Annals of Mathematics **174** (2011), 1189–1217
http://dx.doi.org/10.4007/annals.2011.174.2.10

The single ring theorem

By Alice Guionnet, Manjunath Krishnapur, and Ofer Zeitouni

Abstract

We study the empirical measure L_{A_n} of the eigenvalues of nonnormal square matrices of the form $A_n = U_n T_n V_n$ with U_n, V_n independent Haar distributed on the unitary group and T_n real diagonal. We show that when the empirical measure of the eigenvalues of T_n converges, and T_n satisfies some technical conditions, L_{A_n} converges towards a rotationally invariant measure μ on the complex plane whose support is a single ring. In particular, we provide a complete proof of the Feinberg-Zee single ring theorem [6]. We also consider the case where U_n, V_n are independently Haar distributed on the orthogonal group.

1. The problem

Horn [16] asked the question of how to describe the eigenvalues of a square matrix with prescribed singular values. If A is a $n \times n$ matrix with singular values $s_1 \geq \cdots \geq s_n \geq 0$ and eigenvalues $\lambda_1, \ldots, \lambda_n$ in decreasing order of absolute values, then the inequalities

$$\prod_{j=1}^{k} |\lambda_j| \leq \prod_{j=1}^{k} s_j, \text{ if } k < n \qquad \text{and} \qquad \prod_{j=1}^{n} |\lambda_j| = \prod_{j=1}^{n} s_j$$

hold as shown by Weyl [29]. Horn [16] established that these were all the relationships between singular values and eigenvalues.

In this paper we study the natural probabilistic version of this problem and show that for "typical matrices", the singular values almost determine the eigenvalues. To frame the problem precisely, fix $s_1 \geq \cdots \geq s_n \geq 0$ and consider $n \times n$ matrices having these singular values. They are of the form $A = PTQ$, where T is diagonal with entries s_j on the diagonal, and P, Q are arbitrary unitary matrices.

We make A into a random matrix by choosing P and Q independently from the Haar measure on $\mathcal{U}(n)$, the unitary group of $n \times n$ matrices, and

A.G. was partially supported by the ANR project ANR-08-BLAN-0311-01. M.K. was supported, in part, by DST and UGC through DSA-SAP-PhaseIV. O.Z. was partially supported by NSF grant DMS-0804133 and by a grant from the Israel Science Foundation.

independent of T. Let $\lambda_1, \dots, \lambda_n$ be the (random) eigenvalues of A. The following natural questions arise.

1. Are there deterministic or random sets $\{s_j\}$, for which one can find the exact distribution of $\{\lambda_j\}$?
2. Let $L_S = \frac{1}{n}\sum_{j=1}^n \delta_{s_j}$ and $L_\Lambda = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j}$ denote the empirical measures of $S = \{s_j\}$ and $\Lambda = \{\lambda_j\}$. Suppose S_n are sets of size n such that L_{S_n} converges weakly to a probability measure θ supported on $\mathbb{R}_+$. Then, does L_Λ converge to a deterministic measure μ on the complex plane? If so, how is the measure μ determined by θ?
3. For finite n, for fixed S, is L_Λ *concentrated* in the space of probability measures on the plane?

In this paper, we concentrate on the second question and answer it in the affirmative, albeit with some restrictions. In this context, we note that Fyodorov and Wei [28, Th. 2.1] gave a formula for the mean eigenvalues density of A, yet in terms of a large sum which does not offer an easy handle on asymptotic properties (see also [7] for the case where T is a projection). The authors of [28] explicitly state the second question as an open problem.

Of course, questions 1–3 above are not new, and have been studied in various formulations. We now describe a partial and necessarily brief history of what is known concerning questions 1 and 2; partial results concerning question 3 will be discussed elsewhere.

The most famous case of a positive answer to question 1 is the *Ginibre ensemble*, see [8], and its asymmetric variant, see [18]. (There are some pitfalls in the standard derivation of Ginibre's result. We refer to [17] for a discussion.) Another situation is the truncation of random unitary matrices, described in [30].

Concerning question 2, the convergence of the empirical measure of eigenvalues in the Ginibre ensemble (and other ensembles related to question 1) is easy to deduce from the explicit formula for the joint distribution of eigenvalues. Generalizations of this convergence in the absence of such explicit formula, for matrices with iid entries, is covered under *Girko's circular law*, which is described in [9]; the circular law was proved under some conditions in [2] and finally, in full generality, in [10] and [24]. Such matrices, however, do not possess the invariance properties discussed in connection with question 2. The *single ring theorem* of Feinberg and Zee [6] is, to our knowledge, the first example where a partial answer to this question is offered. (Various issues of convergence are glossed over in [6] and, as it turns out, require a significant effort to overcome.) As we will see in Section 3, the asymptotics of the spectral measure appearing in question 2 are described by the Brown measure of R-diagonal operators. (The Brown measure is a continuous analogue of the spectral distribution of nonnormal operators, introduced in [4].)

R-diagonal operators were introduced by Nica and Speicher [19] in the context of free probability; they represent the weak*-limit (or more precisely, the limit in $*$-moments) of operators of the form UT with U unitary with size going to infinity and T diagonal, and were intensively studied in the last decade within the theory of free probability, in particular in connection with the problem of classifying invariant subspaces [13], [14].

2. **Limiting spectral density of a nonnormal matrix**

Throughout, for a probability measure μ supported on $\mathbb{R}$ or on $\mathbb{C}$, we write G_μ for its Stieltjes transform; that is,

$$G_\mu(z) = \int \frac{\mu(dx)}{z-x}\,.$$

G_μ is analytic off the support of μ. We let $\mathcal{H}_n$ denote the Haar measure on the n-dimensional unitary group $\mathcal{U}(n)$. Let $\{P_n, Q_n\}_{n\geq 1}$ denote a sequence of independent, $\mathcal{H}_n$-distributed matrices. Let T_n denote a sequence of diagonal matrices, independent of (P_n, Q_n), with real positive entries $S_n = \{s_i^{(n)}\}$ on the diagonal, and introduce the *empirical measure* of the *symmetrized* version of T_n as

$$L_{S_n} = \frac{1}{2n}\sum_{i=1}^n [\delta_{s_i^{(n)}} + \delta_{-s_i^{(n)}}]\,.$$

We write G_{T_n} for $G_{L_{S_n}}$. For a measure μ supported on $\mathbb{R}_+$, we write $\tilde{\mu}$ for its *symmetrized version*, that is, for any $0 < a < b < \infty$,

$$\tilde{\mu}([-a,-b]) = \tilde{\mu}([a,b]) = \frac{1}{2}\mu([a,b])\,.$$

Let $A_n = P_n T_n Q_n$, let $\Lambda_n = \{\lambda_i^{(n)}\}$ denote the set of eigenvalues of A_n, and set

$$L_{A_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i^{(n)}}\,.$$

We refer to L_{A_n} as the empirical spectral distribution (ESD) of A_n. (Note that the law of L_{A_n} does not change if one considers $P_n T_n$ instead of $P_n T_n Q_n$, since if $P_n T_n Q_n w = \lambda w$ for some (w, λ), then with $\overline{P}_n = Q_n P_n$ and $v = Q_n w$, it holds that $\overline{P}_n T_n v = \lambda v$, and $\overline{P}_n$ is again Haar distributed.) Finally, for any matrix A, we set $\|A\|$ to denote the ℓ^2 operator-norm of A, that is, its largest singular value.

To state our results, we recall the notion of *free convolution* of probability measures on $\mathbb{R}$, introduced by Voiculescu. For a compactly supported probability measure on μ, define the formal power series

$$G_\mu(z) = \sum_{n\geq 0}\int x^n d\mu(x) z^{-(n+1)},$$

and let $K_\mu(z)$ denote its inverse in a neighborhood of infinity, satisfying

$$G_\mu(K_\mu(z)) = z.$$

The *R-transform* of μ is the function $R_\mu(z) = K_\mu(z) - 1/z$. The moments of μ (and therefore μ itself, since it is compactly supported) can be recovered from the knowledge of K_μ, and therefore from R_μ, by a formal inversion of power series. For a pair of compactly supported probability measures μ_1, μ_2, introduce the *free convolution* $\mu_1 \boxplus \mu_2$ as the (compactly supported) probability measure whose R-transform is $R_{\mu_1}(z) + R_{\mu_2}(z)$. (That this defines indeed a probability measure needs a proof; see [1, §5.3] for details and background.)

For $a \in \mathbb{R}_+$, introduce the symmetric Bernoulli measure $\lambda_a = \frac{1}{2}(\delta_a + \delta_{-a})$ with atoms at $\{-a, a\}$. All our main results, Theorem 3 and Propositions 4 and 6, will be derived from the following technical result.

THEOREM 1. *Assume $\{L_{T_n}\}_n$ converges weakly to a probability measure Θ compactly supported on $\mathbb{R}_+$. Assume further*

1. *There exists a constant $M > 0$ so that*

$$\lim_{n\to\infty} P(\|T_n\| > M) = 0\,. \tag{1}$$

2. *There exist a sequence of events $\{\mathcal{G}_n\}$ with $P(\mathcal{G}_n^c) \to 0$ and constants $\delta, \delta' > 0$ so that for Lebesgue almost any $z \in \mathbb{C}$, with σ_n^z the minimal singular value of $zI - A_n$,*

$$E(\mathbf{1}_{\mathcal{G}_n} \mathbf{1}_{\{\sigma_n^z < n^{-\delta}\}} (\log \sigma_n^z)^2) < \delta'\,. \tag{2}$$

3. *There exist constants $\kappa, \kappa_1 > 0$ such that*

$$|\Im G_{T_n}(z)| \le \kappa_1 \quad \text{on} \quad \{z : \Im(z) > n^{-\kappa}\}\,. \tag{3}$$

Then the following hold.

a. *L_{A_n} converges in probability to a limiting probability measure μ_A.*
b. *The measure μ_A possesses a radially-symmetric density ρ_A with respect to the Lebesgue measure on $\mathbb{C}$, satisfying $\rho_A(z) = \frac{1}{2\pi}\Delta_z(\int \log|x| d\nu^z(x))$, where Δ_z denotes the Laplacian with respect to the variable z and $\nu^z := \widetilde{\Theta} \boxplus \lambda_{|z|}$.*
c. *The support of μ_A is a single ring: there exist constants $0 \le a < b < \infty$ so that*

$$\text{supp}\mu_A = \{re^{i\theta} : a \le r \le b\}\,.$$

Further, $a = 0$ if and only if $\int x^{-2} d\Theta(x) = \infty$.

See Remark 7 for an explicit characterization of the free convolution appearing in Theorem 1, and [1, Chap. 5] for general background. A different characterization of ρ_A, borrowed from [12] and instrumental in the proof of part (c) of Theorem 1, is provided in Remark 8 in Section 3.1.

Remark 2. We do not believe that the conditions in Theorem 1 are sharp. In particular, we do not know whether Condition (3), which prevents the existence of an atom in the support of $\widetilde{\Theta}$, can be dispensed with; the example $T_n = I$ shows that it is certainly not necessary.

Theorem 1 is generalized to the case where P_n, Q_n follow the Haar measure on the orthogonal group in Theorem 18. Note that, since for Lebesgue almost every $x \in \mathbb{R}$, the imaginary part of the Stieltjes transform of an absolutely continuous probability measure converges, as $z \to x$, towards the density of this measure at x, (3) is verified as soon as $\widetilde{\Theta}$ has a bounded continuous density.

As a corollary of Theorem 1, we prove the Feinberg-Zee "single ring theorem."

