The single ring theorem

Alice Guionnet*, Manjunath Krishnapur† and Ofer Zeitouni‡

September 10, 2009

Keywords Random matrices, non-commutative measure, Schwinger-Dyson equation.
Mathematics Subject of Classification : 15A52 (46L50,46L54)

Abstract

We study the empirical measure L_{A_n} of the eigenvalues of non-normal square matrices of the form $A_n = U_n D_n V_n$ with U_n, V_n independent Haar distributed on the unitary group and D_n real diagonal. We show that when the empirical measure of the eigenvalues of D_n converges, and D_n satisfies some technical conditions, L_{A_n} converges towards a rotationally invariant measure on the complex plane whose support is a single ring. In particular, we provide a complete proof of Feinberg-Zee single ring theorem [5]. We also consider the case where U_n, V_n are independent Haar distributed on the orthogonal group.

1 The problem

Horn [15] asked the question of describing the eigenvalues of a square matrix with prescribed singular values. If A is a $n \times n$ matrix with singular values $s_1 \geq \ldots \geq$

*UMPA, CNRS UMR 5669, ENS Lyon, 46 allée d'Italie, 69007 Lyon, France. aguionne@umpa.ens-lyon.fr. This work was partially supported by the ANR project ANR-08-BLAN-0311-01.

†Department of Mathematics, Indian Institute of Science, Bangalore - 560012, India. manju@math.iisc.ernet.in

‡School of Mathematics, University of Minnesota and Faculty of Mathematics, Weizmann Institute, POB 26, Rehovot 76100, Israel. zeitouni@math.umn.edu. The work of this author was partially supported by NSF grant DMS-0804133 and by a grant from the Israel Science Foundation.

$s_n \geq 0$ and eigenvalues $\lambda_1, \ldots, \lambda_n$ in decreasing order of absolute values, then the inequalities

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

were shown by Weyl [25] to hold. Horn 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 $s_1 \geq \ldots \geq s_n \geq 0$ and consider $n \times n$ matrices with these singular values. They are of the form $A = PDQ$, where D 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 Haar measure on $\mathcal{U}(n)$, the unitary group of $n \times n$ matrices. Let $\lambda_1, \ldots, \lambda_n$ 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 [7, Theorem 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 [6] for the case where D is a projection). The authors of [7] explicitely 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 asymetric variant, see [17]. (There are some pitfalls in the standard derivation of Ginibre's result. We refer to [16] for a discussion.) Another situation is the truncation of random unitary matrices, described in [26].

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 [22]. Such matrices, however, do not possess the invariance properties discussed in connection of question 2. The *single ring theorem* of Feinberg and Zee [5] is, to our knowledge, the first example where a partial answer to this question is offered. (Various issues of convergence are glossed over in [5] 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 non-normal operators, introduced in [3].) R-diagonal operators were introduced by Nica and Speicher [18] in the context of free probability; they represent the weak*-limit (or more precisely, the limit in $*$-moments) of operators of the form UD with U unitary with size going to infinity and D 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 [12, 13].

2 Limiting spectral density of a non-normal 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 D_n denote a sequence of (possibly random) diagonal matrices with real positive entries $S_n = \{s_i^{(n)}\}$ on the diagonal, and introduce the *empirical measure* of the *symmetrized* version of D_n as

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

We write G_{D_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 $A_n = P_n D_n Q_n$, 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. Finally, for any matrix A, we set $\|A\|$ to denote the ℓ^2 operator-norm of A, that is, its largest singular value.

The main result of this paper is the following.

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

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

$$\lim_{n\to\infty} P(\|D_n\| > M) = 0. \tag{2}$$

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 any $z \in \mathbb{C}$, with σ_n^z the minimal singular value of $zI - A_n$,*

$$E(\mathbf{1}_{\mathcal{G}_n}\mathbf{1}_{\{\sigma_n^z < n^{-\delta}\}}(\log\sigma_n^z)^2) < \delta'. \tag{3}$$

3. *There exist constants $K, \kappa, \kappa' > 0$ such that, for all n large,*

$$|G_{\tilde{\Theta}}(z) - G_{D_n}(z)| \le \frac{K}{n^{\kappa}\Im(z)}, \quad \text{if } \Im(z) > n^{-\kappa'}. \tag{4}$$

4. *There exists a constant κ_1 such that*

$$|G_{\tilde{\Theta}}(z)| \le \kappa_1 \quad \text{on} \quad \mathbb{C}^+. \tag{5}$$

Then the following holds.

a. *L_{A_n} converges in probability to a limiting measure μ_A, rotationally invariant in $\mathbb{C}$.*

b. *The measure μ_A possesses a radially-symmetric density ρ_A with respect to the Lebesgue measure on $\mathbb{C}$, which is characterized as follows. For $z \in \mathbb{C}$, let $\nu^z := \tilde{\Theta} \boxplus \lambda_{|z|}$, where $\lambda_{|z|} = \frac{1}{2}(\delta_{|z|} + \delta_{-|z|})$ is the symmetric Bernoulli measure with atoms at $\{-|z|, |z|\}$, and $\boxplus$ denotes free convolution. Then, $\rho_A(z) = \frac{1}{2\pi}\Delta_z(\int \log|x| d\nu^z(x))$, where Δ_z denotes the Laplacian with respect to the variable z.*

c. *The support of μ_A is a single ring: there exist constants $0 \le a < b < \infty$ so that*
$$\operatorname{supp}\mu_A = \{re^{i\theta} : a \le r \le b\}\,.$$
Further, $a = 0$ if and only if $\int x^{-2} d\Theta(x) = \infty$.

