)^*Q + \text{Etr } Y_1 Q(\tilde{\xi}_1^N - z\tilde{e}_1)(\tilde{\xi}_1^N - z\tilde{e}_1)^*Q
\]
\[
1 + (\tilde{\xi}_1 - z\tilde{e}_1)^*Q(\tilde{\xi}_1 - z\tilde{e}_1) 1 + (\tilde{\xi}_1^N - z\tilde{e}_1)^*Q(\tilde{\xi}_1^N - z\tilde{e}_1)
\]
We introduce the following notations
\[
\Upsilon := \begin{pmatrix} \tilde{\xi}_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad Y_{2:n} := \begin{pmatrix} 0 \\ \tilde{\xi}_2 \\ \vdots \\ \tilde{\xi}_n \end{pmatrix}, \quad Y^N := \begin{pmatrix} (\xi^N_1)^* \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\]
and write
\[
Y = \Upsilon + Y_{2:n} \quad \text{and} \quad Y_1 = \Upsilon^N + Y_{2:n}.
\]
Hence
\[
\Delta_1 = -\text{Etr } \Upsilon Q(\tilde{\xi}_1 - z\tilde{e}_1)(\tilde{\xi}_1 - z\tilde{e}_1)^*Q 1 + (\tilde{\xi}_1 - z\tilde{e}_1)^*Q(\tilde{\xi}_1 - z\tilde{e}_1) + \text{Etr } Y^N Q(\tilde{\xi}_1^N - z\tilde{e}_1)(\tilde{\xi}_1^N - z\tilde{e}_1)^*Q 1 + (\tilde{\xi}_1^N - z\tilde{e}_1)^*Q(\tilde{\xi}_1^N - z\tilde{e}_1)
\]
\[
-\text{Etr } Y_{2:n} Q(\tilde{\xi}_1 - z\tilde{e}_1)(\tilde{\xi}_1 - z\tilde{e}_1)^*Q 1 + (\tilde{\xi}_1 - z\tilde{e}_1)^*Q(\tilde{\xi}_1 - z\tilde{e}_1) + \text{Etr } Y_{2:n} Q(\tilde{\xi}_1^N - z\tilde{e}_1)(\tilde{\xi}_1^N - z\tilde{e}_1)^*Q 1 + (\tilde{\xi}_1^N - z\tilde{e}_1)^*Q(\tilde{\xi}_1^N - z\tilde{e}_1),
\]
\[
:= \Delta_2 + \Delta_3.
\]
We first focus on the term \(\Delta_2\).
\[
\Delta_2 = -E \frac{\tilde{\xi}_1 Q(\tilde{\xi}_1 - z\tilde{e}_1)(\tilde{\xi}_1 - z\tilde{e}_1)^*Q}{1 + (\tilde{\xi}_1 - z\tilde{e}_1)^*Q(\tilde{\xi}_1 - z\tilde{e}_1)} + E \frac{(\xi^N_1)^* Q(\tilde{\xi}_1^N - z\tilde{e}_1)(\tilde{\xi}_1^N - z\tilde{e}_1)^*Q}{1 + (\tilde{\xi}_1^N - z\tilde{e}_1)^*Q(\tilde{\xi}_1^N - z\tilde{e}_1)}
\]
\[
:= -E \left( \frac{A}{B} \right) + E \left( \frac{A^N}{B^N} \right).
\]
At this point, we argue as previously. Notice however that the variance of $A$ and $A^N$ is of order $O_{\eta}(1)$. Since the variance of $B^{-1}$ and $(B^N)^{-1}$ is of order $O_{\eta}(n^{-1})$, see for instance Corollary B.3, we obtain:

$$\left| \mathbb{E}\left( \frac{A^{(N)}}{B} \right) - \mathbb{E}A^{(N)}\mathbb{E}(B^{-1}) \right| \leq \sqrt{\text{var}(A^{(N)})} \sqrt{\text{var}(B^{-1})} = O_{\eta}\left( \frac{1}{\sqrt{n}} \right)$$

and

$$\Delta_2 = -\mathbb{E}\left( \frac{A}{B} - \frac{A^N}{B} + \frac{A^N}{B^N} \right),$$

$$= -\mathbb{E}\left( \frac{1}{B} (A - A^N) \right) - \mathbb{E}A^N\left( \frac{1}{B} - \frac{1}{B^N} \right),$$

$$= -\mathbb{E}\left( \frac{1}{B} \right) \mathbb{E}(A - A^N) - \mathbb{E}A^N\left( \frac{1}{B} - \frac{1}{B^N} \right) + O_{\eta}\left( \frac{1}{\sqrt{n}} \right),$$

$$= O_{\eta}(\mathbb{E}(A - A^N)) + O_{\eta}\left( \frac{1}{\sqrt{n}} \right).$$

It remains to compute $\mathbb{E}(A - A^N)$. We denote by $\delta_{ij}$ the Kronecker symbol (with value 1 if $i = j$ and 0 else).

$$\mathbb{E}(A - A^N) = \mathbb{E} \sum_{i=1}^{n} \bar{\xi}_i \left[ Q(\bar{\xi}_i - z\bar{\epsilon}_1)(\bar{\xi}_i - \bar{z}\epsilon_1)^*Q \right]_{ij} - \mathbb{E} \sum_{i=1}^{n} \bar{\xi}_i^N \left[ Q(\bar{\xi}_i^N - z\bar{\epsilon}_1)(\bar{\xi}_i^N - \bar{z}\epsilon_1)^*Q \right]_{ij},$$

$$= \mathbb{E} \sum_{i,j,k=1}^{n} \bar{\xi}_{i1} Q_{ij} (\xi_{j1} - z\delta_{1j}) (\xi_{k1}^N - z\delta_{1k}) Q_{k1} - \mathbb{E} \sum_{i,j,k=1}^{n} \bar{\xi}_{i1}^N Q_{ij} (\xi_{j1}^N - \bar{z}\delta_{1j})(\xi_{k1}^N - \bar{z}\delta_{1k}) Q_{k1}.$$

