



**HAL**  
open science

# Circular law for random matrices with exchangeable entries

Radoslaw Adamczak, Djalil Chafaï, Pawel Wolff

► **To cite this version:**

Radoslaw Adamczak, Djalil Chafaï, Pawel Wolff. Circular law for random matrices with exchangeable entries. *Random Structures and Algorithms*, 2016, 48 (3), pp.454-479. 10.1002/rsa.20599 . hal-00947145

**HAL Id: hal-00947145**

**<https://hal.science/hal-00947145v1>**

Submitted on 14 Feb 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# CIRCULAR LAW FOR RANDOM MATRICES WITH EXCHANGEABLE ENTRIES

RADOSŁAW ADAMCZAK, DJALIL CHAFAÏ, AND PAWEŁ WOLFF

ABSTRACT. An exchangeable random matrix is a random matrix with distribution invariant under any permutation of the entries. For such random matrices, we show, as the dimension tends to infinity, that the empirical spectral distribution tends to the uniform law on the unit disc. This is an instance of the universality phenomenon known as the circular law, for a model of random matrices with dependent entries, rows, and columns. It is also a non-Hermitian counterpart of a result of Chatterjee on the semi-circular law for random Hermitian matrices with exchangeable entries. The proof relies in particular on a reduction to a simpler model given by a random shuffle of a rigid deterministic matrix, on Hermitization, and also on combinatorial concentration of measure and combinatorial Central Limit Theorem. A crucial step is a polynomial bound on the smallest singular value of exchangeable random matrices, which may be of independent interest.

## CONTENTS

1.	Introduction	1
	Main results	3
	Discussion and open problems	5
	Outline	5
	Notations.	6
2.	Limiting distribution of the singular values of shifts	6
3.	Combinatorial Talagrand's concentration inequality	9
4.	Proof of Theorem 1.1 (smallest singular value)	10
5.	Proof of Theorem 1.2 (circular law for the first model)	17
6.	Proof of Theorem 1.3 (circular law for the second model)	20
	References	21

## 1. INTRODUCTION

The spectral measure of an  $n \times n$  matrix  $A$  with real or complex entries is defined by

$$\nu_A = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k},$$

where  $\lambda_1, \dots, \lambda_n$  are the eigenvalues of  $A$ , in other words the roots in  $\mathbb{C}$  of the characteristic polynomial, counted with their algebraic multiplicities, and where  $\delta_x$  is the Dirac mass at the point  $x$ . Thus  $\nu_A$  is a Borel probability measure on  $\mathbb{C}$ , supported on  $\mathbb{R}$  if  $A$  is Hermitian. If  $A$  is a random matrix, then  $\nu_A$  is a random discrete probability measure.

An extensive body of work has been devoted to the study of the behavior of the spectral measure of large dimensional random matrices. While Wigner [33] considered Hermitian

---

*Date:* Winter 2014, compiled on February 14, 2014.

*Key words and phrases.* Random permutations; Random matrices; Combinatorial Central Limit Theorem; exchangeable distributions; concentration of measure; smallest singular value; spectral analysis.

Support: Polish Ministry of Science and Higher Education Iuventus Plus Grant no. IP 2011 000171, and French ANR-2011-BS01-007-01 GeMeCoD and ANR-08-BLAN-0311-01 Granma.

matrices, the non-Hermitian case has also attracted attention, starting with the early work by Mehta [22], who proved, by using explicit formulas due to Ginibre [16], that the average spectral measure of  $n \times n$  matrices with i.i.d. standard complex Gaussian entries scaled by  $\sqrt{n}$  converges to the uniform measure on the unit disc. We will call this measure the *circular law*. In a series of subsequent developments [17, 18, 5, 26, 19], culminating with the work of Tao and Vu [32], this result has been strengthened to almost sure convergence and to sequences of random matrices with i.i.d. entries of unit variance. The Tao-Vu theorem is an instance of the *universality phenomenon*, which has become one of the main themes of Random Matrix Theory. Namely the limiting spectral object depends on the actual distribution of the entries of the random matrix only via a global scale parameter such as the variance. It is natural to ask whether this universal behavior extends to models of random matrices with heavy-tailed entries or with dependent entries. It has been recently shown [7, 10] that the spectral measure of matrices with independent entries in the domain of attraction of an  $\alpha$ -stable distribution with  $0 < \alpha < 2$  converges to a measure which depends only on  $\alpha$  and is supported on the whole complex plane. In particular the limiting distribution is not the circular law. Regarding the independence assumption, several classes of random matrices have been discussed in the literature, starting from models with i.i.d. rows such as random Markov matrices [8], random matrices with i.i.d. log-concave rows [1, 2], and random  $\pm 1$  matrices with i.i.d. rows of given sum [25]. The circular law still holds for models with dependent rows, for instance for blocs of Haar unitary matrices [20, 15], for random doubly stochastic matrices following the uniform distribution on the Birkhoff polytope [23], for random matrices with log-concave unconditional distribution [3], and for random matrices with i.i.d. quaternionic entries [24] (seen as  $2 \times 2$  i.i.d. real blocks with possible dependencies inside the blocks).

The present work introduces first a model of random matrices with exchangeable entries, obtained by shuffling a globally constrained deterministic matrix using a random uniform permutation. Such random matrices have dependent entries, rows, and columns. We show that the asymptotic behavior of the spectral measure is governed by the circular law. The precise formulation is given in Theorem 1.2 below. This first result is then used to deduce the circular law for a larger class of random matrices with exchangeable entries, as stated in Theorem 1.3 below. This second result can be seen as a counterpart of the one of Chatterjee [13], who proved that the spectral measure of symmetric random matrices with exchangeable entries in the upper triangle converges to Wigner’s semi-circular law, which is the universal limit for random matrices with i.i.d. entries of finite variance. Our model of random matrices with exchangeable entries, like all the available models belonging to the universality class of the circular law, has symmetries related to the canonical basis. On the other hand, we stress that the sole invariance under permutation of rows and columns, which is weaker than full exchangeability, is not sufficient, as shown by the simple example of Haar unitary random matrices for which the limiting spectral distribution is the uniform distribution on the unit circle (not the unit disc!) often referred to as the arc law [15]. On the other hand, note that the uniform law on the Birkhoff polytope of doubly stochastic matrices is invariant by permutations of rows or columns, but is not exchangeable globally.

A main tool in our proof is a lower bound on the smallest singular value of random matrices with exchangeable entries, formulated in Theorem 1.1. In recent years, an extensive amount of work has been devoted to the analysis of the smallest singular value of random matrices with (partially) independent entries or rows, in connection with the circular law, but also with numerical analysis or geometry, which makes our result of independent interest. Our approach follows the path introduced by Rudelson and Vershynin in [27] for random matrices with i.i.d. entries, more precisely what they call the “soft approach”

(nothing more is needed for the purpose of the circular law). However several technical problems related to the lack of independence among the entries have to be overcome.

**Main results.** For any  $n \geq 1$ , let  $\mathbf{x}^{(n)} = (\mathbf{x}_{ij}^{(n)})_{1 \leq i, j \leq n}$  be a deterministic real matrix, with

$$\begin{aligned} \text{(A1)} \quad & \sum_{i,j=1}^n \mathbf{x}_{ij}^{(n)} = 0; \\ \text{(A2)} \quad & \sum_{i,j=1}^n |\mathbf{x}_{ij}^{(n)}|^2 = n^2. \end{aligned}$$

Thanks to assumption (A2) we have

$$K_n := \max_{1 \leq i, j \leq n} |\mathbf{x}_{ij}^{(n)}| \geq 1. \quad (1)$$

Of course, the entries of  $\mathbf{x}^{(n)}$  depend typically on  $n$ , hence the notation. The assumption (A2) puts also a constraint on the sparsity of the vector  $\mathbf{x}^{(n)}$ , depending on  $K_n$ .

Let now  $\pi_n$  be a uniform random permutation of the set  $[n] \times [n] = \{(i, j) : 1 \leq i, j \leq n\}$ , where  $[n] = \{1, \dots, n\}$ . We consider the random matrix

$$X^{(n)} = (X_{ij}^{(n)})_{1 \leq i, j \leq n} \quad \text{where} \quad X_{ij}^{(n)} := \mathbf{x}_{\pi_n(i, j)}^{(n)}. \quad (2)$$

The matrix  $X^{(n)}$  inherits the structure of  $\mathbf{x}^{(n)}$ . Namely, from (A1-A2) we get

$$\sum_{i,j=1}^n X_{ij}^{(n)} = 0 \quad \text{and} \quad \sum_{i,j=1}^n |X_{ij}^{(n)}|^2 = n^2 \quad \text{and} \quad \max_{1 \leq i, j \leq n} |X_{ij}^{(n)}| = K_n.$$

Since the uniform law on the symmetric group is invariant under translation (it is a normalized Haar measure), it follows that the law of  $X^{(n)}$  is invariant under any deterministic permutation of its  $n^2$  coordinates, i.e. the random vector  $(X_{ij}^{(n)} : (i, j) \in [n] \times [n])$  of  $\mathbb{R}^{n^2}$  has exchangeable coordinates. In other words, for every permutation  $\sigma$  of  $[n] \times [n]$ ,

$$(X_{\sigma(i, j)}^{(n)})_{1 \leq i, j \leq n} = (\mathbf{x}_{(\pi_n \circ \sigma)(i, j)}^{(n)})_{1 \leq i, j \leq n} \stackrel{d}{=} (\mathbf{x}_{\pi_n(i, j)}^{(n)})_{1 \leq i, j \leq n} = (X_{i, j}^{(n)})_{1 \leq i, j \leq n}.$$

It follows that for every  $(i, j) \in [n] \times [n]$ , and for every  $(k, l) \in [n] \times [n]$  with  $(i, j) \neq (k, l)$ ,

$$\mathbb{E}(X_{ij}^{(n)}) = 0 \quad \text{and} \quad \mathbb{E}(|X_{ij}^{(n)}|^2) = 1 \quad \text{and} \quad \mathbb{E}(X_{ij}^{(n)} X_{kl}^{(n)}) = -\frac{1}{n^2 - 1}. \quad (3)$$

Except for special choices of  $\mathbf{x}^{(n)}$ , the random matrix  $X^{(n)}$  is non-Hermitian, and even non-normal with high probability. Normality means here commutation with the transpose-conjugate. The random matrix  $X^{(n)}$  has exchangeable entries obtained by shuffling the globally constrained deterministic matrix  $\mathbf{x}^{(n)}$  using the random uniform permutation  $\pi_n$ . Such random matrices have dependent entries, dependent rows, and dependent columns. The simulation of a random uniform permutation such as  $\pi_n$  amounts to approximately  $2n^2 \log_2(n)$  bits. The randomness in  $X^{(n)}$  is the one of  $\pi_n$ , which should be compared with the  $n^2$  bits needed for a standard random  $n \times n$  symmetric Bernoulli  $\pm 1$  matrix.

Our first main result is a bound on the smallest singular value of the matrix  $X^{(n)}$ . For an  $n \times n$  real or complex matrix  $A$  we denote by

$$s_1(A) \geq \dots \geq s_n(A)$$

the singular values of  $A$ , i.e. eigenvalues of  $\sqrt{AA^*}$  where  $A^* = \overline{A}^\top$  (transpose-conjugate). The operator norm of  $A$  is  $\|A\| = \max_{|x|=1} |Ax| = s_1(A)$ , where  $|x|$  denotes the Euclidean norm in  $\mathbb{C}^n$  or in  $\mathbb{R}^n$ . We have  $s_n(A) = \min_{|x|=1} |Ax|$ . The matrix  $A$  is invertible iff  $s_n(A) > 0$  and in this case  $s_n(A) = \|A^{-1}\|^{-1}$ . The Hilbert-Schmidt norm or the Frobenius norm of  $A$  is  $\|A\|_{\text{HS}} = \sqrt{\text{tr}(AA^*)} = \sqrt{s_1(A)^2 + \dots + s_n(A)^2}$ .

