On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries

Radosław Adamczak[*]

Institute of Mathematics
University of Warsaw
ul. Banacha 2
02-097 Warszawa
Poland

e-mail:R.Adamczak@mimuw.edu.pl

Abstract

In the first part of the article we prove limit theorems of Marchenko-Pastur type for the average spectral distribution of random matrices with dependent entries satisfying a weak law of large numbers, uniform bounds on moments and a martingale like condition investigated previously by Götze and Tikhomirov. Examples include log-concave unconditional distributions on the space of matrices.

In the second part we specialize to random matrices with independent isotropic unconditional log-concave rows for which (using the Tao–Vu replacement principle) we prove the circular law.

---

1 Introduction

One of the main points of interest of the theory of random matrices is universality, i.e. the question to what extent the limiting objects appearing in the theory are common for various random matrix models. Recently a lot of progress has been achieved in this direction including e.g. the proof of the circular law in a general form [36] as well as significant weakening of the assumptions leading to the Tracy-Widom distribution for the operator norm [33, 30] or the sine law for local eigenvalues statistics (see e.g. [15, 34]). Most of these results have been obtained for classical models of random matrices, i.e. models in which all the entries (or all the entries above the diagonal) of the matrix are independent random variables. Such models may be considered one of the cornerstones of the theory of random matrices (the other one being the family of invariant ensembles, which we are not going to discuss here).

There has been also some results concerning models in which the independence assumption is weakened or completely abandoned. Already in the seminal paper by Marchenko and Pastur [26] one considers rather independent rows than independent entries, compensating for the lack of independence with some moment assumptions. One of important examples considered in [26] is a random matrix with independent rows distributed uniformly on the unit sphere. This was generalized by Yin and Krishnaiah [37] to spherically symmetric distributions. Quite recently Aubrun [6] obtained the Marchenko-Pastur law for matrices with independent rows distributed uniformly on the ℓ_p^n ball, which was subsequently generalized by Pajor and Pastur [27] to matrices with independent rows distributed according to an arbitrary isotropic[1] log-concave measure.

Among other interesting results of Wigner and Marchenko-Pastur type there are those by Götze and Tikhomirov, dealing with matrices satisfying certain martingale like conditions, without any assumptions on the independence of the entries ([18, 20]). On the other hand Anderson and Zeitouni [4] constructed models of random matrices with locally dependent entries for which the limiting spectral distribution is different from that of classical random matrices. Similar results have been also obtained for structured random matrices (like Toeplitz or Hankel Matrices, see [12]).

As for the circular law, it is not our aim to list all the historical developments concerning this problem, so we will only mention that most of existing results concern random matrices with independent entries (see e.g. [16, 17, 8, 7, 28, 19, 36]) and to our best knowledge, the only article concerning nonindependent entries is [9] examining the case of random Markov matrices.

The aim of this paper is to provide some rather simple examples of random matrices with nonindependent entries for which classical limit theorems concerning the limiting spectral distribution still hold. We will focus on the Marchenko-Pastur and circular laws.

The amount of dependence we allow for varies from one result to another. Sometimes, when dealing only with the expected spectral distribution, we will not assume any independence, but only martingale like conditions in the spirit of Götze and Tikhomirov, a weak law of large numbers and sufficiently strong integrability. The proofs of these results follow the classical moment approach and are completely elementary.

For the almost sure limit theorems we will assume independence of rows or columns of the matrix (these types of results will be easy corollaries from the results for the expected spectral distribution and well known facts on the concentration of the Stieltjes transform).

---

[1]The term isotropic is used in this paper in the meaning of Definition 2.1, i.e. mean zero, covariance identity, which should not be confused with the meaning in [37], where this term denotes rotational invariance

Finally for the circular law we will assume that the rows are independent and distributed according to an unconditional isotropic log-concave measure (which is needed in the proof to ensure good concentration and bounds on the smallest singular value of the matrix). In this case again our results will follow easily from the results on the expected spectral distribution with use of the Tao–Vu replacement principle and recent general results on log-concave measures.

The organisation of the paper is as follows. In Section 2 we discuss limit theorems for symmetric nonnegative definite random matrices. The main results of this section are gathered in Subsection 2.2 (Theorems 2.3, 2.4).

In Section 3 we apply some of the results of Section 2 to prove the circular law for random matrices with independent log-concave unconditional rows (Theorem 3.4).

For simplicity we restrict our attention to matrices with real-valued entries (even for the circular law). However straightforward modifications of our arguments lead to counterparts of all the results in the complex case.

2 Symmetric matrices

2.1 Probabilistic assumptions

Let $(N_n)_{n\geq 1}$ be a sequence of positive integers such that $\lim_{n\to\infty} n/N_n = y \in (0,\infty)$ and for each $n \geq 1$ consider random matrices $A_n = [X_{ij}^{(n)}]_{1\leq i\leq n, 1\leq j\leq N_n}$. Let us assume that the following assumptions are satisfied

(A1) for every $k \in N$, $\sup_n \max_{i\leq n, j\leq N_n} \mathbb{E}|X_{ij}^{(n)}|^k < \infty$,

(A2) for every n, i, j, $\mathbb{E}(X_{ij}^{(n)}|\mathscr{F}_{ij}) = 0$, where $\mathscr{F}_{ij}$ is the σ-field generated by $\{X_{kl}^{(n)} : (k,l) \neq (i,j)\}$,

(A3) for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\leq n} \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{ij}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right) = 0,$$

and

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{j\leq N_n} \mathbb{P}\left(\left|\sum_{i=1}^{n}(X_{ij}^{(n)})^2 - n\right| \geq \varepsilon n\right) = 0.$$

Remark Note that the assumption (A3) may be read as 'the Euclidean norm of a random row of the matrix A_n, normalized by $\sqrt{N_n}$ and the Euclidean norm of a random column, normalized by $\sqrt{n}$ both converge in probability to 1'. It is obviously implied by the condition

(A3')

$$\lim_{n \to \infty} \max_{i \le n} \mathbb{P}\left(\left| \sum_{j=1}^{N_n} (X_{ij}^{(n)})^2 - N_n \right| \ge \varepsilon N_n \right) = 0,$$

and

$$\lim_{n \to \infty} \max_{j \le N_n} \mathbb{P}\left(\left| \sum_{i=1}^{n} (X_{ij}^{(n)})^2 - n \right| \ge \varepsilon n \right) = 0.$$

Examples and discussion of the assumptions   We would now like to list some examples of random matrices satisfying assumptions (A1)–(A3) and to relate these conditions to other assumptions considered in the literature.

Let us first recall the following definition.

Definition 2.1. *An n-dimensional random vector X is called isotropic if it is centered and its covariance matrix is equal to identity.*

1.   Obviously if the entries of the matrix are independent with mean zero, variance one and uniformly bounded moments of all orders, then the assumptions (A1)–(A3) are satisfied (the assumption (A3) follows e.g. from Chebyshev's inequality). This example is not of particular interest from our point of view as the convergence of spectral distributions of matrices generated by independent random variables is well known under much weaker integrability assumptions.

2.   If the law of A_n in $\mathbb{R}^{nN_n}$ (which we will identify with the space of $n \times N_n$ matrices in a natural way) is log-concave (see Definition 3.2 below) and isotropic (with respect to the standard basis), then the assumptions (A1), (A3) are satisfied. If one also assumes unconditionality (see Definition 3.1) with respect to the standard basis, one obtains assumption (A2), although obviously (A2) is weaker than unconditionality.

Assumption (A1) follows from the so called Borell's lemma (see [11]), which states that log-concave variables are exponentially integrable, Assumption (A3) (or even the stronger condition (A3')) from Klartag's concentration results (see Theorem 3.5 in Section 3.2) and the fact that marginals of isotropic log-concave measures are also isotropic and log-concave.

In particular this class of examples includes matrices sampled from the ℓ_p balls in $\mathbb{R}^{N_n n}$ in isotropic position, i.e. from the sets $B_p = \{M = (x_{ij})_{i \le n, j \le N_n} : (\sum_{ij} |x_{ij}|^p)^{1/p} \le c_{p,N_n,n}\}$. Here $c_{p,k,n}$ are constants of order $(kn)^{1/p}$. Another class of matrices covered by this case is matrices with independent log-concave unconditional rows.

We would like to note that the Marchenko-Pastur theorem has been recently proven by Pajor and Pastur in the case of general matrices with independent log-concave isotropic rows or columns (not necessarily unconditional). With our approach we will be able to show that the expected spectral measure of $\frac{1}{N_n} A_n A_n^T$ converges to the Marchenko-Pastur distribution even if the rows are not independent (however we additionally have to assume (A2)).

3. The above examples of vectors with independent entries and with independent log-concave rows/columns have one common feature. Namely, the convergence of spectral distribution of $\frac{1}{N_n}A_nA_n^T$ can be proven by means of the Stieltjes transform and the bound of the form

$$\mathrm{Var}\left(\langle MX, X\rangle\right) \leq o(n^2), \tag{1}$$

where X is any row of A_n and M is any matrix with $\|M\|_{\ell_2 \to \ell_2} \leq 1$. In the case of independent entries this bound is straightforward, for log-concave rows it has been proven by Pajor and Pastur in [27]. It lies also at the core of the original proof by Marchenko and Pastur. The importance of concentration of measure for quadratic forms for the limit theorems for spectra of random matrices of sample covariance type has been also recently emphasized by El Karoui [14].

Here we would like to present a simple class of random matrices with independent rows, which satisfy assumptions (A1)–(A3) but do not satisfy the above bound.

Choose $k \in \mathbb{N}$ and assume for simplicity that for all n, N_n is divisible by k. Divide the set $\{1, \ldots, N_n\}$ into k disjoint sets of equal cardinality (say $I_1, \ldots, I_k$). Let μ_n, $n = 1, 2, \ldots$ be isotropic probability measures on $\mathbb{R}^{N_n/k}$, satisfying (A1) and (A2), with Euclidean norm concentrated around $\sqrt{N_n/k}$ (i.e. $\mu_n(\{x : \|x\| - \sqrt{N_n/k}\| > \varepsilon\}) \to 0$ for all $\varepsilon > 0$). Let us also identify $\mathbb{R}^{N_n}$ with $\mathbb{R}^{I_1} \times \ldots \times \mathbb{R}^{I_k}$. Let δ_n be a random variable distributed uniformly on $\{1, \ldots, k\}$ and Y_n a random vector independent of δ_n, distributed according to μ_n. Finally let $X^{(n)} = \sqrt{k}(X\mathbf{1}_{\{\delta_n=1\}}, X\mathbf{1}_{\{\delta_n=2\}}, \ldots, X\mathbf{1}_{\{\delta_n=k\}})$. In other words we select the set of nonzero coordinates of $X^{(n)}$ uniformly among $I_1, \ldots, I_k$, distribute the nonzero coordinates according to μ_n and finally rescale the vector to make it isotropic in $\mathbb{R}^{N_n}$. Now define the rows of A_n, $X_1^{(n)}, \ldots, X_n^{(n)}$ as independent copies of $X^{(n)}$.