We first go through all the terms featuring the variables $\xi_{11}$ and $\xi_{11}^N$ and end up with seven different summations:

$$T_1^{(1)} = \sum_{i=1, j,k \geq 2}^{n}, \quad T_2^{(1)} = \sum_{j,k \geq 2}^{n}, \quad T_3^{(1)} = \sum_{k=1}^{n}, \quad T_4^{(1)} = \sum_{i,j=1, k \geq 2}^{n}, \quad T_5^{(1)} = \sum_{j,k=1, i \geq 2}^{n},$$

$$T_6^{(1)} = \sum_{i,k=1, j \geq 2}^{n}, \quad T_7^{(1)} = \sum_{i,j,k=1}^{n}.$$

To be more specific, the term $T_1^{(1)}$ above is a shortcut for

$$T_1^{(1)} = \sum_{i=1, j,k \geq 2}^{n} (\xi_{i1} Q_{1j} \bar{\xi}_{j1} \bar{\xi}_{k1} Q_{k1} - \bar{\xi}_{i1}^N Q_{1j} \xi_{j1}^N \bar{\xi}_{k1}^N Q_{k1})$$

and so on, while $T_7^{(1)} = \xi_{i1} Q_{11} |\xi_{i1} - z|^2 - \bar{\xi}_{i1}^N Q_{11} |\xi_{i1}^N - \bar{z}|^2$. One can readily prove that

$$\mathbb{E}T_1^{(1)} = \mathbb{E}T_3^{(1)} = \mathbb{E}T_4^{(1)} = \mathbb{E}T_5^{(1)} = \mathbb{E}T_6^{(1)} = 0, \quad \mathbb{E}T_7^{(1)} = O_{\eta}\left( \frac{1}{n^{3/2}} \right)$$

because $\mathbb{E}\xi_{i1} = \mathbb{E}\xi_{i1}^N = 0$ and $\mathbb{E}|\xi_{i1}|^2 = \mathbb{E}|\xi_{i1}^N|^2 = O(n^{-1})$. Notice however that the second non-absolute moment of $\xi_{11}$ does not match a priori with the one of $\xi_{11}^N$: $\mathbb{E}|\xi_{11}|^2 \neq \mathbb{E}|\xi_{11}^N|^2 = 0$; in

---

\[\text{Footnote:} \] The exact computation is lengthy and thus omitted but an heuristic argument is the following: simply replace vector $[Q(\xi_1 - z\epsilon_1)(\xi_1 - \bar{z}\epsilon_1)^*Q]_{11}$ by a deterministic vector $\bar{u}$ with bounded norm, then the variance of $\xi_{11}^*\bar{u}$ is clearly of order $O(1)$. 

particular, the variable $\xi_{11}$ can be real. Hence the term $T_{2}^{(1)}$ is written

$$E T_{2}^{(1)} = -zE \sum_{i=1}^{n} \xi_{i1}^{2} Q_{i1} Q_{i1} = O_{\eta} \left( \frac{\|X_{1}^{2}\|_{\infty}}{n} \right).$$

We now repeat the same strategy and collect all the remaining terms featuring variables $\xi_{21}$ and $\xi_{21}^{N}$:

$$T_{1}^{(2)} = \sum_{i=2, j=2, k=2}^{\infty}, \ T_{2}^{(2)} = \sum_{i=2, j=2, k=2}^{\infty}, \ T_{3}^{(2)} = \sum_{i,j,k=2}^{\infty}, \ T_{4}^{(2)} = \sum_{i,j,k=2}^{\infty}, \ T_{5}^{(2)} = \sum_{i,j,k=2}^{\infty},$$

and prove the following

$$E T_{\ell}^{(2)} = 0 \text{ for } 1 \leq \ell \leq 6, \ E T_{7}^{(2)} = O_{\eta} \left( \frac{1}{n^{3/2}} \right).$$

Similarly, one can prove that

$$E T_{\ell}^{(m)} = 0 \text{ for } 1 \leq \ell \leq 6, \ E T_{7}^{(m)} = O_{\eta} \left( \frac{1}{n^{3/2}} \right)$$

holds true for $3 \leq m \leq n$. Gathering all our estimates, we obtain

$$E(A - A^{\wedge}) = \sum_{m=1}^{n} E T_{2}^{(1)} + \sum_{m=1}^{n} E T_{7}^{(m)} = O_{\eta} \left( \frac{1}{n} \right) + O_{\eta} \left( \frac{1}{\sqrt{n}} \right) = O_{\eta} \left( \frac{1}{\sqrt{n}} \right).$$

This yields an estimate for $\Delta_{2}$.

We now briefly explain how to handle the term

$$\Delta_{3} = -E r_{Y_{2,n}} Q(\xi_{1} - z\bar{e}_{1})(\xi_{1} - z\bar{e}_{1})^{*}Q \frac{1}{1 + (\xi_{1} - z\bar{e}_{1})^{*}Q(\xi_{1} - z\bar{e}_{1})}. \ E r_{Y_{2,n}} Q(\xi_{1}^{N} - z\bar{e}_{1})(\xi_{1}^{N} - z\bar{e}_{1})^{*}Q \frac{1}{1 + (\xi_{1}^{N} - z\bar{e}_{1})^{*}Q(\xi_{1}^{N} - z\bar{e}_{1})}.$$  

We mainly need three extra arguments.