THEOREM 3. *Let V denote a polynomial with positive leading coefficient. Let the n-by-n complex matrix X_n be distributed according to the law*

$$\frac{1}{Z_n}\exp(-n \operatorname{tr} V(XX^*))dX\,,$$

where Z_n is a normalization constant and dX the Lebesgue measure on n-by-n complex matrices. Let L_{X_n} be the ESD *of X_n. Then $\{L_{X_n}\}_n$ satisfies the conclusions of Theorem* 1 *with Θ the unique minimizer of the functional*

$$\mathcal{J}(\mu) := \int V(x^2)d\mu(x) - \iint \log|x^2 - y^2| d\mu(x)d\mu(x)$$

on the set of probability measures on $\mathbb{R}^+$.

Theorem 3 will follow by checking that the assumptions of Theorem 1 are satisfied for the spectral decomposition $X_n = P_n T_n Q_n$; see Section 6.

The second hypothesis in Theorem 1 may seem difficult to verify in general; we show in Proposition 4 that adding a small Gaussian matrix guarantees it.

PROPOSITION 4. *Let $(T_n)_{n\geq 0}$ be a sequence of matrices satisfying the assumptions of Theorem* 1 *except for* (2) *and assume that $\|T_n^{-1}\|$ is uniformly bounded. Let N_n be a $n \times n$ matrix with independent* (*complex*) *Gaussian entries of zero mean and covariance equal to the identity. Let U_n, V_n follow the Haar measure on unitary $n \times n$ matrices, independently of T_n, N_n. Then, the empirical measure of the eigenvalues of $Y_n := U_n T_n V_n + n^{-\gamma} N_n$ converges weakly in probability to μ_A as in Theorem* 1 *for any $\gamma \in (\frac{1}{2}, \infty)$.*

In a general framework, P. Śniady [23, Th. 7] has shown that there exists a sequence ε_n going to zero at infinity so that the spectral measure of $U_n T_n V_n + \varepsilon_n n^{-1/2} N_n$ converges to μ_A. The above proposition thus insures that this is true for any polynomially decaying sequence ε_n. Note that in the earlier unpublished notes [11], U. Haagerup proved a similar regularization of the Brown measure by Cauchy-type matrices instead of Gaussian ones.

Example 5. An example of a sequence $(T_n)_{n\geq 0}$ satisfying the hypotheses of Proposition 4 is given as follows: take $T_n = \mathrm{diag}(s_1^n, \ldots, s_n^n)$ with $s_i^n \in [\delta, M]$, for $0 < \delta < M < \infty$ independent of n, so that

- L_{T_n} converges weakly towards a probability measure μ on $[\delta, M]$ which is absolutely continuous with respect to the Lebesgue measure;
- there exist $\kappa > 0$ and C finite so that for all $E \in [\delta, M]$, all $\delta \geq n^{-\kappa}$,

$$\sharp\{i : |s_i^n - E| \leq \delta\} \leq C\delta n\,.$$

A rather straightforward generalization of Theorem 1 concerns the limiting spectral measure of P_n+B_n, where P_n is $\mathcal{H}_n$ distributed and the sequence of $n\times n$ matrices B_n converges in $*$-moments to an operator b in a noncommutative probability space $(\mathcal{A}, \tau)$. (The latter means that for all polynomials P in two noncommutative variables $\lim_{n\to\infty} \frac{1}{n}\,\mathrm{tr}(P(B_n, B_n^*)) = \tau(P(b, b^*))$, which is the case if e.g. B_n is self-adjoint, with spectral measure converging to a probability measure Θ, which is the law of a self-adjoint operator b.) In particular, for any $w \in \mathbb{C}$, the spectral measure of $T_n(w) = |wI - B_n| = \sqrt{(wI - B_n)(wI - B_n)^*}$ converges to the law Θ_w of $|wI-b|$. By Voiculescu's theorem [26, Th. 3.8], if the operator norm of B_n is uniformly bounded, then the couple (B_n, P_n) converges in $*$-moments towards (b, u), a pair of operators living in a noncommutative probability space $(\mathcal{A}, \tau)$ which are free, u being unitary. The Brown measure μ_{b+u} is studied in [3, §4].

PROPOSITION 6. *Assume that $T_n(0)$ satisfies* (1) *and that there exists a set $\Omega \subset \mathbb{C}$ with full Lebesgue measure so that for all $w \in \Omega$, $T_n(w)$ satisfies* (3). *Let N_n be an $n\times n$ matrix with independent* (*complex*) *Gaussian entries of zero mean and covariance equal identity. Then, for any $\gamma > \frac{1}{2}$, the spectral measure of $B_n+n^{-\gamma}N_n+P_n$ converges in probability to the Brown measure μ_{b+u} of $b+u$.*

An example of a sequence of matrices B_n which satisfy the hypotheses of Proposition 6 is given by the diagonal matrices $B_n = \mathrm{diag}(s_1^n, \ldots, s_n^n)$ with entries s_i^n satisfying the hypotheses of Example 5. This is easily verified from the fact that the eigenvalues of $D_n(w)$ are given by $(|w - s_1^n|, \ldots, |w - s_n^n|)$.

2.1. *Background and description of the proof.* The main difficulty in studying the ESD L_{A_n} is that A_n is not a normal matrix; that is, $A_nA_n^* \neq A_n^*A_n$, almost surely. For normal matrices, the limit of ESDs can be found by the method of moments or by the method of Stieltjes' transforms. For nonnormal matrices, the only known method of proof is more indirect and follows an idea of Girko [9] that we describe now (the details are a little different from what is presented in Girko [9] or Bai [2]).

From Green's formula, for any polynomial $P(z) = \prod_{j=1}^n (z - \lambda_j)$, we have

$$\frac{1}{2\pi}\int \Delta\psi(z)\log|P(z)|dm(z) = \sum_{j=1}^n \psi(\lambda_j), \qquad \text{for any } \psi \in C_c^2(\mathbb{C})\,,$$

where $m(\cdot)$ denotes the Lebesgue measure on $\mathbb{C}$. Applied to the characteristic polynomial of A_n, this gives

$$\begin{aligned}\int \psi(z) dL_{A_n}(z) &= \frac{1}{2\pi n} \int\limits_{\mathbb{C}} \Delta\psi(z) \log|\det(zI - A_n)| dm(z) \\ &= \frac{1}{4\pi n} \int\limits_{\mathbb{C}} \Delta\psi(z) \log\det(zI - A_n)(zI - A_n)^* dm(z)\,.\end{aligned}$$

It will be convenient for us to introduce the $2n \times 2n$ matrix

$$H_n^z := \begin{bmatrix} 0 & zI - A_n \\ (zI - A_n)^* & 0 \end{bmatrix}. \tag{4}$$

It may be checked easily that eigenvalues of H_n^z are the positive and negative of the singular values of $zI - A_n$. Therefore, if we let ν_n^z denote the ESD of H_n^z,

$$\int \frac{1}{y - x} d\nu_n^z(x) = \frac{1}{2n} \operatorname{tr}((y - H_n^z)^{-1});$$

then

$$\frac{1}{n} \log\det(zI - A_n)(zI - A_n)^* = \frac{1}{n} \log\det|H_n^z| = 2 \int_{\mathbb{R}} \log|x| d\nu_n^z(x)\,.$$

Thus we arrive at the formula

$$\int \psi(z) dL_{A_n}(z) = \frac{1}{2\pi} \int\limits_{\mathbb{C}} \Delta\psi(z) \int_{\mathbb{R}} \log|x| d\nu_n^z(x) dm(z)\,. \tag{5}$$

This is Girko's formula in a different form and its utility lies in the following attack on finding the limit of L_{A_n}.

1. Show that for (Lebesgue almost) every $z \in \mathbb{C}$, the measures ν_n^z converge weakly in probability to a measure ν^z as $n \to \infty$, and identify the limit. Since H_n^z are Hermitian matrices, there is hope of doing this by Hermitian techniques.
2. Justify that $\int \log|x| d\nu_n^z(x) \to \int \log|x| d\nu^z(x)$ for (almost) every z. But for the fact that "log" is not a bounded function, this would have followed from the weak convergence of ν_n^z to ν^z. As it stands, this is the hardest technical part of the proof.
3. A standard uniform integrability argument is then used in order to convert the convergence for (almost) every z of ν_n^z to a convergence of integrals over z. Indeed, setting $h(z) := \int \log|x| d\nu^z(x)$, we will get from (5) that

$$\int \psi(z) dL_{A_n}(z) \to \frac{1}{2\pi} \int\limits_{\mathbb{C}} \Delta\psi(z)\ h(z) dm(z)\,. \tag{6}$$

4. Show that h is smooth enough so that one can integrate the previous equation by parts to get

$$\int \psi(z) dL_{A_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \psi(z)\, \Delta h(z) dm(z)\,, \tag{7}$$

which identifies $\Delta h(z)/2\pi$ as the density (with respect to Lebesgue measure) of the limit of L_{A_n}.

5. Identify the function h sufficiently precisely to be able to deduce properties of $\Delta h(z)$. In particular, show the *single ring phenomenon*, which states that the support of the limiting spectral measure is a single annulus (the surprising part being that it cannot consist of several disjoint annuli).

Girko's equation (5) and these five steps give a general recipe for finding limiting spectral measures of nonnormal random matrices. Whether one can overcome the technical difficulties depends on the model of random matrix one chooses. For the model of random matrices with i.i.d. entries having zero mean and finite variance, this has been achieved in stages by Bai [2], Götze and Tikhomirov [10], Pan and Zhou [20] and Tao and Vu [24]. While we borrow extensively from that sequence, a major difficulty in the problem considered here is that there is no independence between entries of the matrix A_n. Instead, we will rely on properties of the Haar measure, and in particular on considerations borrowed from free probability and the so called *Schwinger-Dyson* (or *master-loop*) equations. Such equations were already the key to obtaining fine estimates on the Stieltjes transform of Gaussian generalized band matrices in [15]. In [5], they were used to study the asymptotics of matrix models on the unitary group. Our approach combines ideas of [15] to estimate Stieltjes transforms and the necessary adaptations to unitary matrices as developed in [5]. The main observation is that one can reduce attention to the study of the ESD of matrices of the form $(T+U)(T+U)^*$ where T is real diagonal and U is Haar distributed. In the limit (i.e., when T and U are replaced by operators in a C^*-algebra that are freely independent, with T bounded and self-adjoint and U unitary), the limit ESD has been identified by Haagerup and Larsen [12]. The Schwinger-Dyson equations give both a characterization of the limit and, more important to us, a finite approximation that can be used to estimate the discrepancy between the pre-limit ESD and its limit. These estimates play a crucial role in integrating the singularity of the log in Step 2 above, but only once an *a priori* (polynomial) estimate on the minimal singular value has been obtained. The latter is deduced from assumption 2. In the context of the Feinberg-Zee single ring theorem, the latter assumption holds due to an adaptation of the analysis of [22].

Notation. We describe our convention concerning constants. Throughout, by the word *constant* we mean quantities that are independent of n (or of the complex variables z, z_1). Generic constants denoted by the letters C,c or R, have values that may change from line to line, and they may depend on other parameters. Constants denoted by C_i, K, κ and κ' are fixed and do not change from line to line.

