Electron. Commun. Probab. **24** (2019), no. 78, 1–15.
https://doi.org/10.1214/19-ECP277
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Concentration of the empirical spectral distribution of random matrices with dependent entries*

Bartłomiej Polaczyk†

Abstract

We investigate concentration properties of spectral measures of Hermitian random matrices with partially dependent entries. More precisely, let X_n be a Hermitian random matrix of the size $n \times n$ that can be split into independent blocks of the size at most $d_n = o(n^2)$. We prove that under some mild conditions on the distribution of the entries of X_n, the empirical spectral measure of X_n concentrates around its mean.

The main theorem is a strengthening of the recent result by Kemp and Zimmerman, where the size of the blocks grows as $o(\log n)$. As an application, we are able to upgrade the results of Schenker and Schulz on the convergence in expectation to the semicircle law of a class of random matrices with dependent entries to weak convergence in probability. Other applications include patterned random matrices, e.g. matrices of Toeplitz, Hankel or circulant type and matrices with heavy tailed entries in the domain of attraction of the Gaussian distribution.

Keywords: random matrix theory; empirical spectral distribution; concentration of measure.
AMS MSC 2010: 60B20.
Submitted to ECP on September 14, 2018, final version accepted on November 19, 2019.
Supersedes arXiv:1809.05393.

1 Introduction

Throughout this paper we denote by M_n the space of $n \times n$ matrices over the scalar field $\mathbb{C}$ equipped with the Hilbert–Schmidt norm $\|A\|_{HS} = \sqrt{\operatorname{tr} AA^*}$, where $(A^*)_{ij} = \overline{A_{ji}}$. We set M_n^{sa} to be the vector subspace of M_n consisting of Hermitian matrices (i.e. matrices satisfying the condition $A^* = A$). A (general) random matrix is a random variable taking values in the space M_n.

Let X be a random $n \times n$ Hermitian matrix. Its all eigenvalues lie on the real line and thus we may consider its empirical spectral distribution (ESD)

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

where $\lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of X. It is worth remarking that since X is random then so is L_n^X as a distribution on the real line. One can thus consider its

*Research partially supported by the National Science Centre, Poland, grants no. 2015/18/E/ST1/00214.
†Institute of Mathematics, University of Warsaw. E-mail: b.polaczyk@mimuw.edu.pl

expected value, which is now a deterministic probability measure $\mathbb{E}L_n^X$ s.t. for every Borel set A

$$\mathbb{E}L_n^X(A) = \mathbb{E}(L_n^X(A))\,.$$

Studying the asymptotic properties of such distributions was first motivated by questions that arose in various models of quantum physics (cf. [40, 14, 37]). Since then, random matrix theory has evolved significantly, becoming an independent and influential branch of mathematics (we refer to [4, 8, 32, 39] for a detailed exposition of the subject). While the first results in this theory considered matrices with independent entries (up to a symmetry condition), over the last 20 years more and more attention has been directed towards investigation of matrices with dependencies between entries (see e.g. [1, 13, 15, 28, 31, 42]).

This paper is organized in the following way. We begin with presenting the results in Section 2. In Section 3 we show some applications, which cover Wigner-type theorems, matrices with heavy tailed entries in the domain of attraction of the Gaussian distribution, patterned and band matrices. Section 4 contains some facts from linear algebra and concentration of measure theory that are used in Section 5 to prove the results.

2 Main results

In this paper we restrict our attention to matrices with the block dependency structure, i.e. matrices whose entries can be divided into (not necessarily rectangular) blocks which form independent random vectors. The following definition makes this notion precise.

Definition 2.1. *A random matrix X of the size $n \times m$ satisfies the property $\mathcal{S}(d)$ whenever there exists a partition $\Pi = \{P_1, \ldots, P_k\}$ of the set $\{1, \ldots, n\} \times \{1, \ldots, m\}$ such that the vectors $Y_r = \{X_{ij}\}_{(i,j)\in P_r}$ are stochastically independent and the size of each partition set P_r does not exceed d, i.e. $|P_r| \le d$ for all $r = 1, \ldots, k$. We write shortly $X \in \mathcal{S}(d)$.*

The starting point of our considerations is the main theorem of [23], stated below.

Theorem 2.2. *Let $X_n \in M_n^{sa}$ be a sequence of random matrices such that $X_n \in \mathcal{S}(d_n)$ for every n with $d_n = o(\log n)$. If the family $\{|(X_n)_{ij}|^2\}_{1\le i,j\le n\in\mathbb{N}}$ is uniformly integrable, then*

$$\forall\, f \in C_L(\mathbb{R}) \quad \int f dL_n^{\frac{1}{\sqrt{n}}X_n} - \mathbb{E}\int f dL_n^{\frac{1}{\sqrt{n}}X_n} \to_{\mathbb{P}} 0\,,$$

with $C_L(\mathbb{R})$ denoting the set of all real 1-Lipschitz functions on $\mathbb{R}$ and $(X_n)_{ij}$ denoting the entries of the matrix X_n.

The proof is based on the concentration argument by Guionnet and Zeitouni [17] and log-Sobolev inequalities for compactly supported measures convolved with the standard Gaussian distribution derived by Kemp and Zimmerman.

The main result of this work may be seen as a stronger version of Theorem 2.2 where more dependency is allowed, i.e. $d_n = o(n^2)$. Before stating it, let us clarify some notation.

Definition 2.3. *A sequence of probability measures μ_n converges weakly to some measure μ if $\int f\, d\mu_n \to \int f\, d\mu$ for all continuous bounded functions f.*

A sequence of random probability measures μ_n converges weakly in probability to a (possibly random) measure μ if $d(\mu_n, \mu) \to_{\mathbb{P}} 0$ for some (equivalently for all) metric d that metrizes the above notion of weak convergence.

We denote these facts by $\mu_n \Rightarrow \mu$ and $\mu_n \Rightarrow_{\mathbb{P}} \mu$ respectively.

Definition 2.4. *A sequence of random matrices $X_n \in M_n$ has the property $\mathcal{L}$, $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$ if it satisfies the following Lindeberg-type condition*

$$\forall\, \epsilon > 0 \quad \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{n^2} \sum_{i,j=1}^{n} |(X_n)_{ij}|^2 \mathbb{1}_{\{|(X_n)_{ij}|>M\}} > \varepsilon \right) = 0\,.$$

Remark 2.5. By Markov's inequality, if the family $\{|(X_n)_{ij}|^2\}_{1\le i,j\le n\in\mathbb{N}}$ is uniformly integrable, then $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$.

The main result of this work is the following theorem.