See Remark 6 for an explicit characterization of the free convolution appearing in Theorem 1, and [1, Ch. 5] for general background. A different characterization of ρ_A, borrowed from [11], is provided next.

Remark 2. We provide, following [11, Theorem 4.4 and Corollary 4.5], an alternative characterization of ρ_A and its support. Recall that $\Theta(\{0\}) = 0$ by Assumption 5. 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 $\{L_{D_n^2}\}$. Let $\mathcal{S}$ denote the S-transform of $\Theta^{\sharp 2}$ (see [11, Section 2] for the definition of the S transform of a probability measure on $\mathbb{R}$ and its relation to the R transform). Define $F(t) = \mathcal{S}(1/\sqrt{t-1})$ on $\mathcal{D} = (0,1]$. Then, F maps $\mathcal{D}$ to the interval
$$(a,b] = (\frac{1}{(\int x^{-2} d\Theta(x))^{1/2}}, (\int x^2 d\Theta(x))^{1/2}]\,,$$
and has an analytic continuation to a neighborhood of $\mathcal{D}$, and $F' > 0$ on $\mathcal{D}$. Further, with $\rho_A(re^{i\theta}) = \rho_A(r)$, it holds that
$$\rho_A(r) = \begin{cases} \frac{1}{2\pi r F'(F^{-1}(r))}\,, & r \in (a,b]\,, \\ 0\,, & otherwise. \end{cases} \tag{6}$$
Finally, ρ_A has an analytic continuation to a neighborhood of $(a,b]$.

Theorem 1 is generalized to the case where U_n, V_n follow the Haar measure on the orthogonal group in Theorem 16.

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

Corollary 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*

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

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 minizer of the functional

$$\int V(x^2)d\mu(x) - \int\int \log|x^2-y^2|d\mu(x)d\mu(x)$$

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

Corollary 3 will follow by checking that the assumptions of Theorem 1 are satisfied for the spectral decomposition $X_n = U_nD_nV_n$, see Section 6.

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

Corollary 4. *Let $(D_n)_{n\geq 0}$ be a sequence of matrices satisfying the assumptions of Theorem 1 except for* (3) *and assume that $\|D_n^{-1}\|$ is uniformly bounded. Let N_n be a $n\times n$ matrix with independent (complex) Gaussian entries of zero mean and covariance equal identity. Let U_n, V_n follow the Haar measure on unitary $n\times n$ matrices. Then, the empirical measure of the eigenvalues of $Y_n := U_nD_nV_n + n^{-\gamma}N_n$ converges weakly in probability to μ_A as in Theorem 1 for any $\gamma\in(\frac{1}{2},\infty)$.*

Example 5. An exemple of sequence $(D_n)_{n\geq 0}$ satisfying the hypotheses of Corollary 4 is given as follows: take μ a compactly supported probability measure on $\mathbb{R}^{+*}$. Assume the inverse F^{-1} of the distribution function $F(x) = \mu([0,x])$ is Hölder continuous and that μ has a uniformly bounded Stieltjes transform on $\mathbb{C}^+$. Then the diagonal matrix D_n with entries

$$s_i^n = \inf\{s : \mu([0,s]) \geq \frac{i}{n}\}, \quad 1\leq i\leq n\,,$$

satisfies the hypotheses of Corollary 4.

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 non-normal 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

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

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

$$H_n^z := \begin{bmatrix} 0 & zI-A_n \\ (zI-A_n)^* & 0 \end{bmatrix}. \tag{7}$$

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,

$$\int \frac{1}{y-x}d\nu_n^z(x) = \frac{1}{2n}\mathrm{tr}\left((y-H_n^z)^{-1}\right),$$

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

$$\int \psi(z)dL_{A_n}(z) = \frac{1}{2\pi}\int_{\mathbb{C}} \Delta\psi(z)\left[\int_{\mathbb{R}}\log|x|d\nu_n^z(x)\right]dm(z). \tag{8}$$

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 weak convergence 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 (8) that

$$\int \psi(z) dL_{A_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z)\, h(z) dm(z) . \tag{9}$$

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

$$\int \psi(z) dL_{A_n}(z) \to \frac{1}{2\pi} \int_{\mathbb{C}} \psi(z)\, \Delta h(z) dm(z) , \tag{10}$$

   which identifies $\Delta h(z)$ 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 (8) and these five steps give a general recipe for finding limiting spectral measures of non-normal 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 [19] and Tao and Vu [22]. While we heavily borrow from that sequence, a major difficulty in the problem considered here is that no independence between entries of the matrix A_n is present here. 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 obtain fine estimates on the Stieltjes transform of Gaussian generalized band matrices in [14]. In [4], they were used to study the asymptotics of matrix models on the unitary group. Our approach combines ideas of [14] to estimate Stieltjes transform and the necessary adaptations to unitary matrices as developped in [4]. 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 [11]. The Schwinger–Dyson equations give both a characterization of the limit and, more important to us, a discrete 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 two above, but only once an a-priori (polynomial) estimate on the minimal singular value has been obtained. The latter is deduced from assumption 3. In the context of the Feinberg–Zee single ring theorem, the latter assumption holds due to an adaptation of the analysis of [21].