**Theorem 1.1** (Smallest singular value). *If  $X^{(n)}$  is as in (2) then for every  $z \in \mathbb{C}$ , there exists a constant  $C_{K_n, z} > 0$  depending only on  $K_n$  (defined in (1)) and  $z$  such that for every  $\varepsilon > 0$ ,*

$$\mathbb{P}\left(s_n(X^{(n)} - z\sqrt{n}\text{Id}) \leq \varepsilon n^{-1/2}\right) \leq C_{K_n, z}\left(\varepsilon + \frac{1}{n^{1/2}}\right).$$

An explicit dependence of  $C_{K, z}$  on  $K$  and  $z$  is given in Theorem 4.1.

Let  $(\nu_n)_{n \geq 1}$  be a sequence of random probability measures on  $E$  with  $E = \mathbb{C}$  or  $E = \mathbb{R}$ , and let  $\nu$  be a probability measure on  $E$ . We say that the sequence  $(\nu_n)_{n \geq 1}$  tends as  $n \rightarrow \infty$  weakly in probability to  $\nu$ , and we denote

$$\nu_n \xrightarrow[n \rightarrow \infty]{\text{w.p.}} \nu,$$

if for any  $f \in \mathcal{C}(E, \mathbb{R})$  and  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\int f d\nu_n - \int f d\nu\right| \geq \varepsilon\right) = 0,$$

where  $\mathcal{C}(E, \mathbb{R})$  is the class of all bounded continuous real valued functions on  $E$ . This is equivalent to saying that  $(\nu_n)_{n \geq 1}$ , seen as a sequence of random variables in the space of Borel probability measures on  $E$ , converges in probability to  $\nu$  (see [32, Definition 1.2]).

We denote by  $\nu^{\text{circ}}$  the uniform law on the unit disk of  $\mathbb{C}$ , with density  $z \mapsto \pi^{-1} \mathbf{1}_{\{z \in \mathbb{C}: |z| \leq 1\}}$ . Our second main result concerns the limiting behavior of the spectral measure of  $X^{(n)}$ .

**Theorem 1.2** (Circular law for shuffled matrices). *Let  $A^{(n)} = \frac{1}{\sqrt{n}}X^{(n)}$  where  $X^{(n)}$  is as in (2). If  $K_n = \mathcal{O}(n^{1/(10+\delta)})$  for some  $\delta > 0$ , then  $\nu_{A^{(n)}} \xrightarrow[n \rightarrow \infty]{\text{w.p.}} \nu^{\text{circ}}$ .*

One can check easily that the normalization by  $n^{-1/2}$  ensures tightness, namely

$$\begin{aligned} \int |\lambda|^2 d\nu_{A^{(n)}}(\lambda) &= \frac{1}{n} \sum_{k=1}^n |\lambda_k(A^{(n)})|^2 \\ &\leq \frac{1}{n^2} \sum_{k=1}^n s_k(X^{(n)})^2 = \frac{1}{n^2} \|X^{(n)}\|_{\text{HS}}^2 = \frac{1}{n^2} \sum_{i,j=1}^n |X_{ij}^{(n)}|^2 = 1. \end{aligned}$$

Theorem 1.2 leads to the following result concerning more general random matrices with exchangeable entries, beyond model (2). Namely, for every  $n \gg 1$ , let

$$Y^{(n)} = (Y_{ij}^{(n)})_{1 \leq i, j \leq n} \tag{4}$$

be an  $n \times n$  matrix with exchangeable real entries, and define

$$\mu_n := \frac{1}{n^2} \sum_{i,j=1}^n Y_{ij}^{(n)} \quad \text{and} \quad \sigma_n^2 := \frac{1}{n^2} \sum_{i,j=1}^n (Y_{ij}^{(n)} - \mu_n)^2.$$

**Theorem 1.3** (Circular law for matrices with exchangeable entries). *Let  $Y^{(n)}$  be as in (4). If for  $n \gg 1$ ,  $\sigma_n^2 > 0$  a.s. and for some  $\delta > 0$ ,*

$$\limsup_{n \rightarrow \infty} \mathbb{E}\left[\left(\sigma_n^{-1} |Y_{11}^{(n)} - \mu_n|\right)^{20+\delta}\right] < \infty,$$

then, denoting

$$B^{(n)} = \frac{1}{\sqrt{n}\sigma_n} (Y_{ij}^{(n)} - \mu_n)_{1 \leq i, j \leq n},$$

we have

$$\nu_{B^{(n)}} \xrightarrow[n \rightarrow \infty]{\text{w.p.}} \nu^{\text{circ}}.$$

## Discussion and open problems.

*Assumptions.* It is not enough to assume in Theorem 1.2 that  $K_n = \mathcal{O}(\sqrt{n})$ , since one could construct a matrix with at most  $n/2$  nonzero entries, which would create a substantial atom at zero in the spectral measure. It is tempting to conjecture that  $K_n = \mathcal{O}(n^{1/(2+\delta)})$  is sufficient for the circular law to hold. This would allow to prove Theorem 1.3 under the assumption on moments of order  $4 + \delta$ . On the other hand, Chatterjee's counterpart [13] to Theorem 1.3 in the Hermitian case works under the assumption which when translated to our setting would be  $\mathbb{E}(|Y_{11}^{(n)} - \mu_n|/\sigma_n)^4 = o(n^{2/3})$ .

*Operator norm.* One way to improve our growth/integrability assumptions would be to obtain a better bound on the operator norm of  $X^{(n)}$  defined in (2). The simple bound we use (Lemma 4.2) is  $\mathbb{E}\|X^{(n)}\| \leq CK_n\sqrt{n}$ . By analogy with the available results for matrices with i.i.d. entries, one may ask if for some  $\alpha < 1/2$  and a universal constant  $C$ ,

$$\mathbb{E}\|X^{(n)}\| \stackrel{?}{\leq} \sqrt{2n} + CKn^\alpha.$$

Another way of weakening our assumptions would be to improve the factor  $1/\sqrt{n}$  in the probability bound on the smallest singular value. This factor comes from the Combinatorial Central Limit Theorem and cannot be further improved by our methods. In the i.i.d. case it has been strengthened to  $e^{-cn}$  by relating the bound on the smallest singular value to the Littlewood-Offord problem [27, 30, 31]. While certain ingredients of this approach can be transferred to the exchangeable setting, an important step seems to rely on a Fourier analytic argument, related to Esseen's inequality, involving independence of the entries of the matrix. It would be interesting to develop tools which would allow to use similar ideas in the exchangeable case. Note that since the right hand side in Theorem 1.1 is not summable in  $n$ , one cannot use the first Borel-Cantelli lemma in order to get an almost sure lower bound on  $s_n(X^{(n)} - z\sqrt{n}\text{Id})$ .

*Other models.* One can consider a model of the form  $M_n = n^{-1/2}(X^{(n)} + F_n)$ , where  $F_n$  is deterministic with  $\text{rank } F_n = o(n)$ , satisfying a bound on the operator norm. An inspection of the proof of Theorem 1.2 reveals that if  $\|F_n\| = \mathcal{O}(\sqrt{n})$ , then one can obtain the convergence to the circular law for the spectral measure of  $M_n$  without changing the assumptions on  $K_n$ . On the other hand if  $\|F_n\| = o(n/\log^{3/2} n)$ , then an adaptation of our proof gives the result for  $K_n = \mathcal{O}(1)$ . We remark that for matrices with i.i.d. entries with mean zero and variance one it is enough to assume that  $\|F_n\|_{\text{HS}} = \mathcal{O}(n)$  [11]. Improving this part in the exchangeable case would allow to treat e.g. the adjacency matrices of uniform random directed graphs  $D(n, m_n)$ . More precisely, consider a sequence  $m_n$  of positive integers. Let  $D_n$  be the adjacency matrix of the random graph  $D(n, m_n)$ , i.e. a graph chosen uniformly from the set of all directed random graphs on  $n$  vertices, with  $m_n$  directed edges (including possible loops). One may ask about the limiting spectral behavior of  $D_n$  (depending on the behavior of  $m_n$  as  $n \rightarrow \infty$ ). In some sense, this random graph model lies between the oriented Erdős-Rényi random graph model [9], and the uniform random oriented regular graph model [10]. The singular values of  $D_n$  (as well as its shifts by multiples of identity) may be analyzed using Theorem 2.1 below. More generally, one may think about a Boltzmann-Gibbs probability distribution on the set of oriented graphs with  $n$  vertices, with density with respect to the counting measure proportional to  $\exp(-\beta_n H_n)$  where  $H_n$  is a symmetric functional, and  $\beta_n > 0$ .

**Outline.** In Section 2 we analyze the limiting distribution of the singular values of shifts by  $\sqrt{n}z\text{Id}_n$  for the model (2), which is essential for the proof of Theorem 1.3, and which may be actually of independent interest. Section 3 is devoted to the derivation of a combinatorial version of the Talagrand concentration inequality for convex functions and

product measures, which may be also of independent interest. This inequality is an important tool in the proof of Theorem 1.1 and Theorem 1.3. Theorems 1.1, 1.2, 1.3 are proved in Sections 4, 5, 6 respectively.

**Notations.** We will frequently write  $K, \mathbf{x}, X$  instead of  $K^{(n)}, \mathbf{x}^{(n)}, X^{(n)}$  for convenience, when no confusion is possible. In the whole article by  $C, c$  we denote universal constants and by  $C_a, c_a$  constants depending only on the parameter  $a$ . In both cases the values of constants may differ between occurrences.

## 2. LIMITING DISTRIBUTION OF THE SINGULAR VALUES OF SHIFTS

This section is devoted to the proof of the following result.

**Theorem 2.1** (Limiting singular values distribution of shifted matrices). *For every  $z \in \mathbb{C}$ , there exists a unique probability measure  $\nu_z$  on  $\mathbb{R}_+$ , absolutely continuous with respect to the Lebesgue measure, depending only on  $z$ , such that if  $X^{(n)}$  is as in (2) with  $K_n = o(n^{1/2})$ , and if  $A^{(n)} = \frac{1}{\sqrt{n}}X^{(n)}$ , then, for every  $z \in \mathbb{C}$ ,*

(j)

$$\nu_{\sqrt{(A^{(n)} - z\text{Id}_n)^*(A^{(n)} - z\text{Id}_n)}} = \frac{1}{n} \sum_{k=1}^n \delta_{s_k(A^{(n)} - z\text{Id}_n)} \xrightarrow[n \rightarrow \infty]{\rightsquigarrow} \nu_z;$$

(jj)

$$U(z) = - \int_{\mathbb{R}_+} \log(s) \nu_z(ds) = \begin{cases} -\log|z| & \text{if } |z| > 1; \\ \frac{1-|z|^2}{2} & \text{otherwise.} \end{cases}$$

Furthermore, for  $z = 0$ , the probability distribution  $\nu_0$  is the so-called Marchenko-Pastur quarter circular law of Lebesgue density  $x \mapsto \frac{1}{\pi} \sqrt{4 - x^2} \mathbf{1}_{[0,2]}(x)$ .

Theorem 2.1 is not valid if we only assume that  $K_n = \mathcal{O}(n^{1/2})$ . Indeed, one can construct  $\mathbf{x}^{(n)}$  in (2) in such a way that  $K_n = \mathcal{O}(n^{1/2})$  while 0 is an eigenvalue of  $X^{(n)}$  of multiplicity at least  $cn$ , which would contradict the fact that  $\nu_z$  is absolutely continuous.

*Proof of Theorem 2.1.* First of all, note that since  $K_n = o(n^{1/2})$ , we have

$$\mathbb{E}|X_{11}^{(n)}|^4 = o(n) \quad \text{and} \quad \mathbb{E}|X_{11}^{(n)}|^3 = o(n^{1/2}). \quad (5)$$

If  $K_n = \mathcal{O}(1)$ , the result follows from [2, Theorem 2.11 and Proposition 2.12]. In fact one can check that the proof given there works for  $K_n$  being a small power of  $n$ .