Theorem 2.6. *Let $X_n \in M_n^{sa}$ be a sequence of random matrices such that $X_n \in \mathcal{S}(d_n)$ for every n with $d_n = o(n^2)$. If $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$, then for any metric d that metrizes weak convergence of probability measures*

$$d(L_n^{\frac{1}{\sqrt{n}}X_n}, \mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n}) \to_{\mathbb{P}} 0\,.$$

In particular, if $\mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n} \Rightarrow \mu$, then $L_n^{\frac{1}{\sqrt{n}}X_n} \Rightarrow_{\mathbb{P}} \mu$.

The following observations show connection between Theorems 2.2 and 2.6.

Proposition 2.7. *If $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$, then the sequence $\mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n}$ is tight.*

Proposition 2.8. *If the sequence $\mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n}$ is tight, then the following conditions are equivalent:*

i) $d(L_n^{\frac{1}{\sqrt{n}}X_n}, \mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n}) \to_{\mathbb{P}} 0$ *for any metric d that metrizes weak convergence of probability measures,*

ii) $\int f dL_n^{\frac{1}{\sqrt{n}}X_n} - \mathbb{E}\int f dL_n^{\frac{1}{\sqrt{n}}X_n} \to_{\mathbb{P}} 0$ *for all $f \in C_b(\mathbb{R})$,*

iii) $\int f dL_n^{\frac{1}{\sqrt{n}}X_n} - \mathbb{E}\int f dL_n^{\frac{1}{\sqrt{n}}X_n} \to_{\mathbb{P}} 0$ *for all $f \in C_c(\mathbb{R}) \cap C_L(\mathbb{R})$,*

where $C_b(\mathbb{R})$ (resp. $C_c(\mathbb{R})$) denotes the set of all bounded (resp. compactly supported) real continuous functions on $\mathbb{R}$.

If additionally the family $\{|(X_n)_{ij}|^2\}_{1\le i,j\le n\in\mathbb{N}}$ is uniformly integrable, then all the above conditions become equivalent to

iv) $\int f dL_n^{\frac{1}{\sqrt{n}}X_n} - \mathbb{E}\int f dL_n^{\frac{1}{\sqrt{n}}X_n} \to_{\mathbb{P}} 0$ *for all $f \in C_L(\mathbb{R})$.*

Combining the above observations and Remark 2.5 asserts that Theorem 2.6 strengthens Theorem 2.2.

Remark 2.9. The assumption $d_n = o(n^2)$ in Theorem 2.6 is optimal for the convergence in probability. To see that, consider two random matrix ensembles $X_n, Y_n \in M_n$, whose ESDs converge a.s. to distinct limits μ and ν. Set

$$Z_n = \begin{bmatrix} \epsilon X_{nt_n} + (1-\epsilon)Y_{nt_n} & 0 \\ 0 & I_{n(1-t_n)} \end{bmatrix},$$

where $I_k \in M_k$ is the identity matrix, ϵ is a Bernoulli random variable ($\mathbb{P}(\epsilon = 1) = \mathbb{P}(\epsilon = 0) = 1/2$) independent of all X_n and Y_n and $t_n \in (0,1)$ is a sequence converging to some $t \in (0,1)$, s.t. nt_n and $n(1-t_n)$ are integers for every n. One can see that $Z_n \in \mathcal{S}(d_n)$ with $d_n = (t_n n)^2$. On the other hand $L_n^{Z_n} \Rightarrow t(\epsilon\mu + (1-\epsilon)\nu) + (1-t)\delta_1$ a.s., whence $d(L_n^{Z_n}, \mathbb{E}L_n^{Z_n})$ cannot converge in probability to zero for any metric d that metrizes weak convergence of probability measures.

Using the standard hermitization technique one can extend the result of Theorem 2.6 onto the convergence of the distribution of singular values of not necessarily Hermitian ensembles.

Theorem 2.10. *Let X_n be a sequence of $n \times N$ random matrices (with $N = N(n)$) such that $X_n \in \mathcal{S}(d_n)$ for every n with $d_n = o(n^2)$ and set $Y_n = \sqrt{X_n X_n^*}$. Assume $n/N \to c \in (0, \infty)$. If the family $\{X_n\}_{n\in\mathbb{N}}$ satisfies the Lindeberg-type condition*

$$\forall\, \varepsilon > 0 \quad \lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{nN} \sum_{i=1}^{n} \sum_{j=1}^{N} |(X_n)_{ij}|^2 \mathbb{1}_{\{|(X_n)_{ij}|^2 > M\}} > \varepsilon \right) = 0\,,$$

then for any metric d that metrizes weak convergence of probability measures

$$d(L_n^{\frac{1}{\sqrt{n}}Y_n}, \mathbb{E}L_n^{\frac{1}{\sqrt{n}}Y_n}) \to_{\mathbb{P}} 0\,.$$

Finally, we deduce Theorem 2.6 from another, more general result, stated below.

Definition 2.11. *A sequence of random matrices $X_n \in M_n$ has the property $\mathcal{L}(a_n)$, $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}(a_n)$, for some sequence $a_n \to \infty$ if it satisfies the following Lindeberg-type condition*

$$\forall\, \epsilon > 0 \quad \limsup_{n\to\infty} \mathbb{P}\left(\frac{1}{n^2} \sum_{i,j=1}^{n} |(X_n)_{ij}|^2 \mathbb{1}_{\{|(X_n)_{ij}| > \varepsilon a_n\}} > \varepsilon \right) = 0\,.$$

Theorem 2.12. *Let $X_n \in M_n^{sa}$ be a sequence of random matrices such that $X_n \in \mathcal{S}(d_n)$ for every n with $d_n = O(n^2/a_n^2)$ for some sequence $a_n \to \infty$. If $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}(a_n)$, then*

$$\forall\, f \in C_c(\mathbb{R}) \cap C_L(\mathbb{R}) \quad \int f dL_n^{\frac{1}{\sqrt{n}}X_n} - \mathbb{E}\int f dL_n^{\frac{1}{\sqrt{n}}X_n} \to_{\mathbb{P}} 0\,.$$

3 Consequences and examples

3.1 Wigner-type theorems

One of the most classical results in the theory of asymptotic behaviour of the spectrum of random matrices dates back to [40]. It states that if (X_n) is a sequence of real Hermitian random matrices with i.i.d. entries (up to the symmetry constraint) with zero mean and variance equal to one, then $L_n^{\frac{1}{\sqrt{n}}X_n}$ converges almost surely to the semicircular distribution σ, that is

$$\mathbb{P}\left(L_n^{\frac{1}{\sqrt{n}}X_n} \Rightarrow \sigma \right) = 1\,,$$

where

$$\sigma(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\,\mathbb{1}_{|x|\le 2}$$

is the Wigner semicircular distribution playing the analogous role in free probability as the Gaussian distribution plays in classical probability.

Recently Schenker and Schultz-Baldes [35] proved a version of Wigner Theorem in which one allows some degree of dependence between the entries of a matrix and the price paid is the weaker notion of convergence obtained, i.e. convergence in expectation instead of probability.