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 (7). Note that for $z \neq 0$, L_n^z is also the ESD of the matrix $\tilde{H}_n^z$ given by

$$\begin{aligned}\tilde{H}_n^z &:= |z| \begin{bmatrix} 0 & W_n^z - D_n/|z| \\ (W_n^z - D_n/|z|)^* & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & Q_n \\ P_n^* & 0 \end{bmatrix} \begin{bmatrix} 0 & zI - A_n \\ \bar{z}I - A_n^* & 0 \end{bmatrix} \begin{bmatrix} 0 & P_n \\ Q_n^* & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & Q_n \\ P_n^* & 0 \end{bmatrix} H_n^z \begin{bmatrix} 0 & P_n \\ Q_n^* & 0 \end{bmatrix}, \end{aligned} \tag{11}$$

where $W_n^z = \bar{z}Q_nP_n/|z|$ is unitary and $\mathcal{H}_n$ distributed. We are thus led to the study of the ESD for a sequence of matrices of the form

$$\mathbf{Y}_n = \begin{pmatrix} 0 & B_n \\ B_n^* & 0 \end{pmatrix} \tag{12}$$

with $B_n = U_n + T_n$, T_n being a real, diagonal matrix, and U_n a $\mathcal{H}_n$ unitary matrix.

We denote in short

$$\mathbf{U}_n = \begin{pmatrix} 0 & U_n \\ 0 & 0 \end{pmatrix}, \quad \mathbf{U}_n^* = \begin{pmatrix} 0 & 0 \\ U_n^* & 0 \end{pmatrix}, \quad \mathbf{T}_n = \begin{pmatrix} 0 & T_n \\ T_n & 0 \end{pmatrix}. \tag{13}$$

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 non-commutative probability space $(\mathcal{A}, *, \mu)$ on which variable a variable U lives and where μ is a tracial state satisfying the relations $\mu((UU^* - 1)^2) = 0$, $\mu(U^a) = 0$ for $a \in \mathbb{Z} \setminus \{0\}$. μ is the unique non-commutative law of bounded variables which is invariant under unitary conjuguation, and therefore corresponds to the asymptotics of the Haar measure. In the sequel, 1 will denote the identity in $\mathcal{A}$. We refer to [1, Section 5.2] for definitions.

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

$$\partial PQ = \partial P \times 1 \otimes Q + P \otimes 1 \times \partial Q, \tag{14}$$

$$\partial U = U \otimes 1, \partial U^* = -1 \otimes U^*, \partial T = 0 \otimes 0.$$

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

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

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

By the invariance of μ under unitary conjuguation, see [24, Proposition 5.17] or [1, (5.4.31)], we have the Schwinger-Dyson equation

$$\mu\otimes\mu(\partial P)=0\,. \tag{15}$$

We continue to use the notation $\mathbf{Y}$, $\mathbf{U},\mathbf{U}^*$ and $\mathbf{T}$ in a way similar to (12) and (13). So, we let $\mathbf{Y}=\mathbf{U}+\mathbf{U}^*+\mathbf{T}$ with

$$\mathbf{U}=\begin{pmatrix}0&U\\0&0\end{pmatrix},\quad \mathbf{U}^*=\begin{pmatrix}0&0\\U^*&0\end{pmatrix},\quad \mathbf{T}=\begin{pmatrix}0&T\\T&0\end{pmatrix}. \tag{16}$$

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 non-commutative derivative ∂ extends naturally to the algebra generated by the matrix-valued $\mathbf{U},\mathbf{U}^*,\mathbf{T}$, using the Leibniz rule (14) together with the relations

$$\partial\mathbf{U}=\mathbf{U}\otimes p\,,\quad \partial\mathbf{U}^*=-p\otimes\mathbf{U}^*\,, \tag{17}$$

where we denoted $p=\begin{pmatrix}0&0\\0&1\end{pmatrix}$. In the sequel we shall 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

$$\partial(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}=$$
$$b(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}(\mathbf{U}\otimes p-p\otimes\mathbf{U}^*)(z-b\mathbf{U}-b\mathbf{U}^*-a\mathbf{T})^{-1}$$

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

$$\begin{aligned}
G(z_1,z_2) &= \mu\big((z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\big)\,, \\
G_U(z_1,z_2) &= \mu\big(\mathbf{U}(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\big)\,, \\
G_U(z_1) &= \mu\big(\mathbf{U}(z_1-\mathbf{Y})^{-1}\big)\,, \\
G_{U^*}(z_1,z_2) &= \mu\big(\mathbf{U}^*(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\big)\,, \\
G_T(z_1,z_2) &= \mu\big(\mathbf{T}(z_1-\mathbf{Y})^{-1}(z_2-\mathbf{T})^{-1}\big)\,, \\
G(z_1) &= \mu\big((z_1-\mathbf{Y})^{-1}\big)\,, \\
G_T(z_2) &= \mu\big((z_2-\mathbf{T})^{-1}\big)\,.
\end{aligned} \tag{18}$$