However, to get larger range of  $K_n$  we will adapt the proof by Chatterjee [13], who obtained the semi-circular law for symmetric random matrices with exchangeable entries, by using the Cauchy-Stieltjes trace-resolvent transform, and a generalization of the Lindeberg principle for the Central Limit Theorem.

For a probability measure  $\mu$  supported on  $\mathbb{R}$ , we denote by  $m_\mu$  its Cauchy-Stieltjes transform, defined for  $\xi \in \mathbb{C}_+ = \{z \in \mathbb{C} : \Im(z) > 0\}$  as (see Theorem 2.4.4 in [4])

$$m_\mu(\xi) = \int_{\mathbb{R}} \frac{1}{\lambda - \xi} \mu(d\lambda).$$

If  $M$  is a  $n \times n$  matrix with real spectrum and  $\xi \in \mathbb{C}_+$ , then  $m_{\nu_M}(\xi)$  is the normalized trace of the resolvent  $G = (M - \xi\text{Id}_n)^{-1}$  of  $M$  at point  $\xi$ , namely

$$m_{\nu_M}(\xi) = \frac{1}{n} \sum_{k=1}^n (\lambda_k(M) - \xi)^{-1} = \frac{1}{n} \text{tr}(G).$$

Fix  $z \in \mathbb{C}$ . We start by the usual linearization trick, which consists in constructing a Hermitian matrix, depending linearly on  $A^{(n)} - z\text{Id}_n$ , and for which the eigenvalues are the singular values of  $A^{(n)} - z\text{Id}_n$ . Namely, we consider the  $2n \times 2n$  Hermitian matrix

$$B_n = \begin{pmatrix} 0 & A^{(n)} - z\text{Id}_n \\ (A^{(n)} - z\text{Id}_n)^* & 0 \end{pmatrix}.$$

The eigenvalues of  $B_n$  are the singular values of  $A^{(n)} - z\text{Id}$  multiplied with  $\pm 1$ . More precisely, if  $\lambda > 0$  is a singular value of  $A^{(n)} - z\text{Id}_n$  of multiplicity  $m$ , then  $\pm\lambda$  are both eigenvalues of  $B_n$  of multiplicity  $m$ , and if zero is a singular value of  $A^{(n)} - z\text{Id}_n$  of multiplicity  $m$  then it is also an eigenvalue of  $B_n$  of multiplicity  $2m$ . Thus, to prove the proposition it is enough to show that the spectral measure of  $B_n$  converges weakly in probability to the same measure as the spectral measure of the matrix

$$\tilde{D}_n = \begin{pmatrix} 0 & \frac{1}{\sqrt{n}}G_n - z\text{Id}_n \\ (\frac{1}{\sqrt{n}}G_n - z\text{Id}_n)^* & 0 \end{pmatrix},$$

where  $G_n = (g_{ij})_{1 \leq i, j \leq n}$  is an  $n \times n$  matrix with i.i.d.  $\mathcal{N}(0, 1)$  coefficients, as it is known [6, 10] that  $\nu_{\sqrt{(n^{-1/2}G_n - z\text{Id}_n)^*(n^{-1/2}G_n - z\text{Id}_n)}}$  converges to the measure  $\nu_z$  satisfying (jj).

Let  $b = n^{-2} \sum_{1 \leq i, j \leq n} g_{ij}$  and let  $\mathbf{1}_n \otimes \mathbf{1}_n$  be the  $n \times n$  matrix with all elements equal to one. Define now the matrix

$$D_n = \begin{pmatrix} 0 & \frac{1}{\sqrt{n}}G_n - b\mathbf{1}_n \otimes \mathbf{1}_n - z\text{Id}_n \\ (\frac{1}{\sqrt{n}}G_n - b\mathbf{1}_n \otimes \mathbf{1}_n - z\text{Id}_n)^* & 0 \end{pmatrix}.$$

Since  $D_n$  is a rank two additive perturbation of  $\tilde{D}_n$ , it is well known by using interlacing inequalities [12] that the Kolmogorov distance between  $\nu_{\tilde{D}_n}$  and  $\nu_{D_n}$  is at most  $2/n$ . Thus it is enough to prove that  $\nu_{B_n} - \nu_{D_n} \rightarrow 0$  weakly in probability, which will follow if we prove that for each  $\xi \in \mathbb{C}_+$ ,

$$m_{\nu_{B_n}}(\xi) - m_{\nu_{D_n}}(\xi) \rightarrow 0$$

in probability as  $n \rightarrow \infty$ . Since  $m_{\nu_{D_n}}(\xi)$  converges in probability to the deterministic quantity  $m_{\nu_z}(\xi)$ , it is enough to show that for every smooth function  $g$  with compact support we have

$$|\mathbb{E}g(\Re m_{\nu_{B_n}}(\xi)) - \mathbb{E}g(\Re m_{\nu_{D_n}}(\xi))| \rightarrow 0$$

and that an analogous statement holds for the imaginary parts of  $m_{\nu_{B_n}}$  and  $m_{\nu_{D_n}}$ .

To this end we will use the following Theorem, which is [13, Theorem 1.2].

**Theorem 2.2** (Chatterjee-Lindeberg principle). *Suppose  $X$  is a random vector in  $\mathbb{R}^n$  with exchangeable components of finite fourth moments and let  $Z$  be a standard Gaussian vector in  $\mathbb{R}^n$ , independent of  $X$ . Define*

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu})^2, \quad \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad Y_i = \hat{\mu} + \hat{\sigma}(Z_i - \bar{Z}).$$

Let  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be a  $\mathcal{C}^3$  function, and let  $L'_r(f)$  be a uniform bound on all  $r$ -th partial derivatives of  $f$ , including mixed partials. For each  $p$ , let  $w_p = \mathbb{E}|X_1 - \hat{\mu}|^p$ . Then

$$|\mathbb{E}f(X) - \mathbb{E}f(Y)| \leq 9.5w_4^{1/2}L'_2(f)n^{1/2} + 13w_3L'_3(f)n.$$

It is easy to check that for any matrix valued differentiable function  $M$  of  $x \in \mathbb{R}$  with values in the space of Hermitian matrices and  $G(x) = (M(x) - \xi\text{Id})^{-1}$  ( $\xi \in \mathbb{C}_+$ ), we have

$$\frac{d}{dx}G = -G \frac{dM}{dx} G. \quad (6)$$



Consider now for  $x \in \mathbb{R}^{n \times n}$  the  $2n \times 2n$  matrix  $M(x) = (M_{ij}(x))_{1 \leq i, j \leq 2n}$ , where  $M_{ij}(x) = 0$  if  $i, j \leq n$  or  $i, j > n$ ,  $M_{ij}(x) = \frac{1}{\sqrt{n}}x_{i, j-n} - z\mathbf{1}_{\{i=j\}}$  for  $i \leq n, j > n$  and  $M_{ij}(x) = \frac{1}{\sqrt{n}}x_{i-n, j} - \bar{z}\mathbf{1}_{\{i=j\}}$  for  $i > n, j \leq n$ .

As in [13] we will write  $\partial_\alpha$  for  $\partial/\partial x_\alpha$  for  $\alpha = (i, j)$ ,  $1 \leq i, j \leq n$ . From (6) we get

$$\partial_\alpha \frac{1}{2n} \operatorname{tr} G = -\frac{1}{2n} \operatorname{tr} \left( G(\partial_\alpha M)G \right).$$

Differentiating two more times and using again (6) together with the observation that for any  $\alpha_1, \alpha_2$ ,  $\partial_{\alpha_1} \partial_{\alpha_2} M = 0$ , we get

$$\begin{aligned} \partial_{\alpha_1} \partial_{\alpha_2} \frac{1}{2n} \operatorname{tr} G &= \frac{1}{2n} \sum_{1 \leq i \neq j \leq 2} \operatorname{tr} \left( G(\partial_{\alpha_i} M)G(\partial_{\alpha_j} M)G \right), \\ \partial_{\alpha_1} \partial_{\alpha_2} \partial_{\alpha_3} \frac{1}{2n} \operatorname{tr} G &= -\frac{1}{2n} \sum_{1 \leq i \neq j \neq k \leq 3} \operatorname{tr} \left( G(\partial_{\alpha_i} M)G(\partial_{\alpha_j} M)G(\partial_{\alpha_k} M)G \right). \end{aligned}$$

Note that  $\partial_\alpha M$  has two nonzero entries, each equal to  $n^{-1/2}$ . Moreover  $\|G\| \leq (\Im \xi)^{-1}$ . Thus

$$\left| \partial_\alpha \frac{1}{2n} \operatorname{tr} G \right| = \left| \frac{1}{2n} \operatorname{tr} \left( (\partial_\alpha M)G^2 \right) \right| \leq \frac{1}{n^{3/2}} (\Im \xi)^{-2}.$$

As for higher order derivatives, using the inequalities  $|\operatorname{tr} AB| \leq \|A\|_{\text{HS}} \|B\|_{\text{HS}}$  and  $\|AB\|_{\text{HS}} \leq \|A\| \|B\|_{\text{HS}}$ , we get

$$\begin{aligned} |\partial_{\alpha_1} \partial_{\alpha_2} \frac{1}{2n} \operatorname{tr} G| &\leq \frac{1}{2n} \sum_{1 \leq i \neq j \leq 2} \|G \partial_{\alpha_i} M\|_{\text{HS}} \|G(\partial_{\alpha_j} M)G\|_{\text{HS}} \\ &\leq \frac{1}{2n} \sum_{1 \leq i \neq j \leq 2} \|G\|^3 \|\partial_{\alpha_i} M\|_{\text{HS}} \|\partial_{\alpha_j} M\|_{\text{HS}} \\ &\leq \frac{2}{n^2} (\Im \xi)^{-3} \end{aligned}$$

and similarly

$$|\partial_{\alpha_1} \partial_{\alpha_2} \partial_{\alpha_3} \frac{1}{2n} \operatorname{tr} G| \leq \frac{6\sqrt{2}}{n} \|G\|^4 n^{-3/2} \leq 9n^{-5/2} (\Im \xi)^{-4}.$$

Moreover, since  $x_{ij}$  are real, we have  $\partial_\alpha \Re G = \Re(\partial_\alpha G)$  and a similar equality for the imaginary parts. Using this together with the above estimates on the derivatives, we get that for any smooth function  $g$  on  $\mathbb{R}$ , with compact support, the function  $f: \mathbb{R}^{n^2} \rightarrow \mathbb{R}$ , defined as  $f(x) = g(\frac{1}{2n} \Re \operatorname{tr} G(x))$ , satisfies

$$L'_2(f) \leq C_{g, \xi} n^{-2}, L'_3(f) \leq C_{g, \xi} n^{-5/2}.$$

Thus, by Theorem 2.2 (applied with  $n^2$  instead of  $n$ ) and (5),

$$|\mathbb{E}g(\Re m_{\nu_{B_n}}(\xi)) - \mathbb{E}g(\Re m_{\nu_{D_n}}(\xi))| \leq C_{g, \xi} (n^{-1} (\mathbb{E}|X_{11}^{(n)}|^4)^{1/2} + n^{-1/2} \mathbb{E}|X_{11}^{(n)}|^3) \rightarrow 0$$

as  $n \rightarrow \infty$ . Since an analogous convergence holds for the imaginary parts, this ends the proof of Theorem 2.1.  $\square$

**Remark 2.3** (Assumptions). *One can see from the above proof that the theorem remains true under an assumption weaker than  $K_n = o(n^{1/2})$ , since it is enough to assume (5).*

## 3. COMBINATORIAL TALAGRAND'S CONCENTRATION INEQUALITY

The following result is a combinatorial analogue of the Talagrand concentration of measure inequality for convex functions under product measures. It plays a crucial role in our approach, and may be of independent interest. We give a proof for completeness.