3. An auxiliary problem: evaluation of ν^z and convergence rates

Recall from the proof sketch described above that we are interested in evaluating the limit ν^z of the ESD L_n^z of the matrix H_n^z; see (4). Note that L_n^z is also the ESD of the matrix $\widetilde{H}_n^z$ given by

$$
\begin{aligned}
\widetilde{H}_n^z &:= \begin{bmatrix} 0 & Q_n \\ P_n^* & 0 \end{bmatrix} H_n^z \begin{bmatrix} 0 & P_n \\ Q_n^* & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & |z|W_n^z - T_n \\ (|z|W_n^z - T_n)^* & 0 \end{bmatrix},
\end{aligned} \tag{8}
$$

where $W_n^z = \overline{z} Q_n P_n / |z|$ is unitary and $\mathcal{H}_n$ distributed. Throughout, we will write $\rho = |z|$. We also will assume in this section that the sequence T_n is deterministic. We are thus led to the study of the ESD for a sequence of matrices of the form

$$
\mathbf{Y}_n = \begin{pmatrix} 0 & B_n \\ B_n^* & 0 \end{pmatrix} \tag{9}
$$

where $B_n = \rho U_n + T_n$, T_n is a real, diagonal matrix of uniformly bounded norm, and U_n a $\mathcal{H}_n$ unitary matrix. Because $\|T_n\|$ is uniformly bounded, it will be enough to consider ρ, throughout, uniformly bounded.

We denote

$$
\mathbf{U}_n = \begin{pmatrix} 0 & U_n \\ 0 & 0 \end{pmatrix}, \quad \mathbf{U}_n^* = \begin{pmatrix} 0 & 0 \\ U_n^* & 0 \end{pmatrix}, \quad \mathbf{T}_n = \begin{pmatrix} 0 & T_n \\ T_n & 0 \end{pmatrix}. \tag{10}
$$

3.1. *Limit equations.* We begin by deriving the limiting Schwinger-Dyson equations for the ESD of $\mathbf{Y}_n$. Throughout this subsection, we consider a noncommutative probability space $(\mathcal{A}, *, \mu)$ on which a variable U lives and where μ is a tracial state satisfying the relations $\mu((UU^* - 1)^2) = 0$, $\mu(U^a) = 0$ for $a \in \mathbb{Z} \setminus \{0\}$. In the sequel, 1 will denote the identity in $\mathcal{A}$. We refer to [1, §5.2] for definitions.

Let T be a self-adjoint (bounded) element in $\mathcal{A}$, with T freely independent of U. Recall the noncommutative derivative ∂, defined on elements of $\mathbb{C}\langle T, U, U^* \rangle$ as satisfying the Leibniz rules

$$
\begin{gathered}
\partial(PQ) = \partial P \times (1 \otimes Q) + (P \otimes 1) \times \partial Q\,, \\
\partial U = U \otimes 1, \quad \partial U^* = -1 \otimes U^*, \quad \partial T = 0 \otimes 0\,.
\end{gathered} \tag{11}
$$

(Here, $\otimes$ denotes the tensor product and we write $(A\otimes B)\times(C\otimes D) = (AC)\otimes(BD)$.) Now, ∂ is defined so that for any $B\in\mathcal{A}$ satisfying $B^*=-B$, any $P\in\mathbb{C}\langle U,U^*,T\rangle$,

$$P(Ue^{\epsilon B}, e^{-\epsilon B}U^*, T) = P(U,U^*,T) + \epsilon\partial P(U,U^*,T)\sharp B + o(\epsilon)\,, \tag{12}$$

where we used the notation $A\otimes B\sharp C = ACB$.

By the invariance of μ under unitary conjugation, see [27, Prop. 5.17] or [1, (5.4.31)], we have the Schwinger-Dyson equation

$$\mu\otimes\mu(\partial P) = 0\,. \tag{13}$$

We continue to use the notation $\mathbf{Y}$, $\mathbf{U},\mathbf{U}^*$ and $\mathbf{T}$ in a way similar to (9) and (10). So, we let $\mathbf{Y}=\rho(\mathbf{U}+\mathbf{U}^*)+\mathbf{T}$ with

$$\mathbf{U} = \begin{pmatrix} 0 & U \\ 0 & 0 \end{pmatrix}, \quad \mathbf{U}^* = \begin{pmatrix} 0 & 0 \\ U^* & 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} 0 & T \\ T & 0 \end{pmatrix}. \tag{14}$$

We extend μ to the algebra generated by $\mathbf{U},\mathbf{U}^*$ and $\mathbf{T}$ by putting for any $A,B,C,D\in\mathcal{A}$,

$$\mu\left(\begin{pmatrix} A & B \\ C & D \end{pmatrix}\right) := \frac{1}{2}\mu(A) + \frac{1}{2}\mu(D)\,.$$

Observe that this extension is still tracial.

The noncommutative derivative ∂ in (12) extends naturally to the algebra generated by the matrix-valued $\mathbf{U},\mathbf{U}^*,\mathbf{T}$, using the Leibniz rule (11) together with the relations

$$\partial\mathbf{U} = \mathbf{U}\otimes p\,, \quad \partial\mathbf{U}^* = -p\otimes\mathbf{U}^*\,, \quad \partial\mathbf{T} = 0\otimes 0\,, \tag{15}$$

where we denoted $p=\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1\end{smallmatrix}\right)$. In the sequel we will apply ∂ to analytic functions of $\mathbf{U}+\mathbf{U}^*$ and $\mathbf{T}$ such as products of Stieltjes functionals of the form $(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}$ with $z\in\mathbb{C}\backslash\mathbb{R}$ and $a,b\in\mathbb{R}$. Such an extension is straightforward; ∂ continues to satisfy the Leibniz rule and, by the resolvent identity

$$\begin{aligned}&\partial\,(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1} \\ &\qquad = b\,(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}\,(\mathbf{U}\otimes p - p\otimes\mathbf{U}^*)\,(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}\,,\end{aligned}$$

where $A(B\otimes C)D = (AB)\otimes(CD)$. Further, (13) extends also in this context.

Introduce the notation, for $z_1,z_2\in\mathbb{C}^+$,

$$\begin{aligned} G(z_1,z_2) &= \mu\left((z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\right)\,, \\ G_U(z_1,z_2) &= \mu\left(\mathbf{U}(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\right)\,, \\ G_U(z_1) &= \mu\left(\mathbf{U}(z_1-\mathbf{Y})^{-1}\right)\,, \\ G_{U^*}(z_1,z_2) &= \mu\left(\mathbf{U}^*(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\right)\,, \end{aligned} \tag{16}$$

$$\begin{aligned} G_T(z_1,z_2) &= \mu\left(\mathbf{T}(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\right)\,, \\ G(z_1) &= \mu\left((z_1-\mathbf{Y})^{-1}\right)\,, \\ G_T(z_2) &= \mu\left((z_2-\mathbf{T})^{-1}\right)\,. \end{aligned}$$

We apply the derivative ∂ to the analytic function $P=(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\mathbf{U}$, while noticing that, by (11) and (15),

$$\partial P = P\otimes p+\rho(z_1-\mathbf{Y})^{-1}\mathbf{U}\otimes pP-\rho(z_1-\mathbf{Y})^{-1}p\otimes \mathbf{U}^*P. \tag{17}$$

Applying (13), with $\mu(P)=G_U(z_1,z_2)$ and $\mu(p)=1/2$, we find

$$\frac{1}{2}G_U(z_1,z_2)=\rho\mu\left((z_1-\mathbf{Y})^{-1}p\right)\mu(\mathbf{U}^*P)-\rho\mu\left((z_1-\mathbf{Y})^{-1}\mathbf{U}\right)\mu(pP)\,. \tag{18}$$

Note that $Pp=P$ and thus $\mu(pP)=\mu(P)$. Further, for any smooth function Q, $\mu(\mathbf{U}^*Q\mathbf{U})$ equals $\mu((1-p)Q)$ due to the traciality of μ and $\mathbf{U}\mathbf{U}^*=1-p$. By symmetry (note that $(1-p)(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}$ and $p(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}$ are given by the same formula up to replacing $(\mathbf{U},\mathbf{U}^*)$ by $(\mathbf{U}^*,\mathbf{U})$, which has the same law) we get $\mu(\mathbf{U}^*P)$ equals

$$\mu((1-p)(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1})=\frac{1}{2}\mu((z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1})=\frac{1}{2}G(z_1,z_2)\,. \tag{19}$$

The first equality holds without the last factor $(z_2-\mathbf{T})^{-1}$, thus implying that $\mu((z_1-\mathbf{Y})^{-1}p)=\mu((z_1-\mathbf{Y})^{-1})/2=G(z_1)/2$ and so we get from (18) that

$$\frac{1}{2}G_U(z_1,z_2)=\frac{\rho}{4}G(z_1,z_2)G(z_1)-\rho G_U(z_1,z_2)G_U(z_1)\,. \tag{20}$$

Noticing that $G_U(z_1)$ is the limit of $z_2G_U(z_1,z_2)$ as $z_2\to\infty$, we find by (20) that

$$\frac{1}{2}G_U(z_1)=-\rho G_U(z_1)^2+\frac{\rho}{4}G(z_1)^2\,,$$

and therefore, as $G_U(z_1)$ goes to zero as $z_1\to\infty$,

$$G_U(z_1)=\frac{1}{2\rho}\left(-\frac{1}{2}+\sqrt{\frac{1}{4}+\rho^2G(z_1)^2}\right)=\frac{1}{4\rho}\left(-1+\sqrt{1+4\rho^2G(z_1)^2}\right)\,. \tag{21}$$

Here, the choice of the branch of the square root is determined by the expansion of $G_U(z)$ at infinity and the fact that both $G(z)$ and $G_U(z)$ are analytic in $\mathbb{C}^+$. This equation is then true for all $z_1\in\mathbb{C}^+$.

Moreover, by (20) and (21), we get

$$G_U(z_1,z_2)=\frac{\rho}{2}\frac{G(z_1,z_2)G(z_1)}{1+2\rho G_U(z_1)}=\frac{\rho G(z_1,z_2)G(z_1)}{1+\sqrt{1+4\rho^2G(z_1)^2}}\,. \tag{22}$$

(Again, here and in the rest of this subsection, the proper branch of the square root is determined by analyticity.) Let R_ρ denote the R-transform of the

Bernoulli law $\lambda_\rho := (\delta_{-\rho} + \delta_{+\rho})/2$; that is,

$$R_\rho(z) = \frac{\sqrt{1+4\rho^2 z^2} - 1}{2\rho z} = \frac{2z\rho}{\sqrt{1+4\rho^2 z^2}+1}\,;$$

see [1, Def. 5.3.22 and Ex. 5.3.27], so that we have

$$G_U(z_1, z_2) = \frac{1}{2} G(z_1, z_2) R_\rho(G(z_1))\,. \tag{23}$$

Repeating the computation with G_{U^*}, we have $G_{U^*} = G_U$. Algebraic manipulations yield

$$G_T(z_1, z_2) = z_2 G(z_1, z_2) - G(z_1)\,, \tag{24}$$

$$2\rho G_U(z_1, z_2) + G_T(z_1, z_2) = z_1 G(z_1, z_2) - G_T(z_2)\,. \tag{25}$$

Therefore, we get by substituting (23) and (24) into (25) that

$$\rho G(z_1, z_2) R_\rho(G(z_1)) + z_2 G(z_1, z_2) - G(z_1) = z_1 G(z_1, z_2) - G_T(z_2)\,, \tag{26}$$

which in turns gives, for any $z_1, z_2 \in \mathbb{C}^+$,

$$G(z_1, z_2)\left(\rho R_\rho(G(z_1)) + z_2 - z_1\right) = G(z_1) - G_T(z_2)\,. \tag{27}$$

Thus,

$$G_T(z_2) = G(z_1) \quad \text{when} \quad z_2 = z_1 - \rho R_\rho(G(z_1))\,. \tag{28}$$

The choice of z_2 as in (28) is allowed for any $z_1 \in \mathbb{C}^+$ because $G\colon \mathbb{C}^+ \to \mathbb{C}^-$ and we can see that $R\colon \mathbb{C}^- \to \mathbb{C}^-$. Thus $\Im(z_2) \geq \Im(z_1) > 0$, implying that such z_2 belong to the domain of G_T.