More precisely, let $\sim_n$ denote an equivalence relation on $\{1, 2, \ldots, n\}^2 = [n]^2$ and let X_n be a sequence of Hermitian random matrices s.t. random vectors made of entries of X_n belonging to distinct equivalence classes are independent (and the dependence between elements of the same class can be arbitrary). We impose the following conditions on X_n and $\sim_n$:

(C0) $\sup\{\,\mathbb{E}\,|(X_n)_{ij}|^k : 1 \le i,j \le n \in \mathbb{N}\,\} < \infty$ for all $k \in \mathbb{N}$,

(C1) $\max_{i\in[n]} \#\{\,(j,i',j') \in [n]^3 : (i,j) \sim_n (i',j')\,\} = o(n^2)$,

(C2) $\max_{(i,j,i')\in[n]^3} \#\{\,j' \in [n] : (i,j) \sim_n (i',j')\,\} \le B$ for some $B > 0$ and all $n \in \mathbb{N}$,

(C3) $\#\{\,(i,j,i') \in [n]^3 : (i,j) \sim_n (j,i'),\;\; i' \ne i\,\} = o(n^2)$.

The main theorem of [35] can be stated as follows.

Theorem 3.1. *If a sequence $(X_n,\sim_n)_{n\in\mathbb{N}}$ satisfies conditions (C0)–(C3), then*

$$\mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n} \Rightarrow \sigma\,.$$

The above result was highly motivated by applications, in particular by the analysis of the Anderson model. The standard Anderson model is given by the following random Hamiltonian acting on the space $\ell^2(\mathbb{Z}^d)$:

$$H\psi(x) = \sum_{|y-x|=1} \psi(y) + V(x)\psi(x)\,,$$

where $\{V(x)\}_{x\in\mathbb{Z}^d}$ is a family of standard Gaussian i.i.d. random variables. In [33], it was shown that the above model at small disorder can be analyzed by dividing the space into small cubes Λ. If V_Λ is the restriction of H to one such cube Λ, then it can be effectively approximated by a finite random matrix whose coefficients are centered complex Gaussian random variables with a given dependency structure (see [9, 35] for more details). Finding the limiting spectral distribution of such matrices for $d = 2$ was solved in [10]. The case $d \ge 3$ was dealt with in [35], where the authors showed that V_Λ falls into the regime of Theorem 3.1 and thus the limiting distribution of V_Λ under appropriate normalization is semicircular.

Note that the condition $d_n = o(n^2)$ from Theorem 2.6 can be written in the above language as

$$\max_{(i,j)\in[n]^2} \#\{\,(i',j') \in [n]^2 : (i,j) \sim_n (i',j')\,\} = o(n^2)\,,$$

which is clearly implied by (C1). Moreover, (C0) implies uniform integrability of $\{|(X_n)_{ij}|^2\}_{1\le i,j\le n\in\mathbb{N}}$, which by Remark 2.5 implies that $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$. Theorem 2.6 gives therefore the following strengthening of Theorem 3.1.

Corollary 3.2. *If a sequence $(X_n,\sim_n)_{n\in\mathbb{N}}$ satisfies conditions (C0)–(C3), then*

$$L_n^{\frac{1}{\sqrt{n}}X_n} \Rightarrow_{\mathbb{P}} \sigma\,.$$

The above may be seen as a special case of application of Theorem 2.6, which in general allows (whenever the assumptions are met) to strengthen convergence in expectation to convergence in probability. The very same scheme may be applied to strengthen Theorem 5.1 from [20] where the authors develop further the method of Schenker and Schultz to deal with matrices of the form

$$A = \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix} \text{ for } X \in \mathbb{C}^{s\times t}\,,$$

where the limiting measure is the Marchenko-Pastur distribution.

3.2 Matrices with heavy tailed entries

Let us recall that a mean zero random variable x is in the domain of attraction of the Gaussian distribution if there exists a sequence b_n s.t.

$$\mathrm{Law}\left(\frac{\sum_{i=1}^n x_i}{b_n}\right) \Rightarrow \mathcal{N}(0,1)\,,$$

where x_i's are i.i.d. copies of x. It can be shown that this is the case if and only if the function

$$l(t) = \mathbb{E}x^2 \mathbb{1}_{\{|x|\le t\}}$$

is slowly varying at infinity (see e.g. [6, 21]).

Assume that x has infinite variance, define

$$b = \inf\{\, t > 0 : l(t) > 0\,\}\,, \quad b_n = \inf\{\, t > b+1 : nl(t) \le t^2\,\} \tag{3.1}$$

and consider a matrix $X_n \in M_n^{sa}$, whose entries are i.i.d. copies of x. It was shown in [42] (see also [2]) that $L_n^{X_n/b_n} \Rightarrow \sigma$ almost surely. The results of this paper allow us to prove the convergence in probability in case of dependent entries, yielding the following proposition.

Proposition 3.3. *Let $X_n \in M_n^{sa}$ be a sequence of random matrices satisfying $X_n \in \mathcal{S}(d_n)$ with $d_n = O(n)$, whose entries have the same distribution as a random variable x with zero mean, infinite variance and in the domain of attraction of the Gaussian distribution. Then for any metric d that metrizes weak convergence of probability measures*

$$d(L_n^{X_n/b_n}, \mathbb{E}L_n^{X_n/b_n}) \to_{\mathbb{P}} 0\,.$$

The proof is moved to the last section. In the above we can observe a drop in the size d_n of the blocks compared to Theorem 2.6. It is not straightforward to see if this result can be improved.

3.3 Patterned matrices

Many ensembles of random matrices considered in the literature can be seen as special cases of the so-called patterned matrices. Following [12], let us consider a family of functions $\mathcal{G} = \{\, I_n : \{1,\ldots,n\}^2 \to \mathbb{Z}^d\,\}_{n\in\mathbb{N}}$, which we will call a link family. A patterned matrix X_n is a matrix of the form $\overline{(X_n)_{ji}} = (X_n)_{ij} = [Z_{I_n(i,j)}]$ for $i \le j$ where $\mathcal{Z} = \{Z_z\}_{z\in\mathbb{Z}^d}$ is a family of independent random variables (note that by construction we demand that $X_n \in M_n^{sa}$, which gives some constrains on $\mathcal{G}$ and $\mathcal{Z}$). We say that the sequence X_n is associated with the link family $\mathcal{G}$. Theorems 2.6 and 2.10 yield the following corollary.

Corollary 3.4. *If a sequence $\{X_n\}_{n\in\mathbb{N}}$ associated with $\mathcal{G}$ belongs to the class $\mathcal{L}$, $(X_n)_{n\in\mathbb{N}} \in \mathcal{L}$, and $\left|I_n^{-1}(z)\right| = o(n^2)$ for every $z \in \mathbb{Z}^d$, then for any metric d that metrizes weak convergence of probability measures*

$$d(L_n^{\frac{1}{\sqrt{n}}X_n}, \mathbb{E}L_n^{\frac{1}{\sqrt{n}}X_n}) \to_{\mathbb{P}} 0\,.$$

Setting $I_n(i,j) = (\min(i,j), \max(i,j))$ restores the generic Wigner ensemble.