By construction, the matrix A_n with rows $X_1^{(n)}, \ldots, X_n^{(n)}$ satisfies (A1)–(A3) (the condition (A1) follows from the fact that k is fixed). However if we let M be the matrix of the orthogonal projection on (say) I_1, we get

$$\mathrm{Var}\left(\langle MX^{(n)}, X^{(n)}\rangle\right) = \mathbb{E}\left(\sum_{i \in I_1}(X_i^{(n)})^2\right)^2 - \left(\mathbb{E}\sum_{i \in I_1}(X_i^{(n)})^2\right)^2$$

$$\simeq \frac{1}{k}(k(N_n/k))^2 - \left(\frac{1}{k}kN_n/k\right)^2 = N_n^2/k - N_n^2/k^2,$$

which is of the order n^2. Thus in this case (1) is not satisfied.

Let us also mention that matrices with non-necessarily independent entries, satisfying assumption (A2) have recently been considered by Götze and Tikhomirov (see [18, 20]). Their results, concerning convergence of the expected spectral measure were obtained with use of the Stein method and work under the assumption of finiteness of the second moments of the entries, which is much milder than our assumption (A1). Nevertheless, the other assumptions introduced in their paper are much more technical, martingale like conditions (although still rather natural in view of the CLT theory for martingales). In particular it may be checked that the example constructed above does not satisfy the assumptions (1.11) and (1.12) of Theorem 1.1. in [20]. It is also relatively easy to construct examples of matrices which satisfy (A1)-(A3) but fail the assumption (1.10) of [20]. From this point of view, some of the results we present may be seen as compliments of the theorem by Götze and Tikhomirov. It would be interesting to find a theorem covering both results.

4. Obviously assumptions of the type (A3'), which are just a quantitative version of the weak law of large numbers, have been studied in the literature. It is classical and goes back to Khintchine that it is enough to assume that $(X_{ij}^{(n)})^2$ are negatively correlated and have uniformly bounded second moment (which would follow from (A1)), one can also consider matrices satisfying all sorts of mixing conditions (see e.g. [22, 32]). Let us finally notice that if we have random matrices $A_n = [X_{ij}]$ satisfying (A1) and (A3), then the matrices $(A_{ij}^{(n)} \varepsilon_{ij})$, where ε_{ij} are independent Rademacher variables, independent of A_n will satisfy (A1)–(A3). If many cases, under stronger regularity of A_n, we may multiply $X_{ij}^{(n)}$ by more general sequences of independent mean zero variables. It is not our aim to go into details here, but rather to argue that assumptions (A1)–(A3) are satisfied by many models when independence is replaced by weaker conditions of partial independence, so we will leave the details to the Reader.

2.2 Main results for the symmetric case

Marchenko-Pastur Law   Recall that the spectral measure of an $n \times n$ symmetric matrix H is a probability measure on the real line defined as

$$L_H = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i},$$

where $\lambda_1 \leq \ldots \leq \lambda_n$ are eigenvalues of A and δ_x stands for the Dirac mass at x.

Recall also the following

Definition 2.2. *The Marchenko-Pastur distribution with parameter $y > 0$ is the probability measure on $[0, \infty)$, given by the density*

$$p_y(x) = \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)} \mathbf{1}_{[a,b]}(x)$$

with an additional point mass of $1 - 1/y$ at 0 for $y > 1$, where $a = (1 - \sqrt{y})^2$, $b = (1 + \sqrt{y})^2$.

Theorem 2.3. *Let A_n be a sequence of random matrices satisfying the assumptions (A1)–(A3). Assume also that $\lim_{n \to \infty} n/N_n = y \in (0, \infty)$. Let $M_n = \frac{1}{N_n} A_n A_n^T$. Then for any $k \in \mathbb{N}$,*

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E} \mathrm{tr} M_n^k = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} y^r. \tag{2}$$

In consequence, if L_n is the spectral measure of the matrix M_n, then the (non-random) measure $\mathbb{E} L_n$ converges weakly as $n \to \infty$ to the Marchenko-Pastur law with parameter y.

Remark   If one additionally assumes that the rows (or columns) of the matrix are independent, by concentration properties of the Stieltjes transform of the spectral distribution (which for the Reader's convenience we formulate in the Appendix) one can strengthen the above results to an almost sure convergence of L_n. In the sequel we will not need the stronger version, however we will need a corresponding statement for a related model of random matrices described below, for which we will present a detailed proof.

Square matrices shifted by a multiple of identity We will now consider the case $N_n = n$ and the empirical spectral distribution of the matrix

$$H_n = H_n(z) = \left(\frac{1}{\sqrt{n}}A_n - z\mathrm{Id}\right)\left(\frac{1}{\sqrt{n}}A_n - z\mathrm{Id}\right)^*,$$

where z is a complex number (note that for $z = 0$ the problem of spectral distribution of H_n reduces to the Marchenko-Pastur theorem).

The interest in this type of random matrices stems from the fact that they play an important role in the proof of the circular law. In the second part of the article we will use the facts established in this section to prove the circular law for random matrices with independent log-concave unconditional rows.

Theorem 2.4. *Assume that $N_n = n$ and a sequence of random matrices A_n satisfies assumptions (A1) – (A3). Then for any $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\mathrm{tr}H_n^k = \mu_k(|z|^2),$$

where $\mu_k(|z|^2)$ is a function depending only on $|z|^2$ and not on the distribution of H_n.

Corollary 2.5. *Assume that $N_n = n$, and A_n is a sequence of random matrices with independent rows satisfying assumptions (A1) – (A3). Let $L_n(z)$ be the spectral measure of the matrix $H_n(z)$. For every $z \in \mathbb{C}$, with probability one, $L_n(z)$ converges weakly to a non-random measure which does not depend on the distribution of the rows of A_n (in particular is the same as for the case of matrices with independent standard Gaussian entries).*

Remark Note that (as already mentioned) one half of the assumption (A3) follows from the independence of the rows and boundedness of moments of the entries.

2.3 Combinatorial facts

2.3.1 Trees

Let $T = (V, E, r)$ be a rooted tree. Divide the set V into two disjoint classes U and D, where U is the set of vertices whose distance from the root is even, while D is the set of vertices whose distance from the root is odd. Thus each edge $e \in E$ joins a vertex from D (call it v) with a vertex from U (call it u). We will denote such an edge by $e = (u \to v)$.

Consider the set I_T^n of functions $\mathbf{i} = (i_v)_{v\in V} : V \to \{1, \ldots \max(n, N_n)\}$ such that

- if A is one of the sets D, U then

$$\forall_{u,v\in A} \, u \neq v \Rightarrow i_u \neq i_v,$$

- if $v \in D$ then $i_v \in \{1, \ldots, N_n\}$,

- if $v \in U$ then $i_v \in \{1, \ldots, n\}$.

Finally define

$$\zeta_n(T) = n^{-|U|}N_n^{-|D|}\mathbb{E}\sum_{\mathbf{i}\in I_T^n}\prod_{e=(u\to v)\in E}(X_{i_u i_v}^{(n)})^2.$$

We will prove the following easy proposition.

Proposition 2.6. *Assume that* $\lim_{n\to\infty} n/N_n = y \in (0,\infty)$. *Then for every rooted tree* $T = (V, E, r)$ *and every sequence of random matrices satisfying (A1) and (A3), one has*

$$\lim_{n\to\infty}\zeta_n(T) = 1.$$

Proof. In the course of the proof we will allow all the constants to depend on the constants in assumption (A1) and y without stating it explicitly. Thus if we write e.g. $O_a(1)$, we mean that the implicit constant depends only on y, the constants in (A1) and the parameter a.

We will proceed by induction with respect to the size of the tree. If $|V| = 1$ then clearly $\zeta_n(T) = 1$ for all n. Let us thus assume that $|V| > 1$ and that the proposition holds for all trees of size smaller than $|V|$. Let us consider an arbitrary leaf w of the tree T, distinct from the root. Let x be the unique neighbour of w. We will consider in detail only the case when $w \in D$, the other case is similar. Let $\tilde{T} = (\tilde{V}, \tilde{E}, r)$ be the tree obtained from T by deleting the vertex w together with the adjacent edge $e = (x \to w)$. Let $\tilde{I}^n$ be the set of multi-indices $\mathbf{i}_{\tilde{V}} : \tilde{V} \to \{1, \ldots, \max(n, N_n)\}$ which can be extended to a multi-index $\mathbf{i}_V = (\mathbf{i}_{\tilde{V}}, i_w) \in I_T^n$. Denote also $U(\mathbf{i}_{\tilde{V}}) = \prod_{e=(u\to v)\in\tilde{E}}(X_{i_u i_v}^{(n)})^2$.
We have

$$\zeta_n(T) = n^{-|U|}N_n^{-|D|}\sum_{\mathbf{i}_{\tilde{V}}\in\tilde{I}^n}\sum_{\substack{i_w:\\(\mathbf{i}_{\tilde{V}},i_w)\in I_T^n}}\mathbb{E}((X_{i_x i_w}^{(n)})^2 U(\mathbf{i}_{\tilde{V}})).$$

For n large enough (depending on T) $I_{\tilde{T}}^n = \tilde{I}^n$, and there are only $|D| - 1$ choices of i_w such that $(\mathbf{i}_{\tilde{V}}, i_w) \notin I_T^n$. Moreover, by the generalized Hölder inequality and the assumption (A1), for every such i_w, $\mathbb{E}((X_{i_x i_w}^{(n)})^2 U(\mathbf{i}_{\tilde{V}}))$ is bounded by a number independent of n. Therefore for large n,

$$\zeta_n(T) = n^{-|U|}N_n^{-|D|}\sum_{\mathbf{i}_{\tilde{V}}\in I_{\tilde{T}}^n}\left(\sum_{i_w=1}^{N_n}\mathbb{E}\left((X_{i_x i_w}^{(n)})^2 U(\mathbf{i}_{\tilde{V}})\right) + O_T(1)\right)$$

$$= n^{-|U|}N_n^{-|D|}\sum_{\mathbf{i}_{\tilde{V}}\in I_{\tilde{T}}^n}\sum_{i_w=1}^{N_n}\mathbb{E}\left((X_{i_x i_w}^{(n)})^2 U(\mathbf{i}_{\tilde{V}})\right) + o_T(1), \tag{3}$$

where in the last inequality we used the fact that $|I_{\tilde{T}}^n| = n(n-1)\cdots(n-|U|+1)N_n\cdots(N_n-|D|+2) = O_T(n^{|U|+|D|-1})$.
Now

$$\left|n^{-|U|}N_n^{-|D|}\sum_{\mathbf{i}_{\tilde{V}}\in I_{\tilde{T}}^n}\sum_{i_w=1}^{N_n}\mathbb{E}\left((X_{i_x i_w}^{(n)})^2 U(\mathbf{i}_{\tilde{V}})\right) - \zeta_n(\tilde{T})\right|$$

$$\le n^{-|U|}N_n^{-|D|+1}\sum_{\mathbf{i}_{\tilde{V}}\in I_{\tilde{T}}^n}\mathbb{E}\left(|N_n^{-1}\sum_{i_w=1}^{N_n}(X_{i_x i_w}^{(n)})^2 - 1|U(\mathbf{i}_{\tilde{V}})\right). \tag{4}$$