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 (14) and (17),

$$\partial P = P \otimes p + (z_1-\mathbf{Y})^{-1}\mathbf{U} \otimes pP - (z_1-\mathbf{Y})^{-1} p \otimes \mathbf{U}^* P. \tag{19}$$

For any smooth function Q,

$$\mu(\mathbf{U}^* Q \mathbf{U}) = \mu((1-p)Q)$$

due to the traciality of μ and $\mathbf{U}\mathbf{U}^* = 1-p$. Further, $Pp = P$ and thus $\mu(pP) = \mu(P)$, and 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 (U,U^*) by (U^*,U), which has the same law)

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

The same equality holds without the last factor $(z_2-\mathbf{T})^{-1}$, and so we get from (19)

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

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

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

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

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

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

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

(Again, here and in the rest of this subsection, the proper branch of the square root is determined by analyticity.) Let R denote the R-transform of the Bernoulli law $\lambda_1 := (\delta_{-1}+\delta_{+1})/2$, that is,

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

see [1, Definition 5.3.22 and Exercise 5.3.27], so that we have

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

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

$$\begin{aligned} G_T(z_1,z_2) &= z_2G(z_1,z_2) - G(z_1)\,, & (24)\\ 2G_U(z_1,z_2)+G_T(z_1,z_2) &= z_1G(z_1,z_2) - G_T(z_2)\,. & (25)\end{aligned}$$

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

$$G(z_1,z_2)R(G(z_1)) + z_2G(z_1,z_2) - G(z_1) = z_1G(z_1,z_2) - G_T(z_2)\,, \tag{26}$$

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

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

Thus,

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

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

The relation (28) is the Schwinger-Dyson equation in our setup. It 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 $z \mapsto z - R(G(z))$ possesses a non-vanishing 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 6. 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 the law of 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_1, i.e. $\mu_Y = \tilde{\mu}_T \boxplus \lambda_1$, where $\lambda_z = (\delta_{-|z|} + \delta_{|z|})/2$ is the Bernoulli law that puts mass $\frac{1}{2}$ at $\pm|z|$.

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

Lemma 7. *If* $|G_T(\cdot)| \leq \kappa_1$ *on* $\mathbb{C}^+$ *then* $|G(\cdot)| \leq \kappa_1$ *on* $\mathbb{C}^+$.

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

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 (4) and (5) are satisfied with T_n and the spectral distribution of T replacing D_n and Θ. Recall first, see [1, (5.4.29)], that by invariance of the Haar measure under unitary conjuguation, with $P \in \mathbb{C}\langle T, U, U^* \rangle$ a noncommutative polynomial (or a product of Stieltjes functionals),

$$E[\frac{1}{2n}\mathrm{tr} \otimes \frac{1}{2n}\mathrm{tr}(\partial P(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*))] = 0 \,. \tag{29}$$

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,

$$0 = \partial_t E[(P(\mathbf{T}_n, \mathbf{U}_n(t), \mathbf{U}_n^*(t)))_{k,\ell}] = E[(\partial P(\mathbf{T}_n, \mathbf{U}_n, \mathbf{U}_n^*)\sharp\mathbf{B})_{k,\ell}] \tag{30}$$

with $\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}$. 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 (29).

We denote by G^n the quantities as defined in (18), but with $E[\frac{1}{2n}\mathrm{tr}]$ replacing μ and the subscript n attached to all variables, so that for instance

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

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

$$\frac{1}{2}G^n_U(z_1,z_2) = -G^n_U(z_1,z_2)G^n_U(z_1) + \frac{1}{4}G^n(z_1,z_2)G^n(z_1) + O(n,z_1,z_2)\,, \qquad (31)$$

with

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

Further, by the standard concentration inequality for $\mathcal{H}_n$, see [1, Corollary 4.4.31], for any smooth function $P : \mathcal{U}(n) \to \mathbb{C}$,

$$\left| E\left[\left(\frac{1}{2n}\mathrm{tr}(P) - E[\frac{1}{2n}\mathrm{tr}](P)\right)^2\right] \right| \le \frac{1}{n^2}\|P\|_L^2\,, \qquad (32)$$

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 $m(A \otimes B) = BA$. Applying (32) to each term of ∂P (recall formula (19)), we get that for $\Im(z_1) > 0$,

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

Multiplying by z_2 and taking the limit as $z_2 \to \infty$ we deduce from (31) that

$$(G^n(z_1))^2 = 2G^n_U(z_1)(1 + 2G^n_U(z_1)) - O_1(n,z_1)\,, \qquad (33)$$

where

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

In particular,

$$G_U^n(z_1) = \frac{1}{4}(-1 + \sqrt{1 + 4G^n(z_1)^2 + 4O_1(n, z_1)}), \tag{34}$$

with again the choice of the square root determined by analyticity.

