

# The single ring theorem

Alice Guionnet\*, Manjunath Krishnapur<sup>†</sup> and Ofer Zeitouni<sup>‡</sup>

September 10, 2009. Revised June 15 and October 16, 2010 and January 13, 2011

**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  $A_n = U_n T_n V_n$  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  $\mu$  on the complex plane whose support is a single ring. In particular, we provide a complete proof of Feinberg-Zee single ring theorem [6]. We also consider the case where  $U_n, V_n$  are independent Haar distributed on the orthogonal group.

## 1 The problem

Horn [18] 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  $s_1 \geq \dots \geq$

---

\*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.

<sup>†</sup>Department of Mathematics, Indian Institute of Science, Bangalore - 560012, India. manju@math.iisc.ernet.in. This author's work was supported, in part, by DST and UGC through DSA-SAP-PhaseIV.

<sup>‡</sup>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  $\lambda_1, \dots, \lambda_n$  in decreasing order of absolute values, then the inequalities

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

were shown by Weyl [30] 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  $s_1 \geq \dots \geq s_n \geq 0$  and consider  $n \times n$  matrices with 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 Haar measure on  $\mathcal{U}(n)$ , the unitary group of  $n \times n$  matrices, and independent from  $T$ . 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  $\theta$  supported on  $\mathbb{R}_+$ . Then, does  $L_\Lambda$  converge to a deterministic measure  $\mu$  on the complex plane? If so, how is the measure  $\mu$  determined by  $\theta$ ?
3. For finite  $n$ , for fixed  $S$ , is  $L_\Lambda$  *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 [8, 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 [7] for the case where  $T$  is a projection). The authors of [8] 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 [9], and its asymmetric variant, see [20]. (There are some pitfalls in the standard derivation of Ginibre's result. We refer to [19] for a discussion.) Another situation is the truncation of random unitary matrices, described in [31].

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 [10]; the circular law was proved under some conditions in [2] and finally, in full generality, in [11] and [26]. Such matrices, however, do not possess the invariance properties discussed in connection of 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 non-normal operators, introduced in [4].)  $R$ -diagonal operators were introduced by Nica and Speicher [21] 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 [15, 16].

## 2 Limiting spectral density of a non-normal matrix

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

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

$G_\mu$  is analytic off the support of  $\mu$ . 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  $\mu$  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  $(w, \lambda)$  then, with  $\bar{P}_n = Q_n P_n$  and  $v = Q_n w$ , it holds that  $\bar{P}_n T_n v = \lambda v$ , and  $\bar{P}_n$  is again Haar distributed.) Finally, for any matrix  $A$ , we set  $\|A\|$  to denote the  $\ell^2$  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  $\mu$ , 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  $\mu$  is the the function  $R_\mu(z) = K_\mu(z) - 1/z$ . The moments of  $\mu$  (and therefore  $\mu$  itself, since it is compactly supported) can be recovered from the knowledge of  $K_\mu$ , and therefore from  $R_\mu$ , by a formal inversion of power series. For a pair of compactly supported probability measures  $\mu_1, \mu_2$ , 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, Section 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  $\Theta$  compactly supported on  $\mathbb{R}_+$ . Assume further*

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

$$\lim_{n \rightarrow \infty} P(\|T_n\| > M) = 0. \quad (1)$$

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

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

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

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

Then the following hold.

- a.  $L_{A_n}$  converges in probability to a limiting probability measure  $\mu_A$ .
- b. The measure  $\mu_A$  possesses a radially-symmetric density  $\rho_A$  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  $\Delta_z$  denotes the Laplacian with respect to the variable  $z$  and  $\nu^z := \tilde{\Theta} \boxplus \lambda_{|z|}$ .
- c. The support of  $\mu_A$  is a single ring: there exist constants  $0 \leq a < b < \infty$  so that

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

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, Ch. 5] for general background. A different characterization of  $\rho_A$ , borrowed from [14] 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 (3) which prevents the existence of an atom in the support of  $\tilde{\Theta}$ , can be dispensed of; 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  $z \rightarrow x$ , towards the density of this measure at  $x$ , (3) is verified as soon as  $\tilde{\Theta}$  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*