**Theorem 3.1** (Concentration inequality for convex functions). *Let  $x_1, \dots, x_n \in [0, 1]$  and let  $\varphi: [0, 1]^n \rightarrow \mathbb{R}$  be an  $L$ -Lipschitz convex function. Let  $\pi$  be a random uniform permutation of the set  $[n]$  and let  $Z = \varphi(x_{\pi(1)}, \dots, x_{\pi(n)})$ . Then for all  $t > 0$ ,*

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t) \leq 2 \exp(-ct^2/L^2). \quad (7)$$

*Proof of Theorem 3.1.* The randomness of  $Z$  comes entirely from  $\pi$ . Let  $P_n$  be the uniform measure on the symmetric group  $S_n$ . The proof follows closely the original argument given by Talagrand in the case of product measures [29] and is based on the following theorem.

**Theorem 3.2** (Talagrand [28]). *For  $\pi \in S_n$  and  $A \subseteq S_n$  define*

$$U_A(\pi) = \{s \in \{0, 1\}^n : \exists \tau \in A \forall l \leq n s_l = 0 \implies \tau(l) = \pi(l)\}.$$

*Let  $V_A(\pi) = \text{ConvexHull}(U_A(\pi))$  and  $f(A, \pi) = \inf\{|s|^2 : s \in V_A(\pi)\}$ . Then*

$$\int_{S_n} \exp\left(\frac{1}{16}f(A, \pi)\right) dP_n(\pi) \leq \frac{1}{P_n(A)}.$$

In what follows, for  $\tau \in S_n$  and  $x \in \mathbb{R}^n$ , we will denote  $x_\tau = (x_{\tau(1)}, \dots, x_{\tau(n)})$ . It is well known that up to universal constants it is enough to prove (7) with the median instead of the mean [21]. Let  $Z(\tau) = \varphi(x_\tau)$  and  $A = \{\tau \in S_n : Z(\tau) \leq \text{Med } Z\}$ . We have  $P_n(A) \geq 1/2$ , so by Theorem 3.2 and Chebyshev's inequality we get

$$P_n(\pi : f(A, \pi) \geq t^2) \leq 2 \exp(-ct^2).$$

If  $f(A, \pi) < t^2$ , then there exist  $s^1, \dots, s^m \in U_A(\pi)$ ,  $p_1, \dots, p_m \geq 0$ ,  $p_1 + \dots + p_m = 1$ , such that

$$\left| \sum_{i=1}^m s^i p_i \right| < t.$$

Let  $\tau_i \in A$  be such that  $s_j^i = 0 \implies \tau_i(j) = \pi(j)$ . We have for  $j = 1, \dots, n$ ,

$$\left| x_{\pi(j)} - \sum_{i=1}^m x_{\tau_i(j)} p_i \right| \leq \sum_{i=1}^m |x_{\pi(j)} - x_{\tau_i(j)}| p_i \leq 2 \sum_{i=1}^m s_j^i p_i,$$

so

$$\left| x_\pi - \sum_{i=1}^m p_i x_{\tau_i} \right| \leq 2 \left| \sum_{i=1}^m s^i p_i \right| < 2t.$$

Moreover, by convexity  $\varphi(\sum_i p_i x_{\tau_i}) \leq \sum_{i=1}^m p_i \varphi(x_{\tau_i}) \leq \text{Med } Z$  and so by the Lipschitz condition  $\varphi(x_\pi) < \text{Med } Z + 2Lt$ . This shows that

$$P_n(\pi : \varphi(x_\pi) \geq \text{Med } Z + 2Lt) \leq P_n(\pi : f(A, \pi) \geq t^2) \leq 2 \exp(-ct^2).$$

To prove the bound on the lower tail let us now denote

$$A = \{\tau \in S_n : Z(\tau) \leq \text{Med } Z - 2Lt\}.$$

By Theorem 3.2 we have

$$\int_{S_n} \exp\left(cf(A, \pi)\right) dP_n(\pi) \leq \frac{1}{P_n(A)}$$

But if  $f(A, \pi) < t^2$ , then by a similar argument as above  $\varphi(x_\pi) < \text{Med } Z$ , so the left hand side above is bounded from below by  $\frac{1}{2} \exp(ct^2)$ , which gives  $P_n(A) \leq 2 \exp(-ct^2)$ .  $\square$

Using integration by parts and the triangle inequality in  $L_p$ , we obtain the following.

**Corollary 3.3** (Moments). *In the setting of Theorem 3.1 we have for any  $p \geq 2$ ,*

$$\|Z\|_p \leq \|Z\|_1 + CL\sqrt{p}.$$

#### 4. PROOF OF THEOREM 1.1 (SMALLEST SINGULAR VALUE)

Throughout this section,  $X^{(n)}$  is as in (2). We will prove the following more precise version of Theorem 1.1, which is actually used in the proof of Theorem 1.2.

**Theorem 4.1.** *Under the assumptions of Theorem 1.1, for any  $\varepsilon \in (0, 1)$ ,*

$$\begin{aligned} \mathbb{P}(s_n(X^{(n)} - \sqrt{n}z\text{Id}) \leq \frac{1}{K_n + |z|}\varepsilon n^{-1/2}) \\ \leq CK_n^2(K_n + |z|)\log(1 + K_n + |z|)\varepsilon \\ + C\frac{K_n^4(K_n + |z|)\log^{3/2}(1 + K_n + |z|)}{n^{1/2}}. \end{aligned}$$

Let us first introduce some tools we will need in the proof of Theorem 4.1. One of our main tools is a version of Talagrand's concentration inequality for convex functions stated in Theorem 3.1, and Corollary 3.3. The next two lemmas are rather standard corollaries. We denote by  $S_{\mathbb{C}}^{n-1} = \{z \in \mathbb{C}^n : |z| = 1\}$  the unit Euclidean ball in  $\mathbb{C}^n$ .

**Lemma 4.2** (Operator norm). *For all  $t > 0$  with probability at least  $1 - \exp(-ct^2/K_n^2)$ ,*

$$\|X^{(n)}\| \leq CK_n\sqrt{n} + t.$$

*Proof of Lemma 4.2.* Let us consider a  $(1/4)$ -net  $\mathcal{N}$  in  $S_{\mathbb{C}}^{n-1}$  of cardinality  $9^{2n}$  (it exists by standard volumetric estimates). We have

$$\|X^{(n)}\| \leq \frac{16}{7} \sup_{x, y \in \mathcal{N}} |\langle X^{(n)}x, y \rangle|.$$

Note that  $\mathbb{E}\langle X^{(n)}x, y \rangle = 0$ . Since for  $x, y \in S_{\mathbb{C}}^{n-1}$ ,  $A \mapsto |\langle Ax, y \rangle|$  is a 1-Lipschitz function with respect to the Hilbert-Schmidt norm, we get by Theorem 3.1 and the union bound

$$\forall t \geq CK_n, \quad \mathbb{P}(\|X^{(n)}\| \geq t\sqrt{n}) \leq 2 \cdot 9^{4n} \exp(-ct^2n/K_n^2) \leq \exp(-ct^2n/K_n^2).$$

Integrating the above inequality by parts, we get  $\mathbb{E}\|X^{(n)}\| \leq CK_n\sqrt{n}$ . The lemma now follows by another application of Theorem 3.1, this time to the function  $X \mapsto \|X\|$ .  $\square$

Let  $\text{Sparse}(\delta) = \{x \in \mathbb{C}^n : \sum_{k=1}^n \mathbf{1}_{x_k \neq 0} \leq \delta n\}$  be the set of  $\delta n$  sparse vectors in  $\mathbb{C}^n$ . Similarly as in [27] we partition the unit sphere  $S_{\mathbb{C}}^{n-1}$  into compressible vectors and incompressible vectors:

$$\text{Comp}(\delta, \rho) = \{x \in S_{\mathbb{C}}^{n-1} : \text{dist}(x, \text{Sparse}(\delta)) \leq \rho\} \quad \text{and} \quad \text{Incomp}(\delta, \varepsilon) = S_{\mathbb{C}}^{n-1} \setminus \text{Comp}(\delta, \varepsilon).$$

If  $M = (M_{ij})_{1 \leq i \leq n, 1 \leq j \leq k}$  and  $I \subseteq [n]$ , then we denote by  $M|_I$  the submatrix  $(M_{ij})_{i \in I, 1 \leq j \leq k}$  of  $M$  formed by the rows of  $M$  labeled by  $I$ .

**Lemma 4.3** (Compressible vectors). *For all  $\alpha \in (0, 1)$ , for all  $I \subseteq [n]$ , such that  $\#I \geq \alpha n$ , for every  $z \in \mathbb{C}$ , if  $n \geq C_\alpha K_n^2$ , then*

a) *for any  $x \in S_{\mathbb{C}}^{n-1}$ , with probability at least  $1 - \exp(-c_\alpha n/K_n^2)$ ,*

$$|X|_I^{(n)}x - \sqrt{n}zx|_I| \geq c_\alpha\sqrt{n},$$

b) with probability at least  $1 - \exp(-c_\alpha n/K_n^2)$ , for all  $x \in \text{Comp}(\delta, \rho)$ ,

$$|X_{|I}^{(n)}x - \sqrt{n}zx_{|I}| \geq c_\alpha\sqrt{n},$$

where

$$\rho = \frac{c'_\alpha}{K_n + |z|} \quad \text{and} \quad \delta = \frac{c''_\alpha}{K_n^2 \log(1 + K_n + |z|)}. \quad (8)$$

*Proof of Lemma 4.3.* We abridge  $K_n$  and  $X^{(n)}$  into  $X$  and  $K$ . Consider first an arbitrary  $x \in S_{\mathbb{C}}^{n-1}$ . Using (3) and  $|x|_1 := |x_1| + \dots + |x_n| \leq \sqrt{n}$ , we get

$$\begin{aligned} \mathbb{E}|X_{|I}x - \sqrt{n}zx_{|I}|^2 &= \sum_{i \in I} \mathbb{E} \left| \sum_{j=1}^n X_{ij}x_j - \sqrt{n}zx_i \right|^2 \\ &\geq \sum_{i \in I} \mathbb{E} \left| \sum_{j=1}^n X_{ij}x_j \right|^2 \\ &= \sum_{i \in I} \sum_{j=1}^n \mathbb{E}X_{ij}^2|x_j|^2 + \sum_{i \in I} \sum_{1 \leq j \neq k \leq n} \mathbb{E}X_{ij}X_{ik}x_j\bar{x}_k \\ &\geq \#I - \frac{\#I}{n^2 - 1}|x|_1^2 \geq \#I/3. \end{aligned}$$

Since the function  $X \mapsto |X_{|I}x - zx_{|I}|$  is convex and 1-Lipschitz with respect to the Hilbert-Schmidt norm, by Corollary 3.3 and the assumption on  $I$  we get, for  $n \geq C_\alpha K^2$ ,

$$\mathbb{E}|X_{|I}x - \sqrt{n}zx_{|I}| \geq c\sqrt{\alpha n} - CK \geq c_\alpha\sqrt{n}.$$

By Theorem 3.1 we get

$$\mathbb{P}(|X_{|I}x - \sqrt{n}zx_{|I}| \leq c_\alpha\sqrt{n}) \leq 2 \exp(-c_\alpha n/K^2), \quad (9)$$

which proves the first part of the lemma. Now, for each  $\delta, \rho \in (0, 1]$ , the set  $\text{Sparse}(\delta) \cap S_{\mathbb{C}}^{n-1}$  admits a  $\rho$ -net of cardinality

$$\binom{n}{\lfloor \delta n \rfloor} (3/\rho)^{2\delta n} \leq \left(\frac{C}{\rho^2 \delta}\right)^{\delta n}.$$

Thus for  $\rho = \frac{c'_\alpha}{K+|z|}$  and  $\delta = \frac{c''_\alpha}{K^2 \log(1+K+|z|)}$  by the union bound, the inequality (9), Lemma 4.2, the estimate  $K \geq 1$ , and the triangle inequality, we get with probability at least

$$1 - \exp(-c_\alpha n/K^2),$$

for all  $x \in \text{Comp}(\delta, \rho)$ ,

$$|X_{|I}x - \sqrt{n}zx_{|I}| \geq c_\alpha\sqrt{n} - C\rho(CK + |z|)\sqrt{n} \geq \frac{1}{2}c_\alpha\sqrt{n}.$$

□

The final ingredient is a Berry-Esseen type estimate for linear combinations of exchangeable random variables.