Recalling that (24) and (25) remain true when we add the subscript n and combining (31) and (24), we get

$$G^n(z_1, z_2)\left(\frac{G^n(z_1)}{(1+2G_U^n(z_1))} + z_2 - z_1\right) = G^n(z_1) - G_T^n(z_2) + \tilde{O}(n, z_1, z_2), \tag{35}$$

with

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

Hence, if we define

$$z_2 = \psi_n(z_1) := z_1 - \frac{G^n(z_1)}{(1+2G_U^n(z_1))}, \tag{36}$$

then

$$G^n(z_1) = G_T^n(z_2) + \tilde{O}(n, z_1, z_2),$$

and therefore

$$G^n(z_1) = G_T^n(\psi_n(z_1)) + \tilde{O}(n, z_1, \psi_n(z_1)). \tag{37}$$

Equation (37) holds at least when $\Im z_2 > 0$ for z_2 as in (36). 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 6).

Lemma 8. *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 $\mu_Y = \tilde{\mu}_T \boxplus \lambda_1$. In particular, if L_{D_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 $\tilde{\Theta} \boxplus \lambda_{|z|}$.*

(Recall that $\tilde{\Theta}$ is the symmetrized version of Θ and note that for $z = 0$, the statement of the lemma is trivial.)

Lemma 8 completes the proof of Step one in our program. To be able to complete Step two, we need to obtain quantitative information from the (finite n) Schwinger-Dyson equations (37): 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 constant C and $\tilde{O}(n, z_1, \psi_n(z_1))$ is analytic and bounded there. This will imply that (37) 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 $G_U^n(\cdot)$ are analytic in $\mathbb{C}^+$. We have the following.

Lemma 9. *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*

$$|1 + 2G_U^n(z)| > C_2[\Im(z)^3 \wedge 1]. \tag{38}$$

Proof Since $G_U^n(z)$ is asymptotic to $1/z$ at infinity, we may and will restrict attention to some fixed ball $B_R \subset \mathbb{C}$, whose interior contains the support of $\mathbf{Y}$. 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 $z, x \in B(0, R)$

$$|G^n(z)| \geq |\Im(G^n(z))| \geq \frac{|\Im(z)|}{4R^2}. \tag{39}$$

Moreover, since $|G_U^n(z)| \leq 1/|\Im(z)|$, for some constant c independent of n and all n large, we deduce from (33) that

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

Combining this estimate and (39), we get that

$$\frac{2|1 + 2G_U^n(z)|}{|\Im(z)|} \geq \frac{|\Im(z)|^2}{16R^4} - \frac{c}{n^2 \Im(z)^2 (\Im(z) \wedge 1)} \geq \frac{|\Im(z)|^2}{32R^4}, \tag{40}$$

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 9 and the analyticity of G^n and G^n_U 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 $z \to \tilde{O}(n, z, \psi_n(z))$ 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 10. *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, we may and will restrict attention to $\Im(z) \leq R$ for some fixed R. We divide the proof to two cases, as follows. Let $\mathbf{e}_n = n^{-1/2}$, and set $\Delta_n = \{z \in \mathbb{C}^+ : |G^n(z) + i/2| \geq \mathbf{e}_n\}$.

Then, for any $z \in \Delta_n$, and whatever choice of branch of the square root made in (34), 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 $r(w) = \frac{1}{4}(-1+\sqrt{1+4w^2})$ is Lipschitz in the sense that

$$|G^n_U(z) - r(G^n(z))| \leq C \mathbf{e}_n^{-\frac{1}{2}} O_1(n,z) . \tag{41}$$

On the other hand, again from (34),

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

Combining the last display with the relation $R(\theta) = 2r(\theta)/\theta$, (41) and (39), one obtains that for $z \in \Delta_n$,

$$\begin{aligned}
\left| \frac{G^n(z)}{1+2G^n_U(z)} - R(G^n(z)) \right| &\leq \left| \frac{2r(G^n(z))}{G^n(z)} - \frac{2G^n_U(z)}{G^n(z)} \right| + \left| \frac{G^n(z)}{1+2G^n_U(z)} - \frac{2G^n_U(z)}{G^n(z)} \right| \\
&\leq C \frac{|O_1(n,z)|}{|G^n(z)(1+2G^n_U(z))|} + C \frac{|O_1(n,z)|}{\mathbf{e}_n^{\frac{1}{2}} |G^n(z)|} \\
&\leq C \frac{|O_1(n,z)|}{\mathbf{e}_n^{1/2} |\Im(z)|} + C \frac{|O_1(n,z)|}{\Im(z)^4} \\
&\leq \frac{C}{n^2 |\Im(z)|^4} \left(\frac{1}{\mathbf{e}_n^{1/2}} + \frac{1}{|\Im(z)|^3} \right) \\
&\leq \frac{C}{n^2 |\Im(z)|^4} \left(n^{1/4} + \frac{1}{|\Im(z)|^3} \right) .
\end{aligned} \tag{42}$$

Since the above right hand side is smaller than $\Im(z)/2$ for $\Im(z) > n^{-1/4}$ and $\Im(R(G^n(z))) \leq 0$, we conclude that for $z \in \Delta_n \cap \{\Im(z) > n^{-1/4}\}$

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

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

On the other hand, when $z \in \mathbb{C}^+ \setminus \Delta_n$ and $\Im(z) > n^{-1/4}$, then we have from (34) that

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

Thus, under these conditions,

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

where we finally used that on Δ_n, $G^n(z)$ is uniformly bounded and so that $\Im(G^n(z)) \leq -1/4$ for $z \in \mathbb{C}^+$ and $n \geq 2$. We thus conclude from the last display and (42) the existence of a constant C_3 such that if $\Im(z) > C_3 n^{-1/4}$ then

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

as claimed. $\square$

From Lemma 10 we thus conclude the analyticity of $z \to \tilde{O}(n,z,\psi_n(z))$ in $\{z : \Im(z) \geq C_3 n^{-1/4}\}$, and thus $G^n(z)/(1+2G_U^n(z))$ is also analytic there. In particular, the equality (37) extends by analyticity to this region.