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

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  $\Theta$  the unique minimizer of the functional

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

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 the next proposition 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 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  $\mu_A$  as in Theorem 1 for any  $\gamma \in (\frac{1}{2}, \infty)$ .*

In a general framework, P. Śniady [25, Theorem 7] has shown that there exists a sequence  $\varepsilon_n$  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  $\mu_A$ . The above proposition thus insures that this is true for any polynomially decaying sequence  $\varepsilon_n$ . Note that in the earlier unpublished notes [13], U. Haagerup had proven similar regularization of the Brown measure by Cauchy type matrices instead of Gaussian ones.

**Example 5.** An example of sequence  $(T_n)_{n \geq 0}$  satisfying the hypotheses of Proposition 4 is given as follows: take  $T_n = \text{diag}(s_1^n, \dots, s_n^n)$  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  $\mu$  on  $[\delta, M]$  which is absolutely continuous with respect to the Lebesgue measure,
- there exists  $\kappa > 0$  and  $C$  finite so that for all  $E \in [\delta, M]$ , all  $\delta \geq n^{-\kappa}$ ,

$$\#\{i : |s_i - E| \leq \delta\} \leq C\delta n.$$

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 non-commutative probability space  $(\mathcal{A}, \tau)$ . (The latter means that for all polynomial  $P$  in two non-commutative variables,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(P(B_n, B_n^*)) = \tau(P(b, b^*)),$$

which is the case if e.g  $B_n$  is self-adjoint, with spectral measure converging to a probability measure  $\Theta$ , 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  $\Theta_w$  of  $|wI - b|$ . By Voiculescu's theorem [28, Theorem 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 couple of operators living in a non-commutative probability space  $(\mathcal{A}, \tau)$  which are free,  $u$  being unitary. The Brown measure  $\mu_{b+u}$  is studied in [3, Section 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 a  $n \times n$  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  $\mu_{b+u}$  of  $b + u$ .*

An example of matrices  $B_n$  which satisfy the hypotheses of Proposition 6 is given by the diagonal matrices  $B_n = \text{diag}(s_1^n, \dots, s_n^n)$  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  $(|w - s_1^n|, \dots, |w - s_n^n|)$ .

## 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  $A_n A_n^* \neq A_n^* A_n$ , 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 [10] that we describe now (the details are a little different from what is presented in Girko [10] or Bai [2]).

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

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

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

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

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  $\nu_n^z$  denote the ESD of  $H_n^z$ ,

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

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

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

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  $\nu_n^z$  converge weakly in probability to a measure  $\nu^z$  as  $n \rightarrow \infty$ , and identify the limit. Since  $H_n^z$  are Hermitian matrices, there is hope of doing this by Hermitian techniques.
2. Justify that  $\int \log |x| d\nu_n^z(x) \rightarrow \int \log |x| d\nu^z(x)$  for (almost every)  $z$ . But for the fact that “log” is not a bounded function, this would have followed from the weak convergence of  $\nu_n^z$  to  $\nu^z$ . 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  $\nu_n^z$  to a convergence of integrals over  $z$ . Indeed, setting  $h(z) := \int \log |x| d\nu^z(x)$ , we will get from (5) that

$$\int \psi(z) dL_{A_n}(z) \rightarrow \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\psi(z) h(z) dm(z). \quad (6)$$

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

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

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 (5) 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 [11], Pan and Zhou [22] and Tao and Vu [26]. While we heavily borrow 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 [17]. In [5], they were used to study the asymptotics of matrix models on the unitary group. Our approach combines ideas of [17] 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 [14]. 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 2. In the context of the Feinberg–Zee single ring theorem, the latter assumption holds due to an adaptation of the analysis of [24].

## 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  $C_i, K, \kappa$  and  $\kappa'$  are fixed and do not change from line to line.