**Lemma 4.4** (Berry-Esseen type estimate for exchangeable variables). *Consider two sequences of real numbers  $x = (x_1, \dots, x_n)$  and  $a = (a_1, \dots, a_n)$ , such that for some constants  $K, L > 0$ ,*

- $\sum_{i=1}^n x_i = 0$ ,  $\sum_{i=1}^n x_i^2 = n$ ,  $|x_i| \leq K$ ,
- $|a_i| \leq \frac{L|a|}{\sqrt{n}}$ ,  $a \neq \pm n^{-1/2}(|a|, \dots, |a|)$ .

Let  $\pi$  be a random (uniform) permutation of  $[n]$  and define

$$W := \sum_{i=1}^n a_i x_{\pi(i)}.$$

Then  $\sigma^2 := \mathbb{E}W^2 = \frac{1}{n-1} (n \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2) > 0$  and

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(W \leq t) - \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^t e^{-x^2/2\sigma^2} dx \right| \leq \frac{34LK|a|}{\sigma\sqrt{n}}. \quad (10)$$

*Proof of Lemma 4.4.* The result follows from the combinatorial Central Limit Theorem (CLT) [14, equation (4.105) and Theorem 6.1]. Namely, let  $(c_{ij})_{i,j \leq n}$  be an array of real numbers and let

$$W = \sum_{i=1}^n c_{i\pi(i)},$$

where  $\pi$  is a random (uniform) permutation of  $[n]$ . Then

$$\sigma^2 := \text{Var } W = \frac{1}{n-1} \sum_{i,j=1}^n \left( c_{ij} - \frac{1}{n} \sum_{k=1}^n c_{ik} - \frac{1}{n} \sum_{k=1}^n c_{kj} + \frac{1}{n^2} \sum_{k,l=1}^n c_{kl} \right)^2.$$

Moreover, if  $\sigma > 0$ , then for every  $\varepsilon > 0$ ,

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\sigma^{-1}\left(W - \frac{1}{n} \sum_{i,j=1}^n c_{ij}\right) \leq t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp(-x^2/2) dx \right| \leq \frac{16.3A}{\sigma},$$

where

$$A = \max_{1 \leq i,j \leq n} \left| c_{ij} - \frac{1}{n} \sum_{k=1}^n c_{ik} - \frac{1}{n} \sum_{k=1}^n c_{kj} + \frac{1}{n^2} \sum_{k,l=1}^n c_{kl} \right|.$$

The lemma follows by setting  $c_{ij} = a_i x_j$  and elementary calculations.  $\square$

**Remark 4.5** (Conditional expectation). *Before we proceed with the proof let us make a comment concerning conditional expectation. Namely, in the proof of Theorem 4.1 as well as in other proofs in the article we will encounter a situation in which a deterministic set  $E \subset [n] \times [n]$  is given and in addition one considers a random set  $F \subseteq ([n] \times [n]) \setminus E$ , measurable with respect to  $\mathcal{F}' = \sigma((\pi(i, j))_{(i,j) \in E})$ . One can then consider the  $\sigma$ -field  $\mathcal{F} = \sigma(\mathcal{F}', (\pi(i, j))_{(i,j) \in F})$  and the random vector  $\pi' = (\pi(i, j))_{(i,j) \in (E \cup F)^c}$  (this notation is slightly informal as the set  $F$  itself is random, but this should not lead to misunderstanding). One can then see that conditionally on  $\mathcal{F}$ ,  $(\pi(i, j))_{(i,j) \in (E \cup F)^c}$  is distributed as a random (uniform) bijection from the set  $(E \cup F)^c$  to the set  $([n] \times [n]) \setminus \{\pi(i, j) : (i, j) \in E \cup F\}$ . In particular the vector  $(X_{ij})_{(i,j) \in (E \cup F)^c}$  is distributed as a random permutation of the sequence obtained from  $(\mathbf{x}_{ij})_{i,j=1}^n$  after removing the elements  $(X_{ij})_{(i,j) \in E \cup F}$  and one can apply to it conditionally e.g. Theorem 3.1 or Lemma 4.4.*

*Proof of Theorem 4.1.* Again we will write  $X$  instead of  $X^{(n)}$  and  $K$  instead of  $K_n$ . We may and will assume that

$$n \geq CK^2 \log(1 + K + |z|) \quad (11)$$

(otherwise the bounds of the theorem become trivial as the right-hand side exceeds one).

Let  $B = X - \sqrt{n}z\text{Id}$  and  $s_n = s_n(B)$ . Since  $s_n = \inf_{x \in S_{\mathbb{C}}^{n-1}} |Bx|$ , it is enough to bound from below the quantities

$$\beta = \inf_{x \in \text{Comp}(\delta, \rho)} |Bx|, \quad \zeta = \inf_{x \in \text{Incomp}(\delta, \rho)} |Bx|$$

for some  $\delta = \delta_{K,z}$ ,  $\rho = \rho_{K,z}$ . By Lemma 4.3 for

$$\rho = \frac{c}{K + |z|} \text{ and } \delta = \frac{c}{K^2 \log(1 + K + |z|)}, \quad (12)$$

we have for  $n \geq CK^2$ ,

$$\beta \geq c\sqrt{n} \quad (13)$$

with probability at least  $1 - \exp(-cn/K^2)$ . By Lemma 3.5 from [27],

$$\mathbb{P}(\zeta \leq \rho\varepsilon n^{-1/2}) \leq \frac{1}{\delta n} \sum_{i=1}^n \mathbb{P}(\text{dist}(X_i - z\sqrt{n}e_i, H_i) \leq \varepsilon),$$

where  $X_i$ 's are the columns of  $X$  and  $H_i = \text{span}(X_k : k \neq i)$ .

Let now  $\eta^i$  be a random (unit) normal to  $H_i$  defined as a measurable function of  $(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ . We have  $\text{dist}(X_i - z\sqrt{n}e_i, H_i) \geq |\langle X_i - z\sqrt{n}e_i, \eta^i \rangle|$ , so by exchangeability of the entries of  $X$ ,

$$\mathbb{P}(\zeta \leq \rho\varepsilon n^{-1/2}) \leq \frac{1}{\delta} \mathbb{P}(|\langle X_n, \eta^n \rangle - z\sqrt{n}\langle e_n, \eta^n \rangle| \leq \varepsilon). \quad (14)$$

To estimate the right hand side we will use Lemma 4.4. To be able to apply it, we have to prove certain properties of the vector  $\eta = \eta^n$ . Let us define for  $\theta, r \in [0, 1]$ ,

$$\mathcal{G}(\theta, r) = \bigcup_{I \subseteq [n], \#I \geq (1-\theta)n} \mathcal{G}_I(\theta, r),$$

where

$$\mathcal{G}_I(\theta, r) = \{x \in S_{\mathbb{C}}^{n-1} : \exists \xi \in \mathbb{C}, |\xi| \leq 1 |x|_I - \frac{\xi}{\sqrt{\#I}} \mathbf{1}_I| \leq r\}$$

(here  $\mathbf{1}_I = (u_i)_{i=1}^{\#I}$  with  $u_i = 1$  for all  $i$ ). We will show that with high probability  $\eta \notin \mathcal{G}(\theta, r)$  for certain  $r = r_{K,z}$  and  $\theta = \theta_{K,z}$ . Note that if  $\#I \geq (1-\theta)n$ , then  $\mathcal{G}_I(\theta, r)$  admits a  $3r$ -net  $\mathcal{N}_I$ , such that

$$\#\mathcal{N}_I \leq \frac{C}{r^2} \left(\frac{3}{r}\right)^{2\theta n}.$$

Indeed, for any  $x \in \mathcal{G}_I(\theta, r)$ ,  $x|_I$  can be approximated up to  $2r$  by one of the vectors  $\frac{\xi}{\sqrt{\#I}} \mathbf{1}_I$ , where  $\xi$  comes from an  $r$ -net in the unit disk in  $\mathbb{C}$ , whereas  $x|_{I^c}$  can be approximated by a vector from an  $r$ -net in the unit ball of  $\mathbb{C}^{I^c}$ . Thus the set  $\mathcal{G}(\theta, r)$  admits a  $3r$  net  $\mathcal{N}$  of cardinality

$$\#\mathcal{N} \leq \sum_{k=0}^{\lfloor \theta n \rfloor} \binom{n}{k} \frac{C}{r^2} \left(\frac{3}{r}\right)^{2\theta n} \leq \frac{C}{r^2} \left(\frac{en}{\lfloor \theta n \rfloor}\right)^{\theta n} \left(\frac{3}{r}\right)^{2\theta n} \leq \exp\left(Cn\theta \log\left(\frac{2}{\theta r}\right)\right)$$

for  $n \geq 1/\theta$ . Let  $B'' = (B')^*$ , where  $B'$  is the matrix obtained by removing the last column of  $B$ . By exchangeability,  $B''$  has the same distribution as  $(X - \bar{z}\sqrt{n}\text{Id})_{|\{1, \dots, n-1\}}$  and so by the first part of Lemma 4.3, the union bound and Lemma 4.2 we get with probability at least  $1 - \exp(-cn/K^2)$ ,

$$\inf_{x \in \mathcal{G}(\theta, r)} |B''x| \geq c\sqrt{n}$$

for

$$r = \frac{c}{K + |z|}, \theta = \frac{c}{K^2 \log(1 + K + |z|)} \quad (15)$$

(note that (11) implies that  $n \geq 1/\theta$ ). We may and will assume that  $\theta \leq 1/2$ . Combining the above estimate with the equality  $B''\eta = 0$ , we get that with probability at least  $1 - \exp(-cn/K^2)$ ,

$$\eta \notin \mathcal{G}(\theta, r). \quad (16)$$

Now let  $I = \{i \in [n]: |\eta_i| \leq \frac{1}{\sqrt{\theta n}}\}$  and note that by Markov's inequality  $\#I \geq n(1 - \theta)$ . Let  $\mathcal{F}'$  be the  $\sigma$ -field generated by  $(\pi(i, j))_{1 \leq i \leq n, 1 \leq j \leq n-1}$ . In particular  $\eta$  is  $\mathcal{F}'$ -measurable. Let also  $\mathcal{F} = \sigma(\mathcal{F}', \{\pi(i, n)\}_{i \notin I})$ .

Note that  $\sum_{i \in I} X_{in}$  and  $\sum_{i \in I} X_{in}^2$  are  $\mathcal{F}$ -measurable. We will use this fact together with Remark 4.5 to estimate the conditional distribution

$$\mathbb{P}(|\langle X_n, \eta \rangle - Y| \leq \varepsilon | \mathcal{F}),$$

where  $Y$  is any  $\mathcal{F}$ -measurable random variable, by another random variable measurable with respect to  $\mathcal{F}$ . To this end we will use Lemma 4.4.

Let us define  $\hat{\mu} = \frac{1}{\#I} \sum_{i \in I} X_{in}$ ,  $\hat{\Sigma}^2 = \frac{1}{\#I} \sum_{i \in I} (X_{in} - \hat{\mu})^2$ .