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

Lemma 11. *There exist positive finite constants C_6, C_7, C_8 such that, for $n > C_6$ and all $z \in \mathcal{E}_n := \{z : \Im(z) > n^{-C_7}\}$,*

$$|G^n(z)| \leq C_8\,. \tag{43}$$

Proof This is immediate from Lemma 9, Lemma 10, the definition of ψ_n, the assumption of G_T^n and the equality (37). $\square$

4 Tail estimates for ν_n^z

Our goal in this short section is to prove the following proposition.

Proposition 12. *(i) Fix $z \in \mathbb{C}$. Under the assumptions of Theorem 1,*

$$\lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} E[1_{\mathcal{G}_n} \int_0^{\varepsilon} \log|x| d\nu_n^z(x)] = 0 . \tag{44}$$

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

$$\int \log|x| d\nu_n^z(x) \to \int \log|x| d\nu^z(x) , \tag{45}$$

in probability.
(ii) For any smooth compactly supported deterministic function φ on $\mathbb{C}$,

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

in probability.

Before bringing the proof of Proposition 12, we recall the following elementary lemma.

Lemma 13. *Let μ be a probability measure on $\mathbb{R}$. For any real $y > 0$, it holds that*

$$\mu((-y,y)) \le 2y|G(iy)| . \tag{47}$$

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 (47) follows. □

We can now provide the
Proof of Proposition 12

Let R be large enough so that $B_R \subset \mathbb{C}$ contains the support of φ. Throughout this proof, we may and will restrict attention to z satisfying $|z| \le R$.

(i) By (3), we can replace the lower limit of integration in (44) with $n^{-\delta}$. Let G_n^z denote the Stieltjes transform of $E[\nu_n^z]$. For $z \neq 0$, by Lemma 11 and Lemma 7, there exist positive constants $c_1(R), c_2(R)$ such that whenever $\Im(u) > n^{-c_1}$, it holds that $|G_n^z(u)| < c_2$. For $z=0$, L_n^0 is simply the ESD of D_n, once symmetrized, and hence G_n^0 is also uniformly bounded by (4) and (5).

Therefore, since G_n^z is the Stieltjes transform of $E[\nu_n^z]$, by Lemma 13, for any $y>0$,

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

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

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

where $2^{J-1}n^{-c_1} < \varepsilon \leq 2^J n^{-c_1}$. Note that by Lemma 13 and the estimate on G_n^z, for $j\geq 0$,

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

It follows that

$$E[\int_{n^{-\delta}}^{\varepsilon}|\log x|^{\alpha} d\nu_n^z(x)] \leq C\varepsilon|\log(\varepsilon)|^{\alpha}, \tag{48}$$

where the constant C does not depend on z. The estimate (44) follows when considering $\alpha=1$.

Moreover, by (3), for $\alpha<2$,

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

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

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

We thus get (44) from (48). By Chebychev's inequality, the convergence in expectation implies the convergence in probability and therefore for any $\delta,\delta' > 0$ there exists $\varepsilon > 0$ small enough so that

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

On the other hand, $\int_{\varepsilon}^{\infty} \log|x| d\nu_n^z(x)$ converges to $\int_{\varepsilon}^{\infty} \log|x| d\nu^z(x)$ by the weak convergence of ν_n^z to ν^z in probability for any $\varepsilon > 0$. Hence, we get (45).

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

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

and set $f_n(z) = f_n^1(z) + f_n^2(z)$. Because ν_n^z is supported in B_{R+M} on $\|D_n\| \le M$, f_n is bounded above. By (48), $E[|f_n^2(\cdot)|^{\alpha}]$ is bounded, uniformly in $z \in B_R$. On the other hand, by (3), again uniformly in z, $E(f_n^1(z)^2) < \delta'$, and therefore

$$E\int_{B_R} (f_n^1(z))^2 dm(z) < \infty.$$

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

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

is bounded in probability. This uniform integrability and the weak convergence (45) are enough to conclude, using dominated convergence (see [23, Lemma 3.1] for a similar argument). $\square$

5 Proof of Theorem 1

By Proposition 12, see (46), we have, with $h(z) := \int \log|x| d\nu^z(x)$, that for any smooth compactly supported function ψ on $\mathbb{C}$,

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

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

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

From Remark 2 (based on [11, Corollary 4.5]), we have that the limit is actually a function. The statement of the theorem follows. ☐

6 Proof of Corollary 3

We let X_n be as in the statement of the corollary and write $X_n = P_n D_n Q_n$ with P_n, Q_n unitary and D_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 D_n possesses a density on $\mathbb{R}_+^n$ which is given by the expression

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

where $\tilde{Z}_n$ is a normalization factor, see e.g. [1, Proposition 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 [20]. By part (i) of Theorem 2.1 there, there exist positive constants M, C_{11} such that $P(\sigma_1 > M-1) \leq e^{-C_{11}n}$, and thus point 1 of the assumptions of Theorem 1 holds. Point 3 of the assumptions (with $\kappa < 1/4$ and $\kappa' = 1/2$) is an immediate consequence of equation (2.26) there. Point 4 of the assumptions is an immediate consequence of equation (2.32) there. 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 (3). We begin with the following proposition, due to [21].