For every $\varepsilon > 0$ and every $\mathbf{i}_{\tilde{V}} \in I_{\tilde{T}}^n$ we have

$$\mathbb{E}\left(|N_n^{-1}\sum_{i_w=1}^{N_n}(X_{i_x i_w}^{(n)})^2 - 1|U(\mathbf{i}_{\tilde{V}})\right)$$

$$\leq \varepsilon\mathbb{E}U(\mathbf{i}_{\tilde{V}}) + \left\||N_n^{-1}\sum_{i_w=1}^{N_n}(X_{i_x i_w}^{(n)})^2 - 1|U(\mathbf{i}_{\tilde{V}})\right\|_2 \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{i_x j}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right)^{1/2}.$$

By assumption (A1), the triangle inequality in L_p and generalized Hölder's inequality

$$\mathbb{E}U(\mathbf{i}_{\tilde{V}}) \quad \text{and} \quad \left\||N_n^{-1}\sum_{i_w=1}^{N_n}(X_{i_x i_w}^{(n)})^2 - 1|U(\mathbf{i}_{\tilde{V}})\right\|_2$$

are bounded by a constant C_T, depending only on T and the constants in (A1). In consequence, by (3) and (4), we get that for large n,

$$|\zeta_n(T) - \zeta_n(\tilde{T})| \leq \varepsilon + C_T n^{-|U|} N_n^{-|D|+1} \sum_{\mathbf{i}_{\tilde{V}} \in I_{\tilde{T}}^n} \left(\varepsilon + \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{i_x j}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right)^{1/2}\right).$$

Now, by the formula for cardinality of $I_{\tilde{T}}^n$ and the fact that for each i_x there are at most $n^{|U|-1}N_n^{|D|-1}$ multi-indices $\mathbf{i}_{\tilde{V}\setminus\{x\}}$, such that $\mathbf{i}_{\tilde{V}} = (\mathbf{i}_{\tilde{V}\setminus\{x\}}, i_x) \in I_{\tilde{T}}^n$, we have

$$|\zeta_n(T) - \zeta_n(\tilde{T})| \leq (C_T + 1)\varepsilon + C_T \frac{1}{n}\sum_{i=1}^n \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{ij}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right)^{1/2}.$$

Note that by the Cauchy-Schwarz inequality,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{ij}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right)^{1/2} \leq \left(\frac{1}{n}\sum_{i=1}^n \mathbb{P}\left(\left|\sum_{j=1}^{N_n}(X_{ij}^{(n)})^2 - N_n\right| \geq \varepsilon N_n\right)\right)^{1/2}$$

and thus by (A3) we get

$$|\zeta_n(T) - \zeta_n(\tilde{T})| = o_T(1),$$

which in combination with the induction assumption proves that $\zeta_n(T) \to 1$ as $n \to \infty$.

The other case to consider, when $w \in U$, is analogous, one simply uses the other part of assumption (A3) to show that $\zeta_n(T)$ and $\zeta_n(\tilde{T})$ are close. $\qquad\square$

2.3.2 Γ-trees

For the purpose of proving Theorem 2.4 it will be convenient to consider a special class of trees, which we introduce below in an abstract form together with a theorem on the asymptotic behaviour of counterparts of quantities ζ_n investigated in the previous section. The proofs will be very similar to those for ζ_n, the main difficulty being a slightly more involved notation.

Definition We define a Γ-tree as a rooted tree $T = (V, E, r)$ possessing the following additional structure

1. The set V is partitioned into two sets S and O. The elements of S (resp. O) will be called special (resp. ordinary) vertices

2. Each edge adjacent to a special vertex is given an orientation in such a way that

   - For any special vertices u, w such that on the path $u = v_0 v_1 \ldots v_m = w$ joining them there are no special vertices besides u and w,
     - if m is odd then the orientations of the first and the last edge on this path are the same (i.e. we have ($u \to v_1$ and $v_{m-1} \to w$) or ($v_1 \to u$ and $w \to v_{m-1}$)),
     - if m is even then the orientation of the first and the last edge on this path are opposite,
   - if $r \in O$ then for any special vertex u such that u is the only special vertex on the path $r = v_0 v_1 \ldots v_m = u$ from r to u, we have $v_{m-1} \to u$ iff m is odd.

Partition The orientation of paths between elements of S allows us to further partition O into two sets U and D in the following way.

Let $u \in O$ and let $r = v_0 v_1 \ldots v_m = u$ be the path from the root to u.

- If $r \in O$ and $v_1, \ldots, v_{m-1} \in O$, then $u \in D$ iff m is odd (otherwise $u \in U$),

- if v_l is the last special vertex on the path then $u \in D$ iff ($m - l$ is odd and $v_l \to v_{l+1}$) or ($m - l$ is even and $v_{l+1} \to v_l$).

Notice that every edge e with both ends in O has one end v in D and the other end u in U. We will assign to such edges the orientation $u \to v$. This way we have given orientation to all the edges of T. Whenever we want to stress the orientation of the edge with ends u, v we will write $e = (u \to v)$.

Definition of $\xi_n(T)$ For a fixed sequence of random $n \times n$ matrices $A_n = [X_{ij}^{(n)}]_{i,j \leq n}$ and any Γ-tree T we define a number $\xi_n(T)$ in the following way.

Consider the set I_T^n of functions $\mathbf{i} = (i_v)_{v \in V} : V \to \{1, \ldots n\}$ such that if A is one of the sets $D \cup S$, $U \cup S$ then

$$\forall_{u,v \in A} \, u \neq v \Rightarrow i_u \neq i_v$$

and for every $e = (u \to v) \in E$, $i_u \neq i_v$.

Set

$$\xi_n(T) = n^{-|E|-1} \mathbb{E} \sum_{\mathbf{i} \in I_T^n} \prod_{e=(u \to v) \in E} (X_{i_u i_v}^{(n)})^2.$$

Example Consider a tree T on four vertices $1, 2, 3, 4$, in which 1 is the root, $S = \{1, 2\}$, $O = \{3, 4\}$ and the edges are $2 \to 1$, $4 \to 2$, $2 \to 3$. Then $D = \{3\}$, $U = \{4\}$ and

$$\xi_n(T) = \frac{1}{n^4} \sum_{\substack{1 \le i_1, i_2, i_3, i_4 \le n \\ i_1 \ne i_2, i_1 \ne i_3, i_1 \ne i_4 \\ i_2 \ne i_4, i_2 \ne i_3}} \mathbb{E}(X_{i_2 i_1}^{(n)})^2 (X_{i_4 i_2}^{(n)})^2 (X_{i_2 i_3}^{(n)})^2.$$

Proposition 2.7. *Assume that $N_n = n$ and the sequence of random matrices A_n satisfies the assumptions (A1) and (A3). Then for every Γ-tree T,*

$$\lim_{n \to \infty} \xi_n(T) = 1.$$

Proof. Since the argument is very similar to the proof of Proposition 2.6, we will present only a sketch. We proceed by induction with respect to the number of vertices in T. For $V = 1$, we have $\xi_n(T) = 1$ for all n. If $|V| > 1$, we consider an arbitrary leaf v of T different from the root. Let us notice that by removing this vertex together with the adjacent edge, we obtain a new tree $\tilde{T}$ endowed with a structure of a Γ-tree, inherited from T. By arguments very similar to those given for Proposition 2.6 one can show that $\gamma_n(T) = \gamma_n(\tilde{T}) + o_T(1)$, which ends the induction step (again one has to consider two cases depending on the orientation of the edge adjacent to v and for each of them use one of the assertions given in assumption (A3)). $\qquad \square$

2.4 Proof of Theorem 2.3

Since the Marchenko-Pastur distribution is determined by its moments, which are given by the right hand side of (2), it is enough to prove the first part of the theorem. A major part of the proof will follow the classical approach for matrices with independent entries, which will be complemented by Proposition 2.6.

Definition of Δ graphs We will work with the class of Δ graphs, following the definition given in [7] but slightly changing the formalism to better suit our needs.

1. For two sequences $\mathbf{i} = (i_1, \ldots, i_k)$ and $\mathbf{j} = (j_1, \ldots, j_k)$ of integers (not necessarily distinct) we define a Δ-graph $G(\mathbf{i}, \mathbf{j})$ as a bipartite graph $(I_\mathbf{i}, I_\mathbf{j}, E)$, such that $I_\mathbf{i} = \{i_1, \ldots, i_k\}$ (*upper indices*), $I_\mathbf{j} = \{j_1, \ldots, j_k\}$ (*lower indices*) and the set E of edges consists of k directed edges from i_u to j_u ($u = 1, \ldots, k$) and k directed edges from j_u to i_{u+1} ($u = 1, \ldots k$), where we set $i_{k+1} = i_1$. We will also label the edges from 1 to $2k$ in the order $(i_1, j_1), (j_1, i_2), (i_2, j_2), \ldots, (i_k, j_k), (j_k, i_1)$ (which clearly allows for the reconstruction of the indices $\mathbf{i}, \mathbf{j}$ from the graph $G(\mathbf{i}, \mathbf{j})$). We stress that $I_\mathbf{i}$ and $I_\mathbf{j}$ may not be disjoint, but their common elements will be nevertheless treated as different objects when considered as upper and lower vertices of the graph.

2. An edge (i_u, j_u) will be called a down-edge, an edge (j_u, i_{u+1}) an up-edge.

Let us also define for any edge e in $G(\mathbf{i}, \mathbf{j})$, $u(e)$ and $d(e)$ to be the upper and lower vertices of e (more formally, if $e = (v, w)$ and e is labeled by an odd number then $u(e) = v, d(e) = w$, whereas if e is labeled by an even number then $u(e) = w, d(e) = v$).

3. Following the classical approach we will now introduce an equivalence relation on the pairs of indices $(\mathbf{i}, \mathbf{j})$.

We will say that two pairs $(\mathbf{i}, \mathbf{j})$ and $(\mathbf{i}', \mathbf{j}')$ are isomorphic if there exist injective functions f, g from $I_{\mathbf{i}}, I_{\mathbf{j}}$ onto $I_{\mathbf{i}'}, I_{\mathbf{j}'}$ respectively such that for $u = 1, \ldots, k$,

$$f(i_u) = i_u', \; g(j_u) = j_u'.$$

4. We will call graphs $G(\mathbf{i}, \mathbf{j})$ and $G(\mathbf{i}', \mathbf{j}')$ isomorphic (which we will denote by $G(\mathbf{i}, \mathbf{j}) \sim G(\mathbf{i}', \mathbf{j}')$) iff $(\mathbf{i}, \mathbf{j})$ and $(\mathbf{i}', \mathbf{j}')$ are isomorphic.