## 3 An auxiliary problem: evaluation of $v^z$ and convergence rates

Recall from the proof sketch described above that we are interested in evaluating the limit  $v^z$  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  $\tilde{H}_n^z$  given by

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

where  $W_n^z = \bar{z}Q_nP_n/|z|$  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

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

with  $B_n = \rho U_n + T_n$ ,  $T_n$  being 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 throughout  $\rho$  uniformly bounded.

We denote in short

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

### 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 non-commutative probability space  $(\mathcal{A}, *, \mu)$  on which a variable  $U$  lives and where  $\mu$  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, 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  $\partial$ , defined on elements of  $\mathbb{C}\langle T, U, U^* \rangle$  as satisfying the Leibniz rules

$$\partial(PQ) = \partial P \times (1 \otimes Q) + (P \otimes 1) \times \partial Q,\tag{11}$$

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

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

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

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

By the invariance of  $\mu$  under unitary conjugation, see [29, Proposition 5.17] or [1, (5.4.31)], we have the Schwinger–Dyson equation

$$\mu \otimes \mu(\partial P) = 0. \quad (13)$$

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

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

We extend  $\mu$  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  $\partial$  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

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

where we denoted  $p = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ . In the sequel we will apply  $\partial$  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  $z \in \mathbb{C} \setminus \mathbb{R}$  and  $a, b \in \mathbb{R}$ . Such an extension is straightforward;  $\partial$  continues to satisfy the Leibniz rule and, using the resolvent identity

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

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

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

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

We apply the derivative  $\partial$  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

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

Note that  $Pp = P$  and thus  $\mu(pP) = \mu(P)$ . Further, for any smooth function  $Q$ ,  $\mu(\mathbf{U}^*QU)$  equals  $\mu((1-p)Q)$  due to the traciality of  $\mu$  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 that  $\mu(\mathbf{U}^*P)$  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  $z_2 G_U(z_1, z_2)$  as  $z_2 \rightarrow \infty$ , 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  $z_1 \rightarrow \infty$ ,

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

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

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

(Again, here and in the rest of this subsection, the proper branch of the square root is determined by analyticity.) Let  $R_\rho$  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, Definition 5.3.22 and Exercise 5.3.27], so that we have

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

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

$$G_T(z_1, z_2) = z_2 G(z_1, z_2) - G(z_1), \quad (24)$$

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

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

$$\rho G(z_1, z_2) R_\rho(G(z_1)) + z_2 G(z_1, z_2) - G(z_1) = z_1 G(z_1, z_2) - G_T(z_2), \quad (26)$$

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

$$G(z_1, z_2) (\rho R_\rho(G(z_1)) + z_2 - z_1) = G(z_1) - G_T(z_2). \quad (27)$$

Thus,

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

The choice of  $z_2$  as in (28) is allowed for any  $z_1 \in \mathbb{C}^+$  because  $G : \mathbb{C}^+ \rightarrow \mathbb{C}^-$  and we can see that  $R : \mathbb{C}^- \rightarrow \mathbb{C}^-$ . 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  $z \mapsto z - \rho R_\rho(G(z))$  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  $\infty$ . 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  $\mu_T$  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  $\mu_T$ ; 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  $\mu_Y$ , equals the *free convolution* of  $\tilde{\mu}_T$  and  $\lambda_\rho$ , i.e.  $\mu_Y = \tilde{\mu}_T \boxplus \lambda_\rho$ .

**Remark 8.** We provide, following [14], an alternative characterization of  $\mu_A$  and its support. We first introduce some terminology from [14]. Consider a tracial non-commutative  $W^*$ -probability space  $(\mathcal{M}, \tau)$ . Let  $u$  be Haar-distributed and let  $h$  be a  $*$ -free from  $u$  self-adjoint element (whose law will be taken to be  $\Theta$ ). Let  $\tilde{\nu}^z$  denote the law of  $|zI - uh|$ . The *Brown measure* for  $uh$  is defined as