Denote by $Y_{2,n}^{\wedge}$ the truncated matrix with entries $[Y_{2,n}^{\wedge}]_{ij} = \frac{\sigma_{i,j}^{2}}{\sqrt{n}} X_{i,j} 1_{\{X_{i,j} \leq \sqrt{n}\delta_{n}\}}$, where $\delta_{n} = n^{-a}, a > 0$ depending on $\varepsilon$ in A0. Applying Latała’s inequality [48], and for an appropriate choice of $a > 0$, we easily obtain

$$E\|Y_{2,n} - Y_{2,n}^{\wedge}\| = O \left( \frac{1}{\sqrt{n}} \right). \quad \text{(B.11)}$$

By [14, Theorem 5.8],

$$P\{\|Y_{2,n}^{\wedge}\|^{2} > 2\sigma_{\max}^{2} + x\} \leq \frac{K(m)}{n^{m}} \quad \text{(B.12)}$$

for any $x > 0$ and any integer $m \geq 1$.

Finally, we will need the following estimate, whose proof is postponed shortly after.

$$\left| \frac{(\xi_{1} - z\bar{e}_{1})^{*}Q(\xi_{1} - z\bar{e}_{1})}{1 + (\xi_{1} - z\bar{e}_{1})^{*}Q(\xi_{1} - z\bar{e}_{1})} \right| \leq \frac{1}{\text{Im}^{2}(\eta)} \quad \text{(B.13)}$$

(notice that the previous inequality holds without the expectation). Denote by

$$B = \{\|Y_{2,n}^{\wedge}\|^{2} \leq 2\sigma_{\max}^{2} + x\} \quad \text{and} \quad B^{c} = \{\|Y_{2,n}^{\wedge}\|^{2} > 2\sigma_{\max}^{2} + x\}, \quad x > 0.$$
We are now in position to prove that $\Delta_3 = O_\eta (n^{-1/2})$. We have

$$\Delta_3 \| \leq \left| \text{E} \text{btr} Y_{2,n}^\wedge \left\{ \frac{Q(\xi_1 - z e_1)(\xi_1 - z e_1)^* Q}{1 + (\xi_1 - z e_1)^* Q(\xi_1 - z e_1)} - \frac{Q(\xi_1^N - z e_1)(\xi_1^N - z e_1)^* Q}{1 + (\xi_1^N - z e_1)^* Q(\xi_1^N - z e_1)} \right\} \right|$$

$$+ \left| \text{E} \text{btr} Y_{2,n}^\wedge \left\{ \frac{Q(\xi_1 - z e_1)(\xi_1 - z e_1)^* Q}{1 + (\xi_1 - z e_1)^* Q(\xi_1 - z e_1)} - \frac{Q(\xi_1^N - z e_1)(\xi_1^N - z e_1)^* Q}{1 + (\xi_1^N - z e_1)^* Q(\xi_1^N - z e_1)} \right\} \right|$$

$$+ \text{E}||Y_{2,n} - Y_{2,n}^\wedge|| \left| \frac{Q(\xi_1 - z e_1)(\xi_1 - z e_1)^* Q}{1 + (\xi_1 - z e_1)^* Q(\xi_1 - z e_1)} + \frac{Q(\xi_1^N - z e_1)(\xi_1^N - z e_1)^* Q}{1 + (\xi_1^N - z e_1)^* Q(\xi_1^N - z e_1)} \right| .$$

Hence, by applying (B.13), we get

$$\Delta_3 \| \leq \left| \text{E} \text{btr} Y_{2,n}^\wedge \left\{ \frac{Q(\xi_1 - z e_1)(\xi_1 - z e_1)^* Q}{1 + (\xi_1 - z e_1)^* Q(\xi_1 - z e_1)} - \frac{Q(\xi_1^N - z e_1)(\xi_1^N - z e_1)^* Q}{1 + (\xi_1^N - z e_1)^* Q(\xi_1^N - z e_1)} \right\} \right|$$

$$+ \left| \frac{2}{\text{Im}^2(\eta)} \text{E}||Y_{2,n}^\wedge|| + \frac{2}{\text{Im}^2(\eta)} \text{E}||Y_{2,n} - Y_{2,n}^\wedge|| \right| .$$

The first term on the r.h.s. can be handled as (B.6) with the property that $\text{btr}||Y_{2,n}^\wedge||$ is bounded; the second term is an $O_\eta (n^{-1/2})$ with the help of (B.12), and the third term is estimated with the help of (B.11).

Gathering the estimates for $\Delta_2$ and $\Delta_3$, we obtain the corresponding estimate (B.10) for $\Delta_1$, and finally for $\frac{1}{n} \text{btr} (F(\eta, z) - F(\eta, z))$ via (B.9). Convergence (B.7) is proved.

**Proof of (B.13).** Recall the definition of $R(\eta)$ in (4.1) and consider an eigen-decomposition of $R(\eta)$:

$$R(\eta) = \sum_{i=1}^{2n} \frac{u_i u_i^*}{\lambda_i R - \eta} ,$$

where the $u_i$'s are $R$'s eigenvectors and the $\lambda_i$'s, its eigenvalues. For a given $n \times 1$ vector $a$, denote by $\tilde{a}_0$ the $2n \times 1$ vector whose $n$ first components are those of $a$, the others being equal to zero. Notice that $G(\eta) = \eta Q(\eta^2)$. We have

$$\text{Im}(\eta + a^* \eta Q(\eta^2) a) = \text{Im}(\eta + \tilde{a}_0^* R(\eta) \tilde{a}_0)$$

$$\overset{(a)}{=} \text{Im}(\eta) + \text{Im}(\eta) \tilde{a}_0^* R(\eta) R^*(\eta) \tilde{a}_0$$

$$\overset{(b)}{\geq} \text{Im}(\eta) a^* G(\eta) G^*(\eta) a$$

$$\geq \text{Im}(\eta) |\eta|^2 |a^* Q(\eta^2) Q(\eta^2) a|$$

where $(a)$ follows from the eigen-decomposition of $R$, and $(b)$ from the fact that

$$\tilde{a}_0^* R R^* \tilde{a}_0 = a^* G G^* a + a^* F(F')^* a \geq a^* G G^* a .$$

Hence the final estimate.