5. Let $\Delta(k)$ be the set of representatives for the isomorphism classes of Δ graphs $G(\mathbf{i}, \mathbf{j})$ such that $\mathbf{i} = (i_1, \ldots, i_k)$, $\mathbf{j} = (j_1, \ldots, j_k)$ and $i_l \in \{1, \ldots, k\}$, $j_l \in \{k+1, \ldots, 2k\}$. Note that any graph based on two sequences of length k is isomorphic to a graph in $\Delta(k)$ (it is enough to properly relabel the vertices).

We will also define for $\Delta \in \Delta(k)$, I_Δ^n to be the set of all indices $\mathbf{i} \colon V(\Delta) \to \{1, \ldots, \max(n, N_n)\}$ (where $V(\Delta)$ is the set of all vertices of Δ)), such that

- for any two upper vertices v, w, $i_v \neq i_w$,

- for any two lower vertices v, w, $i_v \neq i_w$,

- for any upper vertex v, $i_v \in \{1, \ldots, n\}$,

- for any lower vertex v, $i_v \in \{1, \ldots, N_n\}$.

With the above notation we can write

$$
\frac{1}{n} \mathbb{E} \mathrm{tr} M_n^k = \frac{1}{nN_n^k} \sum_{i_1, \ldots, i_k = 1}^{n} \sum_{j_1, \ldots, j_k = 1}^{N_n} \mathbb{E} X_{i_1 j_1}^{(n)} X_{i_2 j_1}^{(n)} X_{i_2 j_2}^{(n)} X_{i_3 j_2}^{(n)} \ldots X_{i_k j_k}^{(n)} X_{i_1 j_k}^{(n)}
$$

$$
= \frac{1}{nN_n^k} \sum_{\Delta \in \Delta(k)} \sum_{\substack{\mathbf{i}, \mathbf{j} \in \{1, \ldots, \max(n, N_n)\}^k : \\ G(\mathbf{i}, \mathbf{j}) \sim \Delta}} \mathbb{E} X_{i_1 j_1}^{(n)} X_{i_2 j_1}^{(n)} X_{i_2 j_2}^{(n)} X_{i_3 j_2}^{(n)} \ldots X_{i_k j_k}^{(n)} X_{i_1 j_k}^{(n)}
$$

$$
= \sum_{\Delta \in \Delta(k)} \frac{1}{nN_n^k} \sum_{\mathbf{i} \in I_\Delta^n} \mathbb{E} \prod_{e \in E(\Delta)} X_{i_{u(e)} i_{d(e)}}^{(n)},
$$

where $E(\Delta)$ is the set of edges of Δ.

Now, still following [7], we can divide $\Delta(k)$ into three classes $\Delta_i(k)$, $i = 1, 2, 3$, where

- $\Delta_1(k)$ is the class of graphs, in which to each down edge there corresponds exactly one up edge with the same vertices and after merging the corresponding up and down edges and disregarding the orientation, one obtains a tree with $k+1$ vertices,

- $\Delta_2(k)$ is the class of graphs in which after disregarding the orientation of edges there is an edge of multiplicity one,

- $\Delta_3(k) = \Delta(k) \setminus (\Delta_1(k) \cup \Delta_2(k))$.

Note that by assumption (A2) for $\Delta \in \Delta_2(k)$ and every $\mathbf{i} \in I_\Delta^n$,

$$\mathbb{E} \prod_{e \in E(\Delta)} X_{i_{u(e)} i_{d(e)}}^{(n)} = 0.$$

Moreover, using connectedness, it is easy to see that each $\Delta \in \Delta_3(k)$ has at most k vertices and therefore $|I_\Delta^n| = O(n^k)$. Since by Hölder's inequality $\mathbb{E} \prod_{e \in E(\Delta)} X_{i_{u(e)} i_{d(e)}}^{(n)}$ is bounded by a constant depending only on k and the constants in (A1), this implies that graphs from $\Delta_3(k)$ have no asymptotic contribution to $\frac{1}{n} \mathbb{E} \mathrm{tr} M_n^k$.

Thus, just as in the classical case of independent entries, we are left with the analysis of the contribution coming from $\Delta_1(k)$. This is where we will apply Proposition 2.6.

For $r = 0, \ldots, k-1$ let $\Delta_1(k,r)$ be the class of those graphs in $\Delta_1(k)$ which have exactly $r+1$ upper vertices (which implies that there are $k - r$ lower vertices). It is well known (see e.g. Lemma 3.3 in [7]) that $|\Delta_1(k,r)| = \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r}$. For each $\Delta \in \Delta_1(k,r)$ let $T(\Delta)$ be the rooted tree obtained from Δ in a way described when introducing the class $\Delta_1(k)$ (we choose the vertex i_1 to be the root). We have

$$
\begin{aligned}
\frac{1}{n} \mathbb{E} \mathrm{tr} M_n^k &= \sum_{r=0}^{k-1} \sum_{\Delta \in \Delta_1(r,k)} \frac{1}{nN_n^k} \sum_{\mathbf{i} \in I_\Delta^n} \mathbb{E} \prod_{e \in E(\Delta)} X_{i_{u(e)} i_{d(e)}}^{(n)} + o(1) \\
&= \sum_{r=0}^{k-1} \sum_{\Delta \in \Delta_1(r,k)} \left(\frac{n}{N_n}\right)^r \frac{1}{n^{r+1} N_n^{k-r}} \sum_{\mathbf{i} \in I_\Delta^n} \mathbb{E} \prod_{e \in E(\Delta)} X_{i_{u(e)} i_{d(e)}}^{(n)} + o(1) \\
&= \sum_{r=0}^{k-1} \sum_{\Delta \in \Delta_1(r,k)} \left(\frac{n}{N_n}\right)^r \zeta_n(T(\Delta)) = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} y^r + o(1),
\end{aligned}
$$

where in the last line we used Proposition 2.6. This ends the proof of Theorem 2.3.

2.5 Proofs of Theorem 2.4 and Corollary 2.5

Theorem 2.4 will also be proved with use of Δ-graphs. Let us remark that the proof of the corresponding theorem for matrices with independent entries given in [16] or [7], as well as its generalization in [13], use Stieltjes transform. However in [7] the authors mention that a combinatorial proof is possible and leave the details to the Reader as an exercise. Many of the formalities introduced below may be seen as a solution to this (rather involved) exercise and one of many possible ways to formalize the underlying combinatorics. We are not aware of any description of this particular problem in the literature (for specific models of random matrices the result may follow from general facts in free probability), but obviously in view of well known proofs of Wigner or Marchenko-Pastur theorems the methodology is rather standard and we do not claim any novelty here. Our main contribution is the observation given in Proposition 2.7 and its implications for the proof of Theorem 2.4 available beyond the case of independent random variables.

The combinatorial construction below will be introduced in full generality, however to illustrate it we provide two concrete examples which, while being relatively simple, capture the essential part of the argument. So as not to obscure the general idea we present these examples after the proof of Theorem 2.4.

1. For the proof of Theorem 2.4 let us again consider the bipartite graphs $G(\mathbf{i}, \mathbf{j})$ as introduced in point 1 of Section 2.4. Although the basic definition of $G(\mathbf{i}, \mathbf{j})$ remains the same, because of the shift of the random matrices by $z\mathrm{Id}$ we are forced to consider different combinatorial structure on the family of graphs, in particular to distinguish a special class of perpendicular edges (as in [7], Chapter 10) and to change the notion of isomorphism (in a way which will preserve perpendicular edges).

2. Additionally to the partition of edges into the classes of up and down edges we will call an edge perpendicular if its two end-vertices are equal (i.e. have equal labels) or skew if they are distinct. We will denote the set of perpendicular up (resp. down) edges by $UP(\Delta)$ (resp. $DP(\Delta)$) and the set of skew edges by $S(\Delta)$.

3. The pairs $(\mathbf{i}, \mathbf{j})$ and $(\mathbf{i}', \mathbf{j}')$ are said to be isomorphic if there exist injective functions f, g from $I_\mathbf{i}, I_\mathbf{j}$ onto $I_{\mathbf{i}'}, I_{\mathbf{j}'}$ respectively such that for $u = 1, \ldots, k$,

- $f(i_u) = i'_u$, $g(j_u) = j'_u$,
- $f(i_u) = g(j_u)$ iff $i_u = j_u$,
- $f(i_{u+1}) = g(j_u)$ iff $i_{u+1} = j_u$.

3. We will call two graphs $G(\mathbf{i}, \mathbf{j})$ and $G(\mathbf{i}', \mathbf{j}')$ isomorphic iff the pairs $(\mathbf{i}, \mathbf{j})$, $(\mathbf{i}', \mathbf{j}')$ are isomorphic in the above sense.

4. Let $\Delta(k)$ be the set of representatives for the isomorphism classes of Δ graphs $G(\mathbf{i}, \mathbf{j})$ where $\mathbf{i} = (i_1, \ldots, i_k)$, $\mathbf{j} = (j_1, \ldots, j_k)$ and $i_l, j_l \in \{1, \ldots, 2k\}$. Similarly as in the previous section any graph based on two sequences of length k is isomorphic to a graph in $\Delta(k)$.

We will also define for $\Delta \in \Delta(k)$, I_Δ^n to be the set of all indices $\mathbf{i} \colon V(\Delta) \to \{1, \ldots, n\}$ such that

- for any two upper indices v, w, $i_v \neq i_w$,
- for any two lower indices v, w, $i_v \neq i_w$,
- for any edge $i_{u(e)} = i_{d(e)}$ iff e is perpendicular.