Setting $I_n(i,j) = |i-j|$ or $I_n(i,j) = i+j$ with $\mathcal{Z}$ being an i.i.d. family results in Toeplitz (T_n) and Hankel (H_n) ensembles respectively, considered firstly in the influential paper [7]. The problem of the convergence of their ESDs was solved in [15], where the authors prove almost sure convergence of $T_n/\sqrt{n}$ and $H_n/\sqrt{n}$ to some deterministic distributions that do not depend on the law of Z_0 and have unbounded support.

Setting $I_n(i,j) = i + j \pmod n$ or $I_n(i,j) = n/2 - |n/2 - |i-j||$ results in reversed circulant and symmetric circulant ensembles respectively. These ensembles (and more general G-circulants) were extensively studied (cf. [3, 12, 28, 29]).

It can be easily seen that, under some mild assumptions on the family $\mathcal{Z}$, all these ensembles satisfy hypothesis of Corollary 3.4. Moreover, whenever all elements of $\mathcal{Z}$ are with zero mean, infinite variance and in the domain of attraction of the Gaussian distribution, then all these ensembles satisfy assumptions of Proposition 3.3 as well.

Additionally, Theorem 2.6 allows us to simplify the proofs from [3] and [29] where the authors strengthen convergence in expectation to convergence in probability (cf. [29, Proof of Theorem 4.1] and [3, Proof of Theorem 1.5]).

3.4 Band and block matrices with correlation structure

A classical band matrix ensemble consists of matrices whose entries are independent and equal zero at far distance from the diagonal. It is known (cf. [5]) that if one assumes some regularity of the distribution of the entries, then the ESD of such matrices converges almost surely to the semicircular measure.

There are few papers however covering the behavior of the ESD of band matrices with dependent entries. Some of the best known results can be found in [36] and [34]. The former work deals with a wider class of block matrices. Such ensembles arise naturally in applications, e.g. in wireless communication theory in the Multiple-Input Multiple-Output (MIMO) systems with Intersymbolic Interference (ISI). The capacity of such system with n transmit antennas and m receive antennas can be described in terms of the ESD of the matrix GG^*, where G is a band random matrix consisting of the finite number of matrices (A_l) of the size $n \times m$. The elements of A_l are independent and the correlation structure between A_i and A_j is given (for a more precise formulation we refer to [24, Chapter 2]). In [34], the authors have proved (in the Gaussian case), that the ESD of (appropriately normalized) GG^* converges almost surely, as n, m tend to infinity, to some deterministic probability measure described in terms of its Cauchy transform. Using Theorem 2.10, one immediately deduces that the empirical spectral measure is concentrated around its expectation, which in combination with their analysis of the expected spectral measure gives a weaker property of weak convergence in probability. However, as can be easily seen from the proof of Theorem 2.12, in this case our argument gives in fact almost sure convergence, since the size of independent blocks remains bounded.

4 Auxiliary lemmas, facts and definitions

We start with recalling some definitions and important results.

Definition 4.1. *We say that a random vector X satisfies the (subgaussian) concentration property with positive constants C and c with respect to the family $\mathcal{F}$ if*

$$\mathbb{P}\left(|f(X) - \mathbb{M}f(X)| > t\right) \le Ce^{-t^2/c} \tag{4.1}$$

for all $t \ge 0$ and for every every $f \in \mathcal{F}$, with $\mathbb{M}f(X)$ denoting some (any) median of $f(X)$.

Remark 4.2. A standard observation is that in (4.1) one can replace the median by the mean (at the cost of enlarging of c by some multiplicative factor $\widetilde{c}$ depending only on C).

Recall that $C_L(\mathbb{R})$ denotes the set of all 1-Lipschitz functions on $\mathbb{R}$. Substituting in the above $\mathcal{F} = C_L(\mathbb{R})$ restores the definition of the classical subgaussian concentration property, to which we refer simply as $CP(C,c)$ and substituting

$$\mathcal{F} = \{ f \in C_L(\mathbb{R}) : f \text{ is convex} \}$$

restores the weaker notion of the convex concentration property (as stated in e.g. [30]), to which we refer as $CCP(C,c)$.

Some convex concentration results concerning the spectral distribution of random matrices were firstly discovered by Guionnet and Zeitouni (cf. [17]). The proofs are mostly based on the famous theorem due to Talagrand (cf. [38, 25]), whose corollary we state below.

Theorem 4.3 ([27, Corollary 4])**.** *Let V be the direct sum of normed vector spaces $(V_i, \|\cdot\|_i)_{1\le i\le N}$, equipped with the norm $\|(v_1,\ldots,v_N)\| = \sqrt{\sum \|v_i\|_i^2}$ and X_i be random variables taking values in V_i s.t. $\operatorname{ess\,sup} \|X_i\|_i \le \rho$. Then the V-valued random vector $(X_1,\ldots,X_N)$ satisfies CCP$(4, 16\rho^2)$.*

The following is the so-called Hoffman–Wielandt lemma and its immediate corollary.

Lemma 4.4 ([19, Theorem 1])**.** *Let $A, B \in M_n^{sa}$ with eigenvalues $\lambda_1^A \le \ldots \le \lambda_n^A$ and $\lambda_1^B \le \ldots \le \lambda_n^B$ respectively. Then*

$$\sum_{i=1}^{n} \left|\lambda_i^A - \lambda_i^B\right|^2 \le \|A - B\|_{HS}^2 .$$

Corollary 4.5. *For every $f \in C_L(\mathbb{R})$ the map $M_n^{sa}(\mathbb{C}) \ni X \to \int f \, dL_n^{\frac{1}{\sqrt{n}}X}$ is $\frac{1}{n}$-Lipschitz with respect to the Hilbert–Schmidt norm.*

We also need the following classical observation (sometimes called Klein's lemma). For a proof we refer to [17, Lemma 1.2] or [26, Theorem 9.G.1].

Lemma 4.6. *If f is a real valued convex function on $\mathbb{R}$, then the mapping $M_n^{sa}(\mathbb{C}) \ni X \to \int f d\, L_n^X$ is convex.*

As a consequence of the above observations, we obtain that whenever the entries of a random matrix $X \in M_n^{sa}$ are compactly supported, then X_n satisfies convex concentration property (w.r.t. the Hilbert–Schmidt norm) which allows to estimate $\int f \, dL_n - \mathbb{E} \int f \, dL_n$ for large n. This fact plays a crucial role in the proof of Theorem 2.12. A similar argument was used to prove convex concentration in [18] (cf. the proof of Theorem 6 therein). For other results concerning convex concentration of random matrices see e.g. [27, 16].

