INTRODUCTION TO LARGE RANDOM MATRICES

FRANÇOIS CHAPON

University of Science, Ho Chi Minh City, Vietnam, 2025

Contents

1. Introduction

Random matrix theory (RMT) is a branch of mathematics that studies matrices whose entries are random variables. It has become an essential tool in various fields, such as physics, statistics, and computer science, especially in situations where data sets are high-dimensional. The theory provides deep insights into the spectral properties of large random matrices and has direct implications for data analysis, particularly when the number of data samples and the dimensionality of data points both become large.

Given an i.i.d. sample $X_1, \dots, X_n$ of random vectors in $\mathbb{R}^p$, then the classical estimators of the mean (which is vector in $\mathbb{R}^p$) and of the covariance (which is a matrix in $M_p(\mathbb{R})$) are respectively given by:

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} X_k,$$

and

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} (X_k - \hat{\mu})(X_k - \hat{\mu})^\intercal,$$

where $A^\intercal$ denotes the transpose of a matrix A. To simplify, assume that the data X_i's are i.i.d. with zero mean and variance one, so that $\hat{\mu} = 0$. The estimator of the covariance can be taken as

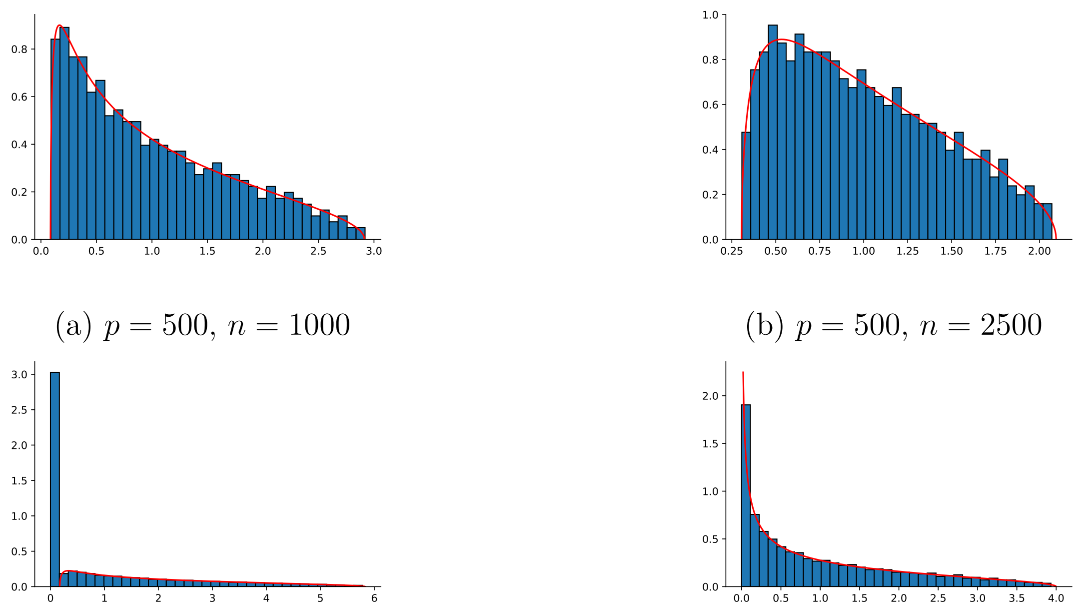
$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^{n} X_k X_k^\intercal = \frac{1}{n} Y Y^\intercal,$$

where Y is the $p \times n$ matrix whose columns are $X_1, \ldots, X_n$. Because of the law of large numbers, for fixed p, $\hat{\Sigma} \to I_p$ almost surely, as $n \to \infty$, so $\hat{\Sigma}$ is a consistent estimator of the true covariance matrix, corresponding to the identity matrix I_p in this case.

Now suppose that p is comparable to n, so that $\frac{p}{n} \to c > 0$. Naively, one would think that the empirical spectral distribution of $\frac{1}{n} Y Y^{\intercal}$,

$$\mu_n = \frac{1}{n} \sum_{\lambda \in \mathrm{Sp}(\frac{1}{n} Y Y^{\intercal})} \delta_\lambda,$$

would converge, as $n \to \infty$, to the Dirac mass δ_1. However, this is not the case, and a plot of the histogram of the eigenvalues of $\frac{1}{n} Y Y^{\intercal}$ gives the following pictures:



(a) $p = 500$, $n = 1000$

(b) $p = 500$, $n = 2500$

(c) $p = 2000$, $n = 1000$

(d) $p = 1000$, $n = 1000$

FIGURE 1. Histogram of the eigenvalues of $\frac{1}{n} Y Y^{\intercal}$ for different values of p and n. Note that when $p > n$, the matrix $\frac{1}{n} Y Y^{\intercal}$ is singular with $p - n$ eigenvalues equal to 0.

This is the content of the Marchenko-Pastur theorem, as we will see in Section 4. This phenomenon is sometimes referred as the "curse of dimensionality" and random matrix theory has developed a broad spectrum of tools to understand this phenomenon.

Random matrix theory was first introduced in multivariate statistics in the thirties by Wishart [22] and in theoretical physics in the fifties by Wigner in his fundamental article [21].

The aim of this lectures is thus to give an introduction to the most standard results in random matrix theory, but also to present different techniques commonly used in the field, such as the combinatorics of the moment method, and Stieltjes transform and resolvent method.

We will first focus on the Wigner theorem as it is more easily proved, but the techniques involved are quite similar in the case of other models.

Let X_n be a $n \times n$ Hermitian matrix with independent coefficients (these matrix models are known as Wigner matrices). The spectral distribution of X_n is defined by

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

where λ_i, $1 \leq i \leq n$ are the (random) eigenvalues of X_n. Then, the Wigner theorem asserts that the measure μ_{X_n} converges, as the size n of the matrix goes to infinity, towards the Wigner semicircular distribution $\frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}(x)dx$. This is the macroscopic regime, that is, we look at the convergence of $\mu_{X_n}(B)$, for a Borel set B of fixed size, and holds on some minimal assumptions on the coefficients.

This lecture is organized as follows. Chapter 2 presents Wigner theorem, that is the convergence of the spectral distribution of Wigner matrices towards the semicircular distribution. The proof is achieved by the computation of the moments of the spectral distribution of Wigner matrices via combinatorics methods. In Chapter 3, we present a second proof of Wigner theorem, only in the case of the Gaussian Unitary Ensemble, using the Stieltjes transform, which is, as the more commonly used characteristic function, a functional of the measure which characterizes the weak convergence of measures. Some standard complex analysis tools will be also used and will be recalled. In chapter 4, we will state the main results concerning sample covariance matrices and the Marchenko-Pastur theorem and chapter 5 will present a few results concerning random perturbations of finite rank matrices. At last, Chapter 6 is an appendix where we recall some complex analysis tools, and some useful matrix inequalities.

Here are some notations that we are going to use in the sequel.

- $\mathcal{H}_n$ is the space of Hermitian $n \times n$ matrices.
- The coefficients of a matrix $A \in \mathrm{M}_n(\mathbb{C})$ are denoted $A(i,j)$ or A_{ij}, for $1 \leq i,j \leq n$.
- To simplify notation, we frequently omit the dependence on the matrix dimension.
- The cardinal of a set A is denoted either $\#A$ or $|A|$.

2. The Wigner theorem

Random matrix theory has been widely developed since Wigner's work in the fifties [21]. In quantum theory, energy levels are given by the eigenvalues of a Hermitian operator on some Hilbert space, the so-called system Hamiltonian. The study of such systems can become very tricky when the dimension becomes large. Wigner's idea was then to modelize such systems by random Hermitian matrices of large dimension. We first describe the matrix models that we are going to study.

Definition 2.1. *Let $X_n \in \mathcal{H}_n$ be a random $n \times n$ Hermitian matrix such that $(X_n(i,j))_{1\leq i\leq j\leq n}$ are independent random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\mathbb{E}(X_n(i,j)) = 0$. Such matrix models with independent coefficients are called* **Wigner matrices***.*

One of the most important model of Wigner matrices is the following.

Definition 2.2. *A Wigner matrix X_n is said to be from the* **Gaussian Unitary Ensemble (GUE)** *if*

$$X_{ii}, i = 1, \dots, n, \sqrt{2}\Re X_{ij}, \sqrt{2}\Im X_{ij}, 1 \leq i < j \leq n$$

are independent random variables, distributed according to the standard normal distribution $\mathcal{N}(0,1)$.

The GUE(n,σ^2) distribution is defined as the Gaussian distribution on $\mathcal{H}_n$ defined by

$$(2\pi\sigma^2)^{-n^2/2}\exp\Big(-\frac{1}{2\sigma^2}\operatorname{Tr}(M^2)\Big)dM$$

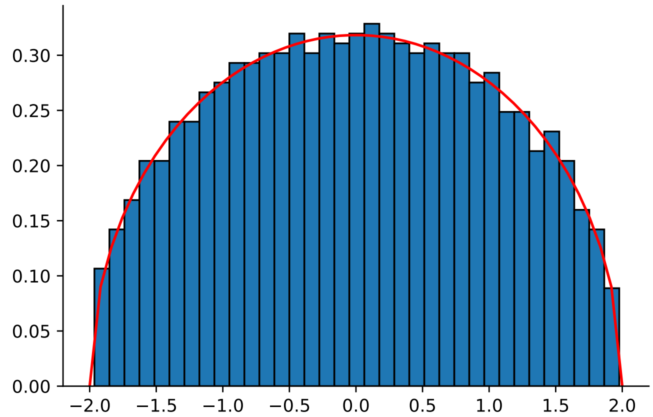


FIGURE 2. Histogram (blue) of the eigenvalues of a 1000×1000 GUE matrix and the semicircular distribution (red).

where dM is Lebesgue measure on $\mathcal{H}_n$ defined by

$$dM = \prod_{i=1}^{n} dM_{ii} \prod_{1 \le i < j \le n} d\sqrt{2}\Re M_{ij} d\sqrt{2}\Im M_{ij}.$$

We will abbreviate GUE(n, σ^2) by GUE when $\sigma^2 = 1$ and when the dimension n is clear from the context.

Note that in the above, we have identify the inner product space $\mathcal{H}_n$ with $\mathbb{R}^{n^2}$, the identification being:

$$X \in \mathcal{H}_n \leftrightarrow (X_{ii}, 1 \le i \le n, \sqrt{2}\Re X_{ij}, \sqrt{2}\Im X_{i,j}, 1 \le i < j \le n) \in \mathbb{R}^{n^2}$$

the inner product on $\mathcal{H}_n$ being:

$$\begin{aligned} \langle A, B \rangle = \operatorname{Tr}(AB^*) &= \sum_{i,j=1}^{n} A_{ij} B_{ji} \\ &= \sum_{i=1}^{n} A_{ii} B_{ii} + \sum_{1 \le i < j \le n} \left(\sqrt{2}\Re A_{ij} \sqrt{2}\Re B_{ij} + \sqrt{2}\Im A_{ij} \sqrt{2}\Im B_{ij} \right) \end{aligned}$$

It is easy to see, using $\operatorname{Tr}(M^2) = \operatorname{Tr}(MM^*) = \sum_{i=1}^{n} M_{ii}^2 + 2\sum_{1 \le i < j \le n} |M_{ij}|^2$, that a Wigner matrix from the GUE is distributed according to the GUE distribution.

Remark 2.3. The GUE distribution is invariant by unitary conjugation, that is if X is distributed according to the GUE then $UXU^* \overset{(d)}{=} X$ for all unitary matrix U. Indeed, we have

$$\operatorname{Tr}(UXU^*UX^*U^*) = \operatorname{Tr}(XX^*),$$

and one can easily see that the determinant of the change of variables $X \mapsto UXU^*$ is equal to 1, since it is an isometry.

Figure 2 shows a simulation of the eigenvalues of a large GUE matrix, where one can see the relationship with the semicircular distribution, which is the following probability measure.

Definition 2.4. *The semicircular distribution μ_{sc,σ^2} is the probability measure on $\mathbb{R}$ given by*

$$\mu_{sc,\sigma^2}(dx) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\mathbb{1}_{[-2\sigma,2\sigma]}(x)dx,$$

where $\sigma > 0$. When $\sigma^2 = 1$, we will abbreviate μ_{sc,σ^2} by μ_{sc}.

In the global regime, we are interested in the convergence of the spectral measure of Wigner matrices which is the following.

Definition 2.5. *Let $A \in \mathcal{H}_n$, with eigenvalues $\lambda_1(A), \dots, \lambda_n(A)$. The spectral mesure of A, denoted μ_A, is the probability measure defined by*

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

that is, for a Borel set $B \subset \mathbb{R}$,

$$\mu_A(B) = \frac{1}{n}\#\{1 \leq i \leq n \,|\, \lambda_i(A) \in B\}.$$

We can now state Wigner theorem.

Wigner theorem. *Let $H_n = \frac{1}{\sqrt{n}}X_n$, where X_n is a Wigner matrix such that such that $(X_n(i,j))_{1\leq i\leq j\leq n}$ are independent and identically distributed centered random variables with variance σ^2. Then, the spectral measure of H_n, μ_{H_n}, converges weakly, as n goes to infinity, towards μ_{sc,σ^2}, almost surely.*

In this section, the proof of Wigner theorem, under some additional assumptions on the moments of the coefficients will be achieved by some combinatorial interpretation of the Catalan numbers, which are, as we will see, the moments of the semicircular distribution.

2.1. **Combinatorics of Catalan numbers.**

Definition 2.6. *The Catalan numbers C_n are the numbers defined by $C_0 = 1$ and for $n \geq 1$,*

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}.$$

The sequence of Catalan numbers is $1, 1, 2, 5, 14, 42, 132, 429, \dots$. It is sequence A000108 in OEIS (the On-Line Encyclopedia of Integer Sequences: oeis.org). The Catalan numbers are ubiquitous in combinatorics. The sequence has the bigger entry in OEIS, and for instance Stanley in his book Enumerative Combinatorics vol. 2 lists 66 different combinatorial interpretations of the Catalan numbers!

First we will see in the next lemma, that the Catalan numbers are the moments of the semicircular distribution.

Lemma 2.7. *Let μ_{sc,σ^2} be the semicircular distribution, i.e.*

$$\mu_{sc,\sigma^2} = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}\mathbb{1}_{[-2\sigma,2\sigma]}(x)dx.$$

The moments of μ_{sc,σ^2} are given by

$$\int_{\mathbb{R}} x^{2n+1}\mu_{sc}(dx) = 0, \quad \int_{\mathbb{R}} x^{2n}\mu_{sc}(dx) = \sigma^{2n}C_n.$$

Proof. By parity, odd moments are clearly zero. Suppose without loss of generality that $\sigma^2 = 1$. Now,

$$m_{2n} := \int_{-2}^{2} x^{2n}\frac{1}{2\pi}\sqrt{4-x^2}dx = \frac{4}{\pi}2^{2n}\int_0^1 x^{2n}\frac{1}{\pi}\sqrt{1-x^2}dx.$$

Using the change of variables $x = \cos(\theta)$, we obtain

$$\int_0^1 x^{2n} \frac{1}{\pi}\sqrt{1-x^2}dx = \int_0^{\pi/2} \cos^{2n}\theta \sin^2\theta d\theta = \int_0^{\pi/2} \cos^{2n}\theta d\theta - \int_0^{\pi/2} \cos^{2n+2}\theta d\theta$$

This is now a classic calculation of Wallis integrals: Define

$$W_{2n} := \int_0^{\pi/2} \cos^{2n}(\theta) d\theta.$$

Using integration by parts with $U = -\frac{\cos^{2n+1}(\theta)}{2n+1}$, $U' = \cos^{2n}(\theta)\sin(\theta)$, $V = \sin(\theta)$, $V' = \cos(\theta)$, we get

$$\int_0^{\pi/2} \cos^{2n}\theta \sin^2\theta d\theta = \frac{1}{2n+1} W_{2n+2},$$

so one obtains the recurrence formula, for $n \geq 2$, ($W_0 = \pi/2$),

$$\begin{aligned} W_{2n} &= \frac{2n-1}{2n} W_{2n-2} = \frac{2n-1}{2n}\frac{2n-3}{2n-2}\cdots\frac{3}{4}\frac{\pi}{2} \\ &= \frac{2n}{2n}\frac{2n-1}{2n}\frac{2n-2}{2n-2}\frac{2n-3}{2n-2}\cdots\frac{3}{4}\frac{2}{2}\frac{\pi}{2} = \frac{(2n)!}{2^{2n}(n!)^2}\frac{\pi}{2}. \end{aligned}$$

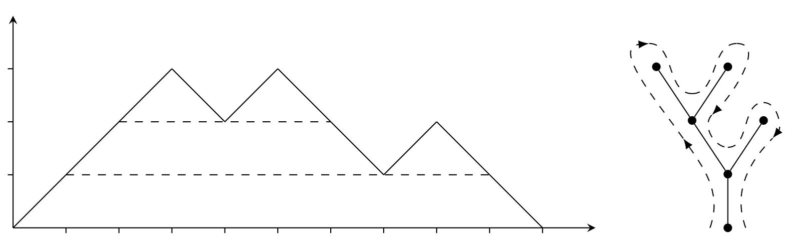
Hence,

$$m_{2n} = \frac{4}{\pi} 2^{2n} \left(\frac{(2n)!}{2^{2n}(n!)^2} - \frac{(2n+2)!}{2^{2n+2}((n+1)!)^2} \right) \frac{\pi}{2} = \frac{(2n)!}{n!(n+1)!}. \qquad \square$$

We are now going to give some well-known combinatorial interpretations of the Catalan numbers.

Definition 2.8. *A Dyck path with $2n$ steps is a nonnegative path in $\mathbb{N}^2$ starting from the origin $(0,0)$, ending at $(2n,0)$, with steps $(1,1)$ or $(1,-1)$ (also coded as UP and DOWN steps or just $+1$ and -1).*

Definition 2.9. *A graph $G = (V,E)$ is a set of vertices V and a set of edges E where an edge "links" two vertices. A tree is a connected graph with no cycles, where a cycle is a path connecting the same vertex. A root is a marked vertex. A tree is oriented if it is embedded in the plane, it inherits the orientation of the plane.*



(a) (b)

FIGURE 3. Fig. (a): a Dyck path with 10 steps. Fig. (b): the corresponding rooted plane tree. The dashed line corresponds to the walk that surrounds the tree.

Lemma 2.10. *The set of Dyck paths with $2n$ steps is in bijection with the set of rooted oriented trees with n edges.*

Proof. It is worth to take a look at Figure 3 while reading the proof. We start by replacing the tree by a "fat tree", that is every edge is replaced by a double edge. The union of these double edges define a path that surrounds the tree. To define a Dyck path, we start from the root, add

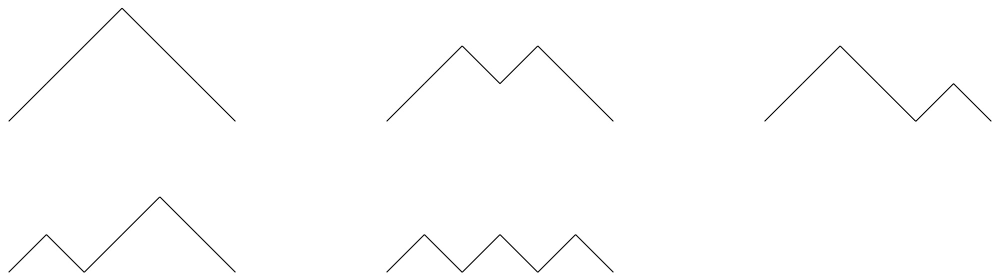


FIGURE 4. The 5 Dyck paths with 6 steps.

a "$+1$" when we meet an edge that has not been visited yet, and a "-1" otherwise. Since to add a -1, we must have already added a $+1$ corresponding to the first visit of the edge, the path is nonnegative, that is above the real axis, and since all edges are visited exactly twice, the path comes back at 0 after $2n$ steps. This defines a Dyck path.

Given a Dyck path, we can recover the rooted oriented tree by first gluing the couples of steps where one step $+1$ is followed by a step -1, and representing each couple of glued steps by one edge. We obtain a path "decorated" with edges. Continuing the same procedure until all steps have been glued two by two provides a rooted oriented tree. $\square$

Figures 4 and 5 show respectively the examples of the 5 Dyck paths of length 6 and the corresponding 5 ordered trees.

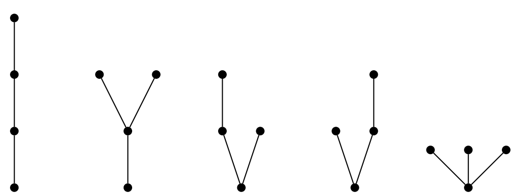


FIGURE 5. The 5 rooted plane trees with 3 edges (the root being the most bottom node).

Lemma 2.11. *The number of Dyck paths with $2n$ steps is equal to the Catalan number C_n.*

Proof. The number of Dyck paths is easily counted using the reflection principle. We have that

$$\#\{\text{all paths from } (0,0) \text{ to } (2n,0) \text{ with steps } \pm 1\} = \#\{\text{"good" paths}\} + \#\{\text{"bad" paths}\},$$

where the "good" paths are the Dyck paths with $2n$ steps, and the "bad" paths are ± 1 paths from $(0,0)$ to $(2n,0)$ that are not Dyck paths. A "bad" path must cross the x-axis, hence must hit the line $y=-1$. That's where we use the reflection principle, see Fig. 6 for an example: we reflect the path after the first hitting time of -1. We obtain a path from $(0,0)$ to $(2n,-2)$, and this gives a bijection between "bad" paths and paths from $(0,0)$ to $(2n,-2)$. Since the number of all paths from $(0,0)$ to $(2n,0)$ is given by $\binom{2n}{n}$ (since there is $\binom{2n}{n}$ choices for the $+1$ steps), and the number of paths from $(0,0)$ to $(2n,-2)$ is $\binom{2n}{n-1}$, we get:

$$\#\{\text{Dyck paths with } 2n \text{ steps}\} = \binom{2n}{n} - \binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}. \qquad \square$$

Proposition 2.12. *The generating function of the Catalan numbers $(C_n)_{n\geq 0}$ is given by*

$$S(x) := \sum_{n\geq 0} C_n x^n = \frac{1-\sqrt{1-4x}}{2x}.$$

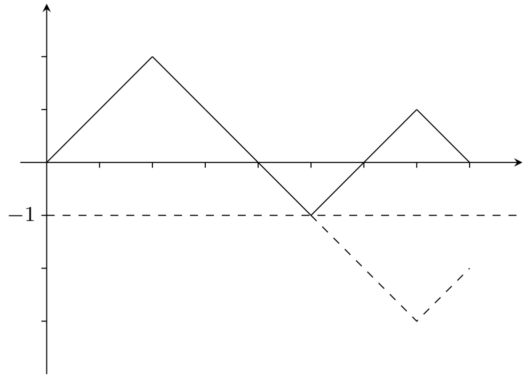


FIGURE 6. The reflection principle. A "bad" path and its reflection (dashed line) after the first hitting time of -1.

Moreover, the Catalan numbers satisfy the recurrence relation

$$C_n = \sum_{l=1}^{n} C_{l-1} C_{n-l}, \quad \text{for all } n \geq 1,$$

with $C_0 = 1$, and this characterizes the Catalan numbers.

Proof. The usual Taylor series (for $|x| < 1$)

$$(1+x)^\alpha = 1 + \sum_{k \geq 1} \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!} x^k$$

yields that

$$\frac{1-\sqrt{1-4x}}{2x} = \sum_{k \geq 0} \frac{(2k)!}{k!(k+1)!} x^k.$$

Let $C_{n,l}$ be the number of Dyck paths with $2n$ steps hitting the real axis for the first time after $2l$ steps. Then obviously we have $C_n = \sum_{l=1}^n C_{n,l}$. But it is easy to see that

$$\begin{aligned} C_{n,l} = \#\{&\text{Dyck paths from } (0,0) \text{ to } (2l,0) \text{ strictly above the real axis (except at both endpoints)}\} \\ &\times \#\{\text{Dyck paths from } (2l,0) \text{ to } (2n,0)\}. \end{aligned}$$

By shifting $2l$ to 0, we have that $\#\{\text{Dyck paths from } (2l,0) \text{ to } (2n,0)\} = C_{n-l}$. Now let a Dyck path from $(0,0)$ to $(2l,0)$ strictly above the real axis. Since the first and last steps are prescribed and equal respectively to $+1$ and -1, by shifting the real axis by $(1,1)$, we get that

$$\#\{\text{Dyck paths from } (0,0) \text{ to } (2l,0) \text{ strictly above the real axis}\} = C_{l-1}.$$

Hence,

$$C_n = \sum_{l=1}^{n} C_{l-1} C_{n-l}.$$

We now show that this recurrence relation characterizes the Catalan numbers. Indeed, suppose that $(D_n)_{n \geq 0}$ are numbers such that $D_0 = 1$ and $D_n = \sum_{l=1}^n D_{l-1} D_{n-l}$. Consider the generating function of the D_n's:

$$G(x) = \sum_{k \geq 0} D_k x^k.$$

Then the recurrence relation gives

$$\begin{aligned} G(x) &= 1 + \sum_{k\geq 1}\sum_{l=1}^{k} D_{l-1}D_{k-l}x^k \\ &= 1 + \sum_{l\geq 1}\left(D_{l-1}x^{l-1}\sum_{k\geq l} D_{k-l}x^{k-l+1}\right) \\ &= 1 + x(G(x))^2. \end{aligned}$$

Thus, we get that $G(x) = \frac{1-\sqrt{1-4x}}{2x}$, the minus branch being determined by the fact that $G(0) = 1$. Hence, we obtain that $G(x) = S(x)$, hence $D_n = C_n$, for all $n \geq 0$. $\square$

The following lemma will be used later in the proof of Wigner theorem.

Lemma 2.13. *Let $G = (V, E)$ a connected graph. Then,*

$$|V| \leq |E| + 1,$$

and equality holds if and only if G is a tree.

Proof. Suppose first that $G = (V, E)$ is a tree. Then it is easy to see that $|E| = |V| - 1$ by induction on $|V|$.

Now let $G = (V, E)$ be a connected graph. A *spanning tree* is a subgraph of G which is a tree and has the same set of vertices V than G (see Fig. 7). Note that it exists (but is non-unique), as a maximal element of the finite partially ordered set (for the inclusion) of subgraphs of G that are trees. So denote by $T = (V, E')$ a spanning tree of G. Then one has $|E'| = |V| - 1$, and since $|E| \geq |E'|$, we get $|E| \geq |V| - 1$.

Now, if $|E| = |V| - 1$, then $|E| = |E'|$ so $E = E'$ and $G = T$, hence a G is a tree. $\square$

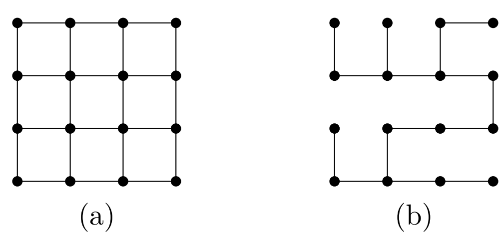


FIGURE 7. A graph (a) and one of its spanning tree (b).

2.2. **Wigner theorem.** Let X_n be a Wigner matrix, that is $X_n = (X_n(i,j))_{1\leq i,j\leq n}$ is a $n \times n$ Hermitian random matrix defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the coefficients $(X_n(i,j))_{1\leq i\leq j\leq n}$ are independent random variables with

$$\mathbb{E}(X_{ij}) = 0, \quad \text{and} \quad \mathbb{E}(|X_{ij}|^2) = \sigma^2.$$

We will prove Wigner theorem, under the additional assumption that the coefficients have bounded moments of all order.

Theorem 2.14. *Assume that for all $k \geq 0$*

$$\sup_n \sup_{1\leq i\leq j\leq n} \mathbb{E}(|X_n(i,j)|^k) < \infty.$$

Let $H_n = \frac{1}{\sqrt{n}}X_n$. Then we have,

$$\lim_{n\to\infty} \frac{1}{n}\operatorname{Tr}(H_n^k) = \begin{cases} 0, & \textit{if } k \textit{ is odd}, \\ \sigma^k C_{k/2}, & \textit{if } k \textit{ is even}, \end{cases}$$

where the convergence holds in expectation and almost surely, and where the C_k's are the Catalan numbers.

Proof. Without loss of generality we can suppose that $\sigma^2 = 1$. We first prove the convergence in expectation. We drop the dependance in n in all matrix notations to simplify the readability. We have that,

$$\mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(H^k)\Big) = \frac{1}{n}\mathbb{E}(\sum_{i_1,\dots,i_k=1}^{n} H_{i_1i_2}H_{i_2i_3}\cdots H_{i_ki_1})$$
$$= \frac{1}{n^{k/2+1}}\sum_{i_1,\dots,i_k=1}^{n}\mathbb{E}(X_{i_1i_2}X_{i_2i_3}\cdots X_{i_ki_1}). \tag{1}$$

Let $I = (i_1,\dots,i_k)$, and put $P(I) = \mathbb{E}(X_{i_1i_2}X_{i_2i_3}\cdots X_{i_ki_1})$. Then, since by assumption

$$\sup_n \sup_{i,j} \mathbb{E}(|X_{ij}|^k) < \infty,$$

we have by Hölder's inequality that

$$|P(I)| \le a_k,$$

where a_k is a constant independent of n.

The sum is indexed by k indices varying from 1 to n, but most of them give a zero contribution in the limit $n \to \infty$. Instead of indexing the sum by k-tuples, we will index the sum by walks on graphs as follows.

To I we associate the graph $G(I) = (V(I), E(I))$, where the vertices $V(I)$ are distincts elements of $i_1,\dots,i_k$, and the edges are distincts pairs among $(i_1,i_2),\dots,(i_k,i_1)$. We thus have $|V(I)| \le k$ and $|E(I)| \le k$. The set I is seen as a walk on $G(I)$ given by $i_1 \to i_2 \to \cdots \to i_k \to i_1$.

First, remark that $G(I)$ is connected since $G(I)$ is explored by the walk I. Moreover, from the independence and centering of the entries, we have

$$P(I) = 0,$$

unless to any edge (i_p, i_{p+1}) (with the convention that $i_{k+1} = i_1$) there exists $l \neq p$ such that $(i_p, i_{p+1}) = (i_l, i_{l+1})$ or (i_{l+1}, i_l), since a single edge gives a zero contribution. We next show that the set of indices I giving a non zero contribution is described by trees.

Let I such that $P(I) > 0$. Then each edge must be visited by the walk at least twice. Hence, we get that $|E(I)| \le \lfloor k/2 \rfloor$ and by lemma 2.13, we have $|V(I)| \le \lfloor k/2 \rfloor + 1$. We may write

$$\mathbb{E}\left(\frac{1}{n}\operatorname{Tr}(H^k)\right) = \frac{1}{n^{k/2+1}}\sum_I P(I)$$

Since indices vary from 1 to n, there are at most $n^{\lfloor k/2 \rfloor + 1}$ indices contributing to the sum (1), so we have, for some constant $c_k > 0$,

$$\mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(H^k)\Big) \le c_k n^{\lfloor k/2 \rfloor - k/2}.$$

In particular, if k is odd, we have

$$\lim_{n\to\infty}\mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(H^k)\Big) = 0.$$

Suppose now k is even. Since the only indices I that contribute to the limit of the above sum are those for which $|V(I)|$ is exactly equal to $\frac{k}{2}+1$, Lemma 2.13 implies that $G(I)$ is a tree and

$|E(I)| = \frac{k}{2}$. We then get that each edge in $G(I)$ is visited by the walk I exactly twice, once in each direction, so for such I, we have, by independence, that

$$P(I) = \prod_{e \in E(I)} \mathbb{E}\left(|X_e|^2\right) = 1.$$

Moreover, the walk induces an unique ordering of the vertices $V(I)$, thus $G(I)$ is a rooted plane tree (the root being given by i_1). Since there are $n(n-1)\cdots(n-k/2)$ choices for the distinct $k/2+1$ vertices for the same geometry of the rooted plane tree, we get that

$$\mathbb{E}\left(\frac{1}{n}\operatorname{Tr}(H^k)\right) = \frac{n(n-1)\cdots(n-k/2)}{n^{k/2+1}} \times \#\{\text{rooted oriented trees with } k/2 \text{ edges}\}.$$

Hence, since $n(n-1)\cdots(n-k/2) \sim n^{k/2+1}$, we deduce that

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{1}{n}\operatorname{Tr}(H^k)\right) = \#\{\text{rooted oriented trees with } k/2 \text{ edges}\} = C_{k/2},$$

which proves the convergence in expectation.

To prove the almost sure convergence, we prove that the variance of $\frac{1}{n}\operatorname{Tr}(H^k)$ is of order n^{-2}, the Borel-Cantelli lemma will thus give the result.

We have,

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\frac{1}{n}\operatorname{Tr}(H^k)\right) &= \mathbb{E}\left(\left(\frac{1}{n}\operatorname{Tr}(H^k)\right)^2\right) - \left(\mathbb{E}\left(\frac{1}{n}\operatorname{Tr}(H^k)\right)\right)^2 \\ &= \frac{1}{n^{k+2}} \sum_{I,I'} \Big(P(I,I') - P(I)P(I')\Big),\end{aligned}$$

where as before $I = \{i_1, \ldots, i_k\}$, $I' = \{i'_1, \ldots, i'_k\}$, and

$$P(I,I') = \mathbb{E}(X_{i_1 i_2} \cdots X_{i_k i_1} X_{i'_1 i'_2} \cdots X_{i'_k i'_1}).$$

We now have two walks I and I', and we denote as before by $G(I)$ and $G(I')$ the corresponding graphs. We also denote by $G(I,I') = (V(I,I'), E(I,I'))$ the union of this two graphs, that is the vertex set $V(I,I')$ is the union of the two vertex sets $V(I)$ and $V(I')$, and the set of edges is the union of $E(I)$ and $E(I')$. Note that $G(I,I')$ may contain multiple edges.

To give a non zero contribution, the graphs $G(I)$ and $G(I')$ must share a common edge, otherwise, $P(I,I') = P(I)P(I')$ by independence. We may thus restrict to graphs $G(I,I')$ which are connected.

Now, $P(I,I') = 0$ unless each edge is visited at least twice by *either* of the two walks. Thus if $P(I,I') > 0$, one has $|E(I,I')| \leq k$, so $|V(I,I')| \leq k+1$ by Lemma 2.13.

This first shows that the variance is at least of order n^{-1}, since $(P(I,I') - P(I)P(I'))$ is bounded by Holder's inequality. Note that this already implies convergence in probability by Bienaymé-Tchebychev inequality.

To obtain almost sure convergence, we want to improve this bound to the order n^{-2}. We must then show that the case where $|V(I,I')| = k+1$ cannot occur. In this case, by Lemma 2.13, $G(I,I')$ is a tree, so $|E(I,I')| = k$ and each edge must be visited exactly twice, by either of the two walks. Denotes by $n_I(e)$ (resp. $n_{I'}(e)$), the number of visits of the edge e by the walk I (resp. I'). Let $e \in E(I,I')$ be the common edge of $G(I)$ and $G(I')$. Then we have $(n_I(e), n_{I'}(e)) = (2,0)$, $(0,2)$ or $(1,1)$. The first two cases are impossible since this edge is visited by I and I'. The third one is also impossible, since in that case, there is a loop in $G(I)$ and $G(I')$, which is not possible since they are trees as subgraphs of $G(I,I')$. This leads to a contradiction, hence the case $|V(I,I')| = k+1$ cannot occur.

Therefore, for all contributing indices we have $|V(I,I')| \leq k$, which implies that $\mathbb{V}\mathrm{ar}(\frac{1}{n}\operatorname{Tr}(H_n^k)) = \mathrm{O}(n^{-2})$.

Thus, Chebyshev's inequality implies that

$$\mathbb{P}\Big(\Big|\frac{1}{n}\operatorname{Tr}(H_n^k) - \mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(H_n^k)\Big)\Big| > \varepsilon\Big) \leq \frac{C}{\varepsilon^2 n^2},$$

for some constant $C > 0$, so Borel-Cantelli lemma implies that

$$\Big|\frac{1}{n}\operatorname{Tr}(H_n^k) - \mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(H_n^k)\Big)\Big| \underset{n\to\infty}{\longrightarrow} 0, \qquad \text{almost surely.}$$

This yields the result using the previous convergence in expectation. □

Theorem 2.15 (Wigner theorem)**.** *Let X_n be a Wigner matrix such that for all $k \geq 0$,*

$$\sup_n \sup_{1\leq i\leq i\leq n} \mathbb{E}(|X_n(i,j)|^k) < \infty,$$

and let $H_n = \frac{1}{\sqrt{n}} X_n$. Then, the spectral distribution of H_n, μ_{H_n}, converges weakly almost surely, as n goes to infinity, towards the semicircular distribution μ_{sc,σ^2}, that is, for all bounded continuous function f, we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\mu_{H_n}(dx) = \int_{\mathbb{R}} f(x)\mu_{sc,\sigma^2}(dx) \quad a.s. \tag{2}$$

Proof. We use a standard Weierstrass polynomial approximation argument to pass from the convergence in moments of Theorem 2.14 to the convergence (2).

Let $B > 2\sigma$ and $\delta > 0$. By Weierstrass approximation theorem, we can find a polynomial P such that

$$\sup_{x\in[-B,B]} |f(x) - P(x)| \leq \delta.$$

Then,

$$\begin{aligned}
&\Big|\int_{\mathbb{R}} f(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} f(x)\mu_{sc,\sigma^2}(dx)\Big| \\
&\qquad\leq \Big|\int_{\mathbb{R}} f(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} P(x)\mu_{H_n}(dx)\Big| \\
&\qquad\quad + \Big|\int_{\mathbb{R}} P(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} P(x)\mu_{sc,\sigma^2}(dx)\Big| \\
&\qquad\quad + \Big|\int_{\mathbb{R}} P(x)\mu_{sc,\sigma^2}(dx) - \int_{\mathbb{R}} f(x)\mu_{sc,\sigma^2}(dx)\Big| \\
&\qquad\leq 2\delta + \Big|\int_{\mathbb{R}} P(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} P(x)\mu_{sc,\sigma^2}(dx)\Big| \\
&\qquad\quad + \Big|\int_{|x|>B} f(x)\mu_{H_n}(dx) - \int_{|x|>B} P(x)\mu_{H_n}(dx)\Big|
\end{aligned}$$

where we use the fact that μ_{sc,σ^2} has support $[-2\sigma, 2\sigma]$ and $B > 2\sigma$. By the convergence in moments of Theorem 2.14, we have

$$\lim_{n\to\infty} \Big|\int_{\mathbb{R}} P(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} P(x)\mu_{sc,\sigma^2}(dx)\Big| = 0$$

Moreover, since f is bounded, if we denote by p the degree of P, we can find a constant K such that

$$\begin{aligned}
\Big|\int_{|x|>B} f(x)\mu_{H_n}(dx) - \int_{|x|>B} P(x)\mu_{H_n}(dx)\Big| &\leq K \int_{|x|>B} |x|^p \mu_{H_n}(dx) \\
&\leq KB^{-p-2q} \int_{\mathbb{R}} |x|^{2(p+q)} \mu_{H_n}(dx),
\end{aligned}$$

writing $p = 2(p+q)-(p+2q)$, for all $q \geq 0$. Hence, since $\int_{\mathbb{R}} |x|^{2(p+q)}\mu_{H_n}(dx) \to_{n\to\infty} \int_{\mathbb{R}} |x|^{2(p+q)}\mu_{sc,\sigma^2}(dx)$ using again Theorem 2.14, we have that

$$\limsup_n \left| \int_{|x|>B} f(x)\mu_{H_n}(dx) - \int_{|x|>B} P(x)\mu_{H_n}(dx) \right| \leq KB^{-p-2q}(2\sigma)^{2(p+q)}.$$

Since $B > 2\sigma$, letting q goes to infinity gives that

$$\limsup_n \left| \int_{|x|>B} f(x)\mu_{H_n}(dx) - \int_{|x|>B} P(x)\mu_{H_n}(dx) \right| = 0.$$

Finally, since δ is arbitrary, we have that

$$\limsup_n \left| \int_{\mathbb{R}} f(x)\mu_{H_n}(dx) - \int_{\mathbb{R}} f(x)\mu_{sc,\sigma^2}(dx) \right| = 0,$$

which proves the theorem. □

The condition of boundedness of the moments in Wigner's theorem can be weakened, as stated in the beginning of this section, and we refer to [1] for the proof. It relies on an approximation of the Wigner matrix H_n by a matrix with bounded coefficients.

2.3. **Noncrossing partitions.** We give in this section the following comment. A standard proof of Wigner's theorem, using the moment approach, can be done via the combinatorics of noncrossing partitions instead of that of Dyck paths and trees. We refer to [11] for a detailed proof, and only present below the definition of noncrossing partitions.

Definition 2.16. *A partition π of the set $\{1,\dots,n\}$ is called crossing if there exists (a,b,c,d) with $1 \leq a < b < c < d \leq n$ such that a,c belong to one block of π while b,d belong to another block. A partition which is not crossing is called a* noncrossing partition*.*

Figure 8 shows an example which enlightens the terminology of noncrossing. We put the points $1,\dots,n$ on the circle and draw for each block of the partition the convex polygon whose vertices are the points of the block. The partition is noncrossing if and only if the polygons do not intersect.

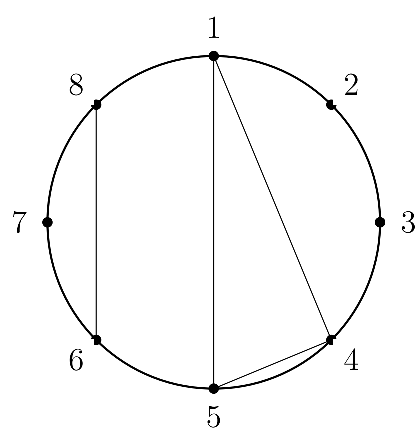


FIGURE 8. The noncrossing partition $\{1,4,5\} \cup \{2\} \cup \{3\} \cup \{6,8\} \cup \{7\}$.

Proposition 2.17. *The number of noncrossing partitions of the set $\{1,\dots,n\}$ is equal to the Catalan number C_n.*

Proof. Denote by NC_n the set of noncrossing partition of $\{1, \dots, n\}$ and let $\pi \in NC_n$. Let j the largest element of the block of π containing 1. Then, since π is noncrossing, it induces a noncrossing partition of the set $\{1, \dots, j-1\}$, and a noncrossing partition of the set $\{j+1, \dots, n\}$. Therefore, we have

$$\#NC_n = \sum_{j=1}^{n} \#NC_{j-1} \times \#NC_{n-j},$$

which characterizes, as we have already seen in the proof of Lemma 2.12, the Catalan numbers. □

3. The Stieltjes transform approach

We present in this section a second proof of Wigner theorem in the case of the Gaussian Unitary Ensemble, following the presentation of [14]. We start by recalling properties of the Stieltjes transform of a measure.

Definition 3.1. *Let m be a probability measure on $\mathbb{R}$. The Stieltjes transform of m is the function*

$$g_m(z) = \int_{\mathbb{R}} \frac{1}{x-z} m(dx),$$

defined for $z \in \mathbb{C} \setminus \mathbb{R}$ (in fact for $z \in \mathbb{C} \setminus supp(m)$).

Note that the Stieltjes transform is well defined on $\mathbb{C} \setminus \mathbb{R}$ since

$$\frac{1}{|x-z|} \leq \frac{1}{|\Im z|},$$

so that $|g_m(z)| \leq \frac{1}{|\Im z|}$, for $z \in \mathbb{C} \setminus \mathbb{R}$.

Proposition 3.2. *Let g_m be the Stieltjes transform of a probability measure m. Then the following holds.*

(i) *The function g_m is analytic on $\mathbb{C} \setminus \mathbb{R}$, and $g_m(\bar{z}) = \overline{g_m(z)}$.*
(ii) *$\Im(z)\Im(g_m(z)) > 0$ for $\Im(z) \neq 0$.*
(iii) $\lim_{y \to \infty} -iyg_m(iy) = 1$.
(iv) *If g is a function satisfying (i)-(iii), then there exists a probability measure μ such that g is the Stieltjes transform of μ.*
(v) *Inversion formula: If I is an interval such that m does not charge both endpoints, then,*

$$m(I) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_I \Im(g_m(x + i\varepsilon)) dx.$$

Proof. Parts (i)-(iii) are easy and left as an exercise.

Part (iv): Since g is analytic from $\mathbb{C}^+$ to $\mathbb{C}^+$, where $\mathbb{C}^+$ is the positive half-plane, we have using Nevanlinna's representation theorem (see Appendix Corollary 6.4),

$$g(z) = az + b + \int_{\mathbb{R}} \frac{1+uz}{u-z} \sigma(du),$$

for some constants $a, b \in \mathbb{R}$, $a \geq 0$, and σ a finite measure. Hence, for $z = iy$, one has

$$-iyg(iy) = ay^2 + \int \frac{y^2(1+u^2)}{u^2+y^2} \sigma(du) - iby - iy \int \frac{u(1-y^2)}{u^2+y^2} \sigma(du).$$

By hypothesis (iii), letting y goes to infinity yields $a = 0$ and $\int_{\mathbb{R}} (1+u^2)\sigma(du) = 1$, and $b = \int_{\mathbb{R}} u\sigma(du)$. Hence,

$$g(z) = \int_{\mathbb{R}} u\sigma(du) + \int_{\mathbb{R}} \frac{1+uz}{u-z} \sigma(du) = \int_{\mathbb{R}} \frac{1+u^2}{u-z} \sigma(du),$$

which yields the result setting $\mu(du) = (1+u^2)\sigma(du)$.

Part (v): Observe that

$$\frac{1}{\pi}\int_I \Im(g_m(x+i\varepsilon))dx = \int_I\int_{\mathbb{R}} \frac{1}{\pi}\frac{\varepsilon}{(x-t)^2+\varepsilon^2} m(dt)dx = \mathbb{E}(\mathbb{1}_{\{\varepsilon Y + T \in I\}}),$$

where Y has Cauchy distribution $\frac{1}{\pi}\frac{1}{1+y^2}dy$, T is distributed according to m and Y and T are independent. The dominated convergence theorem then gives the result. $\square$

The last item in the above proposition allows one to reconstruct a measure from its Stieltjes transform. Moreover, we have the following characterization of convergence.

Proposition 3.3. *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures. One has,*

(i) *If $(\mu_n)_{n\geq 1}$ converges weakly to a probability measure μ, then $g_{\mu_n}(z)$ converges to $g_\mu(z)$ for each $z \in \mathbb{C}\setminus\mathbb{R}$.*
(ii) *If $g_{\mu_n}(z)$ converges for each $z \in \mathbb{C}\setminus\mathbb{R}$ to some limit $g(z)$, then g is the Stieltjes transform of a sub-probability measure μ, and $(\mu_n)_{n\geq 1}$ converges vaguely to μ.*

Recall that a sequence $(\mu_n)_{n\geq 1}$ of bounded measure converges vaguely to μ if for all continuous function f that goes to zero at infinity, one has $\int f d\mu_n \to \int f d\mu$. Vague convergence is slightly weaker than weak convergence, e.g. the sequence of probability measures $(\delta_n)_n$ converges vaguely to the zero measure, but does not converge weakly. For vague convergence, constants are not allowed to be test functions, hence vague convergence does not in general preserves total mass (the mass can escape at infinity). When the μ_n and μ are probability measures, the two notion coincide. Moreover, the set of bounded measures (not probability measures!) is compact for the vague topology.

Proof. Item (i) follows from the definition of the weak convergence of measure and the fact that $x \mapsto \frac{1}{x-z}$ is continuous and bounded since $\left|\frac{1}{x-z}\right| \leq \frac{1}{|\Im z|}$.

For item (ii), let $(n_k)_{k\geq 1}$ be a subsequence on which μ_{n_k} converges vaguely to some sub-probability measure, say μ (recall that the set of bounded measures is compact for the vague topology). Then, since $x \mapsto \frac{1}{x-z}$ is continuous and decays to zero at infinity, one has $g_{n_k}(z) \to g_\mu(z)$. Hence by hypothesis, it follows that $g(z) = g_\mu(z)$ for all $z \in \mathbb{C}\setminus\mathbb{R}$. Applying the inversion formula of Proposition 3.2, one has that every subsequence that converges vaguely converges to the same μ, hence μ_n converges vaguely to μ. $\square$

Remark 3.4. Suppose that m has compact support. Then its Stieltjes transform g_m writes, using the series development of $1/(x-z)$, for $z \in \mathbb{C}\setminus \operatorname{supp}(m)$,

$$g_m(z) = -\frac{1}{z}\int_{\mathbb{R}} \frac{1}{1-x/z} m(dx) = -\frac{1}{z}\sum_{k\geq 0} z^{-k}\int_{\mathbb{R}} x^k m(dx) = -\frac{1}{z}\sum_{k\geq 0} m_k z^{-k},$$

where m_k is the k^{th} moment of m. For the semicircular distribution μ_{sc}, one gets, recalling that odd moments are zero and even moments are given by the Catalan's numbers C_k,

$$g_{\mu_{sc}}(z) = -\frac{1}{z}\sum_{k\geq 0} C_k z^{-2k} = -\frac{1}{z}S(1/z^2),$$

where S is the generating function for the Catalan numbers, as defined in the proof of Lemma 2.12. Hence, we have, for $z \in \mathbb{C}\setminus[-2,2]$,

$$g_{\mu_{sc}}(z) = \frac{1}{2}\Big(-z + \sqrt{z^2-4}\,\Big).$$

Definition 3.5. *Let $M \in \mathcal{H}_n$. The resolvent of M is defined as the matrix $G_M(z) = (M - zI)^{-1}$ for $z \in \mathbb{C}\setminus\mathbb{R}$.*

Note that if $\mu_M = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(M)}$ is the spectral distribution of the matrix M, then for $z \in \mathbb{C}\setminus\mathbb{R}$,

$$g_{\mu_M}(z) = \int_{\mathbb{R}} \frac{1}{x-z}\mu_M(dx) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\lambda_i(M)-z} = \frac{1}{n}\operatorname{Tr} G_M(z).$$

The above remark informally explains why the Stieltjes transform appears naturally in the context of random matrix theory.

The next proposition gives the usual properties of the resolvent. We denote by $||\cdot||$ the operator norm, that is

$$||M|| = \sup\{|Mv|\,;v\in\mathbb{C}^n, |v|=1\},$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbb{C}^n$.

Proposition 3.6. *Let $M \in \mathcal{H}_n$ with resolvent $G_M(z)$. Then, for $z \in \mathbb{C}\setminus\mathbb{R}$,*

(i) $||G_M(z)|| \le \frac{1}{|\Im z|}$,
(ii) $|G_M(i,j)(z)| \le \frac{1}{|\Im z|}$, *for all $i,j = 1,\dots,n$,*
(iii) $\mathrm{d}G_M(z)\cdot H = -G_M(z)HG_M(z)$, *for all $H \in \mathcal{H}_n$, where* d *is the differential with respect to M.*

Proof. (i) This follows from the bound $\left|\frac{1}{x-z}\right| \le \frac{1}{|\Im z|}$.
(ii) Follows from (i).
(iii) Using $(M+H-z)^{-1}(M+H-z) = I$, we have

$$(M+H-z)^{-1}H + (M+H-z)^{-1}(M-z) = I,$$

hence multiplying on the right by $(M-z)^{-1}$, we obtain

$$G_{M+H}(z) = -G_{M+H}(z)HG_M(z) + G_M(z).$$

Thus,

$$G_{M+H}(z) = -G_M(z)HG_M(z) + G_M(z) + G_{M+H}(z)HG_M(z)HG_M(z),$$

and using (i), we obtain

$$G_{M+H}(z) - G_M(z) = -G_M(z)HG_M(z) + \mathrm{O}(||H||^2). \qquad \square$$

We now establish an integration by parts formula for the GUE (Stein's lemma), which generalizes the well known formula for the Gaussian distribution,

$$\mathbb{E}(f'(X)) = \frac{1}{\sigma^2}\mathbb{E}(f(X)X), \qquad \text{where } X \sim \mathcal{N}(0,\sigma^2).$$

Proposition 3.7. *Let X_n be a matrix distributed according to the GUE distribution, and let $H_n = \frac{1}{\sqrt{n}}X_n$. Let Φ be a C^1 function on $\mathcal{H}_n$ with bounded differential. Then for all $A \in \mathcal{H}_n$,*

$$\mathbb{E}(\mathrm{d}\Phi(H_n)\cdot A) = n\mathbb{E}(\Phi(H_n)\operatorname{Tr}(H_nA)).$$

Proof. Since the Lebesgue measure on $\mathcal{H}_n$ is invariant by translation, we have

$$\begin{aligned} I &= \int_{\mathcal{H}_n} \Phi(M)\exp\Big(-\frac{n}{2}\operatorname{Tr}(M^2)\Big)dM \\ &= \int_{\mathcal{H}_n} \Phi(M+\varepsilon A)\exp\Big(-\frac{n}{2}\operatorname{Tr}((M+\varepsilon A)^2)\Big)dM. \end{aligned}$$

Hence, $\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} I = 0$, and since $\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \operatorname{Tr}((M+\varepsilon A)^2) = 2\operatorname{Tr}(MA)$, we have

$$\frac{d}{d\varepsilon}\Big|_{\varepsilon=0} I = \int_{\mathcal{H}_n} \mathrm{d}\Phi(M) \exp\Big(-\frac{n}{2}\operatorname{Tr}(M^2)\Big)dM + \int_{\mathcal{H}_n} \Phi(M) \exp\Big(-\frac{n}{2}\operatorname{Tr}(M^2)\Big)(-n\operatorname{Tr}(MA))dM,$$

which yields the result. ☐

Proposition 3.8. *Let $H_n = \frac{1}{\sqrt{n}}X_n$, where X_n is distributed according to the GUE, and define for $z \in \mathbb{C}\setminus\mathbb{R}$,*

$$g_n(z) = \frac{1}{n}\operatorname{Tr}(G_{H_n}(z)),$$

the Stieltjes transform of the spectral distribution of H_n. Then, we have

$$\mathbb{E}(g_n(z)^2) + z\mathbb{E}(g_n(z)) + 1 = 0.$$

Proof. We apply the integration by parts formula of Proposition 3.7 to the function $\Phi(M) = (G_M(z))_{ij}$. Put $G = G_{H_n}$ for simplicity. Then, using $\mathrm{d}G(z)\cdot A = -G(z)AG(z)$, Proposition 3.7 writes

$$-\mathbb{E}((GAG)_{ij}) = n\mathbb{E}(G_{ij}\operatorname{Tr}(H_nA))$$

for all $A \in \mathrm{M}_n(\mathbb{C})$ by linearity. Take $A = e_{kl}$ the matrix with only 1 at coefficient (k,l), and 0 elsewhere. We get

$$\mathbb{E}(G_{ik}G_{lj} + n\mathbb{E}(G_{ij}(H_n)_{lk}) = 0.$$

Now taking $k=i$, $l=j$, and summing over i,j, we obtain, dividing by n^2,

$$\frac{1}{n^2}\mathbb{E}((\operatorname{Tr}(G))^2) + \frac{1}{n}\mathbb{E}(\operatorname{Tr}(GH_n)) = 0.$$

But, $GH_n = (H_n - zI)^{-1}H_n = (H_n - zI)^{-1}(H_n - zI + zI) = I + zG$, thus

$$\mathbb{E}\Big(\Big(\frac{1}{n}\operatorname{Tr}(G)\Big)^2\Big) + 1 + z\mathbb{E}\Big(\frac{1}{n}\operatorname{Tr}(G)\Big) = 0,$$

that is

$$\mathbb{E}(g_n(z)^2) + z\mathbb{E}(g_n(z)) + 1 = 0.$$ ☐

The next proposition shows that the Gaussian measure on $\mathbb{R}^n$ satisfies a concentration inequality. Informally, this means that "a random variable which depends in a smooth way on many independent random variables (but not too much on any of them) is concentrated around its mean, and therefore is essentially constant" (quote by Talagrand). We refer to the book by Ledoux [13] for a complete treatment of the concentration of measure phenomenon.

Proposition 3.9. *Let γ_{d,σ^2} be the Gaussian measure on $\mathbb{R}^d$, centered, with covariance $\sigma^2 I$ and let X be distributed according to γ_{d,σ^2}. Let f a Lipschitz function on $\mathbb{R}^d$ with constant c. Then, there exists a positive constant κ independent of d such that for all $\delta > 0$,*

$$\mathbb{P}\left(|f(X) - \mathbb{E}f(X)| \geq \delta\right) \leq 2\exp\Big(-\frac{\kappa\delta^2}{c^2\sigma^2}\Big).$$

Note that the above inequality is dimension free.

Proof. Without loss of generality, we can suppose that $\sigma^2 = 1$, and that f is Lipschitz with constant 1. Also, by subtracting a constant from f, we can suppose that $\int f d\gamma_d = 0$, denoting $\gamma_d = \gamma_{d,1}$. By symmetry, it suffices to prove that

$$\mathbb{P}(f(X) \geq \delta) \leq Ce^{-\kappa\delta^2},$$

where X is distributed according to γ_n. Moreover, it suffices to prove that

$$\mathbb{E}\Big(\exp(tf(X))\Big) \leq \exp(Ct^2),$$

since using Markov's inequality and optimizing in t will yield the result. Using some regularization argument, we can also suppose that f is smooth. Now, the Lipschitz bound on f implies the gradient estimate

$$|\nabla f(x)| \leq 1, \quad \text{for all } x \in \mathbb{R}^d,$$

where $|\cdot|$ denotes the Euclidean norm. We use the "duplication trick", following Maurey and Pisier. Let Y be an independent copy of X. Since $\mathbb{E}f(Y) = 0$, by Jensen's inequality, we get that

$$\mathbb{E}\Big(\exp(-tf(Y))\Big) \geq 1,$$

and since X and Y are independent,

$$\mathbb{E}\Big(\exp(tf(X))\Big) \leq \mathbb{E}\Big(\exp(t(f(X) - f(Y)))\Big).$$

Now, write

$$f(X) - f(Y) = \int_0^{\pi/2} \frac{d}{d\theta} f(Y\cos\theta + X\sin\theta)d\theta.$$

Define $G(\theta) = Y\cos\theta + X\sin\theta$ and consider its derivative $G'(\theta) = -Y\sin\theta + X\cos\theta$. It is easy to see that $\big(G(\theta), G'(\theta)\big)$ is a Gaussian vector in $\mathbb{R}^{2d}$ with covariance matrix the identity matrix. Hence, $G(\theta)$ and $G'(\theta)$ are independent Gaussian vectors in $\mathbb{R}^d$.

Using again Jensen's inequality, we get

$$\exp\Big(t(f(X) - f(Y))\Big) \leq \frac{2}{\pi}\int_0^{\pi/2} \exp\left(\frac{\pi t}{2}\frac{d}{d\theta} f(G(\theta))\right)d\theta,$$

and Fubini's theorem gives

$$\mathbb{E}\exp\Big(t(f(X) - f(Y))\Big) \leq \frac{2}{\pi}\int_0^{\pi/2} \mathbb{E}\exp\left(\frac{\pi t}{2}\langle \nabla f(G(\theta)), G'(\theta)\rangle\right)d\theta,$$

since by the chain rule, $\frac{d}{d\theta}f(G(\theta)) = \langle \nabla f(G(\theta)), G'(\theta)\rangle$. Since $G(\theta)$ and $G'(\theta)$ are independent, conditioning by $G(\theta)$ gives that $\frac{\pi t}{2}\langle \nabla f(G(\theta)), G'(\theta)\rangle$ is a Gaussian variable with variance bounded by $\frac{\pi^2 t^2}{4}$ since $|\nabla f(x)| \leq 1$. Thus one obtains

$$\mathbb{E}\left(\exp\left(\frac{\pi t}{2}\langle \nabla f(G(\theta)), G'(\theta)\rangle\right)\right) \leq \exp(Ct^2),$$

for some absolute constant C, and the proposition follows. □

For a function $F\colon \mathbb{R} \to \mathbb{R}$, we define its extension to $\mathcal{H}_n$, still denoted F, by

$$F(M) = U\,\mathrm{Diag}(F(\lambda_1), \ldots, F(\lambda_n))U^*,$$

if $M = U\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)U^*$. We have the following property.

Lemma 3.10. *Let $F\colon \mathbb{R} \to \mathbb{R}$ be a Lipschitz function with constant c. Then its extension to $\mathcal{H}_n$ is Lipschitz with constant c, for the Frobenius norm $||M||_2 = \sqrt{\mathrm{Tr}(M^2)}$. In particular, the function $M \mapsto \frac{1}{n}\mathrm{Tr}(F(M))$ is $\frac{c}{\sqrt{n}}$-Lipschitz.*

Proof. Let $A, B \in \mathcal{H}_n$ with eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$ and $\lambda_1(B), \cdots, \lambda_n(B)$ respectively and consider the spectral decompositions

$$A = U\,\mathrm{Diag}(\lambda_1(A), \ldots, \lambda_n(A))U^*$$
$$B = V\,\mathrm{Diag}(\lambda_1(B), \ldots, \lambda_n(B))V^*,$$

with U, V unitary matrices. Then, we have

$$||A-B||_2^2 = \mathrm{Tr}\left((A-B)^2\right) = \mathrm{Tr}(A^2) + \mathrm{Tr}(B^2) - 2\,\mathrm{Tr}(AB),$$

with $\mathrm{Tr}(A^2) = \sum_{i=1}^n \lambda_i(A)^2$, $\mathrm{Tr}(B^2) = \sum_{i=1}^n \lambda_i(B)^2$, and

$$\mathrm{Tr}(AB) = \sum_{i,j=1}^n \lambda_i(A)\lambda_j(B)|W_{ij}|^2,$$

with $W = U^*V$, which is still a unitary matrix. Using $\sum_{j=1}^n |W_{ij}|^2 = \sum_{i=1}^n |W_{ij}|^2 = 1$, since W is unitary, we obtain

$$||A-B||_2^2 = \sum_{i,j=1}^n \left(\lambda_i(A) - \lambda_j(B)\right)^2 |W_{ij}|^2,$$

and since by definition $F(A)$ and $F(B)$ have spectral decompositions

$$\begin{aligned} A &= U\,\mathrm{Diag}(F(\lambda_1(A)), \ldots, F(\lambda_n(A)))U^* \\ B &= V\,\mathrm{Diag}(F(\lambda_1(B)), \ldots, F(\lambda_n(B)))V^*, \end{aligned}$$

respectively, we get

$$||F(A) - F(B)||_2^2 = \sum_{i,j=1}^n \left(F(\lambda_i(A)) - F(\lambda_j(B))\right)^2 |W_{ij}|^2.$$

Hence, since $F\colon \mathbb{R} \to \mathbb{R}$ is c-Lipschitz, we obtain

$$||F(A) - F(B)||_2^2 \leq c^2 \sum_{i,j=1}^n \left(\lambda_i(A) - \lambda_j(B)\right)^2 |W_{ij}|^2 = c^2||A-B||_2^2,$$

so F is c-Lipschitz. This yields for $M \mapsto \frac{1}{n}\mathrm{Tr}(F(M))$, using Cauchy-Schwarz inequality,

$$|\frac{1}{n}\mathrm{Tr}(F(A)) - \frac{1}{n}\mathrm{Tr}(F(B))| \leq \frac{1}{n}\sqrt{n}||F(A) - F(B)||_2 \leq \frac{c}{\sqrt{n}}||A-B||_2,$$

which proves the second assertion of the lemma. □

We can now prove an estimate on the variance of the Stieltjes transform of the spectral measure of H_n.

Proposition 3.11. *Let $H_n = \frac{1}{\sqrt{n}}X_n$, where X_n is distributed according to the GUE. Let g_n denote the Stieltjes transform of the spectral measure of H_n. Then, there exists a constant K independent of n and z, such that for all $z \in \mathbb{C} \setminus \mathbb{R}$,*

$$\mathbb{V}\mathrm{ar}(g_n(z)) \leq \frac{K}{n^2|\Im z|^4}.$$

Proof. Using the fact that $x \mapsto \frac{1}{x-z}$ is Lipschitz with constant $\frac{1}{|\Im z|^2}$, Lemma 3.10 and the concentration inequality of the Gaussian measure of Proposition 3.9 (identifying $\mathcal{H}_n$ with $\mathbb{R}^{n^2}$ and the distribution of H_n with $\gamma_{n^2,\frac{1}{n}}$), we have

$$\mathbb{P}\Big(|g_n(z) - \mathbb{E}(g_n(z))| \geq \sqrt{\delta}\Big) \leq 2\exp\Big(-\frac{\kappa\delta|\Im z|^4 n}{2/n}\Big) = 2\exp\Big(-\frac{\kappa\delta|\Im z|^4 n^2}{2}\Big),$$

for all $\delta > 0$. Using the formula $\mathbb{V}\mathrm{ar}(Y) = \int_0^{+\infty}\mathbb{P}(|Y - \mathbb{E}(Y)|^2 \geq \delta)d\delta$ (exercice), integrating the above inequality over δ gives the result. □

We can now give an alternative proof of the Wigner theorem for the Gaussian Unitary Ensemble.

Theorem 3.12 (Wigner theorem)**.** *Let X_n be a GUE random matrix, and $H_n = \frac{1}{\sqrt{n}} X_n$. Then, the spectral measure μ_{H_n} of H_n converges weakly almost surely, as n goes to infinity, towards the semicircular distribution.*

Proof of Wigner theorem: Put $f_n(z) = \mathbb{E}(g_n(z))$. We have, since $\mathbb{E}(g_n(z)^2) + z\mathbb{E}(g_n(z)) + 1 = 0$ by proposition 3.8,

$$\left|\mathbb{E}(g_n(z)^2) - (\mathbb{E}(g_n(z)))^2\right| = \left|z\mathbb{E}(g_n(z)) + 1 + (\mathbb{E}(g_n(z)))^2\right| = \left|f_n(z)^2 + zf_n(z) + 1\right|$$

hence,

$$\left|f_n(z)^2 + zf_n(z) + 1\right| \leq \mathbb{E}\left(|g_n(z) - \mathbb{E}(g_n(z))|^2\right) = \mathbb{V}\mathrm{ar}\left(g_n(z)\right).$$

Hence, by the above estimate on the variance of $g_n(z)$, we get

$$|f_n(z)^2 + zf_n(z) + 1| \leq \frac{K}{n^2|\Im z|^4}.$$

Furthermore, we have $|f_n(z)| \leq \frac{1}{|\Im z|}$, thus the sequence $(f_n(z))_{n\geq 1}$ is analytic and uniformly bounded on compact sets of $\mathbb{C}_+ = \{z \in \mathbb{C} \,|\, \Im z > 0\}$. Hence by the classical Montel's theorem, see Theorem 6.1 in the Appendix, the sequence $(f_n(z))_{n\geq 1}$ is normal: for each subsequence of $(f_n(z))_{n\geq 1}$, there exists a sub-subsequence which converges uniformly on compact sets of $\mathbb{C}_+$ to some analytic function f. Passing to the limit in the above bound, we get that f satisfies the equation

$$f(z)^2 + zf(z) + 1 = 0,$$

which is the equation satisfied by the Stieltjes transform of the semicircular distribution. Hence, we get $f(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4})$, the sign before the square root being determined by the fact that for $\Im z > 0$, we have $\Im f_n(z) > 0$. Hence f is the Stieltjes transform of the semicircular distribution, so is uniquely determined, that is does not depend on the choice of the sub-subsequence of $(f_n(z))_{n\geq 1}$. Thus, it implies that $(f_n(z))_{n\geq 1}$ converges to f uniformly on compact sets of $\mathbb{C}_+$. Now, using Bienaymé-Tchebychev inequality and Proposition 3.11, we have

$$\mathbb{P}(|g_n(z) - f_n(z)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2}\,\mathbb{V}\mathrm{ar}(g_n(z)) \leq \frac{K}{n^2\varepsilon^2|\Im z|^4}.$$

Hence, Borel-Cantelli lemma implies that for all $z \in \mathbb{C}_+$, a.s.,

$$f_n(z) - g_n(z) \to_{n\to\infty} 0,$$

so, a.s., $g_n(z) \to f(z)$ as n goes to infinity.

It remains to show that we have that a.s., for all $z \in \mathbb{C}_+$, $g_n(z) \to f(z)$ as $n \to \infty$, that is we need to exchange the "for all z" and the "a.s." in the previous convergence. Let $(z_p)_{p\geq 0}$ be a sequence in $\mathbb{C}_+$ that admits an accumulation point in $\mathbb{C}_+$. For all $p \geq 0$, there exists a measurable set $\mathcal{N}_p$ such that $\mathbb{P}(\mathcal{N}_p) = 0$, and such that $g_n(z_p)$ converges to $f(z_p)$ on $\mathcal{N}_p^c$. Put $\mathcal{N} = \bigcup_{p\geq 0} \mathcal{N}_p$. Then $\mathcal{N}$ is a negligible set, and on $\mathcal{N}^c$, $g_n(z_p)$ converges to $f(z_p)$, as $n \to \infty$, for all $p \geq 0$.

By Vitali's theorem (see Appendix Theorem 6.2), we get that on $\mathcal{N}^c$, g_n converges to f uniformly on compacts of $\mathbb{C}_+$ as $n \to \infty$, and this ends the proof of the theorem using Proposition 3.3. □

Remark 3.13. The Stieltjes transform approach can be used to prove Wigner theorem for the so-called Gaussian Orthogonal Ensemble (GOE), which are Wigner symmetric (instead of Hermitian) matrices with independent real Gaussian coefficients. The term orthogonal comes from the fact that the distribution of such matrices are invariant by conjugation by orthogonal matrices.

This proof of Wigner theorem using Stieltjes transform can also be adapted to more general Wigner matrices. Indeed, the intregration by parts formula can be generalized using a cumulant development for a random variable X,

$$\mathbb{E}(X\Phi(X)) = \sum_{l=0}^{p} \frac{\kappa_{l+1}}{l!}\mathbb{E}(\Phi^{(l)}(X)) + \varepsilon_p,$$

where κ_l are the cumulants of X, defined using the moment-generating function of X as $\log \mathbb{E}(e^{t\cdot X}) = \sum_{l\geq 1} \kappa_l \frac{t^l}{l!}$, and where $|\varepsilon_p| \leq \sup_x \Phi^{(p+1)}(x)\mathbb{E}|X|^{p+2}$ (see for instance [12]). Note also that concentration of measure phenomenon can also be used for Wigner matrices such that the entries are i.i.d. and satisfy a log-Sobolev inequality, using Herbst argument (see [1]).

3.1. **Extremal eigenvalues.** In light of Wigner's theorem, it is natural to inquire about the convergence properties of, for instance, the largest eigenvalue. We have the following:

Theorem 3.14 ([3])**.** *Let $H_n = \frac{1}{\sqrt{n}}X_n$ where X_n is a Wigner matrix such that $\mathbb{E}(|X_n(i,j)|^4) < \infty$. Then, the largest eigenvalue $\lambda_{\max}(H_n)$ converges to 2σ almost surely.*

Let $0 < \varepsilon' < \varepsilon$ and consider a continuous bounded non-negative function f_ε supported on $[2-\varepsilon', 2]$, such that $f_\varepsilon \leq \mathbb{1}_{[2-\varepsilon,2]}$. Using Wigner theorem, and the fact that $\int f_\varepsilon d\mu_{sc} > 0$, one has that a.s., for all $\varepsilon > 0$, there exists n_0 such that for all $n \geq n_0$,

$$\frac{1}{n}\sum_{k=1}^{n} f_\varepsilon(\lambda_k(H_n)) > 0.$$

Hence, since $f_\varepsilon \leq \mathbb{1}_{[2-\varepsilon,2]}$, we get that for n large enough, $\#\{k \,|\, \lambda_k(H_n) \in [2-\varepsilon, 2]\} > 0$. Hence, one has that

$$\liminf_{n\geq 0} \lambda_{\max}(H_n) \geq 2, \qquad \text{almost surely.}$$

The corresponding upper bound on $\limsup_n \lambda_{\max}$ does not follow directly from Wigner theorem, and requires sharp combinatorial techniques. Indeed, Wigner's does not prevent single eigenvalue to detach from the limiting spectrum (such eigenvalues are called *outliers*). In case where the coefficients of the Wigner matrix have a finite moment of order 4, Bai and Yin proved the above result, see [3].

We can give a proof in case of the GUE, which is more easy. As in the proof of the inversion formula for the Stieltjes transform, one has, for any function f which is continuous and bounded,

$$\int f(x)d\mu(x) = \lim_{y\to 0^+} \int f(x)\frac{1}{\pi}\Im g_\mu(x+iy)dx.$$

This implies that

$$\left|\mathbb{E}\left(\int f(x)d\mu_n(x)\right) - \int f(x)d\mu_{sc}(x)\right| \leq \limsup_{y\to 0} \frac{1}{\pi}\left|\int f(x)\varepsilon_n(x+iy)dx\right|,$$

where $\varepsilon_n(z) = f_n(z) - g_{\mu_{sc}}(z)$, where we recall that $f_n(z) = \mathbb{E}g_n(z)$. But, using the equation satisfied by the Stieltjes transform of the semicircular distribution, one has:

$$\begin{aligned} f_n(z)^2 + zf_n(z) + 1 &= f_n(z)^2 + zf_n(z) + 1 - (g_{\mu_{sc}}(z)^2 + zg_{\mu_{sc}}(z) + 1) \\ &= f_n(z)^2 - g_{\mu_{sc}}(z)^2 + z(f_n(z) - g_{\mu_{sc}}(z)) \\ &= (f_n(z) - g_{\mu_{sc}}(z))(f_n(z) + g_{\mu_{sc}}(z) + z). \end{aligned}$$

Hence, we get that

$$\begin{aligned}|f_n(z) - g_{\mu_{sc}}(z)| &= \frac{1}{|f_n(z) + g_{\mu_{sc}}(z) + z|} \left| f_n(z)^2 + zf_n(z) + 1 \right| \\ &\leq \frac{K}{n^2|\Im z|^5},\end{aligned}$$

since $|\Im(f_n(z) + g_{\mu_{sc}}(z) + z)| > |\Im z|$ for all $z \in \mathbb{C} \setminus \mathbb{R}$ (since all the terms have the same sign), and from the concentration inequality of Proposition 3.11.

Now we use the following lemma due to Haagerup and Thorbjornsen, [10].

Lemma 3.15. *Let h be an analytic function on $\mathbb{C}\backslash\mathbb{R}$ such that $|h(z)| \leq P(\frac{1}{|\Im z|}$, for some polynomial P with nonnegative coefficients and $\deg P = k$. Then, there exists a polynomial Q such that for any φ a C^∞ function with compact support,*

$$\limsup_{y \to 0} \left| \int_{\mathbb{R}} \varphi(x) h(x + iy) dx \right| \leq \int_{\mathbb{R}} \int_0^\infty |(1 + \mathrm{d})^{k+1} \varphi(x)| Q(t) e^{-t} dt dx.$$

Applying this lemma to $h(z) = n^2 \varepsilon_n(z)$, we obtain that for any C^∞ function φ with compact support, for n large enough,

$$\limsup_{y \to 0} \left| \int \varphi(x) \varepsilon_n(x + iy) dx \right| \leq \frac{K}{n^2},$$

for some constant K. Hence, one has

$$\left| \mathbb{E}\left(\int \varphi(x) d\mu_n(x) \right) - \int \varphi(x) d\mu_{sc}(x) \right| = O\left(\frac{1}{n^2} \right).$$

Now, consider a smooth function $\varphi \colon \mathbb{R} \to [0, 1]$ which is equal to 1 on $[-2 - \varepsilon, 2 + \varepsilon]$ and positive on $[-2 - 2\varepsilon, 2 + 2\varepsilon]$, and 0 elsewhere. Let $\Psi = 1 - \varphi$. By the previous bound applied to φ, and the fact that $\Psi = 0$ on $[-2 - \varepsilon, 2 + \varepsilon]$ (which contained the support of μ_{sc}), one obtains that

$$\mathbb{E}\left(\frac{1}{n} \Psi\left(\frac{1}{\sqrt{n}} X_n \right) \right) = O\left(\frac{1}{n^2} \right),$$

so,

$$\mathbb{E}\left(\frac{1}{n} \Psi(\lambda_{\max}) \right) \leq E\left(\frac{1}{n} \Psi\left(\frac{1}{\sqrt{n}} X_n \right) \right) = O\left(\frac{1}{n^2} \right).$$

Hence, we get that

$$\mathbb{E}\left(\Psi(\lambda_{\max}) \right) = O\left(\frac{1}{n} \right) \underset{n \to \infty}{\longrightarrow} 0.$$

Moreover, $\lambda_{\max}$ is a Lipschitz function by the Hoffman-Wielandt inequality, so one can again use the concentration inequality for the Gaussian measure. Thus, by Borel-Cantelli lemma, we get that

$$\Psi(\lambda_{\max}) - \mathbb{E}\Psi(\lambda_{\max}) \underset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.},$$

so

$$\Psi(\lambda_{\max}) \underset{n \to \infty}{\longrightarrow} 0 \quad \text{a.s.}$$

But note that $\Psi \geq \mathbb{1}_{\{|x| \geq 2 + 2\varepsilon\}}$, hence, $\mathbb{1}_{\{|\lambda_{\max}| \geq 2 + 2\varepsilon\}} \to 0$ a.s., that is for all $\varepsilon > 0$, a.s.

$$\limsup_n \lambda_{\max} \leq 2 + 2\varepsilon,$$

which gives the upper bound.

4. Sample Covariance Matrices

We first introduce the Marchenko-Pastur distribution.

Definition 4.1. *The Marchenko-Pastur distribution with shape parameter $c > 0$ is the probability measure defined by*

$$\mu_{MP,c}(dx) = \left(1 - \frac{1}{c}\right)_+ \delta_0 + \frac{1}{2\pi} \frac{\sqrt{(x-c_-)(c_+ - x)}}{cx} \mathbb{1}_{[c_-,c_+]}(x)dx,$$

where $(x)_+ = \max\{x, 0\}$, and $c_\pm = (1 \pm \sqrt{c})^2$.

The Marchenko-Pastur distribution has thus a density part supported on $[c_-, c_+]$, and an atom at 0 when $c > 1$. As an exercice, one can easily show that if X is a random variable distributed according to the semi-circular distribution μ_{sc}, then X^2 is distributed according to the Marchenko-Pastur distribution with parameter $c = 1$.

The moments of the Marchenko-Pastur distribution can be computed in a straightforward way:

Proposition 4.2. *For all $n \geq 1$, we have*

$$\int x^n \mu_{MP,c}(dx) = \sum_{k=0}^{n-1} \frac{c^k}{k+1} \binom{n}{k} \binom{n-1}{k}.$$

Proof. Since $c_- + c_+ = 2(1 + c)$ and $c_-c_+ = (1 - c)^2$, we have

$$\sqrt{(x-c_-)(c_+ - x)} = \sqrt{4c - (x - (1 + x))^2},$$

hence the change of variable $y = (x - (1 + c))/\sqrt{c}$ gives

$$\begin{aligned}
\int x^n \mu_{MP,c}(dx) &= \int_{[-2,2]} (\sqrt{c}y + 1 + c)^{n-1} \frac{1}{4\pi} \sqrt{4 - y^2} dy \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} (1 + c)^{n-1-k} c^{k/2} \int_{[-2,2]} y^k \frac{1}{4\pi} \sqrt{4 - y^2} dy.
\end{aligned}$$

Since odd moments of the semicircular distribution are zero and even moments are given by the Catalan numbers, we get

$$\begin{aligned}
\int x^n \mu_{MP,c}(dx) &= \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \binom{n-1}{2k} (1 + c)^{n-1-2k} c^k \frac{1}{k+1} \binom{2k}{k} \\
&= \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \sum_{j=0}^{n-1-2k} c^{k+j} \binom{n-1-2k}{j} \binom{n-1}{2k} \frac{1}{k+1} \binom{2k}{k} \\
&= \sum_{k=0}^{\lfloor (n-1)/2 \rfloor} \sum_{l=k}^{n-1-k} c^l \binom{n-1-2k}{l-k} \binom{n-1}{2k} \frac{1}{k+1} \binom{2k}{k} \\
&= \sum_{l=0}^{n-1} c^l \sum_{k=0}^{\min(l,n-1-l)} \frac{(n-1-2k)!}{(l-k)!(n-1-k-l)!} \frac{(n-1)!}{(2k)!(n-1-2k)!} \frac{(2k)!}{(k+1)!k!} \\
&= \sum_{l=0}^{n-1} c^l \frac{1}{n} \binom{n}{l} \sum_{k=0}^{\min(l,n-1-l)} \binom{l}{k} \binom{n-l}{k+1} \\
&= \sum_{l=0}^{n-1} c^l \frac{1}{n} \binom{n}{l} \binom{n}{l+1},
\end{aligned}$$

using the Vandermonde's identity. Eventually, we get

$$\int x^n \mu_{MP,c}(dx) = \sum_{l=0}^{n-1} c^l \frac{1}{l+1}\binom{n}{l}\binom{n-1}{l}. \qquad \square$$

The Stieltjes transform of the Marchenko-Pastur distribution can also be computed:

Proposition 4.3. *Let g be the Stieltjes transform of the Marchenko-Pastur distribution $\mu_{MP,c}$, with $c > 1$. Then, for all $z \in \mathbb{C}_+$, we have*

$$g(z) = \frac{1 - c - z + \sqrt{(z-c_-)(z-c_+)}}{2cz}.$$

Moreover, g satisfies the fixed point equation on $\mathbb{C}_+$:

$$g(z) = \frac{1}{1 - c - z - czg(z)}.$$

Proof. We have,

$$g(z) = \int_{[c_-,c_+]} \frac{\sqrt{(c_+ - t)(t - c_-)}}{2\pi c(t-z)} dt.$$

We perform the change of variable $t = 1 + 2\sqrt{c}\cos\theta + c^2$, for $\theta \in (0, \pi)$. This gives:

$$g(z) = \frac{1}{\pi}\int_0^{2\pi} \frac{\sin^2\theta}{(1 + 2\sqrt{c}\cos\theta + c)(1 + 2\sqrt{c}\cos\theta + c - z)} dt.$$

Now, we compute this integral using contour integration and residue calculus. We have,

$$g(z) = \frac{1}{\pi i}\oint_\Gamma \frac{(w^2-1)^2}{w(w^2 + aw + 1)(w^2 - bw + 1)} dw$$

where Γ is the unit circle $\{|w| = 1\}$, and $a = (1+c)/\sqrt{c}$ and $b = (z-(1+c))/\sqrt{c}$. The integrand has 5 poles, and residue calculus gives the result. The fixed point equation can be easily verified. $\square$

In 1967, Marchenko and Pastur proved the following fundamental result:

Theorem 4.4. *Let p, n, with p depending on n in such a way that:*

$$\lim_{n\to\infty} \frac{p}{n} = c \in (0, +\infty).$$

Let Y_n be a $p \times n$ rectangular random matrix with i.i.d. centered Gaussian coefficients, with mean 0 and variance 1. Denote by μ_n the empirical spectral distribution of the eigenvalues of $\frac{1}{n}Y_nY_n^\intercal$. Then, almost surely, one has

$$\mu_n \underset{n\to\infty}{\longrightarrow} \mu_{MP,c}, \quad \textit{weakly}.$$

Figure 9 shows a simulation of the eigenvalues of a large random Gaussian matrix of the form $\frac{1}{n}YY^\intercal$ and the density of the Marchenko-Pastur distribution. Note that the eigenvalues of $\frac{1}{n}YY^\intercal$ and $\frac{1}{n}Y^\intercal Y$ only differs by $|n-p|$ zeroes, hence their empirical spectral distributions converge to the same limit (according to the cases $c > 1$ or $c \geq 1$).

Hence, if $X_1, \ldots, X_n$ are independent and identically distributed random (column) vectors of $\mathbb{R}^p$, with Gaussian distribution with zero mean and covariance matrix the identity I_p, then the empirical eigenvalue distribution of the empirical covariance matrix

$$\frac{1}{n}YY^\intercal = \frac{1}{n}\sum_{i=1}^n X_iX_i^\intercal,$$

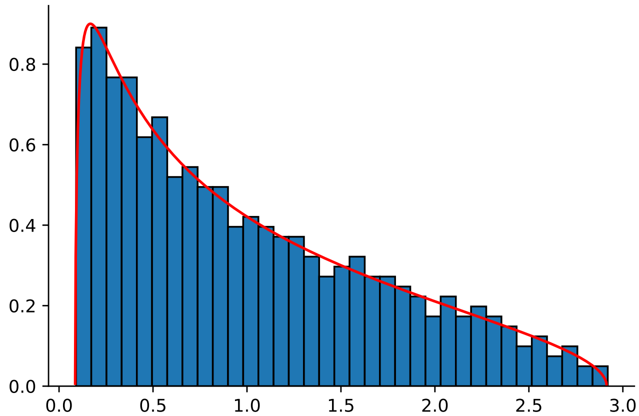


FIGURE 9. Histogram (blue) of the eigenvalues of the sample covariance matrix of a 500×1000 Gaussian random matrix and the Marchenko-Pastur distribution (red).

where Y is the rectangular matrix whose columns are $(X_1, \ldots, X_n)$, converges weakly to the Marchenko-Pastur distribution.

The fundamental result of Marchenko-Pastur have been generalized to other distribution with finite second moment, and we have:

Theorem 4.5 ([4])**.** *Let p, n, with p depending on n in such a way that:*

$$\lim_{n\to\infty} \frac{p}{n} = c \in (0, +\infty).$$

Let Y_n be a $p \times n$ rectangular random matrix with i.i.d. coefficients, with zero mean and variance 1. Denote by μ_n the empirical spectral distribution of the eigenvalues of $\frac{1}{n} Y_n Y_n^\intercal$. Then, almost surely, one has

$$\mu_n \xrightarrow[n\to\infty]{} \mu_{MP,c}, \quad \text{weakly.}$$

The above result also holds in the complex case, that is when Y_n has complex i.i.d. coefficients.

Both the method of moments or the Stieltjes resolvent method can be used to prove the Marchenko-Pastur theorem, but in both cases it is more complicated than in the case of Wigner's theorem. We refer to [4] for the details. Moreover, as for the case of Wigner matrices, under the additional assumption that the entries have finite fourth moment, the largest eigenvalue converges to the edge of the bulk:

Theorem 4.6 ([4])**.** *Let p, n, with p depending on n in such a way that:*

$$\lim_{n\to\infty} \frac{p}{n} = c \in (0, +\infty).$$

Let Y_n be a $p \times n$ rectangular random matrix with i.i.d. coefficients, with zero mean and variance 1, and such that $\mathbb{E}|Y(i,j)|^4 < \infty$. Then, the largest eigenvalue $\lambda_{\max}$ of $\frac{1}{n} Y Y^\intercal$ satisfies:

$$\lambda_{\max} \xrightarrow[n\to\infty]{} (1 + \sqrt{c})^2, \quad \text{a.s.}$$

In real world data analysis, it is expected that the vectors $X_1, \ldots, X_n$ exhibit a correlation structure instead of being i.i.d. The following result takes care of this generalization:

Theorem 4.7 ([4])**.** *Let p, n, with p depending on n in such a way that:*

$$\lim_{n\to\infty} \frac{p}{n} = c \in (0, +\infty).$$

Let Σ be a $p\times p$ deterministic symmetric nonnegative definite matrix, with bounded operator norm. Let X_n be a $p\times n$ rectangular random matrix with i.i.d. coefficients, with zero mean and variance 1, and let $Y_n = \Sigma^{1/2}X_n$. Suppose that the empirical eigenvalue distribution of Σ converges weakly, as $p\to\infty$, to some probability measure ν. Then, almost surely, the empirical eigenvalue distribution

$$\mu_{\frac{1}{n}YY^\intercal} \underset{n\to\infty}{\longrightarrow} \mu, \quad \mu_{\frac{1}{n}Y^\intercal Y} \underset{n\to\infty}{\longrightarrow} \tilde{\mu},$$

where μ and $\tilde{\mu}$ are the unique probability measures having Stieltjes transform m and $\widetilde{m}$ on $\mathbb{C}_+$ respectively, given by

$$m(z) = \frac{1}{c}\tilde{m}(z) + \frac{1-c}{cz}, \quad \tilde{m}(z) = \left(-z + c\int_{\mathbb{R}} \frac{t}{1+\tilde{m}(z)t}\nu(dt)\right)^{-1}.$$

Note that when $\nu=\delta_1$, we recover the fixed point equation of the Marchenko-Pastur distribution.

5. Spiked models

When considering finite rank deformation, Lemma 6.6 implies immediately that finite rank perturbation of a Wigner matrix does not change the limiting distribution. The same holds for sample covariance matrix when considering finite rank perturbation of the idenity. Moreover, we have seen that the convergence of the empirical spectral distribution does not prevent in general some eigenvalues to detach from the support of the limiting distribution. Nevertheless, under a finite fourth moment condition, this does not happen in both cases of the Wigner's theorem and the Marchenko-Pastur's theorem, as the largest eigenvalue converges to the edge of the "bulk". When considering additive perturbation or general sample covariance matrices, if the deformation matrix is structured, then outliers may appear. This was first observed for sample covariance matrices, and is referred now to the BBP phase transition, because of the seminal work of Baik-Ben Arous-Péché [2]. We have, in the rank one case, to simplify:

Theorem 5.1 ([2]). *Let $A_n = \mathrm{diag}(\theta,1,\dots,1)$ for some fixed $\theta>0$ independent of n. Let X_n be a $p\times n$ rectangular random matrix with i.i.d. complex coefficients, with zero mean and variance 1. Consider the model:*

$$M_n = \frac{1}{n}A_n^{1/2}X_nX_n^*A_n^{1/2}.$$

Let $\omega_c = 1+\sqrt{c}$. Then, we have,

(i) *If $\theta\le\omega_c$,*

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} (1+\sqrt{c})^2 \quad a.s.$$

(ii) *If $\theta>\omega_c$,*

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} \theta\left(1+\frac{c}{\theta-1}\right) \quad a.s.$$

Hence, when the "spike" θ of the matrix A_n is less than the threshold $\omega_c = 1+\sqrt{c}$, the largest eigenvalue of the sample covariance matrix M_n converges to the edge of the bulk of the Marchenko-Pastur distribution. On the other hand, when $\theta>\omega_c$, the largest eigenvalue of M_n converges to $\theta\left(1+\frac{c}{\theta-1}\right) > (1+\sqrt{c})^2$, hence $\lambda_{\max}$ is an "outlier" eigenvalue: it converges outside the support of the limiting spectral distribution and does not stick to the bulk.

An analogue of this phase transition was proved for GUE matrices by Péché [15]. Again, to simplify, we state the theorem for rank one perturbation.

Theorem 5.2. *Let $H_n = \frac{1}{\sqrt{n}}X_n$, where X_n is a $GUE(n,\sigma^2)$ random matrix. Consider a rank one matrix $A_n = \mathrm{diag}(\theta,0,\dots,0)$ for some fixed θ independent of n, and let $M_n = H_n + A_n$. Then,*

(i) *If* $\theta \leq \sigma$,

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} 2\sigma \quad a.s.$$

(ii) *If* $\theta > \sigma$,

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} \theta + \frac{\sigma^2}{\theta} \quad a.s.$$

Note that $\theta + \frac{\sigma^2}{\theta} > 2\sigma$, hence again, this exhibits a phase transition: when the value of the "spike" θ of the matrix A_n is strictly greater than σ, the largest eigenvalue detaches from the limiting support of the semicircular distribution, thus one observes an "outlier".

In applications, another model of interest is the so-called "information plus noise" matrix model, defined as:

$$M_n = \left(\frac{\sigma}{\sqrt{n}} X_n + A_n\right)\left(\frac{\sigma}{\sqrt{n}} X_n + A_n\right)^*,$$

where X_n is a rectangular $p \times n$ random matrix with i.i.d. complex Gaussian coefficients with zero mean and variance 1, and $A_n = \mathrm{diag}(\theta, 0, \ldots, 0)$. Then one has,

Theorem 5.3. *Consider the above information plus noise model, such that* $\frac{p}{n} \to c \in (0, 1]$.

(i) *If* $\theta \leq \sigma^2\sqrt{c}$,

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} \sigma^2(1 + \sqrt{c})^2 \quad a.s.$$

(ii) *If* $\theta > \sigma^2\sqrt{c}$,

$$\lambda_{\max}(M_n) \underset{n\to\infty}{\longrightarrow} \frac{(\sigma^2 + \theta)(\sigma^2 c + \theta)}{\theta} \quad a.s.$$

6. Appendix

6.1. **Complex analysis tools.** In what follows, we denote by $\mathbb{C}^+$ the half-plane $\mathbb{C}^+ = \{z \in \mathbb{C} \,|\, \Im(z) > 0\}$, and by $D(0, \rho)$ the disk centered at 0 with radius ρ.

Theorem 6.1 (Montel's theorem)**.** *Let $U \subset \mathbb{C}$ be an open set. Let $\mathcal{F}$ be a family of holomorphic functions on U. Suppose that $\mathcal{F}$ is uniformly bounded on every compact sets of U. Then every sequence of $\mathcal{F}$ admits a subsequence which converges uniformly on compact sets of U.*

Sketch of the Proof (see [16]*):* $\mathcal{F}$ uniformly bounded on every compact sets of U says that for all $K \subset U$ compact, there exists $M(K) > 0$ such that $\forall f \in \mathcal{F}$, $\forall z \in K$, $|f(z)| \leq M(K)$. Let $(K_n)_n$ a sequence of compact sets in U such that $U = \bigcup_n K_n$, and K_n is included in the interior of K_{n+1}, for all n. From this last property, we can find a sequence $(\delta_n)_n$ such that

$$D(z, 2\delta_n) \subset K_{n+1}, \quad \text{for } z \in K_n.$$

Let $x, y \in K_n$ such that $|x - y| < \delta_n$. Denote γ the circle, with positive orientation, centered at x of radius $2\delta_n$. Then, Cauchy's formula gives

$$f(x) - f(y) = \frac{1}{2i\pi}\int_\gamma f(\xi)\Big(\frac{1}{\xi - x} - \frac{1}{\xi - y}\Big) d\xi = \frac{x - y}{2i\pi}\int_\gamma \frac{f(\xi)}{(\xi - x)(\xi - y)} d\xi.$$

For ξ in the image of the contour γ, we have $|\xi - x| = 2\delta_n$, and $|\xi - y| > \delta_n$, hence

$$|f(x) - f(y)| \leq \frac{M(K_{n+1})}{\delta_n}|x - y|,$$

for all $f \in \mathcal{F}$, and all $x, y \in K_n$ such that $|x - y| < \delta_n$. Thus, for all K_n the restrictions of elements of $\mathcal{F}$ to K_n are an uniformly bounded equicontinuous family, and by Ascoli's theorem a pre-compact family in $C(K_n)$. A classical diagonal extraction procedure gives the result. ∎

Theorem 6.2 (Vitali's theorem)**.** *Let $U \in \mathbb{C}$ be a connected open set. Let $(z_p)_{p\geq 0}$ be a sequence in U which admits an accumulation point in U. Let $(f_n)_{n\geq 0}$ be a bounded sequence of the set of analytic functions endowed with the topology of uniform convergence on compact sets and suppose that $(f_n(z_p))_{n\geq 0}$ converges for every $p \geq 0$. Then $(f_n)_{n\geq 0}$ converges uniformly on compact sets of U.*

Proof. Suppose, to the contrary, that there is a compact set $K \subset U$ such that (f_n) is not uniformly Cauchy on K. Then for some $\varepsilon > 0$, we can find subsequences m_j and n_j such that $m_1 < n_1 < m_2 < n_2 < \cdots$ and for each j, $|f_{m_j} - f_{n_j}| \geq \varepsilon$. Put $g_j = f_{m_j}$ and $h_j = f_{n_j}$. By Montel's theorem applied to g_j, one obtains a subsequence g_{j_r} converging uniformly on compact subsets of U to some analytic function g, and the same holds for h_{j_r} denoting the limit by h. Hence we have $|h - g| \geq \varepsilon$. But since $(f_n(z_p))_{n\geq 0}$ converges for every $p \geq 0$, we have $g(z_p) = h(z_p)$, and since $(z_p)_{p\geq 0}$ has an accumulation point in U and U is open and connected, $g = h$ on U which yields a contradiction. □

Theorem 6.3 (Herglotz formula)**.** *Let f be an holomorphic function on the unit disk such that $\Re(f) \geq 0$. Then, there exists a positive measure σ with $\int_{\mathbb{R}} d\sigma = \Re(f(0))$ such that, for $|z| < 1$,*

$$f(z) = i\Im(f(0)) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z}\sigma(d\theta).$$

Proof. Let $0 < R < 1$. Then, one has, for $|z| < R$,

$$\frac{Re^{i\theta} + z}{Re^{i\theta} - z} = 1 + 2\sum_{n=1}^{\infty} \frac{z^n}{R^n e^{in\theta}}.$$

Since f is holomorphic in the disk with radius R, f admits a Taylor series development

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Hence,

$$\frac{Re^{i\theta} + z}{Re^{i\theta} - z}\Re(f(Re^{i\theta})) = \Big(1 + 2\sum_{n=1}^{\infty} R^{-n}e^{-in\theta}z^n\Big)\Big(\Re(a_0) + \sum_{m\geq 1}\frac{1}{2}(a_m R^m e^{im\theta} + \overline{a}_m R^m e^{-im\theta})\Big).$$

Integrating the last expression over θ, using the fact that $\int_{-\pi}^{\pi} e^{ik\theta}d\theta = 2\pi\delta_{k,0}$, gives

$$f(z) = i\Im(f(0)) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{Re^{i\theta} + z}{Re^{i\theta} - z}\Re(f(Re^{i\theta}))d\theta.$$

Letting R going to 1 yields the result. □

Corollary 6.4 (Nevanlinna's representation theorem)**.** *Let f be a holomorphic function on $\mathbb{C}^+$ such that $\Im(f) \geq 0$. There exists a positive finite measure μ and constants $a \geq 0$, $b \in \mathbb{R}$ such that,*

$$f(z) = az + b + \int_{\mathbb{R}} \frac{1 + uz}{u - z}\mu(du).$$

Proof. We consider the conformal mappping (that is holomorphic and bijective)

$$\begin{array}{ccc} \mathbb{C}^+ & \to & D(0,1) \\ z & \mapsto & \dfrac{z-i}{z+i}. \end{array}$$

By Herglotz theorem, we have

$$-if\Big(\frac{z-i}{z+i}\Big) = \Im(f(0)) + \frac{-i}{2\pi}\int_{-\pi}^{\pi} \frac{z(e^{i\theta}+1) + i(e^{i\theta}-1)}{z(e^{i\theta}-1) + i(e^{i\theta}+1)}\sigma(d\theta).$$

Let μ be the pushforward by the map $[-\pi, \pi] \setminus \{0\} \ni \theta \mapsto u = i\frac{1+e^{i\theta}}{1-e^{i\theta}} = -i \cot \frac{\theta}{2} \in \mathbb{R}$, of the restriction to $[-\pi, \pi] \setminus \{0\}$ of σ. Then we get,

$$-if\left(\frac{z-i}{z+i}\right) = \Im(f(0)) + \sigma(\{0\})z + \frac{1}{2\pi}\int_{\mathbb{R}} \frac{1+uz}{u-z}\mu(du),$$

which gives the result letting $a = \sigma(\{0\})$ and $b = \Im(f(0))$. $\square$

6.2. **Matrix inequalities.**

Lemma 6.5 (Hoffman-Wielandt inequality)**.** *Let A, B be two $n \times n$ normal matrices, with eigenvalues $\lambda_1(A) \leq \cdots \leq \lambda_n(A)$ and $\lambda_1(B) \leq \cdots \leq \lambda_n(B)$ respectively. Denote by $||\cdot||$ the Hilbert-Schmidt norm. Then,*

$$\sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)|^2 \leq ||A - B||^2.$$

Proof. Since A and B are nomal matrices, there are diagonalizable, so one can write:

$$A = U^* D_A U \quad B = V^* D_B V,$$

where U, V are unitary matrices, and $D_A = \operatorname{diag}(\lambda_1(A), \ldots, \lambda_n(A))$ and $D_B = \operatorname{diag}(\lambda_1(B), \ldots, \lambda_n(B))$. Now,

$$||A - B|| = ||U^* D_A U - V^* D_B V|| = ||U^*(D_A U V^* - U V^* D_B)V||.$$

Put $W = UV^*$. Then W is unitary, and by the trace property the norm $||\cdot||$ is unitary invariant, hence

$$||A - B||^2 = ||D_A W - W D_B)||^2 = \sum_{i,j=1}^{n} |W_{ij}|^2 |\lambda_i(A) - \lambda_j(B)|^2.$$

Let $P = (|W_{ij}|^2)_{1 \leq i,j \leq n}$. Since W is unitary, P is doubly stochastic, that is, $P_{ij} \geq 0$ and,

$$\sum_{i=1}^{n} |W_{ij}|^2 = 1, \quad \text{for all } j = 1, \ldots, n,$$
$$\sum_{j=1}^{n} |W_{ij}|^2 = 1, \quad \text{for all } i = 1, \ldots, n.$$

Define $\Phi(P) = \sum_{i,j=1}^{n} P_{i,j} |\lambda_i(A) - \lambda_j(B)|^2$, for all $P \in \mathcal{P}$, where $\mathcal{P}$ is the set of doubly stochastic $n \times n$ matrices. Then, we have

$$\inf_{P \in \mathcal{P}} \Phi(P) \leq ||A - B||^2.$$

But Φ is linear in P, and since $\mathcal{P}$ is compact and convex, the minimum of Φ is attained at the extremal points of $\mathcal{P}$. By the Birkhoff-Von Neumann theorem, the extremal points of $\mathcal{P}$ are given by permutation matrices P, that is P has only 0 or 1 entries, with exactly one 1 on each row and on each column. Now one can see that this minimum is attained at the permutation matrix corresponding to the identity. $\square$

Lemma 6.6 (Rank inequalities)**.** *Denote by F^A the cumulative distribution function of the empirical eigenvalue distribution μ_A of a matrix A. Denote also by $||\cdot||_\infty$ the supremum norm (that is $||f||_\infty = \sup_{x \in \mathbb{R}} |f(x)|$). Then we have:*

(i) *Let A, B be two $n \times n$ Hermitian matrices. Then,*

$$||F^A - F^B||_\infty \leq \frac{1}{n} \operatorname{rank}(A - B).$$

(ii) *Let A, B be two $p \times n$ complex matrices. Then,*

$$||F^{AA^*} - F^{BB^*}||_\infty \leq \frac{1}{p}\,\mathrm{rank}(A - B).$$

We refer to [4] for a proof.

References

[1] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010, draft available at `http://www.math.umn.edu/~zeitouni/technion/index.html`.

[2] J. Baik, G. Ben Arous and S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. **33** (2005), no. 5, 1643–1697; MR2165575

[3] Z. D. Bai and Y. Q. Yin, *necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*, Ann. Probab. **16** (1988), no. 4, 1729–1741.

[4] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, new York, 2010.

[5] F. Benaych-Georges and R. R. nadakuditi, *The singular values and vectors of low rank perturbations of large rectangular random matrices*, ArXiv e-prints, 2011.

[6] P. A. Deift, *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, Courant Lecture notes in Mathematics, vol. 3, new York University Courant Institute of Mathematical Sciences, new York, 1999.

[7] Catherine Donati-Martin, *Large random matrices, Lectures notes, Tsinghua University*, available at `http://www.proba.jussieu.fr/dw/doku.php?id=users:donati:index`.

[8] Patrick L. Ferrari, *Dimers and orthogonal polynomials: connections with random matrices, Lectures notes, IHP*, available at `http://www-wt.iam.uni-bonn.de/~ferrari/`.

[9] W. Hachem, P. Loubaton, and J. najim, *Deterministic equivalents for certain functionals of large random matrices*, Ann. Appl. Probab. **17** (2007), no. 3, 875–930.

[10] U. Haagerup and S. Thorbjørnsen, A new application of random matrices: $\mathrm{Ext}(C^*_{\mathrm{red}}(F_2))$ is not a group, Ann. of Math. (2) **162** (2005), no. 2, 711–775; MR2183281

[11] Fumio Hiai and Dénes Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000.

[12] Alexei M. Khorunzhy, Boris A. Khoruzhenko, and Leonid A. Pastur, *Asymptotic properties of large random matrices with independent entries*, J. Math. Phys. **37** (1996), no. 10, 5033–5060.

[13] Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.

[14] Leonid Pastur and Antoine Lejay, *Matrices aléatoires: statistique asymptotique des valeurs propres*, Séminaire de Probabilités, XXXVI, Lecture notes in Math., vol. 1801, Springer, Berlin, 2003, pp. 135–164.

[15] S. Péché, The largest eigenvalue of small rank perturbations of Hermitian random matrices, Probab. Theory Related Fields **134** (2006), no. 1, 127–173; MR2221787

[16] Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., new York, 1987.

[17] Gábor Szegő, *Orthogonal polynomials*, fourth ed., American Mathematical Society, Providence, R.I., 1975, American Mathematical Society, Colloquium Publications, Vol. XXIII.

[18] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012, draft available at `http://www.math.ucla.edu/~tao/`.

[19] Craig A. Tracy and Harold Widom, *Level-spacing distributions and the Airy kernel*, Comm. Math. Phys. **159** (1994), no. 1, 151–174.

[20] A.M. Tulino, S. Verdú, *Random matrix theory and wireless communications*, in Foundations and Trends in Communications and Information theory, vol. 1, Hanover, MA, Now Publishers, 2004.

[21] Eugene P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.

[22] J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika **20** (1928), 35–52.

Université de Toulouse, Institut de Mathématiques de Toulouse, 118 route de narbonne, F-31062 Toulouse Cedex 9

Email address: `francois.chapon@math.univ-toulouse.fr`