Finally let us denote $W_n = n^{-1/2} A_n - z\mathrm{Id} = (w_{ij})_{i,j \leq n}$. Note that to simplify the notation we have suppressed the superscript (n) denoting the dependence of the random coefficients on n. In what follows we will keep the same convention and write X_{ij} instead of $X_{ij}^{(n)}$. We have

$$n^{-1}\mathbb{E}\mathrm{tr}H_n^k = n^{-1}\sum_{i_1,\dots,i_k=1}^{n}\sum_{j_1,\dots,j_k=1}^{n}\mathbb{E}w_{i_1j_1}\bar{w}_{i_2j_1}w_{i_2j_2}\bar{w}_{i_3j_2}\cdots w_{i_kj_k}\bar{w}_{i_1j_k}$$

$$= n^{-1}\sum_{\Delta\in\Delta(k)}\sum_{\substack{\mathbf{i},\mathbf{j}\in\{1,\dots,n\}^k:\\ G(\mathbf{i},\mathbf{j})\sim\Delta}}\mathbb{E}w_{i_1j_1}\bar{w}_{i_2j_1}w_{i_2j_2}\bar{w}_{i_3j_2}\cdots w_{i_kj_k}\bar{w}_{i_1j_k}$$

$$= \sum_{\Delta\in\Delta(k)}n^{-1}\sum_{\mathbf{i}\in I_\Delta^n}\mathbb{E}\Big(\prod_{e\in S(\Delta)}w_{i_{u(e)}i_{d(e)}}\prod_{e\in UP(\Delta)}\bar{w}_{i_{u(e)}i_{d(e)}}\prod_{e\in DP(\Delta)}w_{i_{u(e)}i_{d(e)}}\Big)$$

$$= \sum_{\Delta\in\Delta(k)}n^{-1-|S(\Delta)|/2}\sum_{\mathbf{i}\in I_\Delta^n}\mathbb{E}\Big(\prod_{e\in S(\Delta)}X_{i_{u(e)}i_{d(e)}}\prod_{e\in UP(\Delta)}\bar{w}_{i_{u(e)}i_{d(e)}}\prod_{e\in DP(\Delta)}w_{i_{u(e)}i_{d(e)}}\Big).$$

For a fixed Δ let Δ' be a graph obtained from Δ by replacing each pair of vertices connected with a perpendicular edge by one vertex and removing corresponding perpendicular edges (while keeping all the skew edges). Then Δ' has $|S(\Delta)|$ edges and is connected. For fixed Δ all the summands in the internal sum over $\mathbf{i}$ on the right hand side above are bounded by a constant depending only on k and z. Moreover $|I_\Delta^n| \leq n^{|V(\Delta')|}$. Thus graphs Δ such that Δ' has fewer than $|S(\Delta)|/2+1$ vertices have no asymptotic contribution to $n^{-1}\mathbb{E}\mathrm{tr}H_n^k$. On the other hand let us notice that if for a skew edge $e = (v,w)$ of multiplicity 1, (w,v) is not an edge of Δ, then the corresponding variable $X_{i_{u(e)}i_{v(e)}}$ appears in the product above exactly once and by (A2), the expectation of the product vanishes. Thus the only graphs with nonzero asymptotic contribution are those Δ's for which each skew edge e treated as an undirected edge appears at least twice and Δ' has at least $|S(\Delta)|/2+1$ vertices. But the former condition means that the number a of edges of the graph Δ'' obtained from Δ' by identifying corresponding up and down edges (i.e. edges with the same end points in Δ) is at most $|S(\Delta)|/2$. Thus together with the number b of vertices of Δ'' it satisfies the inequality $b \geq a+1$. Since Δ'' is also connected, this means that it is a tree (and $b = a+1 = |S(\Delta)|/2+1$). Moreover, since the cycle in Δ' inherited from Δ corresponds to a walk in Δ'' which visits all vertices and returns to the vertex of departure, it means that all skew edges in Δ appear exactly twice, once as an up edge and once as a down edge.

One can also see that among the perpendicular edges connecting any two vertices of Δ there are equal numbers of up-edges and down-edges. Therefore we have

$$n^{-1}\mathbb{E}\mathrm{tr}H_n^k = \sum_{\substack{\Delta\in\Delta(k):\\ \Delta'' \text{ is a tree}}}|z|^{2|UP(\Delta)|}n^{-1-|S(\Delta)|/2}\sum_{\mathbf{i}\in I_\Delta^n}\mathbb{E}\prod_{e\in S(\Delta)}X_{i_{u(e)}i_{d(e)}} + o_{k,z}(1),$$

as can be seen by expanding the product over $e \in UP(\Delta) \cup DP(\Delta)$ into a sum.

By assigning to each Δ, such that Δ'' is a tree, a Γ-tree $T(\Delta)$ in a natural way (the special vertices are the vertices obtained by merging vertices of Δ connected by perpendicular edges and the orientation of edges is always from the up-end to the down-end), we may rewrite the above formula as

$$n^{-1}\mathbb{E}\mathrm{tr}H_n^k = \sum_{\substack{\Delta\in\Delta(k):\\ \Delta'' \text{ is a tree}}}|z|^{2|UP(\Delta)|}\xi_n(T(\Delta)) + o_{k,z}(1)$$

$$= \sum_{\substack{\Delta\in\Delta(k):\\ \Delta'' \text{ is a tree}}}|z|^{2|UP(\Delta)|} + o_{k,z}(1),$$

1082

where in the last equality we used Proposition 2.7. Thus we can define $\mu_k(z)$ as

$$\mu_k(|z|^2) = \sum_{\substack{\Delta \in \Delta(k): \\ \Delta'' \text{ is a tree}}} |z|^{2|UP(\Delta)|},$$

which ends the proof of the theorem.

Examples Let us illustrate the above proof with two examples for $k = 5$. Consider first $\mathbf{i} = (1,2,1,2,3)$ and $\mathbf{j} = (1,2,1,2,2)$. Then the corresponding graph $\Delta = G(\mathbf{i},\mathbf{j})$ has upper vertices $1,2,3$ and lower vertices $\underline{1},\underline{2}$ (we use the underline to stress that we are dealing with lower indices labelled by the same numbers as some upper indices). The consecutive edges are $(1,\underline{1}),(\underline{1},2),(2,\underline{2}),(\underline{2},1),(1,\underline{1}),(\underline{1},2),(2,\underline{2}),(\underline{2},3),(3,\underline{2}),(\underline{2},1)$. We thus have two pairs of perpendicular edges and after merging the corresponding vertices (1 with $\underline{1}$ and 2 with $\underline{2}$) we obtain a graph Δ' with vertices $1,2,3$, five up-edges $(1,2),(1,2),(2,1),(2,1),(2,3)$ and one down-edge $(3,2)$. By identifying corresponding edges (according to their end-vertices in the original graph Δ) we obtain a graph Δ'' with three vertices $1,2,3$ and three edges: a double-edge between 1 and 2 and a single edge between 2 and 3. This graph is not a tree and has no asymptotic contribution to $n^{-1}\mathbb{E}\mathrm{tr}H_n^5$ (the number of vertices is too small with respect to the number of edges).

Let us now consider another choice of $\mathbf{i}$ and $\mathbf{j}$ by setting $\mathbf{i} = (1,2,2,4,2)$ and $\mathbf{j} = (1,3,2,2,1)$. Then $\Delta = G(\mathbf{i},\mathbf{j})$ has three upper vertices $1,2,4$ and three lower vertices $\underline{1},\underline{2},\underline{3}$. The consecutive edges are $(1,\underline{1}),(\underline{1},2),(2,\underline{3}),(\underline{3},2),(2,\underline{2}),(\underline{2},4),(4,\underline{2}),(\underline{2},2),(2,\underline{1}),(\underline{1},1)$. After merging vertices joined by perpendicular edges, we obtain a graph Δ' with vertices $1,2,3,4$, up-edges $(1,2),(3,2),(2,4)$ and down-edges $(2,1),(2,3),(4,2)$. By identifying corresponding edges (according to their end-points in Δ) we obtain the graph Δ'' with four vertices $1,2,3,4$ three edges: between 1 and 2, between 2 and 4, between 2 and 3. The graph Δ'' is a tree and it can be endowed with a structure of a Γ-tree induced by the original structure from Δ. Indeed, the vertices 1 and 2 are special as obtained by merging vertices joined by perpendicular edges in Δ. The orientation of edges is $2 \to 1$, $4 \to 2$, $2 \to 3$, which reflects the fact that they correspond respectively to pairs of edges $\{(\underline{1},2),(2,\underline{1})\}$, $\{(\underline{2},4),(4,\underline{2})\}$ and $\{(2,\underline{3}),(\underline{3},2)\}$ in Δ. The Γ-tree $T(\Delta)$ obtained with this procedure is the one that served as an example in Section 2.3.2. Asymptotically Δ contributes to $n^{-1}\mathbb{E}\mathrm{tr}H_n^k$ the value $|z|^4\xi_n(T(\Delta)) \simeq |z|^4$.

Proof of Corollary 2.5. Let $\bar{L}_n = \mathbb{E}L_n$. By Theorem 2.4, for any $k \in \mathbb{N}$ we have the convergence $\int x^k \bar{L}_n(z) \to \mu_k(|z|^2)$. This already implies tightness of $\bar{L}_n$ and existence of a probability measure L_∞ with moments $\mu_k(|z|^2)$ together with a subsequence $\bar{L}_{n_k}$ converging to L_∞. Since in the special case of i.i.d. Gaussian entries the measure $\bar{L}_n$ is known to converge to a compactly supported measure, we conclude that the sequence of moments $\mu_k(|z|^2)$ determines the measure L and in fact the whole sequence $\bar{L}_n$ converges to L_∞.

Recall that the Stieltjes transform of a probability measure ν on $\mathbb{R}$ is the function $S_\nu : \mathbb{C}_+ \to \mathbb{C}$ given by the formula

$$S_\nu(z) = \int_\mathbb{R} \frac{1}{\lambda - z} dL_n(\lambda).$$

It is classical that for probability measures ν, ν_n on $\mathbb{R}$, the pointwise convergence $S_{\nu_n} \to S_\nu$ implies the weak convergence of ν_n to ν (see e.g. [7], Chapter 12).

Let $S_n\colon \mathbb{C}_+ \to \mathbb{C}$ be the Stieltjes transform of L_n. By Lemma 4.1 in the Appendix and the Borel-Cantelli Lemma $|S_n(\alpha) - \mathbb{E}S_n(\alpha)| \to 0$ almost surely. Thus with probability one $S_n(\alpha) - \mathbb{E}S_n(\alpha) \to 0$ for every $\alpha \in D$, where D is a countable dense subset of $\mathbb{C}_+$. But $\mathbb{E}S_n(\alpha)$ is the Stieltjes transform of $\bar{L}_n$ and thus converges to $S_{L_\infty}(\alpha)$ – the Stieltjes transform of L_∞, which shows that with probability one, for every $\alpha \in D$, $S_n(\alpha) \to S_{L_\infty}(\alpha)$. Since S_n's are analytic on $\mathbb{C}_+$ and jointly bounded on compact subsets of $\mathbb{C}_+$, a standard application of Montel's theorem shows that with probability one $S_n(\alpha) \to S(\alpha)$ for all $\alpha \in \mathbb{C}_+$, which implies that with probability one, L_n converges weakly to L_∞. $\qquad\square$

3 The circular law for matrices with independent log-concave unconditional columns

3.1 The main result

Before we formulate the main theorem of this section let us recall the basic definitions.

Definition 3.1. *An n-dimensional random vector $X = (X_1, \ldots, X_n)$ is called unconditional if its distribution is invariant under reflections with respect to coordinate hyperplanes, or equivalently if it has the same distribution as $(\varepsilon_1 X_1, \ldots, \varepsilon_n X_n)$, where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher variables, independent of X.*

Log-concave measures are usually defined in terms of Brunn-Minkowski inequalities. However, as proved by Borell (see [11] for general theory), log-concave measures not-supported on a proper hyperplane can be equivalently characterized in terms of densities. Since we will deal only with isotropic measures (see Definition 2.1) we will therefore use the following

Definition 3.2. *An n-dimensional random vector X is called log-concave if it has density of the form $\exp(-h(x))$ where $h\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex.*

Examples Several concrete examples of log-concave distribution have already appeared in Section 2. They were uniform distribution on properly normalized ℓ_p^n balls. This can be generalized to uniform distributions on general convex bodies. Other important examples include e.g. Gaussian random vectors.