The relation (28) is the Schwinger-Dyson equation in our setup. This gives an implicit equation for $G(\cdot)$ in terms of $G_T(\cdot)$. Further, for z with large modulus, $G(z)$ is small and thus $z \mapsto z - \rho R_\rho(G(z))$ possesses a nonvanishing derivative, and further, is close to z. Because G_T is analytic in the upper half-plane and its derivative behaves like $1/z^2$ at infinity, it follows by the implicit function theorem that (28) uniquely determines $G(\cdot)$ in a neighborhood of ∞. By analyticity, it thus fixes $G(\cdot)$ in the upper half-plane (and in fact, everywhere except in a compact subset of $\mathbb{R}$), and thus determines uniquely the law of $\mathbf{Y}$.

Remark 7. Let μ_T denote the *spectral measure* of T, that is $\int f d\mu_T = \mu(f(T))$ for any $f \in C_b(\mathbb{R})$. We emphasize that G_T is not the Stieltjes transform of μ_T; rather, it is the Stieltjes transform of the symmetrized version of the law of T, that is of the probability measure $\tilde{\mu}_T$. With this convention, (28) is equivalent to the statement that the law of $\mathbf{Y}$, denoted μ_Y, equals the *free convolution* of $\tilde{\mu}_T$ and λ_ρ, i.e., $\mu_Y = \tilde{\mu}_T \boxplus \lambda_\rho$.

Remark 8. We provide, following [12], an alternative characterization of μ_A and its support. We first introduce some terminology from [12]. Consider a tracial noncommutative W^*-probability space $(\mathcal{M}, \tau)$. Let u be Haar-distributed and let h be a self-adjoint element having law Θ and that is $*$-free from u. Let $\tilde{\nu}^z$ denote the law of $|zI - uh|$. The *Brown measure* for uh is defined as

$$\frac{1}{2\pi}\Delta_z \int \log |x| d\tilde{\nu}^z(x)\,;$$

cf. [12, p. 333]. Recall that $\Theta(\{0\}) = 0$ by Assumption 3. By [12, Prop. 3.5] and Remark 7 above, $\tilde{\nu}^z = \nu^z$, and therefore, μ_A in the statement of Theorem 1 is the Brown measure for uh. By [12, Th. 4.4 and Cor. 4.5], the Brown measure μ_A is radially symmetric and possesses a density ρ_A that can be described as follows. Let $\Theta^{\sharp 2}$ denote the push forward of Θ by the map $z \mapsto z^2$; i.e., $\Theta^{\sharp 2}$ is the weak limit of $\{L_{T_n^2}\}$. Let $\mathcal{S}$ denote the S-transform of $\Theta^{\sharp 2}$ (see [12, §2] for the definition of the S-transform of a probability measure on $\mathbb{R}$ and its relation to the R-transform). Define $F(t) = 1/\sqrt{\mathcal{S}(t-1)}$ on $\mathcal{D} = (0, 1]$. Then, F maps $\mathcal{D}$ to the interval

$$(a, b] = \left(\frac{1}{(\int x^{-2} d\Theta(x))^{1/2}}, \left(\int x^2 d\Theta(x) \right)^{1/2} \right],$$

and has an analytic continuation to a neighborhood of $\mathcal{D}$, and $F' > 0$ on $\mathcal{D}$. Further, with μ_A as above, $\rho_A(re^{i\theta}) = \rho_A(r)$ and it holds that

$$\rho_A(r) = \begin{cases} \frac{1}{2\pi r F'(F^{-1}(r))}\,, & r \in (a, b]\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{29}$$

Finally, ρ_A has an analytic continuation to a neighborhood of $(a, b]$, and μ_A is a probability measure; see [12, p. 333].

In the next section, we will need the following estimate.

LEMMA 9. *If* $|\Im G_T(\cdot)| \le \kappa_1$ *on* $\{z : \Im(z) \ge \epsilon\}$ *then* $|\Im G(\cdot)| \le \kappa_1$ *on* $\{z : \Im(z) \ge \epsilon\}$.

Proof. Recall that if $z \in \mathbb{C}^+$, then $G(z) \in \mathbb{C}^-$ and also $R_\rho(G(z)) \in \mathbb{C}^-$ because R_ρ maps $\mathbb{C}^-$ into $\mathbb{C}^-$ (regardless of the branch of the square root taken at each point). Thus, $y = z - R(G(z))$ has $\Im(y) \ge \Im(z)$. Therefore, if $\Im(z) \ge \epsilon$ then $|\Im G(z)| = |\Im G_T(y)| \le \kappa_1$. $\square$

3.2. *Finite n equations and convergence.* We next turn to the evaluation of the law of $\mathbf{Y}_n$. We assume throughout that the sequence T_n is uniformly bounded by some constant M, that $L_{T_n} \to \mu_T$ weakly in probability, and further that (3) is satisfied. All constants in this section are independent of ρ, but depend implicitly on M, the uniform bound on $\|T_n\|$ and on ρ.

Recall first that by invariance of the Haar measure under unitary conjugation (see [1, (5.4.29)]), with $P \in \mathbb{C}\langle T, U, U^*\rangle$ (or a product of Stieltjes functionals), it holds that

$$E\left[\frac{1}{2n}\operatorname{tr}\otimes\frac{1}{2n}\operatorname{tr}(\partial P(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*))\right] = 0\,. \tag{30}$$

This key equality can be proved by noticing that for any $n\times n$ matrix B such that $B^* = -B$, for any $(k,\ell)\in[1,n]$, if we let $U_n(t) = U_n e^{tB}$ and construct $\mathbf{U_n}(t)$ and $\mathbf{U_n}^*(t)$ with this unitary matrix, then

$$0 = \partial_t E[(P(\mathbf{T}_n, \mathbf{U}_n(t), \mathbf{U}_n^*(t)))_{k,\ell}] = E[(\partial P(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*)\sharp\mathbf{B})_{k,\ell}] \tag{31}$$

with $\mathbf{B} = \left(\begin{smallmatrix} 0 & 0\\ 0 & B\end{smallmatrix}\right)$. Letting $\Delta(k,\ell)$ be the $n\times n$ matrix so that $\Delta(k,\ell)_{i,j} = 1_{i=k}1_{j=\ell}$, we can choose in the last equality $B = \Delta(k,\ell) - \Delta(\ell,k)$ or $B = i\,(\Delta(k,\ell)+\Delta(\ell,k))$. Summing the two resulting equalities and then summing over k and ℓ yields (30).

We denote by G^n the quantities as defined in (16), but with $E[\frac{1}{2n}\operatorname{tr}]$ replacing μ and the superscript or subscript n attached to all variables, so that for instance

$$G^n(z) = E\left[\frac{1}{2n}\operatorname{tr}((z-\mathbf{Y}_n)^{-1})\right]\,.$$

We get by taking $P = (z_1 - \mathbf{Y}_n)^{-1}(z_2 - \mathbf{T}_n)^{-1}\mathbf{U}_n$ that

$$\frac{1}{2}G_U^n(z_1,z_2) = -\rho G_U^n(z_1,z_2)G_U^n(z_1) + \frac{\rho}{4}G^n(z_1,z_2)G^n(z_1) + O(n,z_1,z_2)\,, \tag{32}$$

with

$$O(n,z_1,z_2) = E\left[\left(\frac{1}{2n}\operatorname{tr} - E[\frac{1}{2n}\operatorname{tr}]\right)\right.$$
$$\left.\otimes\left(\frac{1}{2n}\operatorname{tr} - E\left[\frac{1}{2n}\operatorname{tr}\right]\right)\partial(z_1-\mathbf{Y}_n)^{-1}(z_2-\mathbf{T}_n)^{-1}\mathbf{U}_n\right]\,.$$

Further, by the standard concentration inequality for $\mathcal{H}_n$, see [1, Cor. 4.4.30], for any smooth function $P\colon \mathcal{U}(n)\to\mathbb{C}$,

$$E\left[\left|\frac{1}{2n}\operatorname{tr}(P) - E\left[\frac{1}{2n}\operatorname{tr}\right](P)\right|^2\right] \le \frac{1}{n^2}\|P\|_L^2\,, \tag{33}$$

with $\|P\|_L$ the Lipschitz constant of P given by

$$\|P\|_L = \|DP\|_\infty$$

if D is the cyclic derivative given by $D = m\circ\partial$ with $m(A\otimes B) = BA$ and if $\|DP\|_\infty$ denotes the operator norm. (The appearance of the cyclic derivative in the evaluation of the Lipschitz constant can be seen by approximating P by polynomials.) Applying (33) to each term of ∂P (recall formula (17)), we get that for $\Im(z_1), \Im(z_2) > 0$, and with $a \wedge b = \min(a, b)$,

$$|O(n, z_1, z_2)| \leq \frac{C\rho^2}{n^2 |\Im(z_2)| \Im(z_1)^2 (\Im(z_1) \wedge 1)} .$$

(The inequality uses the fact that for any Hermitian matrix, $\|(z - H)^{-1}\|_\infty \leq 1/|\Im(z)|$.) Multiplying by z_2 and taking the limit as $z_2 \to \infty$ we deduce from (32) that

$$\rho(G^n(z_1))^2 = 2G_U^n(z_1)(1 + 2\rho G_U^n(z_1)) - O_1(n, z_1) , \tag{34}$$

where

$$\begin{aligned} O_1(n, z_1) &= 4E\left[\left(\frac{1}{2n}\operatorname{tr} - E\left[\frac{1}{2n}\operatorname{tr}\right]\right) \otimes \left(\frac{1}{2n}\operatorname{tr} - E\left[\frac{1}{2n}\operatorname{tr}\right]\right) \partial(z_1 - \mathbf{Y}_n)^{-1}\mathbf{U}_n\right] \\ &= O\left(\frac{\rho^2}{n^2 \Im(z_1)^2 (\Im(z_1) \wedge 1)}\right) . \end{aligned}$$

In particular,

$$G_U^n(z_1) = \frac{1}{4\rho}\left(-1 + \sqrt{1 + 4\rho^2 G^n(z_1)^2 + 4O_1(n, z_1)}\right) , \tag{35}$$

with again the choice of the square root determined by analyticity and behavior at infinity.

Recalling that (24) and (25) remain true when we add the subscript n and combining these with (32), we get

$$G^n(z_1, z_2)\left(\frac{\rho^2 G^n(z_1)}{(1 + 2\rho G_U^n(z_1))} + z_2 - z_1\right) = G^n(z_1) - G_{T_n}(z_2) + \widetilde{O}(n, z_1, z_2) , \tag{36}$$

with

$$\widetilde{O}(n, z_1, z_2) = \frac{2O(n, z_1, z_2)}{(1 + 2\rho G_U^n(z_1))} .$$

Hence, if we define

$$z_2 = \psi_n(z_1) := z_1 - \frac{\rho^2 G^n(z_1)}{(1 + 2\rho G_U^n(z_1))} , \tag{37}$$

then

$$G^n(z_1) = G_{T_n}(z_2) - \widetilde{O}(n, z_1, z_2) ,$$

and therefore

$$G^n(z_1) = G_{T_n}(\psi_n(z_1)) - \widetilde{O}(n, z_1, \psi_n(z_1)) . \tag{38}$$

Equation (38) holds at least when $\Im(z_2) > 0$ for z_2 as in (37). In particular, for $\Im(z_1)$ large (say larger than some M), it holds that $G^n(z_1)$ and $G_U^n(z_1)$ are small, implying that z_2 is well defined with $\Im(z_2) > 0$. Assume L_{T_n} converges towards L_T so that G_{T_n} converges to G_T on $\mathbb{C}^+$. Then, the limit points of the sequence of uniformly continuous functions $(G^n(z), G_U^n(z))$ on $\{z : \Im(z) \geq M\}$

satisfy (21) and (28) and therefore equal $(G(z), G_U(z))$ on $\{z : \Im(z) \geq M\}$ by uniqueness of the solutions to these equations. Hence, taking $n \to \infty$ then implies that $G^n \to G$ in a neighborhood in the upper half-plane close to ∞. Since G^n and G are Stieltjes transforms of probability measures, we have now shown the following (see Remark 7).