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

c.f. [14, Pg. 333]. Recall that  $\Theta(\{0\}) = 0$  by Assumption 3. By [14, Proposition 3.5] and Remark 7 above,  $\tilde{\nu}^z = \nu^z$ , and therefore,  $\mu_A$  in the statement of Theorem 1 is the Brown measure for  $uh$ . By [14, Theorem 4.4 and Corollary 4.5], the Brown measure  $\mu_A$  is radially symmetric and possesses a density  $\rho_A$  that can be described as follows. Let  $\Theta^{\sharp 2}$  denote the push forward of  $\Theta$  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 [14, Section 2] for the definition of the S-transform of a probability measure on  $\mathbb{R}$  and its relation to the R-transform). Define  $F(t) = 1/\sqrt{\mathcal{S}(t-1)}$  on  $\mathcal{D} = (0, 1]$ . Then,  $F$  maps  $\mathcal{D}$  to the interval

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

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

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

Finally,  $\rho_A$  has an analytic continuation to a neighborhood of  $(a, b]$ , and  $\mu_A$  is a probability measure, see [14, Pg 333].

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

**Lemma 9.** *If  $|\Im G_T(\cdot)| \leq \kappa_1$  on  $\{z : \Im(z) \geq \varepsilon\}$  then  $|\Im G(\cdot)| \leq \kappa_1$  on  $\{z : \Im(z) \geq \varepsilon\}$ .*

**Proof** Recall that if  $z \in \mathbb{C}^+$  then  $G(z) \in \mathbb{C}^-$  and also  $R_\rho(G(z)) \in \mathbb{C}^-$  because  $R_\rho$  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  $\Im(y) \geq \Im(z)$ . Therefore, if  $\Im(z) \geq \varepsilon$  then  $|\Im G(z)| = |\Im G_T(y)| \leq \kappa_1$ .  $\square$

### 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  $L_{T_n} \rightarrow \mu_T$  weakly in probability, and further that (3) is satisfied. All constants in this section are independent of  $\rho$ , but depend implicitly on  $M$ , the uniform bound on  $\|T_n\|$  and on  $\rho$ .

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

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

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,

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

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  $\ell$  yields (30).

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

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

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

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

with

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

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

$$E\left[\left|\frac{1}{2n}\text{tr}(P) - E\left[\frac{1}{2n}\text{tr}\right](P)\right|^2\right] \leq \frac{1}{n^2}\|P\|_L^2, \quad (33)$$

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

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

where

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

In particular,

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

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

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

with

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

Hence, if we define

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

then

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

and therefore

$$G^n(z_1) = G_{T_n}(\Psi_n(z_1)) - \tilde{O}(n, z_1, \Psi_n(z_1)). \quad (38)$$

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  $n \rightarrow \infty$  then implies that  $G^n \rightarrow G$  in a neighborhood in the upper half plane close to  $\infty$ . 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  $\mu_T$ . 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  $\Theta$ , then for any  $z \in \mathbb{C}$ ,  $v_n^z$  converges weakly in probability to  $\tilde{\Theta} \boxplus \lambda_{|z|}$ .*

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

Lemma 10 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 (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,  $\Psi_n$  is analytic,  $\Im \Psi_n(z) > (\Im(z)/2) \wedge C$  for some positive constant  $C$  and  $\tilde{O}(n, z_1, \Psi_n(z_1))$  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  $G_U^n(\cdot)$  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*

$$|1 + 2\rho G_U^n(z)| > C_2 \rho [\Im(z)^3 \wedge 1]. \quad (39)$$

**Proof** Since  $G_U^n(z)$  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  $\mathbf{Y}$ . 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  $z, x \in B(0, R)$

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

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

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

Combining this estimate and (40), we get that

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

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  $\psi_n$  is analytic in  $\{z : \Im(z) > C_1 n^{-1/3}\}$ , for all  $n$  large.

Our next goal is to check the analyticity of  $z \rightarrow \tilde{O}(n, z, \psi_n(z))$  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 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  $G_U^n(z)$  tend to 0 at infinity, 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  $\Delta_n = \{z \in \mathbb{C}^+ : |\rho G^n(z) + i/2| \geq \mathbf{e}_n\}$ .