Let  $\gamma$  be a small constant which will be fixed later on. Note that by Theorem 3.1 we have

$$\mathbb{P}(|\sum_{i=1}^n X_{in}| \geq \gamma n) \leq 2 \exp(-c\gamma^2 n/K^2). \quad (17)$$

Now, by exchangeability of  $X_{in}, i = 1, \dots, n$  with respect to  $\mathbb{P}(\cdot | \mathcal{F}')$ , we get that on the set  $\Delta_1 = \{|\sum_{i=1}^n X_{in}| \leq \gamma n\} \in \mathcal{F}'$ ,

$$|\mathbb{E}(\sum_{i \in I} X_{in} | \mathcal{F}')| = |\frac{\#I}{n} \sum_{i=1}^n X_{in}| \leq \gamma n.$$

Thus, again by Theorem 3.1 we get on  $\Delta_1$ ,

$$\mathbb{P}(|\sum_{i \in I} X_{in}| \geq 2\gamma n | \mathcal{F}') \leq 2 \exp(-c\gamma^2 n^2 / (K^2 \#I)) \leq 2 \exp(-c\gamma^2 n/K^2). \quad (18)$$

Let now  $\Delta_2 = \{|\hat{\mu}| \leq 4\gamma\} \in \mathcal{F}$ . Combining (17), (18) and the inequality  $\#I \geq (1 - \theta)n \geq n/2$ , we get

$$\mathbb{P}(\Delta_2) \geq 1 - 4 \exp(-c\gamma^2 n/K^2). \quad (19)$$

Now, by Corollary 3.3 applied to  $Z = \sqrt{\sum_{i=1}^n X_{in}^2}$ , for  $n \geq 4C^2 K^2$  we get

$$\sqrt{n} = \sqrt{\mathbb{E} \sum_{i=1}^n X_{in}^2} \geq \mathbb{E} \sqrt{\sum_{i=1}^n X_{in}^2} \geq \sqrt{n} - CK \geq \frac{1}{2} \sqrt{n}$$

and by Theorem 3.1 the set  $\Delta_3 = \{Cn \geq \sum_{i=1}^n X_{in}^2 \geq cn\} \in \mathcal{F}'$  satisfies

$$\mathbb{P}(\Delta_3) \geq 1 - 2 \exp(-cn/K^2). \quad (20)$$

On  $\Delta_3$ ,  $\mathbb{E}(\sum_{i \in I} X_{in}^2 | \mathcal{F}') \geq c(1 - \theta)n$  and by Corollary 3.3,

$$\mathbb{E}(\sqrt{\sum_{i \in I} X_{in}^2} | \mathcal{F}') \geq \sqrt{c(1 - \theta)n} - CK \geq c\sqrt{n},$$

for  $n \geq C'K^2$ , so yet another application of Theorem 3.1 yields that on  $\Delta_3$ ,

$$\mathbb{P}(\sum_{i \in I} X_{in}^2 \leq cn | \mathcal{F}') \leq 2 \exp(-cn/K^2).$$

Combining this with (20) we get

$$\mathbb{P}(Cn \geq \sum_{i \in I} X_{in}^2 \geq cn) \geq 1 - 2 \exp(-cn/K^2). \quad (21)$$

Since  $\hat{\Sigma}^2 = \frac{1}{\#I} \sum_{i \in I} X_{in}^2 - \hat{\mu}^2$ , by the above inequality and (19) (applied with a sufficiently small universal constant  $\gamma$  adjusted to the constant  $c$  in (21)) we get

$$\mathbb{P}(\hat{\Sigma}^2 \geq c) \geq 1 - 2 \exp(-cn/K^2).$$

Let

$$\Delta = \left\{ \sum_{i \in I} X_{ni}^2 \leq Cn, \hat{\Sigma}^2 \geq c, \hat{\mu} \leq 1 \text{ and } \eta \notin \mathcal{G}(\theta, r) \right\} \in \mathcal{F}.$$

By the above inequality together with (16),(19),(21) we have

$$\mathbb{P}(\Delta) \geq 1 - 2 \exp(-cn/K^2). \quad (22)$$

On  $\Delta$  we have  $\text{dist}(\eta|_I, \{\frac{\xi}{\sqrt{\#I}} \mathbf{1}_I : \xi \in \mathbb{C}, |\xi| \leq 1\}) \geq r$ . Let  $\eta'$  and  $\eta''$  be resp. the real and imaginary part of  $\eta$ . Thus at least one of the vectors  $\eta'|_I, \eta''|_I$  is at distance at least  $r/\sqrt{2}$  from  $\{\xi \mathbf{1}_I : \xi \in \mathbb{R}\}$ . Indeed, otherwise we would have  $|\eta'|_I - \frac{\xi'}{\sqrt{\#I}} \mathbf{1}_I|, |\eta''|_I - \frac{\xi''}{\sqrt{\#I}} \mathbf{1}_I| \leq r/\sqrt{2}$  for some  $\xi', \xi'' \in \mathbb{R}$ , and so

$$|\eta|_I - \frac{\xi' + i\xi''}{\sqrt{\#I}} \mathbf{1}_I| \leq r,$$

which implies that  $\text{dist}(\eta|_I, y) \leq r$ , where  $y$  is the orthogonal projection of  $\eta|_I$  onto  $\text{span}(\mathbf{1}_I)$ . Since  $|\eta|_I| \leq 1$  we have  $|y| \leq 1$ , so  $y = \frac{\xi}{\sqrt{\#I}} \mathbf{1}_I$  for some  $\xi$  with  $|\xi| \leq 1$ , which gives a contradiction.

We will consider the case

$$\text{dist}(\eta'|_I, \{\xi \mathbf{1}_I : \xi \in \mathbb{R}\}) \geq r/\sqrt{2}, \quad (23)$$

the other one is analogous.

Assume now that for some  $\mathcal{F}$ -measurable complex random variable  $Y = Y' + iY''$ ,  $|\langle \eta, X_n \rangle - Y| \leq \varepsilon$ . Then  $|\langle \eta', X_n \rangle - Y'| \leq \varepsilon$ . Define  $x_i = (X_{in} - \hat{\mu}) \hat{\Sigma}^{-1}$  for  $i \in I$ .

Since  $\eta', X_{in}, i \notin I$  and  $\hat{\mu}$  are  $\mathcal{F}$ -measurable we have on  $\Delta$ ,

$$\mathbb{P}(|\langle \eta, X_n \rangle - Y| \leq \varepsilon | \mathcal{F}) \leq \sup_{u \in \mathbb{R}} \mathbb{P}(|\sum_{i \in I} \eta'_i x_i - u| \leq \varepsilon \hat{\Sigma}^{-1} | \mathcal{F}) \leq \sup_{u \in \mathbb{R}} \mathbb{P}(|\sum_{i \in I} \eta'_i x_i - u| \leq C\varepsilon | \mathcal{F}), \quad (24)$$

where in the second inequality we used the fact that on  $\Delta$ ,  $\hat{\Sigma}^2 \geq c$ . Moreover, by Remark 4.5, the right-hand side above equals

$$\sup_{u \in \mathbb{R}} \mathbb{P}(|\sum_{i \in I} \eta'_i x_{\tau(i)} - u| \leq C\varepsilon | \mathcal{F}),$$

where  $\tau$  is a random permutation of  $I$ , distributed (conditionally on  $\mathcal{F}$ ) uniformly, and so we are in position to use Lemma 4.4. Denote  $W = \sum_{i \in I} \eta'_i x_{\tau(i)}$ ,  $\sigma^2 = \mathbb{E}(W^2 | \mathcal{F})$ . Note that  $\sum_{i \in I} x_i = 0$  and  $\sum_{i \in I} x_i^2 = \#I$ . Moreover on  $\Delta$  we have  $|x_i| \leq CK$ . Using the fact that the density of a Gaussian distribution with variance  $\sigma^2$  is bounded from above by  $\sigma^{-1}$  and  $\#I \geq n/2$ , we get by Lemma 4.4 that on  $\Delta$ ,

$$\mathbb{P}(|\sum_{i \in I} \eta'_i x_{\tau(i)} - u| \leq \varepsilon | \mathcal{F}) \leq C\varepsilon \sigma^{-1} + C \frac{KL|\eta'_I|}{\sigma n^{1/2}},$$



where  $L = \max_{i \in I} \sqrt{\#I} |\eta'_i| / |\eta'_I| \leq \max_{i \in I} \sqrt{n} |\eta'_i| / |\eta'_I|$ . Note that by the definition of  $I$ , we have  $|\eta'_i| \leq 1/\sqrt{\theta n}$ , so  $L \leq \frac{1}{\sqrt{\theta} |\eta'_I|}$  and the above bound implies that on  $\Delta$ ,

$$\mathbb{P}(|\sum_{i \in I} \eta'_i x_{\tau(i)} - u| \leq \varepsilon | \mathcal{F}) \leq C\varepsilon\sigma^{-1} + C\frac{K}{\sigma\theta^{1/2}n^{1/2}}. \quad (25)$$

It remains to estimate  $\sigma$  from below on the set  $\Delta$ .

By Lemma 4.4 we have

$$\sigma^2 = \frac{\#I}{\#I - 1} \left( \sum_{i \in I} \eta_i'^2 - \frac{1}{\#I} \left( \sum_{i \in I} \eta_i' \right)^2 \right) \geq \sum_{i \in I} \eta_i'^2 - \frac{1}{\#I} \left( \sum_{i \in I} \eta_i' \right)^2.$$

The function  $t \mapsto \sqrt{t}$  is  $C/|\eta'_I|$  Lipschitz on  $(|\eta'_I|^2/2, \infty)$ , so we get that if  $\sum_{i \in I} (\eta_i')^2 - \frac{1}{\#I} (\sum_{i \in I} \eta_i')^2 \leq \kappa$  with  $\kappa < |\eta'_I|^2/4$ , then

$$\left( \sum_{i \in I} (\eta_i')^2 \right)^{1/2} - \frac{1}{\sqrt{\#I}} \left| \sum_{i \in I} \eta_i' \right| \leq C \frac{\kappa}{|\eta'_I|},$$

which can be rewritten as

$$|\eta'_I| - \left| \langle \eta'_I, \frac{1}{\sqrt{\#I}} \mathbf{1}_I \rangle \right| \leq C \frac{\kappa}{|\eta'_I|}.$$

Multiplying both sides by  $2|\eta'_I|$  and using that  $|\frac{1}{\sqrt{\#I}} \mathbf{1}_I| = 1$ , we get

$$|\eta'_I - \frac{h}{\sqrt{\#I}} \mathbf{1}_I|^2 \leq C\kappa \quad (26)$$

where  $h = |\eta'_I| \operatorname{sgn}(\langle \eta'_I, \frac{1}{\sqrt{\#I}} \mathbf{1}_I \rangle)$ . From (23) it follows, that  $|\eta'_I|^2 \geq r^2/2$ . For  $\kappa = cr^2$  with a sufficiently small absolute constant  $c$  (in particular we want to assure that  $\kappa \leq |\eta'_I|^2/4$ ), the right hand side of (26) is smaller than  $r^2/2$  and so (again by (23)) the inequality (26) cannot hold on  $\Delta$ .