Finally, the following facts are important for the proof of Proposition 3.3.

Lemma 4.7 ([8, Theorem A.43])**.** *Let X and Y be two $n \times n$ Hermitian matrices. Then*

$$\left\|L_n^A - L_n^B\right\| \le \frac{1}{n} \operatorname{rank}(A - B) ,$$

where $\|\cdot\|$ denotes the Kolomogorov distance between probability measures.

Lemma 4.8 ([41, Proof of Theorem 3.32])**.** *Let $X \in M_n^{sa}$ with rows $x_1,\ldots,x_n$ and eigenvalues $\lambda_1,\ldots,\lambda_n$. Then for every $0 < r \le 2$*

$$\sum_{i=1}^{n} |\lambda_i|^r \le \sum_{i=1}^{n} \|x_i\|_2^r .$$

5 Proofs

Proofs of Propositions 2.7 and 2.8

In what follows, we keep the notation $L_n = L_n^{\frac{1}{\sqrt{n}}X_n}$.

Proof of Proposition 2.7. Fix $\varepsilon > 0$, let $C = [-K, K]^c$ for some $K > 0$ and set

$$A_n = \left\{ \frac{1}{n^2} \sum_{i,j=1}^{n} |(X_n)_{ij}|^2 \mathbb{1}_{\{|(X_n)_{ij}|^2 > M\}} < 1 \right\}.$$

Then

$$\begin{aligned}\mathbb{E}L_n(C) &= \mathbb{E}L_n(C)\mathbb{1}_{A_n^c} + \mathbb{E}L_n(C)\mathbb{1}_{A_n} \\ &\le \mathbb{P}(A_n^c) + \mathbb{E}\frac{\int x^2\, dL_n}{K^2}\mathbb{1}_{A_n} \le \mathbb{P}(A_n^c) + \frac{M+1}{K^2},\end{aligned}$$

where the first inequality is an application of Markov's inequality and the second uses the fact that $\int x^2\, dL_n = \frac{1}{n^2}\|X_n\|_{HS}^2$. Since $X_n \in \mathcal{L}$, choosing appropriate M asserts that $\mathbb{P}(A_n^c) < \varepsilon/2$ for n large enough. Taking then K large enough yields $\mathbb{E}L_n(C) < \varepsilon$ for every n and the result follows. □

Proof of Proposition 2.8. Recall that a sequence of random elements with values in a metric space converges in probability to some random element if and only if from each of its subsequences one can choose a further subsequence that converges almost surely to that element. Moreover, by Prokhorov's theorem, a family of measures on a Polish space is tight if and only if it is sequentially weakly compact (cf. [22, Lemma 3.2 and Proposition 4.21]).

$i) \Rightarrow ii)$. Take any sequence $N \subset \mathbb{N}$. We will find a further subsequence along which $\int f\, dL_n - \mathbb{E}\int f\, dL_n$ converges a.s. to zero for any $f \in C_b(\mathbb{R})$. By tightness, there exist $N' \subset N$ and a probability measure μ s.t. $\mathbb{E}L_n \Rightarrow \mu$ along N'. By the triangle inequality, $d(L_n, \mu) \to_{\mathbb{P}} 0$ along N' and thus there exists a further subsequence $N'' \subset N'$ s.t. $d(L_n, \mu) \to 0$ a.s. along N'', which yields the result.

Implication $ii) \Rightarrow iii)$ is trivial.

$iii) \Rightarrow i)$. Consider the metric d on the set of all probability measures on $\mathbb{R}$ given by

$$d(\mu, \nu) = \sum_{k \in \mathbb{N}} 2^{-k} \left| \int f_k\, d\mu - \int f_k\, d\nu \right|,$$

where $\{f_k\}_{k\in\mathbb{N}} \subset C_c(\mathbb{R})$ is some dense subset of the unit ball (in the $\sup$ norm) of $C_c(\mathbb{R})$. It is easy to see that d metrizes weak convergence of probability measures. We have

$$\begin{aligned}\mathbb{P}(d(L_n, \mathbb{E}L_n) > \varepsilon) &\le \mathbb{P}\left(\sum_{k=1}^{N} 2^{-k} \left| \int f_k\, dL_n - \mathbb{E}\int f_k\, dL_n \right| > \frac{\varepsilon}{2} \right) \\ &\le \sum_{k=1}^{N} \mathbb{P}\left(\left| \int f_k\, dL_n - \mathbb{E}\int f_k\, dL_n \right| > \frac{2^{k-1}\varepsilon}{N} \right),\end{aligned}$$

for some N large enough, depending only on ε. Now, for every k there is some compactly supported Lipschitz function g_k s.t. $\|f_k - g_k\|_\infty < \frac{2^{k-1}\varepsilon}{3N}$. Application of the triangle inequality yields i).

To prove the second part of Proposition 2.8, assume firstly conditions $i) - iii)$, fix $t, \varepsilon > 0$ and consider some $f \in C_L$. Let f_r be continuous, equal f on the interval $[-r, r]$

and constant beyond it. Since $f_r \in C_b(\mathbb{R})$, then by assumption $\int f_r\, dL_n - \mathbb{E}\int f_r\, dL_n \to_{\mathbb{P}} 0$. We have

$$\mathbb{E}\int |f - f_r|\, dL_n \le \mathbb{E}\int |x|\, \mathbb{1}_{\{|x|>r\}}\, dL_n \le \frac{1}{r}\mathbb{E}\int |x|^2\, dL_n = \frac{1}{r}\mathbb{E}\frac{1}{n^2}\, \|X_n\|_{HS}^2 \le \delta$$

for any δ for r large enough by the uniform integrability of $\{|(X_n)_{ij}|^2\}_{1\le i,j\le n\in\mathbb{N}}$. Take now $\delta \le \min(\frac{t}{3}, \frac{t\varepsilon}{6})$ and choose r such that the above estimate holds. Applying the triangle and Markov inequalities yields

$$\mathbb{P}\left(\left|\int f\, dL_n - \mathbb{E}\int f\, dL_n\right| > t\right) \le \mathbb{P}\left(\left|\int f\, dL_n - \int f_r\, dL_n\right| > \frac{t}{3}\right) + \mathbb{P}\left(\left|\int f_r\, dL_n - \mathbb{E}\int f_r\, dL_n\right| > \frac{t}{3}\right) \le \varepsilon$$

for n large enough.

The proof in the opposite direction is immediate. □

Proofs of Theorems 2.6, 2.10 and 2.12

We start with proving Theorem 2.12. The argument is highly motivated by the work of Guionnet and Zeitouni [17, Theorem 1.3]. The conclusions of Theorem 2.6 and 2.10 follow then easily.