For an $n \times n$ matrix A let μ_A denote the spectral measure of A defined as

$$\mu_A = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i},$$

where λ_i are (complex) eigenvalues of A.

Recall now the classical circular law in its most general form, obtained recently by Tao and Vu [36].

Theorem 3.3. *Let $A_n = [X_{ij}]_{i,j \le n}$, where $(X_{ij})_{i,j<\infty}$ is an infinite array of independent mean zero, variance one random variables. Then with probability one, the spectral measure $\mu_{\frac{1}{\sqrt{n}}A_n}$ converges weakly to the uniform distribution on the unit disc in $\mathbb{C}$.*

The main result of this section is the following version of the circular law which allows to replace the independence assumption on the entries by a geometric condition of log-concavity and unconditionality.

Theorem 3.4. *Let A_n be a sequence of $n \times n$ random matrices with independent rows $X_1^{(n)}, \ldots, X_n^{(n)}$ (defined on the same probability space). Assume that for each n and $i \leq n$, $X_i^{(n)}$ has a log-concave unconditional isotropic distribution. Then, with probability one, the spectral measure $\mu_{\frac{1}{\sqrt{n}} A_n}$ converges weakly to the uniform distribution on the unit disc in $\mathbb{C}$.*

3.2 Preliminary facts on log-concave measures

In this section we will gather the results on log-concave measures which will be used in the proof of Theorem 3.4 (some of them have been already briefly mentioned in Section 2 during the discussion of assumptions (A1)–(A3)).

Let us start with the concentration result for the Euclidean norm (which we will denote by $|\cdot|$) obtained by Klartag [24], which will ensure that the condition (A3) is satisfied, allowing us to use the results of Section 2.

Theorem 3.5. *Let X be an isotropic log-concave random vector in $\mathbb{R}^n$. There exist universal positive constants C and c such that for all $\varepsilon \in (0, 1)$,*

$$\mathbb{P}\left(\left|\frac{|X|^2}{n} - 1\right| \geq \varepsilon\right) \leq C \exp(-c\varepsilon^C n^c).$$

The next theorem is a recent small ball inequality by Paouris [29, Theorem 6.2]

Theorem 3.6. *Let X be an isotropic log-concave random vector in $\mathbb{R}^n$ and let A be an $n \times n$ real nonzero matrix. Then for $y \in \mathbb{R}^n$ and $\varepsilon \in (0, c_1)$,*

$$\mathbb{P}(|AX - y| \leq \varepsilon \|A\|_{HS}) \leq \varepsilon^{c_1(\|A\|_{HS}/\|A\|_{\ell_2 \to \ell_2})},$$

where $c_1 > 0$ is a universal constant.

Recall that singular values of an $m \times n$ matrix A are eigenvalues of $\sqrt{A^*A}$ (denote them by $\sigma_1 \geq \ldots \geq \sigma_n$). In particular $\sigma_1 = \sup_{x \in S^{n-1}} |Ax| = \|A\|_{\ell_2 \to \ell_2}$ (the operator norm of the matrix) and $\sigma_n = \inf_{x \in S_{n-1}} |Ax|$.

Similarly as in the classical case of matrices with i.i.d. entries we will need control on the smallest and the largest singular value of the matrix A_n. To control the latter we will use a recent result from [3] specialized to square matrices (we would like to remark in passing that for the purpose of proving the circular law, a weaker estimate following e.g. from Theorem 3.5 would be enough, we state the strong estimate, since it gives the optimal bound on the operator norm).

Theorem 3.7. *If A_n is an $n \times n$ matrix with independent log-concave isotropic rows, then with probability at least $1 - \exp(-c\sqrt{n})$, $\|A_n\|_{\ell_2 \to \ell_2} \leq C\sqrt{n}$, where C, c are universal constants.*

The smallest singular value of random matrices with independent entries has been thoroughly investigated e.g. in [31, 35, 25]. For random matrices with independent log-concave rows/columns it has been considered in [2, 1]. For the purpose of proving the circular law, it is enough to have a relatively crude bound on the smallest singular value of the matrix $A_n - z\mathrm{Id}$, which we provide below. A similar argument in the case of matrices with independent heavy tailed entries can be found e.g. in [10].

Proposition 3.8. *If A_n is an $n \times n$ matrix with independent log-concave isotropic rows, and M_n is an $n \times n$ deterministic complex matrix, then with probability at least $1 - n^{-2}$, the smallest singular value of $A_n + M_n$ is greater than $cn^{-3.5}$, where c is a universal constant.*

Proof. Denote the smallest singular value of $A_n + M_n$ by σ_n. Denote the rows of A_n by X_i and the rows of M_n by Y_i $(i = 1, \ldots, n)$. It is classical (see e.g. [31, 10]) that

$$\sigma_n \geq \frac{1}{\sqrt{n}} \min_{i \leq n} (\mathrm{dist}(X_i + Y_i, H_i)),$$

where H_i is the linear span of all the vectors $X_j + Y_j$ except for the i-th one. Thus

$$\mathbb{P}(\sigma_n \leq cn^{-3.5}) \leq \mathbb{P}(\exists_{i \leq n} \; \mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3})$$

$$\leq \sum_{i=1}^{n} \mathbb{P}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3}). \tag{5}$$

Consider any $i \leq n$. Since H_i is independent of X_i, by the Fubini theorem we have $\mathbb{P}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3}) = \mathbb{E}_{(X_j)_{j \neq i}} \mathbb{P}_{X_i}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3})$. We will bound the probability $\mathbb{P}_{X_i}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3})$. Let $\eta = x + iy$, where $x, y \in \mathbb{R}^n$, be any unit vector perpendicular to H_i (selected in a measurable way). Then

$$|\langle X_i + Y_i, \eta \rangle| \leq \mathrm{dist}(X_i + Y_i, H_i).$$

Since $|x|^2 + |y|^2 = 1$, at least one of the vectors x, y has the Euclidean norm not smaller than $1/\sqrt{2}$. Without loss of generality assume that $|x| \geq 1/\sqrt{2}$. If $\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3}$ then $|\langle X_i, x \rangle - \mathbf{Re}\langle Y_i, \eta \rangle| \leq cn^{-3}$. But $\langle X_i, x \rangle$ is a log-concave one-dimensional variable of variance at least $1/2$. It is well known that such variables have densities bounded from above by an absolute constant (see [21]). Therefore $\mathbb{P}_{X_i}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3}) \leq \mathbb{P}_{X_i}(|\langle X_i, x \rangle - \mathbf{Re}\langle Y_i, \eta \rangle| \leq cn^{-3}) \leq Ccn^{-3}$ and thus for c sufficiently small $\mathbb{P}_{X_i}(\mathrm{dist}(X_i + Y_i, H_i) \leq cn^{-3}) \leq n^{-3}$, which together with (5) proves the proposition. $\qquad \square$

3.3 Proof of Theorem 3.4

The proof will be based on the following replacement principle due to Tao and Vu [36, Theorem 2.1].

Theorem 3.9. *Suppose for each n that $A_n, B_n \in M_n(\mathbb{C})$ are ensembles of random matrices defined on a common probability space. Assume that*

(i) the expression

$$\frac{1}{n^2} \|A_n\|_{HS}^2 + \frac{1}{n^2} \|B_n\|_{HS}^2$$

is bounded almost surely,

(ii) for almost all complex numbers z,

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - z\mathrm{Id})| - \frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}B_n - z\mathrm{Id})|$$

converges almost surely to zero.

Then $\mu_{\frac{1}{\sqrt{n}}A_n} - \mu_{\frac{1}{\sqrt{n}}B_n}$ converges almost surely to 0.

To apply the above theorem, let $B_n = (g_{ij})_{i,j\leq n}$ where $(g_{ij})_{i,j}$ is an infinite array of independent standard Gaussian variables (we may assume that they are defined on the same probability space as the vectors $X_i^{(n)}$ defining the matrices A_n). We will prove that the hypotheses of Theorem 3.9 are satisfied. To this end we will very closely follow the main steps of the proof given by Tao and Vu for the independent case [36]. Due to lack of independence of all the entries, we cannot use their tools verbatim, however at each step we will be able to replace them by appropriate counterparts available in the log-concave setting (in fact given the 'log-concave toolbox' of Section 3.2, all the steps will be easier in our case) .

Once we check the assumptions of Theorem 3.9, the proof will be concluded since $\mu_{\frac{1}{\sqrt{n}}B_n}$ converges almost surely to the uniform measure on the unit disc.

The aim is thus to prove that for any $z \in \mathbb{C}$, with probability one,

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - z\mathrm{Id})| - \frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}B_n - z\mathrm{Id})| \to 0. \tag{6}$$

To simplify the notation from now on we will suppress the superscript (n) and denote the row vectors of $\frac{1}{\sqrt{n}}A_n - z\mathrm{Id}$ by $Z_1, \ldots, Z_n$. Similarly, let $Y_1, \ldots, Y_n$ be the row vectors of $\frac{1}{\sqrt{n}}B_n - z\mathrm{Id}$. Since

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}A_n - z\mathrm{Id})| = \frac{1}{n}\sum_{i=1}^{n}\log\mathrm{dist}(Z_i, V_i),$$

where V_i is the span of $Z_1, \ldots, Z_{i-1}$ and similarly

$$\frac{1}{n}\log|\det(\frac{1}{\sqrt{n}}B_n - z\mathrm{Id})| = \frac{1}{n}\sum_{i=1}^{n}\log\mathrm{dist}(Y_i, U_i),$$

where U_i is the span of $Y_1, \ldots, Y_{i-1}$, the goal is to prove that

$$\frac{1}{n}\sum_{i=1}^{n}\log\mathrm{dist}(Z_i, V_i) - \frac{1}{n}\sum_{i=1}^{n}\log\mathrm{dist}(Y_i, U_i) \to 0$$

with probability one.