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

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

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  $z \in \Delta_n$ ,

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

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

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

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,

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

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

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

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  $z \rightarrow \tilde{O}(n, z, \psi_n(z))$  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  $z \in \mathcal{E}_n := \{z : \Im(z) > n^{-C_7}\}$ ,*

$$|\Im G^n(z)| \leq C_8. \quad (44)$$

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

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

## 4 Tail estimates for $\mathbf{v}_n^z$

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  $z \in \mathbb{C}$ ,*

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} E[1_{\mathcal{G}_n} \int_0^\varepsilon \log |x| d\mathbf{v}_n^z(x)] = 0. \quad (45)$$

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

$$\int \log |x| d\mathbf{v}_n^z(x) \rightarrow \int \log |x| d\mathbf{v}^z(x), \quad (46)$$

in probability.

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

$$\int \varphi(z) \int \log|x| d\nu_n^z(x) dm(z) \rightarrow \int \varphi(z) \int \log|x| d\nu^z(x) dm(z), \quad (47)$$

in probability.

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

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

$$\mu((-y, y)) \leq 2y |\Im G(iy)|. \quad (48)$$

**Proof** We have

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

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 and will 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

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

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

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

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

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

We conclude that

$$E\left[\int_{n^{-\delta}}^{\varepsilon} |\log x|^\alpha d\mathbf{v}_n^z(x)\right] \leq C\varepsilon |\log(\varepsilon)|^\alpha, \quad (49)$$

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

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

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

$$E\left[\left(\mathbf{v}_n^z([-n^{-\delta}, n^{-\delta}])\right)^{\frac{2-\alpha}{2-\alpha}}\right] \leq E\left[\mathbf{v}_n^z([-n^{-\delta}, n^{-\delta}])\right] \leq 2c_2 n^{-c_1}.$$

By (2), the second factor is bounded by  $(\delta')^{\alpha/2}$ . We thus get (45) from (49). By Chebycheff'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

$$\lim_{n \rightarrow \infty} P\left(\int_0^\varepsilon |\log x| d\mathbf{v}_n^z(x) > \delta\right) < \delta'$$

On the other hand,  $\int_\varepsilon^\infty \log|x| d\mathbf{v}_n^z(x)$  converges to  $\int_\varepsilon^\infty \log|x| d\mathbf{v}^z(x)$  by the weak convergence of  $\mathbf{v}_n^z$  to  $\mathbf{v}^z$  in probability for any  $\varepsilon > 0$ , and  $\int_0^\varepsilon \log|x| d\mathbf{v}^z(x)$  converges to 0 as  $\varepsilon \rightarrow 0$  since  $\mathbf{v}^z$  has a bounded density by Lemma 9. Hence, we get (46).

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

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



and set  $f_n(z) = f_n^1(z) + f_n^2(z)$ . Because  $\nu_n^z$  is supported in  $B_{R+M}$  on  $\|T_n\| \leq M$  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

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

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

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

is bounded in probability. This uniform integrability and the weak convergence (46) are enough to conclude, using dominated convergence (see [27, Lemma 3.1] for a similar argument).  $\square$

## 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  $\psi$  on  $\tilde{B}_R$ ,

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

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 [14, Corollary 4.5]), we have that  $\mu_A$  is a probability measure that possesses a radially symmetric density  $\rho_A$  satisfying the properties stated in parts b and c of the theorem.  $\square$

## 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

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

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 [23]. 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. By equations [23, (2.26) and (2.27)] and Chebycheff's inequality, we get that for  $z$  with  $\Im(z) > n^{-\kappa'}$  where  $\kappa < (1 - \kappa')/2$ ,

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

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

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

which goes to zero for  $\kappa \in (0, (1 - 2\kappa')/4)$ . Together with [23, Equation (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 [24].