**B.4. Proof of Proposition 4.8.** The proof of Proposition 4.8 will rely on the following two results.

**Proposition B.4.** Let $C \succeq 0$ be a $n \times n$ matrix and $u = (u_\ell) \succ 0$ and $v = (v_\ell) \succ 0$ two $n \times 1$ vector. Assume that the following equality holds true:

$$u = C u + v .$$

Then $\rho(C) < 1$, matrix $I - C$ is invertible, $(I - C)^{-1} \succ 0$ and

$$\|(I - C)^{-1}\|_\infty \leq \frac{\max(u_\ell; \ \ell \in [n])}{\min(v_\ell; \ \ell \in [n])} .$$
Moreover, $\det(I - C) > 0$.

Proof. We only prove the last property, for the other properties’ proof can be found in [40, Lemma 5.2]. Let $t \in [0,1]$ and notice that $\rho(tC) = tp(C) < 1$, hence $I - tC$ is invertible for all $t \in [0,1]$. Since this mapping $t \mapsto \det(I - tC)$ is continuous for $t \in [0,1]$ with value 1 at $t = 0$, its sign must remain constant and positive for all $t \in [0,1]$, for otherwise $\det(I - t_0C) = 0$ for some $t_0 \in (0,1)$, which contradicts the invertibility of $I - t_0C$. \hfill $\square$

**Lemma B.5.** Let $A$ and $B$ be two $n \times n$ matrices such that $\rho(A \odot \bar{A}) < 1$ and $\rho(B \odot \bar{B}) < 1$. Let $u, v, w, r$ be $n \times 1$ vectors, then

1. matrix $I - A \odot B$ is invertible,
2. the following inequality holds true:
   \[
   |(u \odot v)^T (I - A \odot B)^{-1} w \odot r| \leq \sqrt{(u \odot \bar{u})^T (I - A \odot \bar{A})^{-1} w \odot \bar{w}} \sqrt{(v \odot \bar{v})^T (I - B \odot \bar{B})^{-1} r \odot \bar{r}}.
   \]
3. the following max-row norm estimate holds true
   \[
   \|(I - A \odot B)^{-1}\|_\infty \leq \sqrt{\|(I - A \odot \bar{A})^{-1}\|_\infty} \sqrt{\|(I - B \odot \bar{B})^{-1}\|_\infty}. \tag{B.14}
   \]

The proof of Lemma B.5 is postponed to Section B.5.

In order to study the properties of matrix $\mathcal{A}(E\bar{\eta}) \odot \mathcal{A}(p)$, we introduce two auxiliary systems. Recall the definitions (4.14) and (4.18) of $\Upsilon(\bar{b})$ and $\Delta(\bar{b})$, $\bar{\Delta}(\bar{b})$.

Since the $p_i$’s satisfy (4.13), we immediately obtain

\[
\text{Im}(p_i) = \frac{[V^T \text{Im}(p)]_i |z|^2}{|z|^2 - (\eta + [V \bar{p}]_i)(\eta + [V^T p]_i)} + \frac{[V \text{Im}(\bar{p})]_i}{|z|^2 - (\eta + [V \bar{p}]_i)(\eta + [V^T p]_i) + \frac{|z|^2}{\eta + [V^T p]_i}^2}
+ \frac{\text{Im}(\eta)}{|z|^2 - (\eta + [V \bar{p}]_i)(\eta + [V^T p]_i) + \frac{|z|^2}{\eta + [V^T p]_i}^2} \left(\frac{|z|^2}{|\eta + [V^T p]_i|} + 1\right)
\]

and its counterpart for $\text{Im}(\bar{p}_i)$. Denote by $v(\bar{p})$ the $2n \times 1$ vector defined by

\[
[v(\bar{p})]_i = \frac{\text{Im}(\bar{\eta})}{|z|^2 - (\eta + [V \bar{p}]_i)(\eta + [V^T p]_i) + \frac{|z|^2}{\eta + [V^T p]_i}^2} \left(\frac{|z|^2}{|\eta + [V^T p]_i|} + 1\right)
\]

for $i \in [n]$ and

\[
[v(\bar{p})]_i = \frac{\text{Im}(\eta)}{|z|^2 - (\eta + [V \bar{p}]_i)(\eta + [V^T p]_i) + \frac{|z|^2}{\eta + [V^T p]_i}^2} \left(\frac{|z|^2}{|\eta + [V^T p]_i|} + 1\right)
\]

for $i \in \{n + 1, \ldots, 2n\}$. Then the system satisfied by $\text{Im}(\bar{p})$ writes

\[
\text{Im}(\bar{p}) = \mathcal{A}(\bar{p}) \odot \mathcal{A}(\bar{p}) \text{Im}(\bar{p}) + v(\bar{p}) \tag{B.15}
\]

where matrix $\mathcal{A}(\bar{p})$ has been defined in (4.19). Since matrix $\mathcal{A}(\bar{p}) \odot \mathcal{A}(\bar{p})$ has nonnegative entries, we will rely on Proposition B.4 to evaluate

\[
\|(I - \mathcal{A}(\bar{p}) \odot \mathcal{A}(\bar{p}))^{-1}\|_\infty.
\]
We need to check that \( \Im(p_\ell), v(\tilde{p}) \succ 0 \), to upper bound \( \Im(p_\ell), \Im(\tilde{p}_\ell) \) and to lower bound \( \left[v(\tilde{\theta})\right]_i \).