Proof of Theorem 2.12. Denote $L_n = L_n^{X_n/\sqrt{n}}$ and let $f \in C_c(\mathbb{R}) \cap C_L(\mathbb{R})$ be supported in the interval $[-M, M]$ (recall that $C_L \cap C_c$ is the set of 1-Lipschitz and compactly supported functions). By Proposition 2.8, it is enough to show that for every $\delta, t > 0$ and n large enough

$$\mathbb{P}\left(\left|\int f\, dL_n - \mathbb{E}\int f\, dL_n\right| > t\right) < \delta\,.$$

Let δ and t be fixed from now on. Take some $\varepsilon > 0$ (to be fixed later), set

$$(X_n^\varepsilon)_{ij} = (X_n)_{ij}\mathbb{1}_{\{|(X_n)_{ij}| \le \varepsilon a_n\}}$$

and denote $L_n^\varepsilon = L_n^{X_n^\varepsilon/\sqrt{n}}$. Firstly, we show that there exists ε s.t.

$$\mathbb{P}\left(\left|\int f\, dL_n^\varepsilon - \mathbb{E}\int f\, dL_n^\varepsilon\right| > \frac{t}{3}\right) < \frac{\delta}{2}$$

for n large enough and then that L_n^ε and L_n do not differ much.

Let $\Pi_n = \{P_1^n, \ldots, P_k^n\}$ be the partition of X_n into independent blocks and $Y_{r,n}^\varepsilon$ be the random vector given by the entries $\{(X_n^\varepsilon)_{ij}\}_{(i,j)\in P_r^n}$. Since $Y_{1,n}^\varepsilon, \ldots, Y_{k,n}^\varepsilon$ are stochastically independent and

$$\text{ess sup}\, \left\|Y_{r,n}^\varepsilon\right\|_2 \le \sqrt{\sum_{(i,j)\in P_r^n} (\text{ess sup}\,|(X_n^\varepsilon)_{ij}|)^2} \le \sqrt{d_n}\varepsilon a_n\,,$$

then Theorem 4.3 implies that X_n^ε satisfies CCP$(4, 16 d_n \varepsilon^2 a_n^2)$.

The aim now is to approximate f with a finite combination of convex functions, which will allow us to exploit CCP of X_n^ε. To that end, let Δ be small enough (to be fixed later) and (following [17]) set

$$g(x) = \begin{cases} 0 & \text{for } x \le 0, \\ x & \text{for } 0 \le x \le \Delta, \\ \Delta & \text{otherwise.} \end{cases}$$

Note that g is a difference of two convex functions. Define recursively $g_0 \equiv 0$,

$$g_{n+1}(x) = \begin{cases} g_n(x) + g(x + M - n\Delta) & \text{if } f(-M + (n+1)\Delta) \geq g_n(-M + n\Delta), \\ g_n(x) - g(x + M - n\Delta) & \text{otherwise} \end{cases}$$

and set $f_\Delta = g_{\lceil 2M/\Delta \rceil}$. Observe that $\|f - f_\Delta\|_\infty \leq \Delta$ and f_Δ can be decomposed into a sum of at most $2\lceil 2M/\Delta \rceil =: \kappa$ different convex and concave 1-Lipschitz functions $\{h_l\}$. Set $\Delta < t/6$. Exploiting Corollary 4.5, Lemma 4.6 and CCP of X_n^ε (replacing median by mean – cf. Remark 4.2) results in

$$\begin{aligned} \mathbb{P}\left(\left| \int f \, dL_n^\varepsilon - \mathbb{E} \int f \, dL_n^\varepsilon \right| > \frac{t}{3} \right) &\leq \mathbb{P}\left(\left| \int f_\Delta \, dL_n^\varepsilon - \mathbb{E} \int f_\Delta \, dL_n^\varepsilon \right| > \frac{t - 6\Delta}{3} \right) \\ &\leq \kappa \sup_l \mathbb{P}\left(\left| \int h_l \, dL_n^\varepsilon - \mathbb{E} \int h_l \, dL_n^\varepsilon \right| > \frac{t - 6\Delta}{3\kappa} \right) \\ &\leq 4\kappa \exp\left(-\frac{(t - 6\Delta)^2}{144\widetilde{c}\kappa^2} \frac{n^2}{d_n a_n^2 \varepsilon^2} \right), \end{aligned}$$

where $\widetilde{c}$ is some universal constant (cf. Remark 4.2). Fix ε s.t. for all n the above quantity is smaller than $\frac{\delta}{2}$ – it is possible to do so since $d_n = O(n^2/a_n)$ implies that $n^2/(d_n a_n)$ is bounded away from zero.

By Corollary 4.5 and since $(X_n)_{n \in \mathbb{N}} \in \mathcal{L}(a_n)$,

$$\begin{aligned} \mathbb{P}\left(\left| \int f \, dL_n - \int f \, dL_n^\varepsilon \right| > \frac{t}{3} \right) &\leq \mathbb{P}\left(\|X_n - X_n^\varepsilon\|_{HS} > \frac{tn}{3} \right) \\ &= \mathbb{P}\left(\frac{1}{n^2} \sum_{i,j=1}^n |(X_n)_{ij}|^2 \mathbb{1}_{\{|(X_n)_{ij}| > \varepsilon a_n\}} > \frac{t^2}{9} \right) \leq \frac{\delta}{2} \end{aligned}$$

for n large enough. Moreover, since the last quantity actually converges to zero with n (for any ε and t), we have $\int f \, dL_n - \int f \, dL_n^\varepsilon \to_{\mathbb{P}} 0$, whence boundedness of f implies that $\mathbb{E} \left| \int f \, dL_n - \int f \, dL_n^\varepsilon \right| \leq t/3$ for n large enough.

For n such that all the above estimates hold, the triangle inequality yields

$$\begin{aligned} \mathbb{P}\left(\left| \int f \, dL_n - \mathbb{E} \int f \, dL_n \right| > t \right) &\leq \mathbb{P}\left(\left| \int f \, dL_n - \int f \, dL_n^\varepsilon \right| > \frac{t}{3} \right) \\ &\quad + \mathbb{P}\left(\left| \int f \, dL_n^\varepsilon - \mathbb{E} \int f \, dL_n^\varepsilon \right| > \frac{t}{3} \right) \leq \delta \, . \end{aligned}$$

Since f was chosen arbitrarily from $C_c \cap C_L$, Proposition 2.8 yields the proof. □

Proof of Theorem 2.6. It can be easily checked that $(X_n)_{n \in \mathbb{N}} \in \mathcal{L}$ implies $(X_n)_{n \in \mathbb{N}} \in \mathcal{L}(a_n)$ for any sequence $a_n \to \infty$. Taking $a_n^2 = n^2/d_n$ we have trivially $a_n \to \infty$ and $d_n = O(n^2/a_n^2)$, whence X_n satisfies the assumptions of Theorem 2.12. Finally, Propositions 2.7 and 2.8 allow us to conclude the proof. □

Proof of Theorem 2.10. The proof boils down to the use of the so-called hermitization technique. Consider the matrix

$$A_n = \begin{bmatrix} 0 & X_n \\ X_n^* & 0 \end{bmatrix}.$$

Clearly, the ESD of Y_n can be inferred from the ESD of A_n. Moreover, A_n meets the assumptions of Theorem 2.6, which yields the result. □

Proof of Proposition 3.3

To prove Proposition 3.3, we need the following auxiliary fact.