LEMMA 10. *Assume L_{T_n} converges weakly in probability to a compactly supported probability measure μ_T. Then, $L_{\mathbf{Y}_n}$ converges weakly, in probability, to $\mu_Y = \tilde{\mu}_T \boxplus \lambda_\rho$. In particular, if L_{T_n} converges weakly in probability to a probability measure Θ, then for any $z \in \mathbb{C}$, ν_n^z converges weakly in probability to $\widetilde{\Theta} \boxplus \lambda_{|z|}$.*

(Recall that $\widetilde{\Theta}$ is the symmetrized version of Θ.)

Lemma 10 completes the proof of Step 1 in our program. To be able to complete Step 2, we need to obtain quantitative information from the (finite n) Schwinger-Dyson equations (38): our goal is to show that the left side remains bounded in a domain of the form $\{z \in \mathbb{C}^+ : \Im(z) > n^{-c}\}$ for some $c > 0$. Toward this end, we will show that in such a region, ψ_n is analytic, $\Im\psi_n(z) > (\Im(z)/2) \wedge C$ for some positive constant C and $\widetilde{O}(n, z_1, \psi_n(z_1))$ is analytic and bounded there. This will imply that (38) extends by analyticity to this region, and our assumption on the boundedness of G_{T_n} will lead to the conclusion.

As a preliminary step, note that $G^n(\cdot)$ and $G^n_U(\cdot)$ are analytic in $\mathbb{C}^+$. We have the following.

LEMMA 11. *There exist constants C_1, C_2 such that for all $z \in \mathbb{C}^+$ with $\Im(z) > C_1 n^{-1/3}$ and all n large, it holds that*

$$|1 + 2\rho G^n_U(z)| > C_2 \rho [\Im(z)^3 \wedge 1] \,. \tag{39}$$

Proof. Since $G^n_U(z)$ is asymptotic to $1/z^2$ at infinity, we may and will restrict attention to some fixed ball $B_R \subset \mathbb{C}$, whose interior contains the support of $\mathbf{Y}_n$ (this is possible by (1)). But

$$\Im(G^n(z)) = -\Im(z) \int \frac{d\mu_{\mathbf{Y}_n}(x)}{(\Re(z) - x)^2 + \Im(z)^2},$$

and therefore, as $(\Re(z) - x)^2 + \Im(z)^2 \leq 4R^2$ for all $z, x \in B(0, R)$,

$$|G^n(z)| \geq |\Im(G^n(z))| \geq \frac{|\Im(z)|}{4R^2} \,. \tag{40}$$

Moreover, since $|G^n_U(z)| \leq 1/|\Im(z)|$, we deduce from (34) that for some constant c independent of n and all n large,

$$|G^n(z)|^2 \leq \frac{2|1 + 2\rho G^n_U(z)|}{\rho|\Im(z)|} + \frac{c\rho}{n^2 \Im(z)^2 (\Im(z) \wedge 1)} \,.$$

Combining this estimate and (40), we get that

$$\frac{2|1+2\rho G_U^n(z)|}{\rho|\Im(z)|} \geq \frac{|\Im(z)|^2}{16R^4} - \frac{c\rho}{n^2\Im(z)^2(\Im(z)\wedge 1)} \geq \frac{|\Im(z)|^2}{32R^4}\,, \tag{41}$$

as soon as $\Im(z) > C_1 n^{-1/3}$ for an appropriate C_1, and $|z| < R$. The conclusion follows. $\square$

As a consequence of Lemma 11 and the analyticity of G^n and G_U^n in $\mathbb{C}^+$, we conclude that ψ_n is analytic in $\{z : \Im(z) > C_1 n^{-1/3}\}$, for all n large.

Our next goal is to check the analyticity of $z \to \widetilde{O}(n, z, \psi_n(z))$ for $z \in \mathbb{C}^+$ with imaginary part bounded away from 0 by a polynomially decaying (in n) factor. Toward this end, we now verify that $\psi_n(z) \in \mathbb{C}^+$ for z up to a small distance from the real axis.

LEMMA 12. *There exists a constant C_3 such that if $\Im(z) > C_3 n^{-1/4}$, then $\Im(\psi_n(z)) \geq \Im(z)/2$.*

Proof. Again, because both $G^n(z)$ and $G_U^n(z_1)$ tend to 0 at infinity, we may and will restrict attention to $\Im(z) \leq R$ for some fixed R. We divide the proof into two cases, as follows. Let $\mathbf{e}_n = n^{-1/2}$, and set $\Delta_n = \{z \in \mathbb{C}^+ : |\rho G^n(z) + i/2| \geq \mathbf{e}_n\}$.

Then, for any $z \in \Delta_n$, and whatever choice of branch of the square root made in (35), if $\mathbf{e}_n^{-1/2} O_1(n, z)$ is small enough (smaller than $\mathbf{e}_n/2$ is fine), then that choice can be extended to include a neighborhood of the point $w = G^n(z)$ such that with this choice, the function $r_\rho(w) = \frac{1}{4\rho}(-1 + \sqrt{1 + 4\rho^2 w^2})$ is Lipschitz in the sense that

$$|G_U^n(z) - r(G^n(z))| \leq C\mathbf{e}_n^{-\frac{1}{2}} O_1(n, z)/\rho\,. \tag{42}$$

On the other hand, again from (34),

$$\left|\frac{\rho G^n(z)}{1 + 2\rho G_U^n(z)} - \frac{2G_U^n(z)}{G^n(z)}\right| \leq C\frac{|O_1(n, z)|}{|G^n(z)(1 + 2\rho G_U^n(z))|}\,.$$

Combining the last display with the relation $R_\rho(\theta) = 2r_\rho(\theta)/\theta$, (42) and (40), one obtains that for $z \in \Delta_n$,

$$\begin{aligned}
\left|\frac{\rho G^n(z)}{1 + 2\rho G_U^n(z)} - \rho R_\rho(G^n(z))\right| &\leq \left|\frac{2r(G^n(z))}{G^n(z)} - \frac{2G_U^n(z)}{G^n(z)}\right| \\
&\quad + \left|\frac{\rho G^n(z)}{1 + 2\rho G_U^n(z)} - \frac{2G_U^n(z)}{G^n(z)}\right| \\
&\leq C\frac{|O_1(n, z)|}{\rho \mathbf{e}_n^{\frac{1}{2}} |G^n(z)|} + C\frac{|O_1(n, z)|}{|G^n(z)(1 + 2\rho G_U^n(z))|} \\
&\leq C\frac{|O_1(n, z)|}{\rho \mathbf{e}_n^{1/2} |\Im(z)|} + C\frac{|O_1(n, z)|}{\rho \Im(z)^4}
\end{aligned}$$

$$\leq \frac{C\rho}{n^2|\Im(z)|^4}\left(\frac{1}{\mathbf{e}_n^{1/2}}+\frac{1}{|\Im(z)|^3}\right)$$
$$\leq \frac{C\rho}{n^2|\Im(z)|^4}\left(n^{1/4}+\frac{1}{|\Im(z)|^3}\right).$$

Since the above right-hand side is smaller than $\Im(z)/2$ for $\Im(z) > n^{-1/4}$, we conclude that for $z \in \Delta_n \cap \{\Im(z) > n^{-1/4}\}$,

$$\Im\left(\frac{\rho G^n(z)}{1+2\rho G_U^n(z)}\right) \leq \frac{1}{2}\Im(z) \tag{43}$$

as, regardless of the branch taken in the definition of $R_\rho(\cdot)$, $\Im R_\rho(G^n(z)) \leq 0$.

On the other hand, when $z \in \mathbb{C}^+ \setminus \Delta_n$ and $\Im(z) > n^{-1/4}$, then we have from (35) that for all n large,

$$|\rho G_U^n(z) + 1/4| \leq \frac{1}{2}\sqrt{\mathbf{e}_n + |O_1(n,z)|} \leq \frac{1}{8}.$$

Thus, under these conditions,

$$\Im\left(\frac{\rho G^n(z)}{1+2\rho G_U^n(z)}\right) = \Im\left(\frac{2\rho G^n(z)}{1+4(\rho G_U^n(z)+1/4)}\right)$$
$$\leq 2\rho\Im(G^n(z)) + 16\rho|G^n(z)||\rho G_U^n(z)+1/4|,$$

where we used that for $|a| \leq 1/2$, we have $|a/(1-a)| \leq 2|a|$. Consequently, since $\rho G^n(z)$ is uniformly bounded on $\mathbb{C}^+ \setminus \Delta_n$ and $\Im(G_n(z)) < 0$ there, we get

$$\Im\left(\frac{\rho G^n(z)}{1+2\rho G_U^n(z)}\right) \leq C\sqrt{\mathbf{e}_n + |O_1(n,z)|} \leq Cn^{-1/4}.$$

We thus conclude from the last display and (43) the existence of a constant C_3 such that if $\Im(z) > C_3 n^{-1/4}$, then

$$\Im(\psi_n(z)) = \Im(z) - \Im\left(\frac{\rho G^n(z)}{1+2\rho G_U^n(z)}\right) \geq \Im(z)/2,$$

as claimed. $\square$

From Lemma 12 we thus conclude the analyticity of $z \to \widetilde{O}(n,z,\psi_n(z))$ in $\{z : \Im(z) \geq C_3 n^{-1/4}\}$, and thus, due to (37) and (38), $\rho G^n(z)/(1+2\rho G_U^n(z))$ is also analytic there (compare with Lemma 11). In particular, the equality (38) extends by analyticity to this region.

We have made all preparatory steps in order to state the main result of this subsection.

LEMMA 13. *There exist positive finite constants C_6, C_7, C_8 such that, for $n > C_6$ and all $z \in \mathcal{E}_n := \{z : \Im(z) > n^{-C_7}\}$,*

$$|\Im G^n(z)| \leq C_8. \tag{44}$$

Moreover, the constants C_6, C_7, C_8 can be chosen uniformly on $\rho \le R$ for any finite R.

Proof. This is immediate from Lemmas 11 and 12, the definition of ψ_n, the assumption (3) on G_{T_n}, and the equality (38). $\square$

4. Tail estimates for ν_n^z

For $R > 0$, let $B_R = \{z \in \mathbb{C} : |z| \in [0, R]\}$. Our goal in this short section is to prove the following proposition.

PROPOSITION 14. (i) *Under the assumptions of Theorem* 1, *for Lebesgue almost every $z \in \mathbb{C}$,*

$$\lim_{\epsilon \downarrow 0} \limsup_{n \to \infty} E[1_{\mathcal{G}_n} \int_0^\epsilon \log|x| d\nu_n^z(x)] = 0\,. \tag{45}$$

Consequently, for any Lebesgue $z \in \mathbb{C}$,

$$\int \log|x| d\nu_n^z(x) \to \int \log|x| d\nu^z(x)\,, \tag{46}$$

in probability.