Since \( p_\ell \) and \( \tilde{p}_\ell \) are Stieltjes transform, we have
\[
|\Im(p_\ell)| \vee |\Im(\tilde{p}_\ell)| \leq \frac{1}{\Im(\eta)}.
\]

Now, if \( i \leq n \),
\[
[v(\tilde{p})]_i \geq \frac{\Im(\eta)}{|-(\eta + [V \tilde{p}]_i) + \frac{|z|^2}{\eta + [V^* p]_i}|^2}
\]

with
\[
\frac{1}{\Im(\eta)}\left[-(\eta + [V \tilde{p}]_i) + \frac{|z|^2}{\eta + [V^* p]_i}\right]^2 \leq 2 \left( \frac{|\eta|^2}{\Im(\eta)} + \frac{\sigma^4}{\Im^3(\eta)} + \frac{|z|^4}{\Im^3(\eta)} \right) = O_\eta(1).
\]

The case where \( n + 1 \leq i \leq 2n \) being handled similarly, we finally get
\[
\min_i \left[v(\tilde{\theta})\right]_i \geq \frac{1}{O_\eta(1)}.
\]  

Hence, \( v(\tilde{\theta}) \succ 0 \) and \( \left[v(\tilde{\theta})\right]_i \) is lower bounded away from zero.

In order to prove \( \tilde{p} \succ 0 \), we argue as follows: the \( p_i \)'s are Stieltjes transforms of probability measures \( \mu_i \). These probability measures are tight, see for instance Proposition 4.1-(v). In particular, there exists a real number \( K \) such that
\[
\mu_i([-K, K]) \geq \frac{1}{2}.
\]

Hence,
\[
\Im(p_i) = \Im(\eta) \int_{\mathbb{R}} \frac{\mu_i(d\lambda)}{|\lambda - \eta|^2} \geq \Im(\eta) \int_{-K}^{K} \frac{\mu_i(d\lambda)}{2(K^2 + |\eta|^2)} \geq \frac{1}{4(K^2 + |\eta|^2)}.
\]

We are now in position to apply Proposition B.4. This proposition yields in particular that \( \rho(A(\tilde{p}) \odot \overline{A(\tilde{p})}) < 1 \) and gathering estimates (B.16) and (B.17), we obtain
\[
\left\| (I - A(\tilde{p}) \odot \overline{A(\tilde{p})})^{-1} \right\|_\infty = O_\eta(1).
\]  

If one considers now the perturbed system (4.11)-(4.12) satisfied by \( \tilde{g} \), one obtains similarly
\[
\Im(\tilde{g}) = A(\tilde{g}) \odot \overline{A(\tilde{g})} \Im(\tilde{g}) + v(\tilde{g}) + O_\eta\left( \frac{1}{n^{3/2}} \right).
\]

By arguing as before (notice in particular that there is no impact of the residual term \( O_\eta(n^{-3/2}) \)), we obtain
\[
\rho\left(A(\tilde{g}) \odot \overline{A(\tilde{g})}\right) < 1 \quad \text{and} \quad \left\| (I - A(\tilde{g}) \odot \overline{A(\tilde{g})})^{-1} \right\|_\infty = O_\eta(1).
\]

It now remains to bound
\[
\left\| (I - A(\tilde{p}) \odot A(\tilde{g}))^{-1} \right\|_\infty = O_\eta(1),
\]
given estimates (B.18) and (B.20). Lemma B.5-(3) provides the appropriate estimate.

Consider now matrices \( A(\tilde{p}) \) and \( A(\tilde{g}) \) as defined in (4.19). We have already proved that \( \rho(A(\tilde{p}) \odot A(\tilde{g})) < 1 \) and \( \rho(A(\tilde{g}) \odot A(\tilde{q})) < 1 \) hence \( I - A(\tilde{p}) \odot A(\tilde{g}) \) is invertible. Plugging estimates (B.18) and (B.20) into (B.14), we obtain:
\[
\left\| (I - A(\tilde{p}) \odot A(\tilde{g}))^{-1} \right\|_\infty = O_\eta(1),
\]
which is the desired result. Proposition 4.8 is proved.
The induction assumption, its verification for \( n = 1 \) and some notations. The induction assumption is the following: let \( A \) and \( B \) be \( n \times n \) matrices with \( \rho(A \odot \tilde{A}) < 1 \) and \( \rho(B \odot \tilde{B}) < 1 \), then \( \det(I - A \odot B) \neq 0 \) (or equivalently \( I - A \odot B \) is invertible) and the inequality of the lemma holds true.

We first verify the induction assumption for \( n = 1 \). In this case, the matrices are scalar.

\[
A = a, \ B = b, \ A \odot B = ab, \ A \odot \tilde{A} = |a|^2, \ B \odot \tilde{B} = |b|^2, \ |a| < 1, \ |b| < 1. \]

Then \( \det(I - A \odot B) = 1 - ab \neq 0 \) and

\[
\frac{|uvwr|}{1 - ab} \leq \sqrt{\frac{|uw|^2}{1 - |a|^2}} \sqrt{\frac{|vs|^2}{1 - |b|^2}},
\]

which is the desired inequality.