Proposition 14. *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*

$$P(\sigma_n(A) < x) \leq C_{12}\sqrt{n}\left(\frac{x}{\sigma}\right)^2. \tag{49}$$

The proof of Proposition 14 is identical to [21, Theorem 3.3], with the required adaptation in moving from real to complex entries. 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 14. 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 (49), with $\sigma_n(A_n)$ denoting the minimal singular value of A_n, we have

$$P(\sigma_n(A_n) < x; \mathcal{G}_n) \leq C_{12}x^2 n^{1/2+2\alpha}. \tag{50}$$

If the estimate (50) 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 (50), 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 (50),

$$P(\sigma_n(\tilde{A}_n) < x; \mathcal{G}_n) \leq C_{12}x^2 n^{1/2+2\alpha} + P((\mathcal{G}'_n)^c) \leq C_{12}[x^2 n^{1/2+2\alpha} + n^2 e^{-n^\alpha/2}]. \tag{51}$$

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

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

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

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

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

$$\begin{aligned}
|\mathrm{tr}(V(W_n) - V(W_n^{(\alpha)}))| &\le \sum_{i=1}^n |V(\theta_i) - V(\mu_i)| \le C_{13} \sum_{i=1}^n |\theta_i - \mu_i| \\
&\le C_{13} n^{1/2} \left(\sum_{i=1}^n |\theta_i - \mu_i|^2 \right)^{\frac{1}{2}} \qquad (52) \\
&\le C_{13} n^{1/2} \left(\mathrm{tr}((W_n - W_n^{(\alpha)})^2) \right)^{\frac{1}{2}} \\
&\le C_{13} n^{1/2-\alpha} \mathbf{1}_{\mathcal{G}_n'} \mathrm{tr}((n^{-\alpha/2} N_n)^2)^{1/2} \le n^{C_{14}-\alpha}, \qquad (53)
\end{aligned}$$

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]). We emphasize that 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 $1 + n^{C_{14}+1-\alpha}$; in particular, it holds that

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

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

$$4n^{C_{14}+1-\alpha}.$$

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

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

Note that this estimate does not depend on z. Combined with (51), we thus deduce that

$$P(\sigma_n(\overline{A}_n) < x; \mathcal{G}_n) \le C_{12} x^2 n^{1/2+2\alpha} + n^{C_{16}-\alpha} \le n^{C_{17}} x^{2/3},$$

where α was chosen as function of x. This yields immediately point 2 of the assumptions of Theorem 1, if $\delta > 3C_{17}/2$.

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

Remark 15. The proof of Corollary 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 [20] still apply. We omit further details.

7 Proof of Corollary 4

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

$$\left| G_{D_n}(z) - G_{\tilde{D}_n}(z) \right| \le \frac{E[\|\tilde{D}_n - D_n\|]}{|\Im z|^2} \le \frac{C(\|D_n^{-1}\|)}{|\Im z|^2} n^{\frac{1}{2}-\gamma}$$

with $C(\|D_n^{-1}\|)$ a finite constant depending only on $\|D_n^{-1}\|$ which we assumed bounded. As a consequence, the third condition is satisfied since

$$\left| G_{\tilde{\Theta}}(z) - G_{\tilde{D}_n}(z) \right| \le \frac{C(\|D_n^{-1}\|)}{|\Im z|^2} n^{\frac{1}{2}-\gamma} + \frac{K}{n^{\kappa} |\Im z|} \le \frac{K'}{n^{\gamma'} |\Im z|}$$

with $\gamma' = \min\{\kappa, \frac{1}{2}(\gamma - \frac{1}{2})\}$ and $\Im z \ge n^{-\max\{\frac{1}{2}(\gamma-\frac{1}{2}), \kappa'\}}$. Hence, the results of Lemma 11 hold and we need only to check as in Proposition 12 that, if ν_n^z 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 $\overline{A}_n = zI - Y_n = zI - U_n D_n V_n + n^{-\gamma}\tilde{N}_n$ with $\tilde{N}_n$ a Gaussian matrix, and therefore we can use Proposition 14 to obtain (49) with $\sigma = n^{-\gamma}$, and the desired estimate on I_n. $\square$

Proof of Example 5

Indeed, the first and the fourth hypotheses of Theorem 1 are verified since μ is compactly supported and we assumed its symmetrized version has a bounded Stieltjes transform. For the third, note that if F^{-1} is Hölder continuous with index α,

$$\left|G_{\tilde{\Theta}}(z) - G_{D_n}(z)\right| \le \sum_{i=1}^{n} \frac{|s_{i+1}^n - s_i^n|}{n|\Im z|^2} = \sum_{i=1}^{n} \frac{|F^{-1}(\frac{i+1}{n}) - F^{-1}(\frac{i}{n})|}{n|\Im z|^2} \le C\frac{n^{-\alpha}}{|\Im z|^2}$$

where we finally used that F^{-1} is Hölder continuous with index α. $\square$

8 Extension to orthogonal conjuguation

In this last section, we generalize Theorem 1 to the case where we conjuguate D_n by orthogonal matrices instead of unitary matrices.