(ii) *Fix $R > 0$. For any smooth compactly supported deterministic function φ on B_R,*

$$\int \varphi(z) \int \log|x| d\nu_n^z(x) dm(z) \to \int \varphi(z) \int \log|x| d\nu^z(x) dm(z)\,, \tag{47}$$

in probability.

Before giving the proof of Proposition 14, we recall the following elementary lemma.

LEMMA 15. *Let μ be a probability measure on $\mathbb{R}$. For any real $y > 0$, it holds that*

$$\mu((-y, y)) \le 2y|\Im G(iy)|\,. \tag{48}$$

Proof. We have

$$-\Im(G(iy)) = \int \frac{y}{y^2 + x^2} \mu(dx) \ge \int_{-y}^{y} \frac{y}{y^2 + x^2} \mu(dx) \ge \frac{1}{2y} \mu((-y, y))\,,$$

from which (48) follows. $\square$

We can now provide the

Proof of Proposition 14. (i) Assume $z \in B_R$ for some $R > 0$. By (2), we can replace the lower limit of integration in (45) with $n^{-\delta}$. Let G_n^z denote the Stieltjes transform of $E[\nu_n^z]$. By Lemma 13 and Lemma 9, there exist positive constants $c_1 = c_1(R), c_2 = c_2(R)$ such that whenever $\Im(u) > n^{-c_1}$, it holds that $|\Im G_n^z(u)| < c_2$. We may assume that $c_1 < \delta$.

Since G_n^z is the Stieltjes transform of $E[\nu_n^z]$, by Lemma 15, we have for any $y > 0$ that

$$E[\nu_n^z((-y,y))] \le E[\nu_n^z((-y \vee n^{-c_1}, y \vee n^{-c_1}))] \le 2c_2 y \vee n^{-c_1}\,.$$

Thus, we get that for any $z \in B_R$ and with $\alpha \in [1,2]$,

$$\begin{aligned}
E&\left[\int_{n^{-\delta}}^{\epsilon} (|\log x|)^\alpha d\nu_n^z(x)\right] \\
&\le E\left[\int_{n^{-\delta}}^{n^{-c_1}} (|\log x|)^\alpha d\nu_n^z(x) + \int_{n^{-c_1}}^{\epsilon} (|\log x|)^\alpha d\nu_n^z(x)\right] \\
&\le (\delta \log n)^\alpha E[\nu_n^z((-n^{-c_1}, n^{-c_1}))] \\
&\quad + \sum_{j=0}^{J} E[\nu_n^z((-2^{(j+1)}n^{-c_1}, 2^{(j+1)}n^{-c_1}))](\log(2^j n^{-c_1}))^\alpha\,,
\end{aligned}$$

where $2^{J-1}n^{-c_1} < \epsilon \le 2^J n^{-c_1}$. Note that by Lemma 15 and the estimate on G_n^z, for $j \ge 0$,

$$E[\nu_n^z((-2^j n^{-c_1}, 2^j n^{-c_1}))] \le 2^{j+1} c_2 n^{-c_1}\,.$$

We conclude that

$$E\left[\int_{n^{-\delta}}^{\epsilon} |\log x|^\alpha d\nu_n^z(x)\right] \le C\epsilon |\log(\epsilon)|^\alpha\,, \tag{49}$$

where the constant $C = C(R)$. To obtain the estimate (45), we will consider $\alpha = 1$ and argue as follows. Due to (2), for $\alpha < 2$ we have

$$\begin{aligned}
E&\left[\mathbf{1}_{\mathcal{G}_n} \int_0^{n^{-\delta}} |\log x|^\alpha d\nu_n^z(x)\right] \\
&\le E[\mathbf{1}_{\mathcal{G}_n} \nu_n^z([-n^{-\delta}, n^{-\delta}]) \mathbf{1}_{\{\sigma_n^z < n^{-\delta}\}} |\log \sigma_n^z|^\alpha] \\
&\le E\left[\left(\nu_n^z([-n^{-\delta}, n^{-\delta}])\right)^{\frac{2}{2-\alpha}}\right]^{\frac{2-\alpha}{2}} E[\mathbf{1}_{\mathcal{G}_n} \mathbf{1}_{\{\sigma_n^z < n^{-\delta}\}} |\log \sigma_n^z|^2]^{\frac{\alpha}{2}}
\end{aligned}$$

by Hölder's inequality. The first factor goes to zero because

$$E\left[\left(\nu_n^z([-n^{-\delta}, n^{-\delta}])\right)^{\frac{2}{2-\alpha}}\right] \le E\left[\nu_n^z([-n^{-\delta}, n^{-\delta}])\right] \le 2c_2 n^{-c_1}.$$

By (2), the second factor is bounded by $(\delta')^{\alpha/2}$. We thus get (45) from (49). By Chebyshev's inequality, the convergence in expectation implies the convergence in probability and therefore for any $\delta, \delta' > 0$ there exists $\epsilon > 0$ small enough so that

$$\lim_{n\to\infty} P\left(\int_0^{\epsilon} |\log x| d\nu_n^z(x) > \delta\right) < \delta'.$$

On the other hand, $\int_\epsilon^\infty \log|x| d\nu_n^z(x)$ converges to $\int_\epsilon^\infty \log|x| d\nu^z(x)$ by the weak convergence of ν_n^z to ν^z in probability for any $\epsilon > 0$, and $\int_0^\epsilon \log|x| d\nu^z(x)$ converges to 0 as $\epsilon \to 0$ since ν^z has a bounded density by Lemma 9. Hence, we get (46).

(ii) Define the functions $f_n^i : B_R \to \mathbb{R}$, $i = 1, 2$, by

$$f_n^1(z) = \mathbf{1}_{\mathcal{G}_n} \mathbf{1}_{\|T_n\| \le M} \int_0^{n^{-\delta}} \log(x) d\nu_n^z(x)\,,$$
$$f_n^2(z) = \mathbf{1}_{\mathcal{G}_n} \mathbf{1}_{\|T_n\| \le M} \int_{n^{-\delta}}^\infty \log(x) d\nu_n^z(x)\,,$$

and set $f_n(z) = f_n^1(z) + f_n^2(z)$. Because ν_n^z is supported in B_{R+M} on $\|T_n\| \le M$ for all $z \in B_R$, f_n is bounded above by $\log(R + M)$. By (49), $E[|f_n^2(\cdot)|^2$ is bounded, uniformly in $z \in B_R$. On the other hand, by (2), again uniformly in $z \in B_R$, $E(f_n^1(z)^2) < \delta'$, and therefore

$$E \int_{\widetilde{B}_R} (f_n^1(z))^2 dm(z) < \infty\,.$$

Thus, $E \int_{\widetilde{B}_R} |f_n(z)|^2 dm(z) < \infty$, and in particular, the sequence of random variables

$$\int_{\widetilde{B}_R} \left| \mathbf{1}_{\mathcal{G}_n} \mathbf{1}_{\|T_n\| \le M} \int \log x d\nu_n^z(x) \right|^2 dm(z)$$

is bounded in probability. This uniform integrability and the weak convergence (46) are enough to conclude the proof by using dominated convergence (see [25, Lemma 3.1] for a similar argument). □

5. **Proof of Theorem 1**

It clearly suffices to prove the theorem for deterministic diagonal matrices T_n. (If T_n is random, use the independence of (U_n, V_n) from T_n to apply the deterministic version, after restricting attention to matrices T_n belonging to a set whose probability approaches 1.) By Proposition 14 (see (47)), we have, with $h(z) := \int \log|x| d\nu^z(x)$, that for any R and any smooth function ψ on $\widetilde{B}_R$,

$$\int \psi(z) dL_{A_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z)\, h(z) dm(z)\,,$$

in probability. Since the sequence L_{A_n} is tight, it thus follows that it converges, in the sense of distribution, to the measure

$$\mu_A := \frac{1}{2\pi} \Delta_z h(z)\,.$$

From Remark 8 (based on [12, Cor. 4.5]), we have that μ_A is a probability measure that possesses a radially symmetric density ρ_A satisfying the properties stated in parts b and c of the theorem. □

6. **Proof of Theorem 3**

We let X_n be as in the statement of the corollary and write $X_n = P_n T_n Q_n$ with P_n, Q_n unitary and T_n diagonal with entries equal to the singular values $\{\sigma_i^n\}$ of X_n. Obviously, $\{P_n, Q_n\}_{n\geq 1}$ is a sequence of independent, $\mathcal{H}_n$-distributed matrices. The joint distribution of the entries of T_n possesses a density on $\mathbb{R}_+^n$ which is given by the expression

$$\widetilde{Z}_n \prod_{i<j} |\sigma_i^2 - \sigma_j^2|^2 e^{-n\sum_{i=1}^n V(\sigma_i^2)} \prod_i \sigma_i d\sigma_i\,,$$

where $\widetilde{Z}_n$ is a normalization factor; see e.g. [1, Prop. 4.1.3]. Therefore, the squares of the singular values possess the joint density

$$\hat{Z}_n \prod_{i<j} |x_i - x_j|^2 e^{-n\sum_{i=1}^n V(x_i)} \prod_i dx_i$$

on $\mathbb{R}_+^n$. In particular, it falls within the framework treated in [21]. By part (i) of Theorem 2.1 there, positive constants M, C_{11} exist such that $P(\sigma_1 > M-1) \leq e^{-C_{11}n}$, and thus point 1 of the assumptions of Theorem 1 holds. By equations [21, (2.26) and (2.27)] and Chebycheff's inequality, we get that for z with $\Im(z) > n^{-\kappa'}$ where $\kappa < (1-\kappa')/2$,

$$P\left(|G_{T_n}(z) - G_{\widetilde{\Theta}}(z)| \geq \frac{1}{2\Im(z)n^\kappa}\right) \leq C|\Im(z)|^{-1} n^{2\kappa-1}\log n\,.$$

As the derivative of $G_{T_n} - G_{\widetilde{\Theta}}$ is bounded by a constant multiple of $1/|\Im(z)|^2$, a covering argument and summation show that for $\kappa' < 1/2$,

$$P\left(\sup_{\substack{z:|z|\leq M\\ \Im(z)\geq n^{-\kappa'}}} |G_{T_n}(z) - G_{\widetilde{\Theta}}(z)| \geq \frac{1}{\Im(z)n^\kappa}\right) \leq M n^{4\kappa+2\kappa'-1}\log n,$$

which goes to zero for $\kappa \in (0, (1-2\kappa')/4)$. Together with [21, eq. (2.32)], this proves point 3 of the assumptions. Thus, it remains only to check point 2 of the assumptions. Toward this end, define $\mathcal{G}_n = \{\sigma_1^n < M+1\}$ and note that we may and will restrict attention to $|z| < M+2$ when checking (2). We begin with the following proposition, due to [22].

PROPOSITION 16. *Let $\overline{A}$ be an arbitrary n-by-n matrix, and let $A = \overline{A} + \sigma N$ where N is a matrix with independent* (*complex*) *Gaussian entries of zero mean and unit variances. Let $\sigma_n(A)$ denote the minimal singular value of A. Then, there exists a constant C_{12} independent of $\overline{A}$, σ or n such that*

$$P(\sigma_n(A) < x) \leq C_{12} n \left(\frac{x}{\sigma}\right)^2\,. \tag{50}$$

The proof of Proposition 16 is identical to [22, Th. 3.3], with the required adaptation in moving from real to complex entries. (Specifically, in the right side of the display in [22, Lemma A.2], $\epsilon\sqrt{2/\pi}/\sigma$ is replaced by its square.) We omit further details.