We now assume the induction assumption at step \( n \) and prove it at step \( n + 1 \). We first introduce some notations. The tilded quantities refer to step \( n + 1 \) (either \((n + 1) \times (n + 1)\) matrices or \((n + 1) \times 1\) vectors). The untilded quantities refer to their counterparts at step \( n \). Plain lowercase letters are scalars.

\[
\tilde{A} = \begin{pmatrix} A & a \\ \alpha^T & a \end{pmatrix}, \ \tilde{B} = \begin{pmatrix} B & b \\ \beta^T & b \end{pmatrix}, \ \tilde{u} = \begin{pmatrix} u \\ u \end{pmatrix}, \ \tilde{v} = \begin{pmatrix} v \\ v \end{pmatrix}, \ \tilde{w} = \begin{pmatrix} w \\ w \end{pmatrix}, \ \tilde{s} = \begin{pmatrix} s \\ s \end{pmatrix}.
\]

We will denote by

\[
\Gamma_{AB} = (I - A \odot B)^{-1}, \ \Gamma_A = (I - A \odot \tilde{A})^{-1}, \ \Gamma_B = (I - B \odot \tilde{B})^{-1}
\]

and consider the following notation for quadratic forms:

\[
Q(u, A, w) = (u \odot \tilde{u})^T \Gamma_A \tilde{w} \odot \tilde{w}, \ Q(v, B, s) = (v \odot \tilde{v})^T \Gamma_B \tilde{s} \odot \tilde{s}, \quad (B.21)
\]

where \( u, v, w \) and \( s \) are generic \( n \times 1 \) vectors.

*Invertibility of \( I - \tilde{A} \odot \tilde{B} \).* By assumption, \( \rho(\tilde{A} \odot \tilde{A}) < 1 \) and \( \rho(\tilde{B} \odot \tilde{B}) < 1 \). Since \( A \odot \tilde{A} \) is a principal submatrix of \( \tilde{A} \odot \tilde{A} \gg 0 \), we have

\[
\rho(A \odot \tilde{A}) \leq \rho(\tilde{A} \odot \tilde{A}) < 1
\]

and similarly, \( \rho(B \odot \tilde{B}) < 1 \). By the induction assumption, \( I - A \odot B \) is invertible. In particular, \( \det(I - A \odot B) \neq 0 \).

Consider now the following decomposition

\[
I - \tilde{A} \odot \tilde{B} = \begin{pmatrix} I - A \odot B & -a \odot b \\ -\alpha \odot \beta & 1 - ab \end{pmatrix}.
\]

The Schur complement of matrix \( I - A \odot B \) in \( I - \tilde{A} \odot \tilde{B} \) writes:

\[
S_{AB} = 1 - ab - (\alpha \odot \beta)^T \Gamma_{AB} a \odot b.
\]

We similarly consider the Schur complement \( S_A \) (resp. \( S_B \)) of matrix \( I - A \odot \tilde{A} \) (resp. matrix \( I - B \odot \tilde{B} \)) in \( I - \tilde{A} \odot \tilde{A} \) (resp. \( I - \tilde{B} \odot \tilde{B} \)):

\[
S_A = 1 - |a|^2 - (\alpha \odot \alpha)^T \Gamma_A a \odot a,
S_B = 1 - |b|^2 - (\beta \odot \beta)^T \Gamma_B b \odot b.
\]
By Proposition B.4, the determinants of $I - \tilde{A} \circ \tilde{A}$ and $I - A \circ \tilde{A}$ are positive. By the determinantal formula [45, Section 0.8.5] involving $S_A$, we obtain:
\[
\det(I - \tilde{A} \circ \tilde{A}) = \det(I - A \circ \tilde{A}) \times S_A .
\]
Hence $S_A > 0$; similarly, $S_B > 0$.

We now prove that $|S_{AB}| \neq 0$. Applying the induction assumption, we get
\[
|S_{AB}| \geq 1 - ab - \sqrt{Q(\alpha, A, a)\sqrt{Q(\beta, B, b)}} \geq \sqrt{1 - |a|^2\sqrt{1 - |b|^2} - \sqrt{Q(\alpha, A, a)\sqrt{Q(\beta, B, b)}}} \geq \sqrt{S_A S_B} > 0 , \tag{B.22}
\]
where the last inequality follows from the elementary inequality $\sqrt{ab} - \sqrt{cd} \geq \sqrt{a-c}\sqrt{b-d}$, valid for $(a-c) \wedge (b-d) \geq 0$.

Using again the determinantal formula, we get
\[
\det(I - \tilde{A} \circ \tilde{B}) = \det(I - A \circ B) \times S_{AB} \neq 0 .
\]
Hence $I - \tilde{A} \circ \tilde{B}$ is invertible.