Theorem 16. *Let D_n be a sequence of diagonal matrices satisfying the assumptions of Theorem 1. Let $O_n, \tilde{O}_n$ be two $n \times n$ independent matrices which follow the Haar measure on the orthogonal group and set $A_n = O_n D_n \tilde{O}_n$. 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 11 which in turn is based on the approximate Schwinger-Dyson equation (35) which is itself a consequence of equation (29) and concentration inequalities. To prove the analogue of (29) when U_n follows the Haar measure on the orthogonal group, observe that (30) 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 $\tilde{m}(A \otimes B) = AB^t$,

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

The right hand side is small as $\tilde{m} \circ \partial P$ 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 (31) 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, Theorem 4.4.27] and [1, Corollary 4.4.28], all the proof of Theorem 1 can be adapted to this set up. $\square$

Acknowledgments: We thank Greg Anderson for many fruitful and encouraging discussions. We thank Yan Fyodorov for pointing out the paper [15].

References

[1] Anderson, G. W., Guionnet, A. and Zeitouni, O., *An introduction to random matrices*, Cambridge University Press, Cambridge (2009). *To appear.* http://www.math.umn.edu/ zeitouni/technion/cupbook.pdf

[2] Bai, Z., *Circular law*, Ann. Probab., **25**, 494–529, (1997).

[3] Brown, L. G., *Lidskii's theorem in the type II case*, in "Proceedings U.S.–Japan, Kyoto/Japan 1983", Pitman Res. Notes. Math Ser. **123**, 1–35, (1983).

[4] Collins, B., Guionnet, A. and Maurel-Segala E. *Asymptotics of unitary and orthogonal matrix integrals*, to appear in Adv. Math., (2009).

[5] Feinberg, J. and Zee, A., *Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism*, Nuclear Phys. B, **501**, 643–669, (1997).

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

[7] Fyodorov, Y.V and Wei, Y., *On the mean density of complex eigenvalues for an ensemble of random matrices with prescribed singular values.* Phys. A., **41**, 502001,(2008).

[8] Ginibre, J., *Statistical ensembles of complex, quaternion, and real matrices*, Jour. Math. Phys **6**, 440–449, (1965)

[9] Girko, V. L., *The circular law*, Teor. Veroyatnost. i Primenen., **29**, 669–679, 1984.

[10] Götze, F. and Tikhomirov, A., *The circular law for random matrices*, arXiv:0709.3995v3 [math.PR] (2007).

[11] Haagerup, U. and Larsen, F., *Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras*, J. Funct. Anal. **2**, 331–367, (2000).

[12] Haagerup, U. and Schultz, A. *The Invariant subspace problem for von Neumann algebras*, arXiv:math/0611256v1 [math.OA] (2006).

[13] Haagerup, U. and Schultz, A., *Brown measures of unbounded operators affiliated with a finite von Neumann algebra*, Math. Scand. **100**, 209–263, (2007).

[14] Haagerup, U. and Thorbjørnsen, S. *A new application of random matrices:* $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ *is not a group*, Ann. of Math. (2) **162**, 711–775, (2005).

[15] Horn, A., *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc., **5**, 4–7, (1954).

[16] J. B. Hough, M. Krishnapur, Y. Peres and B. Virág, *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, Providence, RI, American Mathematical Society, (2009).

[17] N. Lehmann and H.-J. Sommers, *Eigenvalue statistics of random real matrices*, Phys. Rev. Lett. **67**, 941–944, (2001).

[18] Nica, A. and Speicher, R., *$\mathcal{R}$-diagonal pairs – a common approach to Haar unitaries and circular elements*, Fields Inst. Commun., **12**, 149–188 (1997).

[19] Pan, G. and Zhou, W., *Circular law, extreme singular values and potential theory*, arXiv:0705.3773v2 [math.PR], (2007).

[20] Pastur, L. and Shcherbina, M., *Bulk universality and related properties of Hermitian matrix models*, J. Stat. Phys. **130**, 205–250, (2008).

[21] Sankar, A., Spielman, D. A. and Teng, S.-H., *Smoothed analysis of the conditioning number and growth factor of matrices*, SIAM J. Matrix Anal. **28**, 446–476, (2006).

[22] Tao, T. and Vu, V., *Random matrices: the circular law*, Commun. Contemp. Math. **10**, 261–307, (2008).

[23] Tao, T. and Vu, V., with appendix by M. Krishnapur, *Random matrices: Universality of ESD's and the circular law*, arXiv:0807.4898v5 [math.PR] (2008).

[24] Voiculescu, D., *The analogues of entropy and of Fisher's information measure in free probability theory. VI. Liberation and mutual free information*, Adv. Math. **146**, 101–166, (1999).

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

[26] Zyczkowski, K. and H.-J. Sommers, *Truncations of random unitary matrices*, J. phys. A: Math. Gen. **33**, 2045–2057 (2000).