Let us now recall the following identity

Lemma 3.10 (Lemma A.4 in [36]). *Let $1 \leq n' \leq n$ and M be a full rank $n' \times n$ matrix with singular values $\sigma_1(M) \geq \ldots \geq \sigma_{n'}(M) > 0$ and rows $X_1, \ldots, X_{n'} \in \mathbb{C}^n$. For each $1 \leq i \leq n'$ let W_i be the hyperplane generated by the $n' - 1$ vectors $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{n'}$. Then*

$$\sum_{j=1}^{n'}\sigma_j(M)^{-2} = \sum_{j=1}^{n'}\mathrm{dist}(X_j, W_j)^{-2}.$$

Apply the above lemma to $M = M_n = \frac{1}{\sqrt{n}}A_n - z\mathrm{Id}$, $(M = M_n = \frac{1}{\sqrt{n}}B_n - z\mathrm{Id}$ resp.). Note that by Proposition 3.8 and the Borell-Cantelli lemma with probability one $\sigma_i(M_n) \geq cn^{-4}$ for all sufficiently large n and $i \leq n$. Thus with probability one,

$$\log \mathrm{dist}(Z_i, V_i), \log \mathrm{dist}(Y_i, U_i) \geq -C_1 \log n.$$

Moreover, by Theorem 3.5 with probability one for large n the vectors Z_i, Y_i have Euclidean lengths of the order $O_z(1)$, which yields that with probability one for large n and all $i \leq n$,

$$\log \mathrm{dist}(Z_i, V_i), \log \mathrm{dist}(Y_i, U_i) \leq C_2. \tag{7}$$

To obtain (6) it is thus sufficient to prove that with probability one,

$$\frac{1}{n} \sum_{1 \leq i \leq n - n^{0.99}} \log \mathrm{dist}(Z_i, V_i) - \frac{1}{n} \sum_{1 \leq i \leq n - n^{0.99}} \log \mathrm{dist}(Y_i, U_i) \to 0. \tag{8}$$

Here we again follow Tao and Vu and divide the proof of the above convergence into two lemmas.

Lemma 3.11 (High-dimensional contributions)**.** *There exists a constant C, such that for every $\varepsilon \in (0, 1/4)$ and every $\delta \in (0, \varepsilon^2)$, with probability one, for sufficiently large n,*

$$\frac{1}{n} \sum_{(1-\delta)n < i \leq n - n^{0.99}} (|\log \mathrm{dist}(Z_i, V_i)| + |\log \mathrm{dist}(Y_i, U_i)|) \leq C\varepsilon.$$

Lemma 3.12 (Low-dimensional contributions)**.** *There exists a constant C, such that for every $\varepsilon \in (0, 1/4)$ there exists $\delta_0 > 0$ such that for every $\delta \in (0, \delta_0)$ with probability one, for sufficiently large n,*

$$\left| \frac{1}{n} \sum_{1 \leq i \leq (1-\delta)n} (\log \mathrm{dist}(Z_i, V_i) - \log \mathrm{dist}(Y_i, U_i)) \right| \leq C\varepsilon. \tag{9}$$

3.3.1 Proof of Lemma 3.11

It is clearly enough to consider the part of the sum in question corresponding to Z_i's. Let us first notice that (7) implies that with probability one, for $\delta < \varepsilon$,

$$\frac{1}{n} \sum_{(1-\delta)n < i \leq n - n^{0.99}} \max(\log \mathrm{dist}(Z_i, V_i), 0) \leq C_2 \varepsilon.$$

Thus it is sufficient to deal with the negative part of $\log \mathrm{dist}(Z_i, V_i)$. We will show that for any $i \leq n - n^{0.99}$,

$$\mathbb{P}(\mathrm{dist}(\sqrt{n}Z_i, V_i) \leq c\sqrt{n - i + 1}) \leq \exp(-cn^c) \tag{10}$$

for some universal constant $c > 0$.

To this end we will demonstrate the following lemma which is a counterpart of Proposition 5.1. in [36].

1088

Lemma 3.13. *Let $1 \leq d \leq n - n^{0.99}$ and let W be a deterministic d-dimensional subspace of $\mathbb{C}^n$. Then for every $i \leq n$ and every $\varepsilon \in (0, c_1)$ (where c_1 is the constant from Theorem 3.6) we have*

$$\mathbb{P}(\text{dist}(\sqrt{n}Z_i, W) \leq \varepsilon \sqrt{n-d}) \leq \varepsilon^{c_1 \sqrt{n-d}}.$$

Proof. Recall that $\sqrt{n}Z_i = X_i - \sqrt{n}ze_i$. If we denote by P the orthogonal projection onto $W^\perp$, we get

$$\text{dist}(\sqrt{n}Z_i, W) = |PX_i - \sqrt{n}zPe_i|.$$

Since W is of dimension d, we have rank $P = n - d$, hence $\|P\|_{HS} = \sqrt{n-d}$, whereas $\|P\|_{\ell_2 \to \ell_2}$ is clearly equal to 1. The only difficulty in applying Theorem 3.6 with $y = \sqrt{n}zPe_i$ and $A = P$ is that the matrix P and the vector $\sqrt{n}zPe_i$ are complex. This can be however easily overcome by identifying $\mathbb{C}^n$ with R^{2n}, writing $y = (\mathbf{Re}\sqrt{n}zPe_i, \mathbf{Im}\sqrt{n}zPe_i)$ and noticing that $PX_i = A(X_i, \tilde{X}_i)$, where $\tilde{X}_i$ is an independent copy of X_i and

$$A = \begin{bmatrix} \mathbf{Re}P & 0 \\ \mathbf{Im}P & 0 \end{bmatrix}.$$

We have $\|A\|_{HS} = \|P\|_{HS} = \sqrt{n-d}$ and $\|A\|_{\ell_2 \to \ell_2} \leq 1$, which together with Theorem 3.6 ends the proof of the lemma. $\qquad\square$

Since with probability one V_i is of dimension $i - 1$ and is independent of $X_i^{(n)}$, we can apply Lemma 3.13 conditionally on V_i to get (10). Thus by the Borell-Cantelli lemma, with probability one, for n large enough and all $1 \leq i \leq n - n^{0.99}$ we have $\text{dist}(Z_i, V_i) \geq c\sqrt{1 - i/n}$ and so for $\delta < \varepsilon^2$ we get

$$\sum_{(1-\delta)n < i \leq n - n^{0.99}} \max(-\log(\text{dist}(Z_i, V_i)), 0) \leq \delta n \log c^{-1} + \frac{1}{2} \sum_{i=n-\lceil \delta n \rceil}^{n-1} \log \frac{n}{n-i}$$

$$\leq \delta n \log c^{-1} + \frac{1}{2} \log \frac{n^{\lceil \delta n \rceil}}{\lceil \delta n \rceil!} \leq \delta n \log c^{-1} + \delta n \log \frac{C}{\delta} \leq C\varepsilon n,$$

which ends the proof of Lemma 3.11.

Remark   In fact for the purpose of proving the circular law instead of the small ball inequality by Paouris one could use above Klartag's thin shell inequality (however the argument in this case seems to be slightly more technical).

3.3.2   Proof of Lemma 3.12

Let $n' = \lfloor (1 - \delta)n \rfloor$ and let $A_{nn'}$ (resp. $B_{nn'}$) be the matrix with rows $\sqrt{n}Z_1, \ldots, \sqrt{n}Z_{n'}$ (resp. $\sqrt{n}Y_1, \ldots, \sqrt{n}Y_{n'}$). Let $L_{\frac{1}{n}A_{nn'}A_{nn'}^*}$ (resp. $L_{\frac{1}{n}B_{nn'}B_{nn'}^*}$) be the spectral distribution of $\frac{1}{n}A_{nn'}A_{nn'}^*$ (resp. $\frac{1}{n}B_{nn'}B_{nn'}^*$). Similarly as in [36], one can show that (9) is equivalent to

$$\int_0^\infty \log t \, dv_{n,n'}(t) = O(\varepsilon), \tag{11}$$

where $v_{n,n'} = L_{\frac{1}{n}A_{nn'}A_{nn'}^*} - L_{\frac{1}{n}B_{nn'}B_{nn'}^*}$.

The proof of (11) consists in splitting the integral range into several regions.

The region of very large t   Note that

$$\int_0^\infty td|v_{n,n'}|(t) \le \frac{1}{n'}\sum_{i=1}^{n'}\left((\frac{1}{\sqrt{n}}\sigma_i(A_{nn'}))^2 + (\frac{1}{\sqrt{n}}\sigma_i(B_{nn'}))^2\right) = \frac{1}{n'}\sum_{i=1}^{n'}(|Z_i|^2 + |Y_i|^2).$$

By Theorem 3.5 (or Theorem 3.7) and the Borell-Cantelli Lemma with probability one, for sufficiently large n and all i, $|Z_i|, |Y_i| \le |z| + C$ and so the right hand side of the above inequality is bounded by some constant $C(z)$, from which (e.g. by Markov's inequality and integration by parts) it follows that for some R_ε, depending only on z and ε,

$$\int_{R_\varepsilon}^\infty \log t|dv_{n,n'}|(t) \le \varepsilon.$$

The region of intermediate t   Let $k = n - n'$ and recall the Cauchy interlacing property, which says that $\sigma_i(A_n - \sqrt{n}z\mathrm{Id}) \ge \sigma_i(A_{nn'}) \ge \sigma_{i+k}(A_n - \sqrt{n}z\mathrm{Id})$ for $i = 1, \ldots, n - k$ (and similarly for $B_{nn'}$ and $B_n - \sqrt{n}z\mathrm{Id}$).

Let ψ be a continuous function equal to 1 on the interval $[\varepsilon^4, R_\varepsilon]$, vanishing outside $[\varepsilon^4/2, 2R_\varepsilon]$ and linear on the intervals $[\varepsilon^4/2, \varepsilon^4]$ and $[R_\varepsilon, 2R_\varepsilon]$. The interlacing inequalities imply that

$$\left|\int_0^\infty \psi(t)\log t dv_{nn'}(t) - \int_0^\infty \psi(t)\log t dv_{nn}(t)\right| \le C\varepsilon + C(\varepsilon)\delta,$$

for some constant $C(\varepsilon)$ depending only on z and ε. By Corollary 2.5, the integral with respect to v_{nn} converges almost surely to 0. Thus for any $\delta \le \delta_0$ (where δ_0 depends only on ε and z) we have with probability one,

$$\left|\int_0^\infty \psi(t)\log t dv_{nn'}(t)\right| \le C\varepsilon$$

for n large enough.

The regions of moderately small and very small t   Arguing again as in [36] and replacing the Proposition 5.1. therein by Lemma 3.13 we obtain first that the following two inequalities hold with probability one for sufficiently large n,

$$\frac{1}{n'}\sum_{i=1}^{n'}\left((\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'}))^{-2} + (\frac{1}{\sqrt{n}}\sigma_i(B_{n,n'}))^{-2}\right) \le C/\delta, \tag{12}$$

$$\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'}), \frac{1}{\sqrt{n}}\sigma_i(B_{n,n'}) \ge c\frac{n'-i}{n} \text{ for all } 1 \le i \le (1-2\delta)n, \tag{13}$$

where C, c are absolute constants.

Both of the above inequalities are proved in the same way as in [36]. As an example let us demonstrate the proof of (12). By Lemma 3.13, and the Borel-Cantelli lemma with probability one, for n large enough and all $i \le n'$, we have

$$\mathrm{dist}(Z_i, \mathrm{span}(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_{n'})) \ge c\sqrt{\delta},$$

and analogously for Y_i, which via Lemma 3.10 implies (12).

Now we can estimate

$$\int_{\delta^2}^{\varepsilon^4} |\log t| |dv_{n,n'}(t)|$$

$$\leq \frac{2}{n'} \sum_{i=1}^{n'} f(\frac{1}{\sqrt{n}}\sigma_i(A_{nn'})) + \frac{2}{n} \sum_{i=1}^{n'} f(\frac{1}{\sqrt{n}}\sigma_i(B_{nn'})),$$

where $f(t) = |\log t| \mathbf{1}_{\{[\delta^2, \varepsilon^4]\}}(t^2)$.