The inequality at step $n+1$. Using the Schur decomposition
\[
(I - \tilde{A} \circ \tilde{B})^{-1} = \begin{pmatrix} I & \Gamma_{AB} a \circ b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Gamma_{AB} & 0 \\ 0 & S_{AB}^{-1} \end{pmatrix} \begin{pmatrix} (\alpha \circ \beta)^T \Gamma_{AB} & 0 \\ 0 & 1 \end{pmatrix} ,
\]
we obtain
\[
(u \circ \tilde{v})^T (I - \tilde{A} \circ \tilde{B})^{-1} \tilde{w} \circ \tilde{s} = (u \circ v)^T \Gamma_{AB} w \circ s + \frac{(u \circ v)^T \Gamma_{AB} a \circ b \times (\alpha \circ \beta)^T \Gamma_{AB} w \circ s}{S_{AB}} + \frac{uv \alpha \circ \beta}{S_{AB}} + \frac{uvws}{S_{AB}} . \tag{B.23}
\]
A similar computation yields (recall notation (B.21))
\[
(u \circ \tilde{u})^T (I - \tilde{A} \circ \tilde{A})^{-1} \tilde{w} \circ \tilde{w} = Q(u, A, w) + \frac{Q(u, A, a)Q(\alpha, A, w)}{S_A} + |u|^2 \frac{Q(\alpha, A, w)}{S_A} + |w|^2 \frac{Q(u, A, a)}{S_A} + |uw|^2 . \tag{B.24}
\]
Using the induction assumption at step $n$ and the inequality (B.22) over the Schur complements, we can majorize the r.h.s. of (B.23):
\[
\left| (u \circ \tilde{v})^T (I - \tilde{A} \circ \tilde{B})^{-1} \tilde{w} \circ \tilde{s} \right| \leq \sqrt{Q(u, A, w)\sqrt{Q(v, B, s)}} + \frac{\sqrt{Q(u, A, a)\sqrt{Q(v, B, b)\sqrt{Q(\alpha, A, w)\sqrt{Q(\beta, B, s)}}}}}{\sqrt{S_A S_B}}
\]
\[
+ |uv| \frac{\sqrt{Q(\alpha, A, w)\sqrt{Q(\beta, B, s)}}}{\sqrt{S_A S_B}} + \frac{|w|^2}{\sqrt{S_A S_B}} + \frac{|uw|^2}{\sqrt{S_A S_B}} .
\]
Using Cauchy-Schwarz inequality $\sum \sqrt{x_i y_i} \leq (\sum x_i)^{1/2}(\sum y_i)^{1/2}$ together with (B.24) and its counterpart for $(\tilde{v} \circ \tilde{v})^T (I - \tilde{B} \circ \tilde{B})^{-1} \tilde{s} \circ \tilde{s}$, we finally obtain the desired result at step $n+1$. Parts (1) and (2) are proved.
Part (3) is a simple corollary of the previous inequality. Notice that for a $n \times n$ matrix $C$,
\[
\|C\|_\infty = \max_i \left\{ \sum_j |C_{ij}| \right\} = \max_i \left\{ \sum_j C_{ij}x_j ; \|x\|_\infty \leq 1 \right\},
\]
\[
= \max_i \left\{ \sum_j C_{ij}x_jy_j ; \|x\|_\infty, \|y\|_\infty \leq 1 \right\}.
\]
Specializing $u = v = e_i, \|w\|_\infty \leq 1$ and $\|r\|_\infty \leq 1$ in the lemma, we obtain:
\[
\left| \sum_j \left[ (I - A \otimes B)^{-1} \right]_{ij} w_j r_j \right| \leq \sqrt{\sum_j \left[ (I - A \otimes \bar{A})^{-1} \right]_{ij} |w_j|^2 \sqrt{\sum_j \left[ (I - B \otimes B)^{-1} \right]_{ij} |r_j|^2}}
\]
for $i \in [n]$. Noticing that $\|w\|_\infty \leq 1$ implies that $\|w \otimes \bar{w}\|_\infty \leq 1$, it remains to optimize over $w$ and $r$ to conclude.

Lemma B.5 is proved.

Appendix C. Remaining proofs for Section 6

C.1. Proof of Lemma 6.4. Assume without loss of generality that $\zeta$ is a probability measure. Let $\varphi \in C_c(\mathbb{C})$. Since the convergence in probability induces the convergence in distribution, we have
\[
\mathbb{E} \varphi(f_n(\cdot,z) - g(z)) \xrightarrow{n \to \infty} \varphi(0)
\]
for $\zeta$-almost all $z \in \mathbb{C}$. Thus, by the dominated convergence and Fubini’s theorems,
\[
\int_{\Omega \times \mathbb{C}} \varphi(f_n(\omega,z) - g(z)) (P \otimes \zeta)(d\omega \times dz) \xrightarrow{n \to \infty} \varphi(0).
\]
In other words, $f_n - g$ converges to 0 in distribution, hence in probability, for the probability measure $P \otimes \zeta$. As a consequence (see for instance [47, Lemma 3.11]),
\[
\int_{\mathbb{C}} |g(z)|^{1+\alpha} \zeta(dz) = \int_{\Omega \times \mathbb{C}} |g(z)|^{1+\alpha} (P \otimes \zeta)(d\omega \times dz) \leq C.
\]
By (6.8), the sequence $(f_n)$ is $P \otimes \zeta$-uniformly integrable, hence
\[
\int_{\Omega \times \mathbb{C}} |f_n(\omega,z) - g(z)| (P \otimes \zeta)(d\omega \times dz) \xrightarrow{n \to \infty} 0,
\]
see for instance [47, Proposition 3.12]. Convergence (6.9) follows from Markov’s inequality.

C.2. Proof of Lemma 6.5. Let $(\psi_k)_{k \geq 1}$ be a sequence of smooth compactly supported functions, dense in $C_c(\mathbb{C})$ for the supremum norm $\|\psi\|_\infty = \sup_{z \in \mathbb{C}} |\psi(z)|$. By the diagonal extraction procedure, one can find a subsequence $(\zeta_{n'})$ such that with probability one $(\zeta_{n'})$ is tight and
\[
\int \psi_k d\zeta_{n'} \xrightarrow{n' \to \infty} -\frac{1}{2\pi} \int \Delta \psi_k(z) h(z) dz
\]
for all $k \geq 1$. Thus, on this set of probability one, the tight sequence $(\zeta_{n'})$ has a unique non-random limit point $\zeta$, and this limit point satisfies
\[
\zeta = -\frac{1}{2\pi} \Delta h
\]
in $D'(\mathbb{C})$, the set of Schwartz distributions. With this at hand, we get from the assumption that
\[
\int \psi_k(z) \zeta_n(dz) \xrightarrow{P_{n} \to \infty} \int \psi_k(z) \zeta(dz)
\]
for all \( k \geq 1 \). By a density argument, we thus get that

\[
\int \varphi(z) \zeta_n(dz) \xrightarrow{P} \int \varphi(z) \zeta(dz)
\]

for every \( \varphi \in C_c(\mathbb{C}) \).

REFERENCES

[1] Y. Ahmadian, F. Fumarola, and K. D. Miller. Properties of networks with partially structured and partially random connectivity. Physical Review E, 91(1):012820, 2015.

[2] O. Ajanki, L. Erdös, and T. Krüger. Universality for general wigner-type matrices. arXiv preprint arXiv:1506.05098, 2015.

[3] O. Ajanki, L. Erdos, and T. Krüger. Stability of the matrix dyson equation and random matrices with correlations. Preprint arxiv:1604.08188.