On the event $\mathcal{G}_n$, all entries of the matrix X_n are bounded by a constant multiple of $\sqrt{n}$. Let N_n be a Gaussian matrix as in Proposition 16. With $\alpha > 2$ a constant to be determined below, set

$$\mathcal{G}_n' = \{\text{all entries of } n^{-\alpha/2}N_n \text{ are bounded by } 1\}\,.$$

Note that because $\alpha \geq 2$, on $\mathcal{G}_n'$, we have that $\sigma_1(n^{-\alpha}N_n) \leq 1$. Define $\overline{A}_n = zI - X_n$, $\tilde{A}_n = \overline{A}_n + n^{-\alpha}N_n\mathbf{1}_{\mathcal{G}_n'}$ and $A_n = \overline{A}_n + n^{-\alpha}N_n$. Then, by (50), with $\sigma_n(A_n)$ denoting the minimal singular value of A_n, we have

$$P(\sigma_n(A_n) < x; \mathcal{G}_n) \leq C_{12}x^2n^{1+2\alpha}\,. \tag{51}$$

If the estimate (51) concerned $\overline{A}_n$ instead of A_n, it would have been straightforward to check that point 2 of the assumptions of Theorem 1 holds (with an appropriately chosen δ, which would depend on α). Our goal is thus to replace, in (51), A_n by $\overline{A}_n$, at the expense of not too severe degradation in the right side. This will be achieved in two steps: first, we will replace A_n by $\tilde{A}_n$, and then we will construct on the same probability space the matrix X_n and a matrix Y_n so that Y_n is distributed like $X_n + n^{-\alpha}N_n\mathbf{1}_{\mathcal{G}_n'}$ but $P(Y_n \neq X_n)$ is small.

Turning to the construction, observe first that from (51),

$$P(\sigma_n(\tilde{A}_n) < x; \mathcal{G}_n) \leq C_{12}x^2n^{1+2\alpha} + P((\mathcal{G}_n')^c) \leq C_{12}[x^2n^{1+2\alpha} + n^2e^{-n^{\alpha}/2}]\,. \tag{52}$$

Let $X_n^{(\alpha)} = X_n + n^{-\alpha}N_n\mathbf{1}_{\mathcal{G}_n'}$. Let $\{\theta_i\}$ and $\{\mu_i\}$ denote the eigenvalues of $W_n = X_nX_n^*$ and of $W_n^{(\alpha)} = (X_n^{(\alpha)})(X_n^{(\alpha)})^*$, respectively, arranged in decreasing order. Note that the density of X_n is of the form

$$Z_n^{-1}e^{-n\,\mathrm{tr}(V(\mathbf{x}\mathbf{x}^*))}d\mathbf{x}\,,$$

where the variable $\mathbf{x} = \{x_{i,j}\}_{1\leq i,j\leq n}$ is matrix valued and $d\mathbf{x} = \prod_{1\leq i,j\leq n} dx_{i,j}$, while that of $X_n^{(\alpha)}$ is of the form

$$Z_n^{-1}E_N\left[e^{-n\,\mathrm{tr}(V((\mathbf{x}+\mathbf{1}_{\mathcal{G}_n'}n^{-\alpha}N_n)(\mathbf{x}+\mathbf{1}_{\mathcal{G}_n'}n^{-\alpha}N_n)^*))}\right]d\mathbf{x}\,,$$

where E_N denotes expectation with respect to the law of N_n, and Z_n is the same in both expressions. Note that $\sigma_1(X_n^{(\alpha)}) \in [\sigma_1(X_n) - 1, \sigma_1(X_n) + 1]$. Because $V(\cdot)$ is locally Lipschitz, we have that if either $\sigma_1(X_n) \leq M + 1$ or

$\sigma_1(X_n^{(\alpha)}) \le M+1$, then there exists a constant C_{13} independent of α so that

$$\begin{aligned}|\mathrm{tr}(V(W_n)-V(W_n^{(\alpha)}))| &\le \sum_{i=1}^n |V(\theta_i)-V(\mu_i)| \le C_{13}\sum_{i=1}^n |\theta_i-\mu_i| \\ &\le C_{13} n^{1/2}\left(\sum_{i=1}^n |\theta_i-\mu_i|^2\right)^{\frac12} \\ &\le C_{13} n^{1/2}\left(\mathrm{tr}((W_n-W_n^{(\alpha)})^2)\right)^{\frac12},\end{aligned}$$

where the Cauchy-Schwarz inequality was used in the third inequality and the Hoffman-Wielandt inequality in the next (see e.g. [1, Lemma 2.1.19]). On the event $\mathcal{G}_n$, all entries of $W_n - W_n^\alpha$ are bounded by $n^{(3-\alpha)/2}$. Therefore,

$$|\mathrm{tr}(V(W_n)-V(W_n^{(\alpha)}))| \le n^{(C_{14}-\alpha)/2}, \tag{53}$$

where the constant C_{14} does not depend on α. In particular, if $\alpha > (C_{14}+1)\vee 2$, we obtain that on $\mathcal{G}_n$, the ratio of the functions $f_n = e^{-n\,\mathrm{tr}(V(W_n))}$ and $g_n = e^{-n\,\mathrm{tr}(V(W_n^{(\alpha)}))}$ is bounded e.g. by $1+n^{(C_{14}+1-\alpha)/2}$; in particular,

$$\begin{aligned}P(\sigma_1(X_n^{(\alpha)})<M) &\le (1+n^{(C_{14}+1-\alpha)/2})P(\sigma_1(X_n)<M) \\ &\le (1+n^{(C_{14}+1-\alpha)/2})^2 P(\sigma_1(X_n^{(\alpha)})<M).\end{aligned}$$

Therefore, the variational distance between the law of X_n conditioned on $\sigma_1(X_n) < M$ and that of $X_n^{(\alpha)}$ conditioned on $\sigma_1(X_n^{(\alpha)}) < M$, is bounded by

$$4n^{(C_{14}+1-\alpha)/2}.$$

It follows that one can construct a matrix Y_n of law identical to the law of $X_n^{(\alpha)}$ conditioned on $\sigma_1(X_n^\alpha) < M$, together with X_n, on the same probability space so that

$$P(X_n \ne Y_n; \mathcal{G}_n) \le 4n^{(C_{14}+1-\alpha)/2} \le n^{C_{15}-\alpha/2}.$$

Combining this with (52), we thus deduce that

$$P(\sigma_n(\overline{A}_n) < x; \mathcal{G}_n) \le C_{12}x^2 n^{1+2\alpha} + n^{C_{16}-\alpha/2} \le n^{C_{17}} x^{2/5},$$

where α was chosen as function of x. This yields immediately point 2 of the assumptions of Theorem 1, if $\delta > 5C_{17}/2$.

We have checked now that in the setup of Theorem 3, all the assumptions of Theorem 1 hold. Applying now the latter theorem, we complete the proof of Theorem 3. □

Remark 17. The proof of Theorem 3 carries over to more general situations. Indeed, V does not need to be a polynomial; it is enough that its growth at infinity is polynomial and that it is locally Lipschitz, so that the results of [21] still apply. We omit further details.

7. Proof of Proposition 4

We take T_n satisfying the assumptions of Proposition 4 and consider $Y_n = U_n T_n V_n + n^{-\gamma} N_n$, with matrix of singular values $\widetilde{T}_n$. Note that $Y_n = \widetilde{U}_n \widetilde{T}_n \widetilde{V}_n$ with $\widetilde{U}_n, \widetilde{V}_n$ following the Haar measure. We first show that $\widetilde{T}_n$ also satisfies the assumptions of Theorem 1 when $\gamma > \frac{1}{2}$, except for the second one. Since the singular values of N_n follow the joint density of Theorem 3 with $V(x) = \frac{1}{2}x^2$, it follows from the previous section that $P(\|n^{-\frac{1}{2}} N_n\| > M) \leq e^{-C_{11} n}$ and therefore $\|\widetilde{T}_n\| \leq \|T_n\| + n^{-\gamma+\frac{1}{2}} \|n^{-\frac{1}{2}} N_n\|$ is bounded with overwhelming probability. Moreover, since $\widetilde{T}_n = |T_n + n^{-\gamma} U_n^* N_n V_n^*|$, on the event $\|N_n/\sqrt{n}\| \leq M$ we have

$$\left| G_{T_n}(z) - G_{\widetilde{T}_n}(z) \right| \leq \frac{E[\|\widetilde{T}_n - T_n\| \mathbf{1}_{\|N_n/\sqrt{n}\| \leq M}]}{|\Im(z)|^2} \leq \frac{C(\|T_n^{-1}\|, \|T_n\|)}{|\Im(z)|^2} n^{\frac{1}{2}-\gamma}$$

with $C(\|T_n^{-1}\|, \|T_n\|)$ a finite constant depending only on $\|T_n^{-1}\|, \|T_n\|$ which we assumed bounded. (In deriving the last estimate, we used that $\|(I+B)^{1/2} - I\| \leq \|B\|$ when $\|B\| < 1/2$.) As a consequence, the third condition is satisfied since

$$\left| G_{\widetilde{\Theta}}(z) - G_{\widetilde{T}_n}(z) \right| \leq \frac{C(\|T_n^{-1}\|, \|T_n\|)}{|\Im(z)|^2} n^{\frac{1}{2}-\gamma} + \frac{K}{n^{\kappa} |\Im(z)|} \leq \frac{K'}{n^{\gamma'} |\Im(z)|}$$

with $\gamma' = \min\{\kappa, \frac{1}{2}(\gamma - \frac{1}{2})\}$ and $\Im(z) \geq n^{-\max\{\frac{1}{2}(\gamma-\frac{1}{2}), \kappa'\}}$. Hence, the results of Lemma 13 hold and we need only check, as in Proposition 14, that with ν_n^z the empirical measure of the singular values of $zI - Y_n$,

$$I_n := E[1_{\mathcal{G}_n} \int_0^{n^{-\delta}} \log|x| d\nu_n^z(x)]$$

vanishes as n goes to infinity for some $\delta > 0$ and some set $\mathcal{G}_n$ with overwhelming probability. But $\overline{A}_n = zI - Y_n = zI - U_n T_n V_n + n^{-\gamma} \widetilde{N}_n$ with $\widetilde{N}_n$ a Gaussian matrix, and therefore we can use Proposition 16 to obtain (50) with $\sigma = n^{-\gamma}$, and the desired estimate on I_n. □

Proof of Example 5. The only point to check is (3). This follows because if $z = E + i\eta$ and $|\eta| \geq n^{-\kappa}$, then

$$\begin{aligned}
|\Im G_{T_n}(z)| &= \frac{1}{n} \sum_{i=1}^{n} \frac{\eta}{\eta^2 + |E - s_i|^2} \\
&\leq \sum_{k \geq 0} \frac{2\eta}{\eta^2 2^k} \frac{1}{n} \sharp\{i : (2^k - 1)\eta^2 \leq |s_i^n - E|^2 \leq 2^{k+1}\eta^2\} \\
&\leq C \sum_{k \geq 0} \frac{2\eta^2}{\eta^2 2^k} 2^{\frac{k+1}{2}} < \infty\,.
\end{aligned}$$

□

8. **Extension to orthogonal conjugation**

In this section, we generalize Theorem 1 to the case where we conjugate T_n by orthogonal matrices instead of unitary matrices.

THEOREM 18. *Let T_n be a sequence of diagonal matrices satisfying the assumptions of Theorem* 1. *Let $O_n, \widetilde{O}_n$ be two $n \times n$ independent matrices which follow the Haar measure on the orthogonal group and set $A_n = O_n T_n \widetilde{O}_n$. Then, L_{A_n} converges in probability to the probability measure μ_A described in Theorem* 1.