By symmetry it is obviously enough to bound the first summand on the right hand side. Since $f(t) \leq |\log \delta|$, if $\delta \leq \varepsilon^2$ we can discard the contribution of all $i \geq (1-2\delta)n$. Since by (13), for $i < (1-C\varepsilon^2)n$, $\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'}) > \varepsilon^2$, it is enough to estimate

$$\frac{1}{n} \sum_{(1-C\varepsilon^2)n \leq i \leq (1-2\delta)n} |\log(\frac{1}{\sqrt{n}}\sigma_i(A_{n,n'}))|.$$

But again by (13) this is at most

$$Cn^{-1} \sum_{(1-C\varepsilon^2)n \leq i \leq (1-2\delta)n} \log\frac{n}{n'-i}$$

$$\leq Cn^{-1} \sum_{\delta n \leq i \leq C\varepsilon^2 n} \log(\frac{n}{i}) \leq C\varepsilon^2 \log(\frac{1}{\varepsilon^2}) \leq C\varepsilon.$$

It remains to show that with probability one, for n large enough

$$\int_0^{\delta^2} |\log t| |dv_{n,n'}(t)| \leq C\varepsilon.$$

But for $t \leq \delta^2$, we have $|t \log t| \leq C\delta^{3/2}$ and thus by (12),

$$\int_0^{\delta^2} |\log t| d|v_{n,n'}(t)| \leq C\delta^{3/2} \int_0^{\delta^2} t^{-1} d|v_{n,n'}(t)|$$

$$\leq \frac{C\delta^{3/2}}{n'} \sum_{i=1}^{n'} \left(\sigma_i(\frac{1}{\sqrt{n}}A_{nn'})^{-2} + \sigma_i(\frac{1}{\sqrt{n}}B_{nn'})^{-2} \right)$$

$$\leq C\delta^{3/2}\delta^{-1} \leq C\varepsilon$$

for $\delta \leq \varepsilon^2$. This concludes the proof of Lemma 3.12.

3.3.3 Conclusion of the proof of Theorem 3.4

As already mentioned it is enough to check the assumptions of Theorem 3.9.

By Lemmas 3.11 and 3.12 we obtain (8), which implies (6), i.e. the assumption (ii). The assumption (i) follows by Theorem 3.7 and the Borel-Cantelli lemma.

3.4 Final remarks

We would like to point out that the assumption on unconditionality of rows of the matrix A_n was used only in the proof of Corollary 2.5, all the other ingredients in the proof of the circular law work well in the case of general isotropic log-concave vectors (in particular one can replace unconditionality with assumption (A2)).

Finally let us mention that both the small ball inequality (Theorem 3.6), used by us to obtain (10), and the condition (A3), necessary to apply Corollary 2.5, can be replaced e.g. by a Poincaré inequality, i.e. by the assumption that

$$\operatorname{Var} f(X_i^{(n)}) \leq C|\nabla f(X_i^{(n)})|^2$$

for all smooth functions $f : \mathbb{R}^n \to \mathbb{R}$ (see e.g. [5] for a general exposition) or some other sufficiently strong concentration result for Lipschitz functions (it is a long standing conjecture that isotropic log-concave measures satisfy the Poincaré inequality with a universal constant [23]). If we keep the assumption of unconditionality and instead of log-concavity we assume that all one-dimensional marginals of the row-vectors of the matrix A_n have densities bounded by a universal constant (or just a constant depending polynomially on the dimension) we can obtain a suitable version of Proposition 3.8 and repeat the whole proof of the circular law above to get the following

Theorem 3.14. *Let A_n be a sequence of $n \times n$ random matrices with independent isotropic unconditional rows $X_i^{(n)}$, s.t.*

- *the law of $X_i^{(n)}$ satisfies the Poincaré inequality with a constant independent of n,*

- *there exists $\gamma < \infty$ such that for all $x \in S^{n-1}$ and all n, i $\langle X_i^{(n)}, x \rangle$ has density upper-bounded by n^γ.*

Then the empirical spectral distribution of A_n converges almost surely to the uniform measure on the unit disk.

Remark After this paper was accepted for publication the author has been able to remove the assumption of unconditionality from Theorems 3.4 and 3.14. The proof relies heavily on the Stieltjes transform techniques (motivated by [13]) and will be presented in a separate article.

4 Appendix. Concentration of the Stieltjes transform for matrices with independent rows

We will now present a concentration inequality for the Stieltjes transform of the spectral distribution of random matrices of the form AA^*, where A is a random matrix with independent rows. The argument is pretty standard and we include it here just for the sake of completeness. Its various versions can be found e.g. in [7, p. 313] in the case of matrices with independent entries, however (as noted in [14]) in fact just independence of rows is required.

Lemma 4.1. *Let A be any $n \times N$ random matrix with independent rows and let $S \colon \mathbb{C}^+ \to \mathbb{C}$ be the Stieltjes transform of the spectral measure of $H = AA^*$. Then for any $\alpha = x + iy \in \mathbb{C}_+$ and any $\varepsilon > 0$,*

$$\mathbb{P}(|S_n(\alpha) - \mathbb{E}S_n(\alpha)| \geq \varepsilon) \leq C \exp(-c\varepsilon^2 n y^2).$$

Proof. Let $\mathbb{E}_k$ denote the expectation with respect to the last $n - k$ rows of the matrix A_n and let X_k be the k-th row of A, W_k the matrix consisting of the remaining $n - 1$ rows of A and $H_k = W_k W_k^*$. Then, as one can easily check,

$$S_n(\alpha) - \mathbb{E}S_n(\alpha) = \frac{1}{n}\sum_{k=1}^{n}\gamma_k,$$

where

$$\gamma_k = \mathbb{E}_k \mathrm{tr}(H - \alpha \mathrm{Id})^{-1} - \mathbb{E}_{k-1}\mathrm{tr}(H - \alpha \mathrm{Id})^{-1}$$

$$= (\mathbb{E}_k - \mathbb{E}_{k-1})\frac{1 + X_k W_k^*(H_k - \alpha \mathrm{Id})^{-2}W_k X_k^*}{|X_k|^2 - \alpha - X_k W_k^*(H_k - \alpha \mathrm{Id})^{-1}W_k X_k^*}.$$

We have

$$|1 + X_k W_k^*(H_k - \alpha \mathrm{Id})^{-2}W_k X_k^*| \le 1 + X_k W_k^*[(H_k - x\mathrm{Id})^2 + y^2 \mathrm{Id}]^{-1}W_k X_k^*$$

$$= -\frac{1}{y}\mathbf{Im}(|X_k|^2 - \alpha - X_k W_k^*(H_k - \alpha \mathrm{Id})^{-1}W_k X_k^*),$$

which implies that $|\gamma_k| \le 2y^{-1}$.

Note that γ_k form a sequence of martingale differences and thus, by the Azuma inequality, we get

$$\mathbb{P}(|S_n(\alpha) - \mathbb{E}S_n(\alpha)| \ge \varepsilon) \le C\exp(-c\varepsilon^2 n^2/ny^{-2}) = C\exp(-c\varepsilon^2 ny^2).$$

$\square$

References

[1] R. Adamczak, O. Guédon, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Condition number of a square matrix with i.i.d. columns drawn from a convex body. *To appear in Proc. Amer. Math. Soc.*

[2] R. Adamczak, O. Guédon, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Smallest singular value of random matrices with independent columns. *C. R. Math. Acad. Sci. Paris*, 346(15-16):853–856, 2008.

[3] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23(2):535–561, 2010.

[4] G. W. Anderson and O. Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008.

[5] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[6] G. Aubrun. Random points in the unit ball of l_p^n. *Positivity*, 10(4):755–759, 2006.

[7] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[8] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[9] C. Bordenave, P. Caputo, and D. Chafaï. Circular law theorem for random Markov matrices. Available at http://arxiv.org/abs/0808.1502. *To appear in Probab. Theory Related Fields*.

[10] C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of non-hermitian heavy tailed random matrices. Available at http://arxiv.org/abs/1006.1713. *To appear in Comm. Math. Phys.*

[11] C. Borell. Convex measures on locally convex spaces. *Ark. Mat.*, 12:239–252, 1974.

[12] W. Bryc, A. Dembo, and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):1–38, 2006.

[13] R. B. Dozier and J. W. Silverstein. On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.*, 98(4):678–694, 2007.

[14] N. El Karoui. Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond. *Ann. Appl. Probab.*, 19(6):2362–2405, 2009.

[15] L. Erdős, B. Schlein, H.-T. Yau, and J. Yin. The local relaxation flow approach to universality of the local statistics for random matrices. Preprint. Available at http://arxiv.org/abs/0911.3687.

[16] V. L. Girko. The circle law. *Teor. Veroyatnost. i Mat. Statist.*, (28):15–21, 1983.

[17] V. L. Girko. The strong circular law. Twenty years later. I. *Random Oper. Stochastic Equations*, 12(1):49–104, 2004.

[18] F. Götze and A. Tikhomirov. Limit theorems for spectra of positive random matrices under dependence. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 311(Veroyatn. i Stat. 7):92–123, 299, 2004.

[19] F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[20] F. Götze and A. N. Tikhomirov. Limit theorems for spectra of random matrices with martingale structure. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 181–193. Singapore Univ. Press, Singapore, 2005.

[21] D. Hensley. Slicing convex bodies—bounds for slice area in terms of the body's covariance. *Proc. Amer. Math. Soc.*, 79(4):619–625, 1980.

[22] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[23] R. Kannan, L. Lovász, and M. Simonovits. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom.*, 13(3-4):541–559, 1995.

[24] B. Klartag. Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.*, 245(1):284–310, 2007.

[25] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491–523, 2005.

[26] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[27] A. Pajor and L. Pastur. On the limiting empirical measure of eigenvalues of the sum of rank one matrices with log-concave distribution. *Studia Math.*, 195(1):11–29, 2009.

[28] G. Pan and W. Zhou. Circular law, extreme singular values and potential theory. *J. Multivariate Anal.*, 101(3):645–656, 2010.

[29] G. Paouris. Small ball probability estimates for log-concave measures. Preprint. *To appear in Trans. Amer. Math. Soc*.

[30] S. Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields*, 143(3-4):481–516, 2009.

[31] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[32] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000.

[33] A. Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.*, 207(3):697–733, 1999.

[34] T. Tao and V. Vu. Random covariance matrices: Universality of local statistics of eigenvalues. Preprint. Available at http://arxiv.org/abs/0912.0966.

[35] T. Tao and V. Vu. The condition number of a randomly perturbed matrix. In *STOC'07— Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 248–255. ACM, New York, 2007.

[36] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[37] Y. Q. Yin and P. R. Krishnaiah. Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Teor. Veroyatnost. i Primenen.*, 30(4):810–816, 1985.