[4] O. Ajanki, L. Erdos, and T. Krüger. Quadratic vector equations on complex upper half-plane. ArXiv e-prints, June 2015.

[5] J. Aljadeff, D. Renfrew, and M. Stern. Eigenvalues of block structured asymmetric random matrices. J. Math. Phys., 56(10):103502, 14, 2015.

[6] J. Aljadeff, M. Stern, and T. Sharpee. Transition to chaos in random networks with cell-type-specific connectivity. Phys. Rev. Lett., 114:088101, Feb 2015.

[7] 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.

[8] S. Allesina and S. Tang. Stability criteria for complex ecosystems. Nature, 483(7388):205–208, 2012.

[9] S. Allesina and S. Tang. The stability–complexity relationship at age 40: a random matrix perspective. Population Ecology, 57(1):63–75, 2015.

[10] J. Alt, L. Erdős, and T. Krüger. Local law for random Gram matrices. arXiv preprint arXiv:1606.07353, June 2016.

[11] J. Alt, L. Erdős, and T. Krüger. Local inhomogeneous circular law. arXiv preprint arXiv:1612.07776, 2016.

[12] G. W. Anderson and O. Zeitouni. A CLT for a band matrix model. Probab. Theory Related Fields, 134(2):283–338, 2006.

[13] P. M. Anselone and T. W. Palmer. Spectral analysis of collectively compact, strongly convergent operator sequences. Pacific J. Math., 25:423–431, 1968.

[14] Z. Bai and J. W. Silverstein. Spectral analysis of large dimensional random matrices. Springer Series in Statistics. Springer, New York, second edition, 2010.

[15] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. Ann. Probab., 21(2):625–648, 1993.

[16] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. Ann. Probab., 21(2):649–672, 1993.

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

[18] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices. Random Matrices: Theory and Applications, 1(1), 2012.

[19] S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed and covariance random matrices. Comm. Math. Phys., 289(3):1023–1055, 2009.

[20] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of non-Hermitian heavy tailed random matrices. Comm. Math. Phys., 307(2):513–560, 2011.

[21] C. Bordenave, P. Caputo, D. Chafaï, and K. Tikhomirov. On the spectral radius of a random matrix. Preprint arXiv:1607.05484, 07 2016.

[22] C. Bordenave and D. Chafaï. Around the circular law. Probab. Surv., 9:1–89, 2012.

[23] P. Bourgade, H.-T. Yau, and J. Yin. Local circular law for random matrices. Probability Theory and Related Fields, 159(3):545–595, 2014.

[24] N. A. Cook. Lower bounds for the smallest singular value of structured random matrices. arXiv:1608.07347, 2016.

[25] N. A. Cook. Spectral properties of non-Hermitian random matrices. PhD thesis, University of California, Los Angeles, 2016.

[26] K. Deimling. Nonlinear functional analysis. Springer-Verlag, Berlin, 1985.

[27] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach. IEEE Trans. Inform. Theory, 56(3):1048–1069, 2010.

[28] L. Erdös, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. Comm. Math. Phys., 287(2):641–655, 2009.
L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner random matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

K. Fritzschke and H. Grauer. *From holomorphic functions to complex manifolds*, volume 213 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.

L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

K. Fritzschke and H. Grauer. *From holomorphic functions to complex manifolds*, volume 213 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.

L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner random matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

K. Fritzschke and H. Grauer. *From holomorphic functions to complex manifolds*, volume 213 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.

L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

K. Fritzschke and H. Grauer. *From holomorphic functions to complex manifolds*, volume 213 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.

L. Erdős, B. Schlein, and H.-T. Yau. Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.*, 37(3):815–852, 2009.

L. Erdős, H.-T. Yau, and J. Yin. Bulk universality for generalized Wigner random matrices. *Probab. Theory Related Fields*, 154(1-2):341–407, 2012.

J. Feinberg and A. Zee. Non-Hermitian random matrix theory: method of Hermitian reduction. *Nuclear Phys. B*, 504(3):579–608, 1997.

K. Fritzschke and H. Grauer. *From holomorphic functions to complex manifolds*, volume 213 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2002.

J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Mathematical Phys.*, 6:440–449, 1965.
[61] D. Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *Internat. Math. Res. Notices*, (20):1013–1025, 1996.

[62] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.

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

[64] T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields*, 155(1-2):231–263, 2013.

[65] T. Tao, V. Vu, and M. Krishnapur. Random matrices: universality of esds and the circular law. *The Annals of Probability*, 38(5):2023–2065, 2010.

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

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

[68] F. Wegner. Bounds on the density of states in disordered systems. *Z. Phys. B*, 44(1-2):9–15, 1981.

[69] J. Weidmann. *Linear operators in Hilbert spaces*, volume 68 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. Translated from the German by Joseph Szücs.

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

[71] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *arXiv preprint arXiv:1603.04066*, 2016.

**Nicholas Cook**  
Department of Mathematics  
Stanford University  
450 Serra Mall, Bldg. 380  
Stanford, CA 94305  
e-mail: nickcook@stanford.edu

**Walid Hachem**,  
LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay,  
46, rue Barrault,  
75634 Paris Cedex, France  
e-mail: walid.hachem@telecom-paristech.fr

**Jamal Najim**,  
Institut Gaspard Monge LabInfo, UMR 8049  
Université Paris Est Marne-la-Vallée  
5, Boulevard Descartes,  
Champs sur Marne, 77454 Marne-la-Vallée Cedex 2, France  
e-mail: najim@univ-mlv.fr

**David Renfrew**  
Department of Mathematical Sciences  
Binghamton University (SUNY)  
Binghamton, NY 3902-6000  
e-mail: renfrew@math.binghamton.edu