Proof. To prove the theorem, it is enough, following Section 5, to prove the analogue of Lemma 13 which in turn is based on the approximate Schwinger-Dyson equation (36) which is itself a consequence of equation (30) and concentration inequalities. To prove the analogue of (30) when U_n follows the Haar measure on the orthogonal group, observe that (31) remains true with $B^t = -B$ which only leaves the choice $B = \Delta(k,\ell) - \Delta(\ell,k)$ possible. However, taking this choice and summing over k, ℓ, yields, if we denote $\widetilde{m}(A \otimes B) = AB^t$,

$$E\left[\frac{1}{2n}\operatorname{tr}\otimes\frac{1}{2n}\operatorname{tr}(\partial P(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*))\right] = \frac{1}{2n}E\left[\frac{1}{2n}\operatorname{tr}((\widetilde{m}\circ\partial P)(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*))\right].$$

The right-hand side is small as $\widetilde{m} \circ \partial P$ is uniformly bounded. In fact, taking $P = (z_1 - \mathbf{Y}_n)^{-1}(z_2 - \mathbf{T}_n)^{-1}\mathbf{U}_n$, we find that $\tilde{m} \circ \partial P$ is uniformly bounded by $2/(|\Im(z_2)|(|\Im(z_1)| \wedge 1)^2)$ and therefore (32) holds once we add to $O(n, z_1, z_2)$ the above right-hand side which is at most of order $1/n|\Im(z_2)|(|\Im(z_1)| \wedge 1)^2$. Since our arguments did not require a very fine control on the error term, we see that this change will not affect them. Since concentration inequalities also hold under the Haar measure on the orthogonal group, see [1, Th. 4.27] and [1, Cor. 4.4.28], the proof of Theorem 1 can be adapted to this set-up. $\square$

9. **Proof of Proposition 6**

We use again Green's formula, and writing $\hat{B}_n = B_n + n^{-\gamma}N_n$ we have

$$\begin{aligned}\int \psi(z) dL_{\hat{B}_n+P_n}(z) &= \frac{1}{4\pi n}\int_{\mathbb{C}} \Delta\psi(z)\log\det(zI - \hat{B}_n - P_n)(zI - \hat{B}_n - P_n)^* dm(z) \\ &= \frac{1}{4\pi n}\int_{\mathbb{C}} \Delta\psi(z)\log\det(|zI - \hat{B}_n| - P_nU)(|zI - \hat{B}_n| - P_nU)^* dm(z),\end{aligned}$$

where we used the polar decomposition of $zI_n - \hat{B}_n$ to write $zI - \hat{B}_n = |zI - \hat{B}_n|U^*$ with U a unitary matrix. Since P_nU has the same law as P_n, we are back at the same setting as in the proof of Theorem 1, with $|zI - \hat{B}_n|$ replacing T_n. It is then straightforward to check that the same arguments work under our present hypotheses; the symmetrized empirical measure ν_n^z of the singular values of $T_n(z)+P_n$ converges to $\widetilde{\Theta}_z \boxplus \lambda_1$ by Lemma 10, which guarantees the convergence of

$$\int_{\epsilon}^{+\infty} \log|x| d\nu_n^z(x),$$

whereas our hypotheses allow us to bound uniformly the Stieltjes transform of ν_z^n on $\{z_1 : \Im(z_1) \geq n^{-C_7}\}$ as in Lemma 13, hence providing control of the integral on the interval $[n^{-C_7}, \epsilon]$. The control of the integral for $x < n^{-C_7}$ uses a regularization by the Gaussian matrix $n^{-\gamma} N_n$ as in Proposition 4 . $\square$

Acknowledgments. We thank Greg Anderson for many fruitful and encouraging discussions. We thank Yan Fyodorov for pointing out the paper [16] and Philippe Biane for suggesting that our technique could be applied to the examples in [3]. We thank the referee for a careful reading of the manuscript.

References

[1] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, *Cambridge Stud. Adv. Math.* **118**, Cambridge Univ. Press, Cambridge, 2010. MR 2760897. Zbl 1184.15023.

[2] Z. D. Bai, Circular law, *Ann. Probab.* **25** (1997), 494–529. MR 1428519. Zbl 0871.62018. http://dx.doi.org/10.1214/aop/1024404298.

[3] P. Biane and F. Lehner, Computation of some examples of Brown's spectral measure in free probability, *Colloq. Math.* **90** (2001), 181–211. MR 1876844. Zbl 0988.22004. http://dx.doi.org/10.4064/cm90-2-3.

[4] L. G. Brown, Lidskiuı's theorem in the type II case, in *Geometric Methods in Operator Algebras* (Kyoto, 1983), *Pitman Res. Notes Math. Ser.* **123**, Longman Sci. Tech., Harlow, 1986, pp. 1–35. MR 0866489. Zbl 0646.46058.

[5] B. Collins, A. Guionnet, and E. Maurel-Segala, Asymptotics of unitary and orthogonal matrix integrals, *Adv. Math.* **222** (2009), 172–215. MR 2531371. Zbl 1184.15024. http://dx.doi.org/10.1016/j.aim.2009.03.019.

[6] J. Feinberg and A. Zee, Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism, *Nuclear Phys. B* **501** (1997), 643–669. MR 1477381. Zbl 0933.82024. http://dx.doi.org/10.1016/S0550-3213(97)00419-7.

[7] Y. V. Fyodorov and H. J. Sommers, Spectra of random contractions and scattering theory for discrete-time systems, *JETP Lett.* **72** (2000), 422–426.

[8] J. Ginibre, Statistical ensembles of complex, quaternion, and real matrices, *J. Mathematical Phys.* **6** (1965), 440–449. MR 0173726. Zbl 0127.39304. http://dx.doi.org/10.1063/1.1704292.

[9] V. L. Girko, The circular law, *Teor. Veroyatnost. i Primenen.* **29** (1984), 669–679. MR 0773436. Zbl 0565.60034.

[10] F. GÖTZE and A. TIKHOMIROV, The circular law for random matrices, *Ann. Probab.* **38** (2010), 1444–1491. MR 2663633. Zbl 1203.60010. http://dx.doi.org/10.1214/09-AOP522.

[11] U. HAAGERUP, Spectral decomposition of all operators in a II_1 factor, which is embeddable in R^ω, 2001, preprint.

[12] U. HAAGERUP and F. LARSEN, Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras, *J. Funct. Anal.* **176** (2000), 331–367. MR 1784419. Zbl 0984.46042. http://dx.doi.org/10.1006/jfan.2000.3610.

[13] U. HAAGERUP and H. SCHULTZ, Brown measures of unbounded operators affiliated with a finite von Neumann algebra, *Math. Scand.* **100** (2007), 209–263. MR 2339369. Zbl 1168.46039. Available at http://www.mscand.dk/article.php?id=3017.

[14] ———, Invariant subspaces for operators in a general II_1-factor, *Publ. Math. Inst. Hautes Études Sci.* (2009), 19–111. MR 2511586. Zbl 1178.46058. http://dx.doi.org/10.1007/s10240-009-0018-7.

[15] U. HAAGERUP and S. THORBJØRNSEN, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group, *Ann. of Math.* **162** (2005), 711–775. MR 2183281. http://dx.doi.org/10.4007/annals.2005.162.711.

[16] A. HORN, On the eigenvalues of a matrix with prescribed singular values, *Proc. Amer. Math. Soc.* **5** (1954), 4–7. MR 0061573. Zbl 0055.00908. http://dx.doi.org/10.2307/2032094.

[17] J. B. HOUGH, M. KRISHNAPUR, Y. PERES, and B. VIRÁG, *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, *Univ. Lecture Ser.* **51**, Amer. Math. Soc., Providence, RI, 2009. MR 2552864. Zbl 1190.60038.

[18] N. LEHMANN and H.-J. SOMMERS, Eigenvalue statistics of random real matrices, *Phys. Rev. Lett.* **67** (1991), 941–944. MR 1121461. Zbl 0990.82528. http://dx.doi.org/10.1103/PhysRevLett.67.941.

[19] A. NICA and R. SPEICHER, R-diagonal pairs—a common approach to Haar unitaries and circular elements, in *Free Probability Theory* (Waterloo, ON, 1995), *Fields Inst. Commun.* **12**, Amer. Math. Soc., Providence, RI, 1997, pp. 149–188. MR 1426839. Zbl 0889.46053.

[20] G. PAN and W. ZHOU, Circular law, extreme singular values and potential theory, *J. Multivariate Anal.* **101** (2010), 645–656. MR 2575411. Zbl 1203.60011. http://dx.doi.org/10.1016/j.jmva.2009.08.005.

[21] L. PASTUR and M. SHCHERBINA, Bulk universality and related properties of Hermitian matrix models, *J. Stat. Phys.* **130** (2008), 205–250. MR 2375744. Zbl 1136.15015. http://dx.doi.org/10.1007/s10955-007-9434-6.

[22] A. SANKAR, D. A. SPIELMAN, and S.-H. TENG, Smoothed analysis of the condition numbers and growth factors of matrices, *SIAM J. Matrix Anal. Appl.* **28** (2006), 446–476. MR 2255338. Zbl 1179.65033. http://dx.doi.org/10.1137/S0895479803436202.

[23] P. Śniady, Random regularization of Brown spectral measure, *J. Funct. Anal.* **193** (2002), 291–313. MR 1929504. Zbl 1026.46056. http://dx.doi.org/10.1006/jfan.2001.3935.

[24] T. Tao and V. Vu, Random matrices: the circular law, *Commun. Contemp. Math.* **10** (2008), 261–307. MR 2409368. Zbl 1156.15010. http://dx.doi.org/10.1142/S0219199708002788.

[25] ———, Random matrices: universality of ESDs and the circular law, *Ann. Probab.* **38** (2010), 2023–2065, with an appendix by Manjunath Krishnapur. MR 2722794. Zbl 1203.15025. http://dx.doi.org/10.1214/10-AOP534.

[26] D. Voiculescu, Limit laws for random matrices and free products, *Invent. Math.* **104** (1991), 201–220. MR 1094052. Zbl 0736.60007. http://dx.doi.org/10.1007/BF01245072.

[27] ———, The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information, *Adv. Math.* **146** (1999), 101–166. MR 1711843. http://dx.doi.org/10.1006/aima.1998.1819.

[28] Y. Wei and Y. V. Fyodorov, On the mean density of complex eigenvalues for an ensemble of random matrices with prescribed singular values, *J. Phys. A* **41** (2008), 502001. MR 2515905. Zbl 1154.81343. http://dx.doi.org/10.1088/1751-8113/41/50/502001.

[29] H. Weyl, Inequalities between the two kinds of eigenvalues of a linear transformation, *Proc. Nat. Acad. Sci. U. S. A.* **35** (1949), 408–411. MR 0030693. Zbl 0032.38701.

[30] K. Życzkowski and H.-J. Sommers, Truncations of random unitary matrices, *J. Phys. A* **33** (2000), 2045–2057. MR 1748745. Zbl 0957.82017. http://dx.doi.org/10.1088/0305-4470/33/10/307.

(Received: September 15, 2009)
(Revised: October 18, 2010)

UMPA, Ecole Normale Supérieure de Lyon, Lyon, France
E-mail: aguionne@umpa.ens-lyon.fr

Indian Institute of Science, Bangalore, India
E-mail: manju@math.iisc.ernet.in

University of Minnesota, Minneapolis, MN and
Weizmann Institute, Rehovot, Israel
E-mail: zeitouni@math.umn.edu