**Proposition 16.** *Let  $\bar{A}$  be an arbitrary  $n$ -by- $n$  matrix, and let  $A = \bar{A} + \sigma N$  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  $\bar{A}$ ,  $\sigma$  or  $n$  such that*

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

The proof of Proposition 16 is identical to [24, Theorem 3.3], with the required adaptation in moving from real to complex entries. (Specifically, in the right side of the display in [24, Lemma A.2],  $\varepsilon \sqrt{2/\pi}/\sigma$  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  $\bar{A}_n = zI - X_n$ ,  $\tilde{A}_n = \bar{A}_n + n^{-\alpha} N_n \mathbf{1}_{\mathcal{G}'_n}$  and  $A_n = \bar{A}_n + n^{-\alpha} N_n$ . Then, by (50), with  $\sigma_n(A_n)$  denoting the minimal singular value of  $A_n$ , we have

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

If the estimate (51) concerned  $\bar{A}_n$  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  $\delta$ , which would depend on  $\alpha$ ). Our goal is thus to replace, in (51),  $A_n$  by  $\bar{A}_n$ , 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),

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

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  $W_n = X_n X_n^*$  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

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

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

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

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

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

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,

$$|\text{tr}(V(W_n) - V(W_n^{(\alpha)}))| \leq n^{(C_{14}-\alpha)/2}, \quad (53)$$

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

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

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

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

Combined with (52), we thus deduce that

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

where  $\alpha$  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 completes the proof of Theorem 3.  $\square$

**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 [23] 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  $\tilde{T}_n$ . Note that  $Y_n = \tilde{U}_n \tilde{T}_n \tilde{V}_n$  with  $\tilde{U}_n, \tilde{V}_n$  following the Haar measure. We first show that  $\tilde{T}_n$  also satisfies the assumptions of Theorem 1 when  $\gamma > \frac{1}{2}$ , except for the second one. Since the singular values of  $N_n$  follows 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  $\|\tilde{T}_n\| \leq \|T_n\| + n^{-\gamma+\frac{1}{2}}\|n^{-\frac{1}{2}}N_n\|$  is bounded with overwhelming probability. Moreover, since  $\tilde{T}_n = |T_n + n^{-\gamma}U_n^* N_n V_n^*|$ , on the event  $\|N_n/\sqrt{n}\| \leq M$  we have

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

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

$$|G_{\tilde{\Theta}}(z) - G_{\tilde{T}_n}(z)| \leq \frac{C(\|T_n^{-1}\|, \|T_n\|)}{|\mathfrak{S}(z)|^2} n^{\frac{1}{2}-\gamma} + \frac{K}{n^\kappa |\mathfrak{S}(z)|} \leq \frac{K'}{n^{\gamma'} |\mathfrak{S}(z)|}$$

with  $\gamma' = \min\{\kappa, \frac{1}{2}(\gamma - \frac{1}{2})\}$  and  $\mathfrak{S}(z) \geq n^{-\max\{\frac{1}{2}(\gamma - \frac{1}{2}), \kappa'\}}$ . Hence, the results of Lemma 13 hold and we need only check, as in Proposition 14, that with  $v_n^z$  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  $\bar{A}_n = zI - Y_n = zI - U_n T_n V_n + n^{-\gamma} \tilde{N}_n$  with  $\tilde{N}_n$  a Gaussian matrix, and therefore we can use Proposition 16 to obtain (50) with  $\sigma = n^{-\gamma}$ , and the desired estimate on  $I_n$ .  $\square$

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

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

$\square$

## 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  $O_n, \tilde{O}_n$  be two  $n \times n$  independent matrices which follow the Haar measure on the orthogonal group and set  $A_n = O_n T_n \tilde{O}_n$ . Then,  $L_{A_n}$  converges in probability to the probability measure  $\mu_A$  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  $k, \ell$ , yields, if we denote  $\tilde{m}(A \otimes B) = AB^t$ ,

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

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 (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, Theorem 4.4.27] and [1, Corollary 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

$$\begin{aligned} \int \Psi(z) dL_{B_n + P_n}(z) &= \frac{1}{4\pi n} \int_{\mathbb{C}} \Delta\Psi(z) \log \det(zI - B_n - P_n)(zI - B_n - P_n)^* dm(z) \\ &= \frac{1}{4\pi n} \int_{\mathbb{C}} \Delta\Psi(z) \log \det(|zI - B_n| - P_n U)(|zI - B_n| - P_n U)^* dm(z) \end{aligned}$$