Thus, on  $\Delta$  we have

$$\sigma \geq cr,$$

which, when combined with (25) gives

$$\mathbb{P}(|\sum_{i \in I} \eta'_i x_{\tau(i)} - u| \leq \varepsilon | \mathcal{F}) \leq C\frac{\varepsilon}{r} + C\frac{K}{r\theta^{1/2}n^{1/2}}$$

on  $\Delta$ . Going now back to (24), (22) and (14) we get

$$\mathbb{P}(\zeta \leq \rho\varepsilon n^{-1/2}) \leq C\frac{\varepsilon}{\delta r} + C\frac{K}{\delta r\theta^{1/2}n^{1/2}} + \frac{2}{\delta} \exp(-cn/K^2).$$

Together with (13) this gives (after adjusting  $c$ )

$$\mathbb{P}(s_n(B) \leq \rho\varepsilon n^{-1/2}) \leq C\frac{\varepsilon}{\delta r} + C\frac{K}{\delta r\theta^{1/2}n^{1/2}} + 2 \exp(-cn/K^2).$$

Plugging in the values of  $\rho, \delta$  (equation (12)) and  $\theta, r$  (equation (15)) we get

$$\begin{aligned} \mathbb{P}(s_n(B) \leq \frac{1}{K + |z|} \varepsilon n^{-1/2}) \\ \leq 2 \exp(-cn/K^2) + CK^2(K + |z|) \log(1 + K + |z|) \varepsilon + C \frac{K^4(K + |z|) \log^{3/2}(1 + K + |z|)}{n^{1/2}}, \end{aligned}$$

which ends the proof (we again adjust the constants to remove the first term on the right-hand side).  $\square$

## 5. PROOF OF THEOREM 1.2 (CIRCULAR LAW FOR THE FIRST MODEL)

By [7, Lemma A2] or [10], to prove that  $\nu_{A^{(n)}}$  converges weakly in probability to the uniform measure on the unit disc, it is enough to show that the following is true:

- (i) **Singular values of shifts.** For all  $z \in \mathbb{C}$ , there exists a non random probability measure  $\nu_z$  on  $\mathbb{R}_+$ , absolutely continuous with respect to the Lebesgue measure, depending only on  $z$ , and such that

$$\nu_{z,n} := \frac{1}{n} \sum_{k=1}^n \delta_{s_k(A^{(n)} - z\text{Id}_n)} \xrightarrow[n \rightarrow \infty]{\rightsquigarrow} \nu_z.$$

Moreover, for almost all  $z \in \mathbb{C}$ ,

$$U(z) := - \int_{\mathbb{R}_+} \log(s) \nu_z(ds) = \begin{cases} -\log|z| & \text{if } |z| > 1, \\ \frac{1}{2}(1 - |z|^2) & \text{otherwise;} \end{cases}$$

- (ii) **Uniform integrability.** For all  $z \in \mathbb{C}$ , the function  $s \mapsto \log(s)$  is uniformly integrable in probability with respect to the family of measures  $\{\nu_{z,n}\}_{n \geq 1}$ , i.e.

$$\forall \varepsilon > 0, \quad \lim_{t \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \int_{\mathbb{R}_+} |\log s| \mathbf{1}_{\{|\log s| > t\}} d\nu_{z,n}(s) > \varepsilon \right) = 0.$$

The first item (i) is settled by Theorem 2.1. It remains to prove assertion (ii), which is the aim of the rest of this section. We follow the Tao and Vu approach [32]. We will combine estimates on the smallest singular value coming from Section 4 with a rougher bound on intermediate singular values of the matrix, obtained in the following lemmas.

**Lemma 5.1** (Distance to a random subspace). *Let  $R$  be a deterministic  $n \times n$  matrix. Denote the rows of  $X^{(n)} + R$  by  $Z_1, \dots, Z_n$ . Consider  $k \leq n - 1$  and let  $H$  be the random subspace of  $\mathbb{C}^n$  spanned by  $Z_1, \dots, Z_k$ . Then with probability at least  $1 - 2 \exp(-c(n - k)/K_n^2)$ ,*

$$\text{dist}(Z_{k+1}, H) \geq c\sqrt{n - k}.$$

*Proof of Lemma 5.1.* In what follows we will suppress the superscript  $(n)$  and write simply  $X_{ij}$  for the entries of the matrix  $X^{(n)}$ . The rows of  $X^{(n)}$  will be denoted by  $X_1, \dots, X_n$ .

We can assume that  $k \leq n - CK_n^2$  for some absolute constant  $C$ , otherwise the estimate on probability given in the lemma becomes trivial for  $c$  sufficiently small. Consider the  $\sigma$ -field  $\mathcal{F}$  generated by  $(\pi(i, j))_{1 \leq i \leq k, 1 \leq j \leq n}$ .

We will first replace  $H$  by  $\tilde{H} = \text{span}(H, \mathbf{1}, Y)$  where  $\mathbf{1}$  is the vector of ones and  $Y$  is the  $(k + 1)$ -st row of  $R$ . We can assume that with probability one  $\tilde{H}$  is of dimension  $l = k + 2$  (otherwise we may enlarge  $\tilde{H}$  to a subspace of dimension  $l$  in an  $\mathcal{F}$ -measurable way). Let  $X = X_{k+1} - \hat{\mu}\mathbf{1}$ , where  $\hat{\mu} = \frac{1}{n(n-k)} \sum_{i=k+1}^n \sum_{j=1}^n X_{ij}$  (note that  $\hat{\mu}$  is  $\mathcal{F}$ -measurable) and denote the coordinates of  $X$  by  $x_1, \dots, x_n$ . Note that  $\text{dist}(Z_{k+1}, H) \geq \text{dist}(Z_{k+1}, \tilde{H}) = \text{dist}(X, \tilde{H})$ .

We have  $\mathbb{E}(x_i | \mathcal{F}) = 0$ . Moreover, since  $\mathbb{E}\hat{\mu} = 0$ , by Theorem 3.1 we have

$$\mathbb{P}(|\hat{\mu}| \geq \gamma) \leq 2 \exp(-c\gamma^2 n(n - k)/K_n^2). \quad (27)$$

Consider now  $M^2 := \sum_{i=k+1}^n \sum_{j=1}^n X_{ij}^2$  (which is  $\mathcal{F}$ -measurable) and note that by Corollary 3.3,  $\mathbb{E}M \geq \sqrt{\mathbb{E}M^2} - CK_n \geq 2^{-1} \sqrt{n(n - k)}$  (where we used the assumption  $n - k \geq CK_n^2$ ). Thus again by Theorem 3.1 we have

$$\mathbb{P}(M^2 \leq cn(n - k)) \leq 2 \exp(-cn(n - k)/K_n^2). \quad (28)$$

Define  $\hat{\Sigma}^2 = \mathbb{E}(x_j^2|\mathcal{F}) = \frac{1}{n(n-k)} \sum_{i=k+1}^n \sum_{j=1}^n (X_{ij} - \hat{\mu})^2 = \frac{1}{n(n-k)} M^2 - \hat{\mu}^2$  and

$$\Delta = \{\hat{\Sigma} \geq c\}.$$

Combining (28) and (27) with  $\gamma$  small enough we obtain

$$\mathbb{P}(\Delta) \geq 1 - 2 \exp(-cn(n-k)/K_n^2). \quad (29)$$

Let  $P$  be the orthogonal projection on  $\tilde{H}^\perp$  and let  $e_1, \dots, e_{n-l}$  be an orthonormal basis in  $\tilde{H}^\perp$ . Denote  $e_i = (e_{ij})_{1 \leq j \leq n}$ . We have

$$\begin{aligned} \mathbb{E}(\text{dist}(X, \tilde{H})^2|\mathcal{F}) &= \mathbb{E}(|PX|^2|\mathcal{F}) = \sum_{i=1}^{n-l} \mathbb{E}\left(\left|\sum_{j=1}^n x_j \bar{e}_{ij}\right|^2|\mathcal{F}\right) \\ &= \sum_{i=1}^{n-l} \sum_{j=1}^n |e_{ij}|^2 \mathbb{E}(x_j^2|\mathcal{F}) + \sum_{i=1}^{n-l} \sum_{1 \leq j \neq t \leq n} \bar{e}_{ij} e_{it} \mathbb{E}(x_j x_t|\mathcal{F}). \end{aligned}$$

We have

$$\mathbb{E}(x_j^2|\mathcal{F}) = \hat{\Sigma}^2$$

and for  $j \neq t$ ,

$$\begin{aligned} \mathbb{E}(x_j x_t|\mathcal{F}) &= \frac{1}{n(n-k)[n(n-k)-1]} \sum_{\substack{(a,b) \neq (r,s) \\ k+1 \leq a,r \leq n, 1 \leq b,s \leq n}} (X_{ab} - \hat{\mu})(X_{rs} - \hat{\mu}) \\ &= -\frac{1}{n(n-k)[n(n-k)-1]} \sum_{a=k+1}^n \sum_{b=1}^n (X_{ab} - \hat{\mu})^2 = -\frac{1}{n(n-k)-1} \hat{\Sigma}^2. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{E}(\text{dist}(X, \tilde{H})^2|\mathcal{F}) &= \hat{\Sigma}^2 \sum_{i=1}^{n-l} \left( \sum_{j=1}^n |e_{ij}|^2 - \frac{1}{n(n-k)-1} \sum_{1 \leq j \neq t \leq n} \bar{e}_{ij} e_{it} \right) \\ &= \hat{\Sigma}^2 \sum_{i=1}^{n-l} \left( \sum_{j=1}^n |e_{ij}|^2 - \frac{1}{n(n-k)-1} \left| \sum_{j=1}^n \bar{e}_{ij} \right|^2 + \frac{1}{n(n-k)-1} \sum_{j=1}^n |e_{ij}|^2 \right) \\ &\geq \hat{\Sigma}^2 \sum_{i=1}^{n-l} \left( \frac{n(n-k)}{n(n-k)-1} - \frac{n}{n(n-k)-1} \right) \\ &= \hat{\Sigma}^2 (n-l) \frac{n^2 - kn - n}{n^2 - kn - 1}. \end{aligned}$$

Thus, using the assumption that  $k \leq n - C$  we get for  $n \geq C$  that on  $\Delta$ ,

$$\mathbb{E}(\text{dist}(X, \tilde{H})^2|\mathcal{F}) \geq c(n-k).$$

Using now Corollary 3.3, the fact that  $P$  is 1-Lipschitz and the assumption  $n-k \geq CK_n^2$ , we get on  $\Delta$ ,

$$\mathbb{E}(\text{dist}(X, \tilde{H})|\mathcal{F}) \geq c\sqrt{n-k}.$$

Applying Theorem 3.1 we obtain on  $\Delta$ ,

$$\mathbb{P}(\text{dist}(X, \tilde{H}) \geq c\sqrt{n-k}|\mathcal{F}) \geq 1 - 2 \exp(-c(n-k)/K_n^2),$$

which when combined with (29) gives

$$\mathbb{P}(\text{dist}(X, \tilde{H}) \geq c\sqrt{n-k}) \geq 1 - 2 \exp(-c(n-k)/K_n^2).$$

□

**Lemma 5.2** (Lower bound on the intermediate singular values). *Let  $s_1 \geq \dots \geq s_n$  be the singular values of  $A^{(n)} - z\text{Id}_n$ . If  $K_n = \mathcal{O}(n^{1/(2+\delta)})$  for some  $\delta > 0$ , then there exists  $\gamma \in (0, 1)$  such that for every  $z \in \mathbb{C}$  we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\exists_{n^\gamma \leq i \leq n-1} s_{n-i} \leq c \frac{i}{n}\right) = 0.$$

*Proof of Lemma 5.2.* The proof follows an argument due to Tao and Vu [32]. Let  $R = -\sqrt{n}z\text{Id}$  and recall the notation of Lemma 5.1. For some  $\gamma \in (0, 1)$  to be chosen later on, consider  $i \geq n^\gamma$ . Let  $k = n - \lfloor i/2 \rfloor$  and let  $B$  be the  $k \times n$  matrix with rows  $Z_1, \dots, Z_k$ . By Cauchy interlacing inequalities we have  $\sqrt{n}s_{n-j} = s_{n-j}(X^{(n)} + R) \geq s_{n-j}(B)$  for  $j \geq \lfloor i/2 \rfloor$ . Let  $H_j$ ,  $j = 1, \dots, k$  be the subspace of  $\mathbb{C}^n$  spanned by all the rows of  $B$  except for the  $j$ -th one. By [32, Lemma A4],

$$\sum_{j=1}^k s_j(B)^{-2} = \sum_{j=1}^k \text{dist}(Z_j, H_j)^{-2}.$$

By Lemma 5.1, for each  $j \leq k$ ,  $\text{dist}(Z_j, H_j) \geq c\sqrt{n-k+1} \geq c\sqrt{i}$  with probability at least  $1 - 2\exp(-ci/K_n^2) \geq 1 - 2\exp(-cn^\gamma/K_n^2)$ . (Note that we can use the lemma here thanks to exchangeability of the rows of the matrix). By the union bound, with probability at least  $1 - 2n\exp(-cn^\gamma/K_n^2)$ , we get

$$\sum_{j=1}^k s_j(B)^{-2} \leq C \frac{k}{i}.$$

On the other hand, the left-hand side above is at least  $s_{n-i}(B)^{-2}(k-n+i) \geq s_{n-i}(B)^{-2}i/2$ . This gives that with probability at least  $1 - 2n\exp(-cn^\gamma/K_n^2)$ ,

$$ns_{n-i}^2 \geq s_{n-i}(B)^2 \geq c \frac{i^2}{n - \lfloor n^\gamma/2 \rfloor},$$

which implies that  $s_{n-i} \geq c \frac{i}{n}$ . Taking another union bound over all  $i \geq n^\gamma$  we obtain that  $s_{n-i} \geq c \frac{i}{n}$  for all  $n^\gamma \leq i \leq n-1$  with probability at least  $1 - 2n^2\exp(-cn^\gamma/K_n^2)$ . For some  $\gamma \in (0, 1)$ , we have  $\gamma - 2/(2+\delta) > 0$  and so by the assumption on  $K_n$ ,

$$2n^2\exp(-cn^\gamma/K_n^2) \rightarrow 0,$$

which ends the proof.  $\square$

*Conclusion of the proof of Theorem 1.2.* Recall that we have to prove (ii). By Markov's inequality it suffices to show that for some  $\alpha > 0$  and some constant  $C_z$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\int_0^\infty (s^\alpha + s^{-\alpha}) d\nu_{z,n}(s) > C_z\right) = 0.$$