Remark 5.1. If x is in the domain of attraction of the Gaussian distribution, b_n's are defined as in (3.1) and $l(t) = \mathbb{E}\,|x|^2\,\mathbb{1}_{\{|x|\le t\}}$, then

$$\lim_{n\to\infty}\frac{nl(b_n)}{b_n^2} = 1\,, \quad \mathbb{P}\left(|x| > t\right) = o(l(t)/t^2) \quad \text{and} \quad \mathbb{E}\,|x|\,\mathbb{1}_{\{|x|>t\}} = o(l(t)/t)\,.$$

The first equality follows easily from the definitions of b_n and l, while to prove the remaining equalities one has to make use of the fact that l is slowly varying – see [2, Proof of Corollary 2.10].

In what follows, set $(\widetilde{X}_n)_{ij} = (X_n)_{ij}\mathbb{1}_{\{|(X_n)_{ij}|\le b_n\}}$.

Lemma 5.2. *If (X_n) satisfies the assumptions of Proposition 3.3, then $\mathbb{E}L_n^{X_n/b_n}$ is tight.*

Proof of Lemma 5.2. Using Lemma 4.8 with $r = 1$ and denoting rows of X_n by $(X_n)_i$, we arrive at

$$\mathbb{E}\int |x|\;dL_n^{X_n/b_n} \le \frac{1}{nb_n}\mathbb{E}\left[\sum_{1\le i\le n}\left\|(\widetilde{X}_n)_i\right\|_2 + \left\|(X_n)_i - (\widetilde{X}_n)_i\right\|_2\right].$$

Applying arithmetic vs quadratic mean and Jensen's inequalities together with the first item of Remark 5.1 yield

$$\frac{1}{nb_n}\mathbb{E}\sum_{1\le i\le n}\left\|(\widetilde{X}_n)_i\right\|_2 \le \frac{1}{\sqrt{n}b_n}\mathbb{E}\left\|\widetilde{X}_n\right\|_{HS} \le \frac{1}{\sqrt{n}b_n}\sqrt{\mathbb{E}\sum_{ij}\left|(\widetilde{X}_n)_{ij}\right|^2} \le \frac{\sqrt{nl(b_n)}}{b_n} = O(1)\,,$$

whereas the norm inequality $\|\cdot\|_2 \le \|\cdot\|_1$ and Remark 5.1 give

$$\frac{1}{nb_n}\mathbb{E}\sum_{1\le i\le n}\left\|(X_n)_i - (\widetilde{X}_n)_i\right\|_2 \le \frac{1}{nb_n}\mathbb{E}\sum_{1\le i\le n}\left\|(X_n)_i - (\widetilde{X}_n)_i\right\|_1 = \frac{n}{b_n}\mathbb{E}\,|x|\,\mathbb{1}_{\{|x|>b_n\}} = o(1)\,.$$

The above estimates provide a uniform upper bound on the first moment of $\mathbb{E}L_n^{X_n/b_n}$, whence the conclusion follows. □

Proof of Proposition 3.3. Denote by L_n and $\widetilde{L}_n$ the ESDs of $b_n^{-1}X_n$ and $b_n^{-1}\widetilde{X}_n$ respectively. Recall that Kolomogorov's metric defined as the sup distance between cumulative distribution functions dominates Lévy–Prokhorov's metric defined as

$$\pi(\mu,\nu) = \inf\{\,\varepsilon > 0\colon\; \forall\, t\in\mathbb{R}\quad F_\nu(t-\varepsilon) - \varepsilon \le F_\mu(t) \le F_\nu(t+\varepsilon) + \varepsilon\,\}\,,$$

where F_σ denotes the c.d.f. of a measure σ. Note that the latter metrizes weak convergence of probability measures (cf. [11, Theorem 6.8]). By Lemma 4.7 and the second item of Remark 5.1

$$\begin{aligned}\mathbb{E}\pi(L_n,\widetilde{L}_n) \le \mathbb{E}\left\|L_n - \widetilde{L}_n\right\| \le \mathbb{E}\frac{1}{n}\operatorname{rank}(X_n - \widetilde{X}_n) &\le \mathbb{E}\frac{1}{n}\sum_{1\le i,j\le n}\mathbb{1}_{\{|(X_n)_{ij}|>b_n\}}\\ &= n\mathbb{P}\left(|x| > b_n\right) = o(1)\end{aligned}$$

and thus we reduced the problem to proving the convergence of $\widetilde{L}_n$. We achieve this goal by showing that $\frac{\sqrt{n}}{b_n}\widetilde{X}_n$ falls into the regime of Theorem 2.12. Take $a_n = \sqrt{n}$. Applying

Markov's inequality gives

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{n^2}\sum_{i,j=1}^{n}|\frac{\sqrt{n}}{b_n}(\widetilde{X}_n)_{ij}|^2\mathbb{1}_{\{\frac{\sqrt{n}}{b_n}|(\widetilde{X}_n)_{ij}|>\varepsilon a_n\}}>\varepsilon\right)\\
&=\mathbb{P}\left(\frac{1}{n^2}\sum_{i,j=1}^{n}|(X_n)_{ij}|^2\mathbb{1}_{\{b_n\geq|(X_n)_{ij}|>\varepsilon b_n\}}>\frac{\varepsilon b_n^2}{n}\right)\\
&\leq\frac{nl(b_n)}{\varepsilon b_n^2}\cdot\frac{l(b_n)-l(\varepsilon b_n)}{l(b_n)}=o(1)
\end{aligned}$$

by the first item of Remark 5.1 and the fact that l is slowly varying. Whence $(\frac{\sqrt{n}}{b_n}\widetilde{X}_n)_{n\in\mathbb{N}}\in\mathcal{L}(\sqrt{n})$. It remains to apply Theorem 2.12, Lemma 5.2 and Proposition 2.8. □

Remark 5.3. As mentioned previously, Proposition 3.3 provides smaller d_n compared to Theorem 2.12. It seems this cannot be improved via the above reasoning. Indeed, since $d_n=O(n^2/a_n)$, obtaining d_n of greater order than $O(n)$ by the means of the above argument would require an estimate of the form

$$\frac{l(b_n)-l(\varepsilon b_n\frac{a_n}{\sqrt{n}})}{l(b_n)}=o(1)\quad\text{for}\quad a_n=o(\sqrt{n}),$$

which does not need to hold since l can by any slowly varying function. It is unclear to us what is the optimal order of d_n in this case.