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

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

whereas our hypotheses allow us to bound uniformly the Stieltjes transform of  $\nu_n^z$  on  $\{z_1 : \Im(z_1) \geq n^{-C_7}\}$  as in Lemma 13, hence providing a control of the integral on the interval  $[n^{-C_7}, \varepsilon]$ . 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 [18] 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] Anderson, G. W., Guionnet, A. and Zeitouni, O., *An introduction to random matrices*, Cambridge University Press, Cambridge (2010).
- [2] Bai, Z., *Circular law*, Ann. Probab. **25**, 494–529, (1997).
- [3] Biane, P. and Lechner, F., *Computation of some examples of Brown’s spectral measure in free probability*, Colloq. Math. **90**, 181–211, (2001).
- [4] 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).
- [5] Collins, B., Guionnet, A. and Maurel-Segala E. *Asymptotics of unitary and orthogonal matrix integrals*, Adv. Math. **222**, 172–215, (2009).
- [6] Feinberg, J. and Zee, A., *Non-Gaussian non-Hermitian random matrix theory: phase transition and addition formalism*, Nuclear Phys. B **501**, 643–669, (1997).
- [7] Fyodorov, Y.V. and Sommers, H.J., *Spectra of random contractions and scattering theory for discrete-time systems*, JETP Lett. **72**, 422–426, (2000)
- [8] 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).
- [9] Ginibre, J., *Statistical ensembles of complex, quaternion, and real matrices*, Jour. Math. Phys **6**, 440–449, (1965).
- [10] Girko, V. L., *The circular law*, Teor. Veroyatnost. i Primenen. **29**, 669–679, (1984).
- [11] Götze, F. and Tikhomirov, A., *The circular law for random matrices*, Annals Probab. **38**, 1444–1491, (2010).



- [12] Guionnet, A., *Large random matrices: lectures on macroscopic asymptotics* Lecture Notes in Mathematics **1957** Lectures from the 36th Probability Summer School held in Saint-Flour, 2006, Springer-Verlag.
- [13] Haagerup, U. *Spectral Decomposition of all Operators in a  $II_1$  Factor, which is Embedable in  $R^\omega$* , Preprint (2001)
- [14] 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).
- [15] Haagerup, U. and Schultz, A., *Invariant subspaces for operators in a general  $II_1$  factor*, Publ. Math. IHES **109**, 19–111, (2009).
- [16] Haagerup, U. and Schultz, A., *Brown measures of unbounded operators affiliated with a finite von Neumann algebra*, Math. Scand. **100**, 209–263, (2007).
- [17] Haagerup, U. and Thorbjørnsen, S. *A new application of random matrices:  $\text{Ext}(C_{\text{red}}^*(F_2))$  is not a group*, Ann. of Math. (2) **162**, 711–775, (2005).
- [18] Horn, A., *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc., **5**, 4–7, (1954).
- [19] Hough, J.B., Krishnapur, M., Peres, Y. and Virág, B., *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*, Providence, RI, American Mathematical Society, (2009).
- [20] Lehmann, N. and Sommers, H.-J., *Eigenvalue statistics of random real matrices*, Phys. Rev. Lett. **67**, 941–944, (2001).
- [21] 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).
- [22] Pan, G. and Zhou, W., *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101**, 645–656, (2010).
- [23] Pastur, L. and Shcherbina, M., *Bulk universality and related properties of Hermitian matrix models*, J. Stat. Phys. **130**, 205–250, (2008).

- [24] 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).
- [25] Śniady, P., *Random regularization of Brown spectral measure*, J. Funct. Anal. **193**, 291–313 (2002)
- [26] Tao, T. and Vu, V., *Random matrices: the circular law*, Commun. Contemp. Math. **10**, 261–307, (2008).
- [27] 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).
- [28] Voiculescu, D., *Limit laws for random matrices and free products* Inventiones Mathematicae **104**, 201–220, (1991).
- [29] 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).
- [30] Weyl, H., *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35**, 408–411, (1949).
- [31] Zyczkowski, K. and Sommers, H.-J., *Truncations of random unitary matrices*, J. phys. A: Math. Gen. **33**, 2045–2057, (2000).