Note that, using the notation of Lemma 5.2, we have

$$\int_0^\infty (s^\alpha + s^{-\alpha}) d\nu_{z,n}(s) = \frac{1}{n} \sum_{i=1}^n (s_i^\alpha + s_i^{-\alpha}).$$

Note also that for all  $\alpha \in (0, 2]$  we have

$$\begin{aligned}
\left(\frac{1}{n} \sum_{i=1}^n s_i^\alpha\right)^{2/\alpha} &\leq \frac{1}{n} \sum_{i=1}^n s_i^2 \\
&= n^{-1} \|A^{(n)} - z \text{Id}_n\|_{\text{HS}}^2 \\
&\leq 2|z|^2 + 2n^{-2} \sum_{i,j=1}^n |\mathbf{x}_{ij}^{(n)}|^2 \\
&= 2|z|^2 + 2.
\end{aligned} \tag{30}$$

As for the other sum, by the estimate of Theorem 4.1 together with the assumption on  $K_n$  we have with probability tending to one for some finite constant  $\beta > 0$ ,  $s_n \geq n^{-\beta}$ .

Combining this with Lemma 5.2 we get with probability tending to one,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n s_i^{-\alpha} &= \frac{1}{n} \sum_{i=0}^{\lfloor n^\gamma \rfloor} s_{n-i}^{-\alpha} + \frac{1}{n} \sum_{i=\lfloor n^\gamma \rfloor+1}^{n-1} s_{n-i}^{-\alpha} \\
&\leq \frac{1}{n} n^{\beta\alpha} n^\gamma + \frac{1}{n} C \sum_{i=\lfloor n^\gamma \rfloor+1}^{n-1} \left(\frac{n}{i}\right)^\alpha \\
&\leq n^{\beta\alpha+\gamma-1} + C_\alpha n^{\alpha-1} n^{1-\alpha} \leq C_\alpha
\end{aligned}$$

for  $\alpha$  small enough. Together with (30) this gives (ii), and Theorem 1.2 is proved.  $\square$

## 6. PROOF OF THEOREM 1.3 (CIRCULAR LAW FOR THE SECOND MODEL)

*Proof of Theorem 1.3.* Let  $d$  be any distance metrizing the weak convergence of probability measures on  $\mathbb{C}$ , such as the bounded-Lipschitz distance (also referred to as the Fortet-Mourier distance by some authors). Let  $\mathcal{M}_n(\delta)$  be the set of all  $n \times n$  matrices  $\mathbf{x}$ , satisfying (A1-A2) with  $K_n \leq n^{1/(10+\delta/3)}$ . Let us denote by  $A^{(n)}(\mathbf{x}) = n^{-1/2} X^{(n)}(\mathbf{x})$  the matrix constructed from  $\mathbf{x}$  as in (2). Theorem 1.2 implies that for any  $\delta > 0$  and  $\varepsilon > 0$

$$\sup_{\mathbf{x} \in \mathcal{M}_n(\delta)} \mathbb{P}(d(\nu_{A^{(n)}(\mathbf{x})}, \nu^{\text{circ}}) > \varepsilon) \rightarrow 0$$

as  $n \rightarrow \infty$ . Indeed if for some  $\varepsilon$  there exists a sequence  $n_k \rightarrow \infty$  and  $n_k \times n_k$  matrices  $\mathbf{x}_{n_k}$  such that  $\mathbb{P}(d(\nu_{A^{(n_k)}(\mathbf{x}_{n_k})}, \nu^{\text{circ}}) > \varepsilon) > \varepsilon$ , then we can complete this sequence to a sequence  $\mathbf{x}_n$ ,  $n \geq 1$ , violating Theorem 1.2.

Consider now the matrices  $B^{(n)}$  and for any  $n$  define the event

$$\Delta_n = \left\{ \max_{i,j \leq n} \sigma_n^{-1} |Y_{ij}^{(n)} - \mu_n| \leq n^{1/(10+\delta/3)} \right\}.$$

Note that by the union bound, Markov's inequality and exchangeability

$$\mathbb{P}(\Delta'_n) \leq \frac{n^2 \sup_m \mathbb{E} \left( \sigma_m^{-1} |Y_{11}^{(m)} - \mu_m| \right)^{20+\delta}}{n^{\frac{20+\delta}{10+\delta/3}}} \rightarrow 0.$$

Let  $\pi_n$  be a random (uniform) permutation of  $[n] \times [n]$ , independent of  $B^{(n)}$ . Since

$$\bar{B}^{(n)} = \frac{1}{\sqrt{n} \sigma_n} (Y_{\pi_n(i,j)}^{(n)} - \mu_n)_{1 \leq i,j \leq n}$$

has the same distribution as  $B^{(n)}$ , it is enough to show that  $\nu_{\bar{B}^{(n)}}$  converges weakly in probability to the circular law. By the Fubini theorem we have for any  $\varepsilon > 0$ ,

$$\begin{aligned} \mathbb{P}(d(\nu_{\bar{B}^{(n)}}, \nu^{\text{circ}}) > \varepsilon) &\leq \mathbb{P}(\Delta'_n) + \mathbb{E}_B \left( \mathbb{P}_{\pi_n}(d(\nu_{\bar{B}^{(n)}}, \nu^{\text{circ}}) > \varepsilon) \mathbf{1}_{\Delta_n} \right) \\ &\leq \mathbb{P}(\Delta'_n) + \sup_{\mathbf{x} \in \mathcal{M}_n(\delta)} \mathbb{P}(d(\nu_{A^{(n)}(\mathbf{x})}, \nu^{\text{circ}}) > \varepsilon) \rightarrow 0. \end{aligned}$$

□

## REFERENCES

- [1] Radosław Adamczak, *On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries*, Electronic Journal of Probability **16** (2011), 1065–1095. 2
- [2] ———, *Some remarks on the Dozier-Silverstein theorem for random matrices with dependent entries*, To appear in *Random Matrix Theory and Applications*, preprint 2012. 2, 6
- [3] Radosław Adamczak and Djalil Chafaï, *Circular law for random matrices with unconditional log-concave distribution*, preprint arXiv:1303.5838, 2013. 2
- [4] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR-2760897 6
- [5] Zhidong Bai, *Circular law*, Ann. Probab. **25** (1997), no. 1, 494–529. MR-1428519 2
- [6] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR-2567175 7
- [7] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of non-Hermitian heavy tailed random matrices*, Comm. Math. Phys. **307** (2011), no. 2, 513–560. MR-2837123 2, 17
- [8] ———, *Circular law theorem for random Markov matrices*, Probab. Theory Related Fields **152** (2012), no. 3-4, 751–779. MR-2892961 2
- [9] ———, *Spectrum of Markov generators on sparse random graphs*, preprint arXiv:1202.0644, 2012. 5
- [10] Charles Bordenave and Djalil Chafaï, *Around the circular law*, Probab. Surv. **9** (2012), 1–89. MR-2908617 2, 5, 7, 17
- [11] Djalil Chafaï, *Circular law for noncentral random matrices*, J. Theoret. Probab. **23** (2010), no. 4, 945–950. MR-2735731 5
- [12] Djalil Chafaï, Olivier Guédon, Guillaume Lécué, and Alain Pajor, *Interactions between compressed sensing, random matrices, and high dimensional geometry*, Panoramas et Synthèses 37, Société Mathématique de France, to appear, 2012. 7
- [13] Sourav Chatterjee, *A generalization of the Lindeberg principle*, Ann. Probab. **34** (2006), no. 6, 2061–2076. MR-2294976 2, 5, 6, 7, 8
- [14] Louis H. Y. Chen, Larry Goldstein, and Qi-Man Shao, *Normal approximation by Stein's method*, Probability and its Applications (New York), Springer, Heidelberg, 2011. MR-2732624 12
- [15] Zhishan Dong, Tiefeng Jiang, and Danning Li, *Circular law and arc law for truncation of random unitary matrix*, J. Math. Phys. **53** (2012), no. 1, 013301, 14. MR-2919538 2
- [16] Jean Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6** (1965), 440–449. MR-0173726 2
- [17] V. L. Girko, *The circle law*, Teor. Veroyatnost. i Mat. Statist. (1983), no. 28, 15–21. MR-727271 2
- [18] ———, *The strong circular law. Twenty years later. I*, Random Oper. Stochastic Equations **12** (2004), no. 1, 49–104. MR-2046403 2
- [19] Friedrich Götze and Alexander Tikhomirov, *The circular law for random matrices*, Ann. Probab. **38** (2010), no. 4, 1444–1491. MR-2663633 2
- [20] Tiefeng Jiang, *Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles*, Probab. Theory Related Fields **144** (2009), no. 1-2, 221–246. MR-2480790 2
- [21] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347 9
- [22] M. L. Mehta, *Random matrices and the statistical theory of energy levels*, Academic Press, New York, 1967. MR-0220494 2
- [23] H. H. Nguyen, *Random doubly stochastic matrices: the circular law*, preprint arXiv:1205.0843, 2012. 2
- [24] Hoi Nguyen and Sean O'Rourke, *On the concentration of random multilinear forms and the universality of random block matrices*, preprint arXiv:1309.4815, 2013. 2

- [25] Hoi Nguyen and Van Vu, *Circular law for random discrete matrices of given row sum*, preprint arXiv:1203.5941, 2012. 2
- [26] Guangming Pan and Wang Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. **101** (2010), no. 3, 645–656. MR-2575411 2
- [27] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. MR-2407948 2, 5, 10, 13
- [28] Michel Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR-1361756 9
- [29] ———, *A new look at independence*, Ann. Probab. **24** (1996), no. 1, 1–34. MR-1387624 9
- [30] Terence Tao and Van Vu, *The condition number of a randomly perturbed matrix*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 248–255. MR-2402448 5
- [31] ———, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. (N.S.) **46** (2009), no. 3, 377–396. MR-2507275 5
- [32] ———, *Random matrices: universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065, With an appendix by Manjunath Krishnapur. MR-2722794 2, 4, 17, 19
- [33] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327. MR-0095527 1

(RA) INSTITUTE OF MATHEMATICS, UNIVERSITY OF WARSAW, UL. BANACHA 2, 02-097 WARSZAWA, POLAND.

(DC) CEREMADE, UNIVERSITÉ PARIS-DAUPHINE, PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY, 75775 PARIS CEDEX 16, FRANCE; AND INSTITUT UNIVERSITAIRE DE FRANCE, FRANCE.

(PW) INSTITUTE OF MATHEMATICS, UNIVERSITY OF WARSAW, UL. BANACHA 2, 02-097 WARSZAWA, POLAND; AND INSTITUTE OF MATHEMATICS, POLISH ACADEMY OF SCIENCES, UL. ŚNIADECKICH 8, 00-956 WARSZAWA, POLAND.