References

[1] R. Adamczak, *On the Marchenko-Pastur and circular laws for some classes of random matrices with dependent entries*, Electron. J. Probab. **16** (2011), no. 37, 1068–1095. MR-2820070

[2] R. Adamczak, *Some remarks on the Dozier-Silverstein theorem for random matrices with dependent entries*, Random Matrices Theory Appl. **2** (2013), no. 2, 1250017, 46. MR-3077829

[3] R. Adamczak, *Random non-Abelian G-circulant matrices. Spectrum of random convolution operators on large finite groups*, arXiv e-prints (2017), arXiv:1712.04772.

[4] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR-2760897

[5] G. W. Anderson and O. Zeitouni, *A CLT for a band matrix model*, Probab. Theory Related Fields **134** (2006), no. 2, 283–338. MR-2222385

[6] A. Araujo and E. Giné, *The central limit theorem for real and Banach valued random variables*, John Wiley & Sons, New York-Chichester-Brisbane, 1980, Wiley Series in Probability and Mathematical Statistics. MR-0576407

[7] Z. D. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677, With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author. MR-1711663

[8] Z. D. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR-2567175

[9] J. Bellissard, *Random matrix theory and the Anderson model*, J. Statist. Phys. **116** (2004), no. 1-4, 739–754. MR-2082189

[10] J. Bellissard, J. Magnen, and V. Rivasseau, *Supersymmetric analysis of a simplified two-dimensional Anderson model at small disorder*, Markov Process. Related Fields **9** (2003), no. 2, 261–278, Inhomogeneous random systems (Cergy-Pontoise, 2002). MR-2013663

[11] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, Inc., New York, 1999, A Wiley-Interscience Publication. MR-1700749

[12] A. Bose, R. Subhra H., and K. Saha, *Limiting spectral distribution of circulant type matrices with dependent inputs*, Electron. J. Probab. **14** (2009), no. 86, 2463–2491. MR-2563248

[13] A. Bose, R. Subhra H., and K. Saha, *Patterned random matrices and method of moments*, Proceedings of the International Congress of Mathematicians. Volume IV, Hindustan Book Agency, New Delhi, 2010, pp. 2203–2231. MR-2827968

[14] E. Brézin, C. Itzykson, G. Parisi, and J. B. Zuber, *Planar diagrams*, Comm. Math. Phys. **59** (1978), no. 1, 35–51. MR-0471676

[15] W. Bryc, A. Dembo, and T. Jiang, *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, Ann. Probab. **34** (2006), no. 1, 1–38. MR-2206341

[16] B. Delyon, *Concentration inequalities for the spectral measure of random matrices*, Electron. Commun. Probab. **15** (2010), 549–561. MR-2737712

[17] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab. **5** (2000), 119–136. MR-1781846

[18] A. Guntuboyina and H. Leeb, *Concentration of the spectral measure of large Wishart matrices with dependent entries*, Electron. Commun. Probab. **14** (2009), 334–342. MR-2535081

[19] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J. **20** (1953), 37–39. MR-0052379

[20] K. Hofmann-Credner and M. Stolz, *Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices*, Electron. Commun. Probab. **13** (2008), 401–414. MR-2415147

[21] 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. MR-0322926

[22] O. Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002. MR-1876169

[23] T. Kemp and D. Zimmermann, *Random Matrices with Log-Range Correlations, and Log-Sobolev Inequalities*, arXiv e-prints (2014), arXiv:1405.2581.

[24] E. G. Larsson and P. Stoica, *Space-time block coding for wireless communications*, Cambridge University Press, Cambridge, 2003. MR-1992873

[25] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347

[26] A. Marshall, Olkin I., and B. C. Arnold, *Inequalities: theory of majorization and its applications*, second ed., Springer Series in Statistics, Springer, New York, 2011. MR-2759813

[27] M. W. Meckes, *Concentration of norms and eigenvalues of random matrices*, J. Funct. Anal. **211** (2004), no. 2, 508–524. MR-2057479

[28] M. W. Meckes, *Some results on random circulant matrices*, High dimensional probability V: the Luminy volume, Inst. Math. Stat. (IMS) Collect., vol. 5, Inst. Math. Statist., Beachwood, OH, 2009, pp. 213–223. MR-2797949

[29] M. W. Meckes, *The spectra of random abelian G-circulant matrices*, ALEA Lat. Am. J. Probab. Math. Stat. **9** (2012), no. 2, 435–450. MR-3069372

[30] M. W. Meckes and S. J. Szarek, *Concentration for noncommutative polynomials in random matrices*, Proc. Amer. Math. Soc. **140** (2012), no. 5, 1803–1813. MR-2869165

[31] S. O'Rourke, *A note on the Marchenko-Pastur law for a class of random matrices with dependent entries*, Electron. Commun. Probab. **17** (2012), no. 28, 13. MR-2955493

[32] K. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, vol. 171, American Mathematical Society, Providence, RI, 2011. MR-2808038

[33] G. Poirot, *Mean Green's function of the Anderson model at weak disorder with an infra-red cut-off,* Ann. Inst. H. Poincaré Phys. Théor. **70** (1999), no. 1, 101–146. MR-1671211

[34] R. Rashidi F., T. Oraby, W. Bryc, and R. Speicher, *On slow-fading MIMO systems with nonseparable correlation*, IEEE Trans. Inform. Theory **54** (2008), no. 2, 544–553. MR-2444540

[35] J. H. Schenker and H. Schulz-Baldes, *Semicircle law and freeness for random matrices with symmetries or correlations*, Math. Res. Lett. **12** (2005), no. 4, 531–542. MR-2155229

[36] D. Shlyakhtenko, *Random Gaussian band matrices and freeness with amalgamation*, Internat. Math. Res. Notices (1996), no. 20, 1013–1025. MR-1422374

[37] G. 't Hooft, *Magnetic monopoles in unified gauge theories*, Nuclear Phys. **B79** (1974), 276–284. MR-0413809

[38] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR-1361756

[39] T. Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. MR-2906465

[40] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564. MR-0077805

[41] X. Zhan, *Matrix inequalities*, Lecture Notes in Mathematics, vol. 1790, Springer-Verlag, Berlin, 2002. MR-1927396

[42] W. Zhou, *New estimators of spectral distributions of Wigner matrices*, J. Math. Phys. **54** (2013), no. 3, 033503, 10.

Acknowledgments. The results presented in this article are a part of the master's thesis written at the University of Warsaw under supervision of Radosław Adamczak, who suggested the problem, gave the theoretical tools to solve it and guided me patiently through the whole process of creating a publishable paper, for which I am sincerely grateful.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm