

Large Random Matrices

C. Donati-Martin

Tsinghua University, November 2009

1 Introduction

Random matrices were introduced in multivariate statistics, in the thirties by Wishart [Wis] (see also [Mu]) and in theoretical physics by Wigner [Wig] in the fifties. Since then, the theory developed in a wide range of mathematics fields and physical mathematics. These lectures give a brief introduction of one aspect of Random Matrix Theory (RMT): the asymptotic distribution of the eigenvalues of random Hermitian matrices of large size N . The study of some asymptotic regimes leads to interesting results and techniques.

Let A be a $N \times N$ Hermitian matrix with eigenvalues λ_i , $1 \leq i \leq N$. We can define the spectral measure of A by

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

or

$$\mu_A(\Delta) = \frac{1}{N} \#\{1 \leq i \leq N, \lambda_i \in \Delta\}, \Delta \subset \mathbb{R}.$$

One of the main problem in RMT is to investigate the convergence of the random spectral measures (μ_{A^N}) for a given family of random matrices A^N of size N when $N \rightarrow \infty$. Another question of interest is the behavior on extreme eigenvalues.

Two regimes will be considered in the asymptotic $N \rightarrow \infty$:

1) **The global regime:** we consider $\mu_{A^N}(\Delta)$ for a fixed Borel set Δ of size $|\Delta| \sim 1$ and look for the convergence of $\mu_{A^N}(\Delta)$ to $\mu(\Delta)$ where μ is a probability. In general, the distribution μ can be found explicitly (ex:

Wigner distribution) and depends on the distribution of the ensemble of matrices A^N .

2) **The local regime:** we are interested in the microscopic properties of the eigenvalues (ex: spacing distribution). As in the Central Limit Theorem, we make a renormalisation of certain probabilistic quantities to obtain non degenerate limits. We now consider $N\mu_{A^N}(\Delta_N)$ where the size of Δ_N tends to 0, i.e. we make a zoom around u looking at the interval $\Delta_N = [u - \epsilon_N, u + \epsilon_N]$ with $\epsilon_N \rightarrow 0$. The universality conjecture says that, in the local regime, renormalisation gives universal limits (independent of the distribution of the ensemble A^N). This regime is more delicate to study and often requires the distribution of the eigenvalues.

We shall present different techniques in RMT:

1) *Moment method:* in order to study the convergence of μ_{A^N} , we compute its moments

$$\mu_A(x^k) = \frac{1}{N} \sum_{i=1}^N \lambda_i^k = \frac{1}{N} \text{Tr}(A^k)$$

and

$$\mathbb{E}\left[\frac{1}{N} \text{Tr}(A^k)\right] = \frac{1}{N} \sum_{i_1, \dots, i_k=1}^N \mathbb{E}[A_{i_1 i_2} A_{i_2 i_3} \dots A_{i_k i_1}].$$

If A is a matrix with independent coefficients, we can compute the leading term of the right-hand side of the above equation, using some combinatorics. This is the aim of section 2.

2) *The Stieltjes transform:* we can also prove the convergence of μ_{A^N} by proving the convergence of $\mu_{A^N}(f_z)$ for a large class of bounded continuous functions f_z depending on a parameter $z \in \mathbb{C} \setminus \mathbb{R}$. We consider $f_z(x) = \frac{1}{x-z}$. $\mu(f_z)$ is the Stieltjes transform of the measure μ . We present this technique in Section 3 for a Gaussian ensemble of matrices (GUE).

3) *Orthogonal polynomials:* In Section 4, we provide a deep analysis of GUE: density of eigenvalues, correlation functions in order to handle the local regime. The method relies on orthogonal polynomials.

1.1 Notations

- \mathcal{H}_N , resp. \mathcal{S}_N , denotes the set of $N \times N$ Hermitian, resp. symmetric, matrices with complex, resp. real, coefficients. These spaces are equipped with the trace, denoted by Tr .

- If $M \in \mathcal{H}_N$, we denote by $(\lambda_k(A))_{1 \leq k \leq N}$ its eigenvalues (in \mathbb{R}^N).
- For a matrix $M \in \mathcal{H}_N$, we define the spectral measure of M as the probability measure defined by:

$$\mu_M = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(M)}$$

where δ_x is the Dirac mass at point x . If M is a random matrix, then, μ_M is a random probability measure.

- $\mathcal{P}(\mathbb{R})$ is the set of probability measures, equipped with the weak topology so that $\mu \rightarrow \mu(f) := \int f(x)d\mu(x)$ is continuous if f is bounded continuous on \mathbb{R} .
- $N(0, \sigma^2)$ denotes the Gaussian distribution on \mathbb{R} with mean 0 and variance σ^2 .

Part I

Global behavior

2 Wigner matrices

We consider a sequence of random matrices $A^N = (A_{ij}^N)_{1 \leq i, j \leq N}$ of size N in \mathcal{H}_N (or \mathcal{S}_N) defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We assume that for fixed N , $(A_{ij}^N)_{1 \leq i \leq j \leq N}$ are independent random variables and moreover that

$$\mathbb{E}[A_{ij}^N] = 0 \text{ and } \mathbb{E}[|A_{ij}^N|^2] = \frac{\sigma^2}{N}. \quad (2.1)$$

Such matrices, with independent coefficients, are called **Wigner matrices**. The aim of the chapter is to prove the convergence of the spectral measure $\mu_{A^N} = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(A^N)}$ of the eigenvalues of A^N to the semicircular distribution:

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

We first prove the convergence of the moments of μ_{A^N} when the entries of A^N have finite moments.

Theorem 2.1 *Assume (2.1) and for all $k \in \mathbb{N}$*

$$\alpha_k = \sup_N \sup_{1 \leq i \leq j \leq N} \mathbb{E}[|\sqrt{N}A_{ij}^N|^k] < \infty \quad (2.2)$$

Then,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \text{Tr}((A^N)^k) = \begin{cases} 0 & \text{if } k \text{ odd} \\ \sigma^{2p} \frac{(2p)!}{p!(p+1)!} & \text{if } k = 2p \end{cases}$$

where the convergence holds a.s. and in expectation. The numbers $C_p = \frac{(2p)!}{p!(p+1)!}$ are called the Catalan numbers.

Note that $\frac{1}{N} \text{Tr}((A^N)^k) = \int_{\mathbb{R}} x^k d\mu_{A^N}(dx) = \frac{1}{N} \sum_i \lambda_i^k(A^N)$.

Before proving this theorem, we shall give some combinatorial results.

2.1 Some properties of the Catalan numbers

Lemma 2.1 1) *The Catalan numbers C_p satisfy:*

$$\begin{aligned} C_p &= \#\{ \text{rooted oriented trees with } p \text{ edges} \} \\ &= \#\{ \text{Dyck paths with } 2p \text{ steps} \} \end{aligned}$$

2) *Let $\mu_{sc}(dx) = \frac{1}{2\pi}\sqrt{4-x^2}1_{[-2,2]}(x)dx$ the semicircular distribution of variance 1, then*

$$\int_{\mathbb{R}} x^{2p+1} \mu_{sc}(dx) = 0; \quad \int_{\mathbb{R}} x^{2p} \mu_{sc}(dx) = C_p.$$

A **tree** is a connected graph with no cycle. A **root** is a marked vertex. A tree is **oriented** if it is drawn (or embedded) into the plane; it inherits the orientation of the plane.

A **Dyck path** of length $2p$ is a path in \mathbb{Z}^2 starting from $(0,0)$ and ending at $(2p,0)$ with increments of the form $(1,+1)$ or $(1,-1)$ and staying above the real axis.

Property 2.1 *There is a bijection between the set of oriented trees and the set of Dyck paths.*

Sketch of proof: (see [G])

1) We define a walk on the tree with p vertices : we consider the oriented tree as a "fat tree" replacing each edge by a double edge, surrounding the original one. The union of these edges define a path surrounding the tree. The walk on the tree is given by putting the orientation of the plane on this path and starting from the root. To define the Dyck path from the walk on the tree, put an increment $(1,1)$ when one meets an edge that has not been visited and $(1,-1)$ otherwise. It is easy to see that this defines a Dyck path. Now, given a Dyck path, glue the couples of steps when one step $+1$ is followed by a step -1 and replace each glued couple by an edge. We obtain a path decorated with edges. Continue this procedure to obtain a rooted tree. \square

Proof of Lemma 2.1: 1) We verify that the cardinal of the above sets is given by the Catalan numbers. Let $D_p := \#\{ \text{Dyck paths with } 2p \text{ steps} \}$. Let us verify that D_p satisfy the relation:

$$D_0 = 1, \quad D_p = \sum_{l=1}^p D_{l-1} D_{p-l}, \quad (2.3)$$

Let $1 \leq l \leq p$ and $D_{p,l} := \#\{\text{Dyck paths with } 2p \text{ steps hitting the real axis for the first time at time } 2l \text{ (i.e. after } 2l \text{ steps)}\}$. Then, $D_p = \sum_{l=1}^p D_{p,l}$. Now, we have

$$D_{p,l} = \#\{\text{Dyck paths from } (0,0) \text{ to } (2l,0) \text{ strictly above the real axis}\} \\ \times \#\{\text{Dyck paths from } (2l,0) \text{ to } (2p,0)\}.$$

$\#\{\text{Dyck paths from } (2l,0) \text{ to } (2p,0)\} = D_{p-l}$ and $\#\{\text{Dyck paths from } (0,0) \text{ to } (2l,0) \text{ strictly above the real axis}\} = D_{l-1}$ (since the first and last step are prescribed, shift the real axis to $+1$ to have a correspondence with Dyck paths with $2(l-1)$ steps).

Now, it remains to prove that (2.3) characterizes the Catalan numbers. To this end, we introduce the generating function $S(z) = \sum_{k=0}^{\infty} D_k z^k$. Since $D_k \leq 2^{2k}$, the serie is absolutely convergent for $|z| \leq 1/4$. From the recurrence relation (2.3), it is easy to see that $S(z)$ satisfies:

$$S(z) - 1 = z(S(z))^2$$

and therefore

$$S(z) = \frac{1 - \sqrt{1 - 4z}}{2z}.$$

A Taylor development of this function gives $S(z) = \sum \frac{(2k)!}{k!(k+1)!} z^k$ and thus $D_k = C_k$.

2) The computations of the moments of the semicircular distribution follows from standard computations (perform the change of variables $x = 2 \sin(\theta)$ in the integral). \square

Proof of Theorem 2.1 We first prove the convergence in expectation. Put $X = \sqrt{N}A^N$. The entries of X are centered with variance σ^2 .

$$\begin{aligned} \frac{1}{N} \mathbb{E}[\text{Tr}(A^N)^k] &= \frac{1}{N} \mathbb{E}\left[\sum_{i_1, \dots, i_k=1}^N A_{i_1 i_2}^N \dots A_{i_k i_1}^N \right] \\ &= \frac{1}{N^{1+k/2}} \mathbb{E}\left[\sum_{i_1, \dots, i_k=1}^N X_{i_1 i_2} \dots X_{i_k i_1} \right] \end{aligned} \quad (2.4)$$

Let $\mathbf{i} = (i_1, \dots, i_k)$ and set $P(\mathbf{i}) = \mathbb{E}[X_{i_1 i_2} \dots X_{i_k i_1}]$. From the assumption (2.2),

$$P(\mathbf{i}) \leq a_k$$

for some constant a_k independent of N (by Holder's inequality). From the independence and centering of the entries,

$$P(\mathbf{i}) = 0$$

unless to any edge (i_p, i_{p+1}) , there exists $l \neq p$ such that $(i_p, i_{p+1}) = (i_l, i_{l+1})$ or (i_{l+1}, i_l) . A single edge gives a zero contribution.

We now want to characterize the set of indices \mathbf{i} giving a non null contribution in the limit $N \rightarrow \infty$. To \mathbf{i} , we associate a graph $G(\mathbf{i}) = (V(\mathbf{i}), E(\mathbf{i}))$ where the set of vertices is given by $V(\mathbf{i}) = \{i_1, \dots, i_k\}$ and the set of edges $E(\mathbf{i}) = \{(i_1, i_2), \dots, (i_k, i_{k+1})\}$ with by convention $i_{k+1} = i_1$. $G(\mathbf{i})$ is connected. We denote by $\bar{G}(\mathbf{i}) = (\bar{V}(\mathbf{i}), \bar{E}(\mathbf{i}))$ the skeleton of $G(\mathbf{i})$, that is $\bar{V}(\mathbf{i})$ are the distinct points of $V(\mathbf{i})$, $\bar{E}(\mathbf{i})$ the set of edges of $E(\mathbf{i})$ without multiplicities.

Lemma 2.2 *Let $G = (V, E)$ a connected graph, then,*

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

where $|A|$ denotes the cardinal of the distinct elements in the set A . The equality holds iff G is a tree.

Proof: We prove the inequality by recurrence over $|V|$. This is true for $|V| = 1$. Let $|V| = n$. Take a vertex v in V and split G into (v, e_1, \dots, e_l) and (G_1, \dots, G_r) connected graphs where (e_1, \dots, e_l) are the edges containing v . We have $r \leq l$. $G_i := (V_i, E_i)$ and by the induction hypothesis, $|V_i| \leq |E_i| + 1$. Then,

$$\begin{aligned} |V| = 1 + \sum_{i=1}^r |V_i| &\leq 1 + \sum_{i=1}^r (|E_i| + 1) \\ &= 1 + |E| - l + r \leq 1 + |E|. \end{aligned}$$

If $|V| = |E| + 1$, we must have equality in all the previous decompositions. But, if there is a loop in G , we can find a vertex v with $r < l$. \square

Since $P(\mathbf{i}) = 0$ unless each edge appears at least twice, we have $|\bar{E}(\mathbf{i})| \leq \lfloor k/2 \rfloor$ where $\lfloor x \rfloor$ denotes the integer part of x , and from the previous lemma, $|\bar{V}(\mathbf{i})| \leq \lfloor k/2 \rfloor + 1$. Since the indices vary from 1 to N , there are at most $N^{\lfloor k/2 \rfloor + 1}$ indices contributing in the sum (2.4) and

$$\frac{1}{N} \mathbb{E}[\text{Tr}(A^N)^k] \leq a_k N^{\lfloor k/2 \rfloor - k/2}.$$

Therefore, if k is odd,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}[\text{Tr}(A^N)^k] = 0.$$

If k is even, the indices \mathbf{i} which contribute to the limit are those for which $\bar{V}(\mathbf{i}) = k/2 + 1$. From the lemma, it follows that $\bar{G}(\mathbf{i})$ is a tree and $|\bar{E}(\mathbf{i})| = k/2 = |E(\mathbf{i})|/2$, that is each edge appears exactly twice and $G(\mathbf{i})$ appears as a fat tree. For such \mathbf{i} , $\bar{G}(\mathbf{i})$ is an oriented rooted tree: the root is given by the directed edge (i_1, i_2) , the order of the indices induces a cyclic order on the fat tree, that uniquely prescribes an orientation. For these indices \mathbf{i} ,

$$P(\mathbf{i}) = \prod_{e \in \bar{E}(\mathbf{i})} \mathbb{E}[|X_e|^2]^{k/2} = (\sigma^2)^{k/2}.$$

Up to now, we only consider the shape of the tree, without considering the numerotation of the vertices. For the same geometry of the rooted tree, there are $N(N-1) \dots (N - \frac{k}{2}) \sim N^{1+k/2}$ choices for the distinct vertices. Therefore,

$$\begin{aligned} \mathbb{E}\left[\frac{1}{N} \text{Tr}(A^N)^k\right] &= \frac{N(N-1) \dots (N - \frac{k}{2})}{N^{k/2+1}} \\ &\quad \times \#\{\text{rooted oriented trees with } k/2 \text{ edges}\} + o(1) \\ &\longrightarrow C_{k/2}. \end{aligned}$$

This ends the proof of the convergence in expectation. It remains to prove the a.s. convergence. To this end, we prove that the variance of $\frac{1}{N} \text{Tr}(A^N)^k$ is of order $\frac{1}{N^2}$. We refer to [G] for the computation of the variance. We just give an outline .

$$\begin{aligned} \text{Var}\left(\frac{1}{N} \text{Tr}(A^N)^k\right) &= \mathbb{E}\left[\left(\frac{1}{N} \text{Tr}(A^N)^k\right)^2\right] - \mathbb{E}\left[\left(\frac{1}{N} \text{Tr}(A^N)^k\right)\right]^2 \\ &= \frac{1}{N^{2+k}} \sum_{i_1, \dots, i_k, i'_1, \dots, i'_k} [P(\mathbf{i}, \mathbf{i}') - P(\mathbf{i})P(\mathbf{i}')] \end{aligned}$$

where $P(\mathbf{i}, \mathbf{i}') = \mathbb{E}[X_{i_1 i_2} \dots X_{i_k i_1} X_{i'_1 i'_2} \dots X_{i'_k i'_1}]$.

As above, we introduce a graph $G(\mathbf{i}, \mathbf{i}')$ with vertices $V(\mathbf{i}, \mathbf{i}') = \{i_1, \dots, i_k, i'_1, \dots, i'_k\}$ and edges $E(\mathbf{i}, \mathbf{i}') = \{(i_1, i_2), \dots, (i_k, i_{k+1}), (i'_1, i'_2), \dots, (i'_k, i'_{k+1})\}$. To give a contribution to the leading term, the graph must be connected (if $E(\mathbf{i}) \cap$

$E(\mathbf{i}') = \emptyset$, then $P(\mathbf{i}, \mathbf{i}') = P(\mathbf{i})P(\mathbf{i}')$. Moreover, each edge must appear at least twice and thus,

$$|V(\mathbf{i}, \mathbf{i}')| \leq |E(\mathbf{i}, \mathbf{i}')| + 1 \leq k + 1.$$

This shows that the variance is at most of order $\frac{1}{N}$. A finer analysis (see [G]) shows that the case $|V(\mathbf{i}, \mathbf{i}')| = k + 1$ cannot occur and $\text{Var} \left(\frac{1}{N} \text{Tr}(A^N)^k \right) = O\left(\frac{1}{N^2}\right)$. Thus,

$$\mathbb{E}\left[\sum_N \left(\frac{1}{N} \text{Tr}(A^N)^k - \mathbb{E}\left(\frac{1}{N} \text{Tr}(A^N)^k\right)\right)^2\right] < \infty$$

implying that

$$\frac{1}{N} \text{Tr}(A^N)^k - \mathbb{E}\left(\frac{1}{N} \text{Tr}(A^N)^k\right) \longrightarrow 0 \text{ a.s.} \quad \square$$

Theorem 2.2 *Under the assumption (2.1), μ_{A_N} converges to μ_{sc, σ^2} a.s, that is for every bounded continuous function f ,*

$$\lim_N \int_{\mathbb{R}} f(x) d\mu_{A^N}(x) = \int_{\mathbb{R}} f(x) d\mu_{sc, \sigma^2}(x) \text{ a.s.} \quad (2.5)$$

Sketch of Proof: From Theorem 2.1, Lemma 2.1 2) (and a scaling argument to pass from $\sigma = 1$ to general σ), (2.5) is satisfied if f is replaced by a polynomial.

We use Weierstrass theorem to approximate f by a polynomial uniformly on the compact $[-B, B]$ with $B > 2\sigma$. For $\delta > 0$, we choose a polynomial P such that

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

Then,

$$\begin{aligned} \left| \int_{\mathbb{R}} f(x) d\mu_{A^N}(x) - \int_{\mathbb{R}} f(x) d\mu_{sc, \sigma^2}(x) \right| &\leq \left| \int_{\mathbb{R}} P(x) d\mu_{A^N}(x) - \int_{\mathbb{R}} P(x) d\mu_{sc, \sigma^2}(x) \right| \\ &\quad + 2\delta + \int_{|x| > B} |f(x) - P(x)| d\mu_{A^N}(x) \end{aligned}$$

since μ_{sc, σ^2} has a support in $[-2\sigma, 2\sigma]$. The first term in the above equation tends to 0 as $N \rightarrow \infty$ thanks to Theorem 2.1. For the third term, let p

denote the degree of P , then,

$$\begin{aligned} \int_{|x|>B} |f(x) - P(x)| d\mu_{A^N}(x) &\leq C \int_{|x|>B} |x|^p d\mu_{A^N}(x) \\ &\leq CB^{-p+2q} \int x^{2(p+q)} d\mu_{A^N}(x) \quad \forall q \end{aligned}$$

Thus, since $\int x^{2(p+q)} d\mu_{A^N}(x) \rightarrow \int x^{2(p+q)} d\mu_{sc,\sigma^2}(x)$,

$$\limsup_N \int_{|x|>B} |f(x) - P(x)| d\mu_{A^N}(x) \leq CB^{-p+2q} (2\sigma)^{2p+2q} \xrightarrow{q \rightarrow \infty} 0$$

since $B > 2$. Then $\delta \rightarrow 0$ gives the result. \square

We now give a statement of Wigner's theorem without existence of the moments of the coefficients.

Theorem 2.3 *Assume that the Hermitian matrix $A^N = \frac{1}{\sqrt{N}}X$ satisfies $E[X_{ij}] = 0$, $E[|X_{ij}|^2] = 1$ and the diagonal entries are iid real variables, those above the diagonal are iid complex variables, then μ_{A^N} converges to μ_{sc} a.s..*

We refer to Guionnet [G], Bai-Silverstein [BS], Bai [B] for the passage from Theorem 2.1 to Theorem 2.3. It relies on an appropriate truncation of the coefficients of A^N into bounded coefficients and the following lemma:

Lemma 2.3 *Let $\lambda_1(A) \leq \dots \leq \lambda_N(A)$ and $\lambda_1(B) \leq \dots \leq \lambda_N(B)$ the ranked eigenvalues of two Hermitian matrices A and B . Then,*

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

2.2 On the extremal eigenvalues

Proposition 2.1 *[BY], [BS]*

Assume that $A^N = \frac{1}{\sqrt{N}}X^N$ with X_{ii} iid centered with finite variance, $(X_{ij})_{i<j}$ iid with distribution μ with $\int x\mu(dx) = 0$, $\int |x|^2\mu(dx) = \sigma^2$, $\int |x|^4\mu(dx) < \infty$. Then, the largest eigenvalue $\lambda_{max}(A^N)$ converges a.s. to 2σ .

From the convergence of the spectral distribution μ_{A^N} to the semicircular distribution μ_{sc,σ^2} and using that $\mu_{sc,\sigma^2}([2\sigma - \epsilon, 2\sigma]) > 0$, one can easily show that

$$\liminf_{N \rightarrow \infty} \lambda_{max}(A^N) \geq 2.$$

The upper bound requires sharp combinatorial techniques. Bai and Yin ([BS]) shows that one can find $s = s_N$ such that

$$\sum_N P(\text{Tr}(A^N)^{2s} \geq (2\sigma + \epsilon)^{2s}) < \infty.$$

Since $P(\lambda_{\max}(A^N) \geq (2\sigma + \epsilon)) \leq P(\text{Tr}(A^N)^{2s} \geq (2\sigma + \epsilon)^{2s})$, Borel-Cantelli's lemma gives

$$\limsup_{N \rightarrow \infty} \lambda_{\max}(A^N) \leq 2 + \epsilon.$$

2.3 Comments

1. This method is quite general and requires only minimal hypothesis on the moments of the entries of the matrix (we do not need the explicit distribution of the entries). Nevertheless, we cannot obtain sharp informations on the spectrum by this method. The method is also valid for the real symmetric case.
2. The set of Dyck paths with $2k$ steps (or the set of oriented tree with k edges) is also in bijection with the set of non-crossing pair partitions of $\{1; \dots, 2k\}$.

A partition π is said to be crossing if there exists $p < q < p' < q'$ such that p, p' belongs to the same block of π , q, q' belongs to the same block of π , the two blocks being distinct. It is not crossing otherwise. A pair partition is a partition with all blocks of cardinality 2.

We refer to Hiai-Petz [HP] for a proof of Wigner theorem using non crossing pair-partitions.

3 The resolvent approach

In this Section, we prove Wigner's theorem in the particular case of a Gaussian Wigner matrix, using the resolvent approach. We also consider a model of sample covariance matrices and prove the convergence of its spectral measure.

3.1 The Gaussian Unitary Ensemble (GUE)

Definition 3.1 $GUE(N; \sigma^2)$ is the Gaussian distribution on \mathcal{H}_N given by

$$P_{N,\sigma^2}(dM) = \frac{1}{Z_{N,\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} \text{Tr}(M^2)\right) dM \quad (3.1)$$

where dM denotes the Lebesgue measure on \mathcal{H}_N given by

$$dM = \prod_{i=1}^N dM_{ii} \prod_{1 \leq i < j \leq N} d\Re M_{ij} d\Im M_{ij}$$

and Z_{N,σ^2} is a normalizing constant. $Z_{N,\sigma^2} = 2^{N/2}(\pi\sigma^2)^{N^2/2}$.

In the following, in the asymptotics $N \rightarrow \infty$, we shall consider this ensemble with variance $1/N$ i.e. $GUE(N; \frac{1}{N})$. In other words, a random matrix is distributed as $GUE(N; \frac{1}{N})$ if $H_N(i, i)$ is distributed as $N(0, \frac{1}{N})$, $\Re H_N(j, k)$, $\Im H_N(j, k)$, $j < k$ are distributed as $N(0, \frac{1}{2N})$, all the variables being independent.

Remark 3.1 1) $GUE(N; \frac{1}{N})$ is a Wigner matrix (independent entries);
2) The distribution P_{N,σ^2} is invariant under the unitary transformation $T_U : M \rightarrow U^* M U$ where U is a unitary matrix ($U U^* = U^* U = I$), that is for all Borelian B of \mathcal{H}_n :

$$P_{N,\sigma^2}(T_U(B)) = P_{N,\sigma^2}(B).$$

Theorem 3.1 (Wigner's theorem)

Let H_N a random matrix distributed as $GUE(N; \frac{1}{N})$ and μ_{H_N} its spectral measure. Then, a.s.

$$\mu_{H_N} \xrightarrow[N \rightarrow \infty]{\text{weak}} \mu_{sc}$$

where μ_{sc} denotes the semicircular distribution $\mu_{sc}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2} 1_{[-2,2]}(x) dx$.

To prove this theorem, we shall use the Stieltjes transform of μ_N .

Definition 3.2 Let m a probability measure on \mathbb{R} . The function

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

defined for $z \in \mathbb{C} \setminus \mathbb{R}$ is the Stieltjes transform of m .

We present some properties of the Stieltjes transform (see [AG], [BS, Appendix]).

Proposition 3.1 Let g the Stieltjes transform of a probability m . Then,

- i) g is analytic on $\mathbb{C} \setminus \mathbb{R}$ and $g(\bar{z}) = \overline{g(z)}$,
- ii) $\Im z \Im(g(z)) > 0$ for $\Im z \neq 0$,
- iii) $\lim_{y \rightarrow \infty} y|g(iy)| = 1$.
- iv) If i)-iii) are satisfied for a function g , then, there exists a probability m such that g is the Stieltjes transform of m .
- v) If I is an interval such that m does not charge the endpoints, then,

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

(inversion formula).

- vi) $m \rightarrow g_m$ is continuous from $\mathcal{P}(\mathbb{R})$ endowed with the weak topology to the set of analytic functions endowed with the uniform topology on compacts of $\mathbb{C} \setminus \mathbb{R}$.

Note that if $M \in \mathcal{H}_N$ and μ_M denotes its spectral measure, then

$$\begin{aligned} g_{\mu_M}(z) &= \int_{\mathbb{R}} \frac{1}{x - z} d\mu_M(x) \\ &= \frac{1}{N} \sum_{i=1}^N \frac{1}{\lambda_i(M) - z} \\ &= \frac{1}{N} \text{Tr}((M - zI)^{-1}) \end{aligned}$$

$G_M(z) := (M - zI)^{-1}$ is called the resolvent of the matrix M and satisfies:

1) $\|G_M(z)\| \leq \frac{1}{|\Im(z)|}$ where $\|\cdot\|$ denotes the operator norm on the set of matrices.

2) $G'_M(z).A = -G_M(z)AG_M(z)$ where G' is the derivative of G with respect to M , z fixed.

We now establish an integration by part formula for the Gaussian ensemble.

Proposition 3.2 *Let H_N a random matrix defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ distributed as $GUE(N, 1/N)$. Let Φ a C^1 function on \mathcal{H}_N with bounded derivative. Then,*

$$\mathbb{E}[\Phi'(H_N).A] = N\mathbb{E}[\Phi(H_N) \text{Tr}(H_N A)] \quad \forall A \in \mathcal{H}_N. \quad (3.2)$$

This proposition is an extension of the well known integration by part formula for the unidimensional Gaussian distribution: if g is distributed as $N(0, \sigma^2)$, then,

$$\mathbb{E}[f'(g)] = \frac{1}{\sigma^2} \mathbb{E}[f(g)g].$$

Proof: The Lebesgue measure dM on \mathcal{H}_N is invariant by translation, thus

$$I := \int_{\mathcal{H}_N} \Phi(M) \exp\left(-\frac{N}{2} \text{tr}(M^2)\right) dM = Z_{N,1/N} \mathbb{E}[\Phi(H_N)]$$

satisfies

$$I = \int_{\mathcal{H}_N} \Phi(M + \varepsilon A) \exp\left(-\frac{N}{2} \text{tr}((M + \varepsilon A)^2)\right) dM.$$

The derivative in ε of the right-hand side of the above equation is zero and equals

$$\begin{aligned} \frac{dI}{d\varepsilon} \Big|_{\varepsilon=0} &= \int_{\mathcal{H}_N} \Phi'(M).A \exp\left(-\frac{N}{2} \text{tr}(M^2)\right) dM \\ &+ \int_{\mathcal{H}_N} \Phi(M) \exp\left(-\frac{N}{2} \text{tr}(M^2)\right) (-N \text{Tr}(AM)) dM \end{aligned}$$

This gives (3.2). \square

Proposition 3.3 *Let H_N distributed as $GUE(N, 1/N)$ and*

$$g_N(z) = \frac{1}{N} \text{Tr}((H_N - zI)^{-1}) = \int_{\mathbb{R}} \frac{1}{x - z} d\mu_{H_N}(x)$$

with $z \in \mathbb{C} \setminus \mathbb{R}$. Then, $g_N(z)$ satisfies:

$$\mathbb{E}[(g_N(z))^2] + z\mathbb{E}[g_N(z)] + 1 = 0. \quad (3.3)$$

Proof: We apply (3.2) to $\Phi(M) = (G_M(z))_{ij}$, $z \in \mathbb{C} \setminus \mathbb{R}$, $i, j \leq N$. Then,

$$\mathbb{E}[-(G_{H_N}(z)AG_{H_N}(z))_{ij}] = N\mathbb{E}[G_{H_N}(z)(i, j) \operatorname{Tr}(H_N A)]$$

for all $A \in \mathcal{H}_N$ and in fact for all matrix A (by \mathbb{C} -linearity of (3.2) in A). Take $A = E_{kl}$ the element of the canonical basis ($E_{kl}(p, q) = \delta_{kp}\delta_{lq}$) in the above formula to obtain (writing G for $G_{H_N}(z)$ to simplify):

$$\mathbb{E}[G_{ik}G_{lj}] + N\mathbb{E}[G_{ij}H_N(l, k)] = 0.$$

Take $i = k, j = l$ and take the normalized sum in i, j of the above equations ($\frac{1}{N^2} \sum_{i,j} \dots$); we obtain:

$$\frac{1}{N^2}\mathbb{E}[(\operatorname{Tr} G)^2] + \frac{1}{N}\mathbb{E}[\operatorname{Tr}(GH_N)] = 0.$$

Now, $GH_N = G_N(z)(H_N - zI + zI) = I + zG_N(z)$. Thus, we obtain:

$$\mathbb{E}[g_N(z)^2] + 1 + z\mathbb{E}[g_N(z)] = 0. \quad \square$$

Lemma 3.1 *There exists a constant C (independent of N, z) such that:*

$$\mathbf{Var}(g_N(z)) \leq \frac{C}{N^2|\Im z|^2}. \quad (3.4)$$

Proof of Lemma 3.1: We can use an argument of concentration of measure for the Gaussian distribution: let $\gamma_{n,\sigma}$ denote the centered Gaussian distribution on \mathbb{R}^n of covariance $\sigma^2 I$ and F a Lipschitz function on \mathbb{R}^n (for the Euclidian norm) with constant c , then (see [L])

$$\gamma_{n,\sigma}[\{x, |F(x) - \int F d\gamma_{n,\sigma}| \geq \delta\}] \leq 2 \exp\left(-\frac{\kappa\delta^2}{c^2\sigma^2}\right). \quad (3.5)$$

for some (universal) constant κ . For $F : \mathbb{R} \rightarrow \mathbb{R}$, we define its extension on \mathcal{H}_N by $F(M) = U \operatorname{diag}(F(\lambda_1), \dots, F(\lambda_N)) U^*$ if $M = U \operatorname{diag}(\lambda_1, \dots, \lambda_N) U^*$. We have the following property:

If F is c -Lipschitz on \mathbb{R} , its extension to \mathcal{H}_N with the norm $\|M\|_2 = \sqrt{\text{tr}(M^2)}$ (isomorphic to \mathbb{R}^{N^2} with the Euclidian norm) is c -Lipschitz.

It follows that if f is c -Lipschitz on \mathbb{R} , then $F(M) := \frac{1}{N} \text{Tr}(f(M))$ is $\frac{c}{\sqrt{N}}$ -Lipschitz (use $|\text{Tr}(M)| \leq \sqrt{N}\|M\|_2$). Now, using that $g(x) := \frac{1}{x-z}$ is Lipschitz with constant $\frac{1}{|\Im z|}$, we obtain from (3.5):

$$\mathbb{P}[|g_N(z) - \mathbb{E}(g_N(z))| \geq \sqrt{\delta}] \leq 2 \exp\left(-\frac{\delta|\Im z|^2 N}{2/N}\right) = 2 \exp(-\delta|\Im z|^2 N^2/2).$$

Integrating the above inequality in δ gives (3.4). \square

Proof of Theorem 3.1: Set $h_N(z) = \mathbb{E}[g_N(z)]$. From (3.3) and Lemma 3.1, $h_N(z)$ satisfies

$$|h_N^2(z) + zh_N(z) + 1| \leq \frac{C}{N^2|\Im z|^2}$$

and $|h_N(z)| \leq \frac{1}{|\Im z|}$. The sequence $(h_N(z))_N$ is analytic on $\mathbb{C} \setminus \mathbb{R}$, uniformly bounded (as well as the derivatives) on the compact sets of $\mathbb{C} \setminus \mathbb{R}$. From a classical theorem (Montel [He]), $(h_N(z))_N$ is relatively compact. It is enough to prove that any limit point is unique. Let h a limit point. It is easy to see that $h(z)$ satisfies properties i)-iii) of Proposition 3.1 and the quadratic equation

$$h^2(z) + zh(z) + 1 = 0.$$

Therefore, $h(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4})$, the choice of the sign $+$ before the square root follows from iii). Thus, h is determined uniquely and $h_N(z) \rightarrow h(z)$ for $z \in \mathbb{C} \setminus \mathbb{R}$. Moreover, one can verify that h is the Stieltjes transform of the semicircular distribution μ_{sc} .

Now, using again Lemma 3.1 and some tools of complex analysis, we can prove (see [PL] for details) that a.s., g_N converge uniformly to h on compact of $\mathbb{C} \setminus \mathbb{R}$. From vi) of Proposition 3.1, it follows that μ_{H_N} converge to μ_{sc} a.s. \square

3.2 Some generalisations: the Wishart ensemble

Let $A = (A_{ij})$ a $p \times N$ matrix with iid complex Gaussian entries of variance 1 ($\mathbb{E}[|A_{ij}|^2] = 1$). Define $W_N = \frac{1}{N} A^* A$. W_N is a Hermitian positive random matrix of size N called a complex Wishart matrix (or Laguerre matrix).

Proposition 3.4 *Let $p, N \rightarrow \infty$ such that $\frac{p(N)}{N} \rightarrow_{N \rightarrow \infty} c \geq 1$. Then, the spectral measure μ_{W_N} of W_N converges a.s. to the probability measure given by*

$$\mu_{MP}(dx) = \frac{1}{2\pi x} ((x - a_-)(a_+ - x))^{1/2} 1_{[a_-, a_+]}(x) dx$$

where $a_{\pm} = (1 \pm \sqrt{c})^2$. μ_{MP} is called the Marchenko-Pastur distribution

Sketch of Proof: As in the GUE case, we show that $\mathbb{E}[g_{W_N}(z)]$ converges to a function h satisfying a quadratic equation. h is the Stieltjes transform of the Marchenko-Pastur distribution. This relies on:

1) the distribution of the Wishart matrix W_N (see [Mu]):

$$C_{N,p}(\det(M))^{p-N} \exp(-N \operatorname{Tr}(M)) 1_{(M \geq 0)} dM$$

2) an integration by part's formula: for $p > N$, W_N is invertible a.s. and

$$\mathbb{E}[\Phi'(W_N).A] - N\mathbb{E}[\Phi(W_N) \operatorname{Tr}(A)] + (p - N)\mathbb{E}[\Phi(W_N) \operatorname{Tr}(H_N^{-1}A)] = 0.$$

Apply the above formula to $\Phi(M) = (G_M(z)M)_{ij} = (I + zG_M(z))_{ij}$. After some computations as in the GUE case, we obtain:

$$z\mathbb{E}[g_{W_N}^2(z)] + 1 + z\mathbb{E}[g_{W_N}(z)] - \frac{p - N}{N}\mathbb{E}[g_{W_N}(z)] = 0$$

leading to $h(z)$ solution of

$$zh^2(z) + (z + (1 - c))h(z) + 1 = 0. \square$$

3.3 Comments

1) This method also holds for real symmetric matrices, that is for the Gaussian Orthogonal Ensemble (GOE) or for real Wishart matrices. The GOE(N , σ^2) ensemble is the distribution of the symmetric matrix S of size N where S_{ii} is distributed as $N(0, 2\sigma^2)$, S_{ij} , $i < j$ is distributed as $N(0, \sigma^2)$, all these variables being independent.

2) Wigner matrices can also be considered using Stieltjes transform. Instead of the formula $\mathbb{E}[\gamma\Phi(\gamma)] = \sigma^2\mathbb{E}[\Phi'(\gamma)]$ for γ a centered Gaussian distribution of variance σ^2 we can use 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 and $|\varepsilon_p| \leq \sup_{x \in \mathbb{R}} \Phi^{(p+1)}(x) |\mathbb{E}[|X|^{p+2}]$. See [KKP] for the development of this method.

Part II

Local Behavior

4 The Gaussian Unitary Ensemble

In this section, we continue the study of the GUE in a deeper way, in order to obtain the local behavior of the spectrum. Recall that $\text{GUE}(N, \frac{1}{N})$ is the Gaussian distribution on \mathcal{H}_N given by:

$$P_N(dM) = \frac{1}{Z_N} \exp\left(-\frac{1}{2}N \text{Tr}(M^2)\right) dM \quad (4.1)$$

4.1 Distribution of the eigenvalues of $\text{GUE}(N, \frac{1}{N})$

Let H_N a random matrix distributed as $\text{GUE}(N, \frac{1}{N})$ and we denote by $\lambda_1(H_N) \leq \dots \leq \lambda_N(H_N)$ the ranked eigenvalues of H_N .

Proposition 4.1 *The joint distribution of the eigenvalues $\lambda_1(H_N) \leq \dots \leq \lambda_N(H_N)$ has a density with respect to Lebesgue measure equal to*

$$p_N(x) = \frac{1}{Z_N} \prod_{1 \leq i < j \leq N} (x_j - x_i)^2 \exp\left(-\frac{N}{2} \sum_{i=1}^N x_i^2\right) 1_{x_1 \leq \dots \leq x_N} \quad (4.2)$$

$\Delta(x) = \prod_{1 \leq i < j \leq N} (x_j - x_i)$ is called the Vandermonde determinant and equals $\det(x_i^{j-1})_{1 \leq i, j \leq N}$.

We refer to Mehta [Me, Chap. 3] for the proof of this proposition. It relies on the expression of the N^2 components of M in (4.1) in terms of the N eigenvalues (x_i) and $N(N-1)$ independent parameters (p_i) which parametrize the unitary matrix U in the decomposition $M = U \text{diag}(x) U^*$. Heuristically, the term $\exp(-\frac{N}{2} \sum_{i=1}^N x_i^2)$ comes from the $\exp(-\frac{1}{2}N \text{Tr}(M^2))$ in P_N and the square of the Vandermonde determinant comes from the Jacobian of the map $M \mapsto ((x_i), U)$ after integration on U on the unitary group.

Corollary 4.1 *If f is a bounded function of \mathcal{H}_N , invariant by the unitary transformations, that is $f(M) = f(UMU^*)$ for all unitary matrix U then $f(M) = f(\lambda_1(M), \dots, \lambda_N(M))$ is a symmetric function of the eigenvalues*

and

$$\begin{aligned}
\mathbb{E}[f(H_N)] &= \frac{1}{Z_N} \int_{h_N} f(M) \exp\left(-\frac{1}{2}N \operatorname{Tr}(M^2)\right) dM \\
&= \frac{1}{Z_N} \int_{x_1 \leq \dots \leq x_N} f(x_1, \dots, x_N) \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \exp\left(-\frac{N}{2} \sum_{i=1}^N x_i^2\right) d^N x \\
&= \frac{1}{N! Z_N} \int_{\mathbb{R}^N} f(x_1, \dots, x_N) \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \exp\left(-\frac{N}{2} \sum_{i=1}^N x_i^2\right) d^N x
\end{aligned}$$

4.2 k -point correlation functions of the GUE

Let ρ_N a symmetric density distribution on \mathbb{R}^N , considered as the distribution of N particles X_i .

Definition 4.1 *Let $k \leq N$. The k -point correlation functions of (X_i) are defined by*

$$R_{N,k}(x_1, \dots, x_k) = \frac{N!}{(N-k)!} \int_{\mathbb{R}^{N-k}} \rho_N(x_1, \dots, x_N) dx_{k+1} \dots dx_N. \quad (4.3)$$

The correlation functions are, up to a constant, the marginal distributions of ρ_N . Heuristically, $R_{N,k}$ is the probability of finding a particle at x_1, \dots a particle at x_k . The factor $\frac{N!}{(N-k)!}$ comes from the choice of the k particles and the symmetry of ρ_N (see the computation below). We have, using the symmetry of ρ_N ,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i=1}^N (1 + f(X_i))\right] &= \mathbb{E}\left[\sum_{k=0}^N \sum_{i_1 < \dots < i_k} f(X_{i_1}) \dots f(X_{i_k})\right] \\
&= \sum_{k=0}^N \mathbb{E}\left[\sum_{i_1 < \dots < i_k} f(X_{i_1}) \dots f(X_{i_k})\right] \\
&= \sum_{k=0}^N \binom{N}{k} \mathbb{E}[f(X_1) \dots f(X_k)] \\
&= \sum_{k=0}^N \frac{1}{k!} \frac{N!}{(N-k)!} \mathbb{E}[f(X_1) \dots f(X_k)]
\end{aligned}$$

and thus,

$$\mathbb{E}\left[\prod_{i=1}^N (1 + f(X_i))\right] = \sum_{k=0}^N \frac{1}{k!} \int_{\mathbb{R}^k} f(x_1) \cdots f(x_k) R_{N,k}(x_1, \cdots, x_N) dx_1 \cdots dx_k. \quad (4.4)$$

The correlation functions enables to express probabilistic quantities as:

1) *The hole probability:*

Take $f(x) = 1_{\mathbb{R} \setminus I} - 1$ where I is a Borel set of \mathbb{R} . Then, the left-hand side of (4.4) is the probability of having no particles in I . Therefore,

$$\mathbb{P}(\forall i, X_i \notin I) = \sum_{k=0}^N \frac{(-1)^k}{k!} \int_{I^k} R_{N,k}(x_1, \cdots, x_k) dx_1 \cdots dx_k.$$

In particular, for $I =]a, +\infty[$,

$$\mathbb{P}(\max X_i \leq a) = \sum_{k=0}^N \frac{(-1)^k}{k!} \int_{[a, \infty[^k} R_{N,k}(x_1, \cdots, x_k) dx_1 \cdots dx_k. \quad (4.5)$$

2) *the density of state:*

$$\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N f(X_i)\right] = \frac{1}{N} \int_{\mathbb{R}} f(x) R_{N,1}(x) dx$$

that is $\frac{1}{N} R_{N,1}(x) dx$ represents the expectation of the empirical distribution $\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N \delta_{X_i}\right]$.

We now compute the correlation functions associated to the symmetric density of the (unordered) eigenvalues of the GUE

$$\rho_N(x) = \frac{1}{N! \bar{Z}_N} \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \exp\left(-\frac{N}{2} \sum_{i=1}^N x_i^2\right).$$

Proposition 4.2 *The correlation functions of the GUE($N, \frac{1}{N}$) are given by*

$$R_{N,k}(x_1, \cdots, x_k) = \det(K_N(x_i, x_j))_{1 \leq i, j \leq k} \quad (4.6)$$

where the kernel K_N is given by

$$K_N(x, y) = \exp\left(-\frac{N}{4}(x^2 + y^2)\right) \sum_{l=0}^{N-1} q_l(x) q_l(y) \quad (4.7)$$

where

$$q_l(x) = \left(\frac{N}{2\pi}\right)^{1/4} \frac{1}{\sqrt{2^l l!}} h_l(\sqrt{N/2} x) \quad (4.8)$$

where h_l are the Hermite polynomials.

The process of the eigenvalues of GUE is said to be a **determinantal process**.

Proof: Since the value of a determinant does not change if we replace a column by the column + a linear combination of the others, we have that the Vandermonde determinant $\Delta(x) = \det(P_{j-1}(x_i))$ if P_j denotes a polynomial of degree j with higher coefficient equal to 1.

Let $w(x) = \exp(-\frac{N}{2}x^2)$ and define the orthonormal polynomials q_l with respect to w such that:

- q_l is of degree l , $q_l(x) = a_l x^l + \dots$ with $a_l > 0$.

- $\int_{\mathbb{R}} q_l(x) q_p(x) w(x) dx = \delta_{pl}$.

$(q_l)_l$ also depends on N and is up to a scaling factor the family of Hermite polynomials (to be discussed later).

Thus, $\Delta(x) = C_N \det(q_{j-1}(x_i))$ and, using $(\det(A))^2 = \det(A) \det(A^T)$, we have:

$$\begin{aligned} \rho_N(x) &= \frac{1}{\tilde{Z}_N} \prod_{1 \leq i \leq N} w(x_i) (\det(q_{j-1}(x_i)))^2 \\ &= \frac{1}{\tilde{Z}_N} \prod_{1 \leq i \leq N} w(x_i) \det \left(\sum_{l=1}^N q_{l-1}(x_i) q_{l-1}(x_j) \right)_{i,j \leq N} \\ &= \frac{1}{\tilde{Z}_N} \det (K_N(x_i, x_j))_{i,j \leq N} \end{aligned}$$

where

$$K_N(x, y) = \sqrt{w(x)} \sqrt{w(y)} \sum_{l=1}^N q_{l-1}(x) q_{l-1}(y) = \sum_{l=0}^{N-1} \phi_l(x) \phi_l(y)$$

where $\phi_l(x) = \sqrt{w(x)} q_l(x)$. The sequence $(\phi_l)_l$ is orthonormal for the Lebesgue measure dx . From the orthonormality of (ϕ_l) , it is easy to show that the kernel K_N satisfies the properties:

$$\int_{\mathbb{R}} K_N(x, x) dx = N$$

$$\int_{\mathbb{R}} K_N(x, y)K_N(y, z)dy = K_N(x, z).$$

This proves (4.6) for $k = N$ (up to a constant). The general case follows from the Lemma:

Lemma 4.1 *Let $J_N = (J_{ij})$ a matrix of size N of the form $J_{ij} = f(x_i, x_j)$ with f satisfying:*

1. $\int_{\mathbb{R}} f(x, x)dx = C$
2. $\int_{\mathbb{R}} f(x, y)f(y, z)dy = f(x, z)$

Then,

$$\int_{\mathbb{R}} \det(J_N)dx_N = (C - N + 1) \det(J_{N-1})$$

where J_{N-1} is a matrix of size $N - 1$ obtained from J_N by removing the last row and column containing x_N .

Proof of Lemma 4.1:

$$\det(J_N) = \sum_{\sigma \in \Sigma_N} \epsilon(\sigma) \prod_{i=1}^N f(x_i, x_{\sigma(i)})$$

where Σ_N is the set of permutations on $\{1, \dots, N\}$ and ϵ stands for the signature of a permutation. We integrate in dx_N . There are two cases: i) if $\sigma(N) = N$, the integration of $f(x_N, x_N)$ gives C ; the set of such σ is isomorphic to Σ_{N-1} and the signature (in Σ_{N-1}) of the restriction to $\{1, \dots, N - 1\}$ is the same as the signature of σ . This gives the term $C \det(J_{N-1})$.

ii) $\sigma(N) \in \{1, \dots, N - 1\}$. For such σ , the integral in dx_N gives

$$A_\sigma := \prod_{i=1, i \neq \sigma^{-1}(N)}^{N-1} f(x_i, x_{\sigma(i)}) \int f(x_{\sigma^{-1}(N)}, x_N) f(x_N, x_{\sigma(N)}) dx_N$$

$$A_\sigma = \prod_{i=1, i \neq \sigma^{-1}(N)}^{N-1} f(x_i, x_{\sigma(i)}) f(x_{\sigma^{-1}(N)}, x_{\sigma(N)}) = \prod_{i=1}^{N-1} f(x_i, x_{\hat{\sigma}(i)})$$

where $\hat{\sigma}$ is a permutation on Σ_{N-1} which coincides with σ for $i \leq N - 1, i \neq \sigma^{-1}(N)$ and $\hat{\sigma}(\sigma^{-1}(N)) = \sigma(N) \neq N$. Now, it remains to see that $\hat{\sigma}$ can be obtained from $N - 1$ distinct permutations σ or in the other sense, from

$\hat{\sigma}$, we can construct $N - 1$ permutations σ by inserting N between any two indices $i, \hat{\sigma}(i)$. Moreover, since the number of cycles of σ and $\hat{\sigma}$ is the same, $\epsilon(\sigma) = -\epsilon(\hat{\sigma})$. We thus obtain the formula of the lemma. \square

In the case of GUE, $J = (K_N(x_i, x_j))$ satisfies the hypothesis of the lemma with $C = N$.

$$\int_{\mathbb{R}} \det(K_N(x_i, x_j))_{i,j \leq N} dx_N = (N - N - 1) \det(K_N(x_i, x_j))_{i,j \leq N-1}$$

$$\int_{\mathbb{R}} \det(K_N(x_i, x_j))_{i,j \leq N-1} dx_{N-1} = (N - N - 2) \det(K_N(x_i, x_j))_{i,j \leq N-2}.$$

Integrating over all the variables gives:

$$\int_{\mathbb{R}} \det(K_N(x_i, x_j))_{i,j \leq N} dx_1 \dots dx_N = N!$$

and therefore, $\tilde{Z}_N = N!$. Integrating over the $N - k$ variables dx_{k+1}, \dots, dx_N gives:

$$\int_{\mathbb{R}} \det(K_N(x_i, x_j))_{i,j \leq N} dx_1 \dots dx_N = (N - k)! \det(K_N(x_i, x_j))_{i,j \leq k}$$

and

$$\begin{aligned} R_{k,N}(x_1, \dots, x_k) &= \frac{N!}{(N - k)!} \int \rho_N(x_1, \dots, x_N) dx_{k+1} \dots dx_N \\ &= \frac{1}{(N - k)!} \int \det(K_N(x_i, x_j))_{i,j \leq N} dx_{k+1} \dots dx_N \\ &= \det(K_N(x_i, x_j))_{i,j \leq k} \end{aligned}$$

This proves (4.6) and (4.7). It remains to determine the polynomials q_l . Let h_l the Hermite polynomial of degree l defined by:

$$h_l(x) = (-1)^l e^{x^2} \left(\frac{d}{dx} \right)^l (e^{-x^2}).$$

These polynomials (see [S]) are orthogonal with respect to $e^{-x^2} dx$, $\int_{\mathbb{R}} h_l^2(x) e^{-x^2} dx = 2^l l! \sqrt{\pi}$ and the coefficient of x^l in h_l is 2^l . Then, it is easy to see that q_l given by (4.8) are orthonormal with respect to $\exp(-\frac{N}{2}x^2) dx$. \square

Corollary 4.2 Let $\bar{\mu}_{H_N}(dx) = \mathbb{E}[\mu_{H_N}(dx)]$ where μ_{H_N} is the spectral distribution of $GUE(N, \frac{1}{N})$, then $\bar{\mu}_{H_N}(dx)$ is absolutely continuous with respect to Lebesgue measure with density f_N given by:

$$f_N(x) = \frac{1}{N}R_{N,1}(x, x) = \frac{1}{N}K_N(x, x), \quad x \in \mathbb{R}.$$

f_N is called the density of state.

4.3 The local regime

Let us denote, for I a Borel set of \mathbb{R} , $\nu_N(I) = \#\{i \leq N; \lambda_i \in I\} = N\mu_{H_N}(I)$ where λ_i are the eigenvalues of $GUE(N, \frac{1}{N})$. From Wigner's theorem, as $N \rightarrow \infty$, $\nu_N(I) \sim N \int_I f_{sc}(x)dx$ a.s. where f_{sc} is the density of the semi-circular distribution μ_{sc} . The spacing between eigenvalues is of order $1/N$. In the local regime, we consider an interval I_N whose size tends to 0 as $N \rightarrow \infty$. Two cases have to be considered.

a) Inside the bulk: Take $I_N = [u - \varepsilon_N, u + \varepsilon_N]$ with u such that $f_{sc}(u) > 0$ that is $u \in]-2, 2[$. Then, $\nu_N(I_N)$ has the order of a constant for $\varepsilon_N \sim \frac{1}{N}$. This suggest to introduce new random variables (renormalisation) l_i by

$$\lambda_i = u + \frac{l_i}{Nf_{sc}(u)}, \quad i = 1, \dots, N.$$

The mean spacing between the rescaled eigenvalues l_i is 1. Straightforward computations give:

Lemma 4.2 The correlation functions R^{bulk} of the distribution of (l_1, \dots, l_N) are given in terms of the correlation functions of the (λ_i) by

$$R_{N,k}^{bulk}(y_1, \dots, y_k) = \frac{1}{(Nf_{sc}(u))^k} R_{N,k}(u + \frac{y_1}{Nf_{sc}(u)}, \dots, u + \frac{y_k}{Nf_{sc}(u)}). \quad (4.9)$$

We shall see in the next subsection the asymptotic of the correlation functions R^{bulk} (or the kernel K_N).

b) At the edge of the spectrum: $u = 2$ (or -2). $f_{sc}(u) = 0$.

$$\nu_N([2 - \varepsilon_N, 2]) = \frac{N}{2\pi} \int_{2-\varepsilon}^2 \sqrt{4-x^2} dx = \frac{N}{2\pi} \int_0^\varepsilon \sqrt{4y-y^2} dy \sim CN\varepsilon^{3/2}.$$

So the normalisation at the edge is $\varepsilon = \frac{1}{N^{2/3}}$ and we define the rescaled correlation functions by:

$$R_{N,k}^{edge}(y_1, \dots, y_k) = \frac{1}{(N^{2/3})^k} R_{N,k}(2 + \frac{y_1}{N^{2/3}}, \dots, 2 + \frac{y_k}{N^{2/3}}). \quad (4.10)$$

From (4.5) and (4.10),

$$\mathbb{P}[N^{2/3}(\lambda_{max} - 2) \leq a] = \sum_{k=0}^N \frac{(-1)^k}{k!} \int_{[a, \infty[^k} R_{N,k}^{edge}(x_1, \dots, x_k) dx_1 \cdots dx_k. \quad (4.11)$$

where λ_{max} is the maximal eigenvalue of the GUE.

The asymptotic of R^{edge} will be given in the next section.

4.4 Limit kernel

The asymptotic of the correlation functions relies on asymptotic formulas for the orthonormal polynomials q_l for $l \sim N$. We have the following:

Proposition 4.3 (*Plancherel - Rotach formulas, [S]*)

Let $(h_n)_n$ denote the Hermite polynomials.

1) If $x = \sqrt{2n+1} \cos(\Phi)$ with $\varepsilon \leq \Phi \leq \pi - \varepsilon$,

$$\exp(-x^2/2)h_n(x) = b_n(\sin(\Phi))^{-1/2} \left\{ \sin\left[\left(\frac{n}{2} + \frac{1}{4}\right)(\sin(2\Phi) - 2\Phi) + 3\pi/4\right] + O\left(\frac{1}{n}\right) \right\} \quad (4.12)$$

where $b_n = 2^{n/2+1/4}(n!)^{1/2}(\pi n)^{-1/4}$.

2) If $x = \sqrt{2n+1} + 2^{-1/2}n^{-1/6}t$, t bounded in \mathbb{C} ,

$$\exp(-x^2/2)h_n(x) = \pi^{1/4}2^{n/2+1/4}(n!)^{1/2}(n)^{-1/12} \left\{ Ai(t) + O\left(\frac{1}{n}\right) \right\}$$

where Ai is Airy's function, that is the solution of the differential equation

$$y'' = xy \text{ with } y(x) \sim_{x \rightarrow +\infty} \frac{1}{2\sqrt{\pi}} x^{-1/4} \exp\left(-\frac{2}{3}x^{3/2}\right).$$

From these formulas, one can show:

Theorem 4.1

$$\lim_{N \rightarrow \infty} R_{n,k}^{bulk}(y_1, \dots, y_k) = \det(K^{bulk}(y_i, y_j))_{i,j \leq k} \quad (4.13)$$

where

$$K^{bulk}(x, y) = \frac{\sin(\pi(x - y))}{\pi(x - y)} \quad (4.14)$$

$$\lim_{N \rightarrow \infty} R_{n,k}^{edge}(y_1, \dots, y_k) = \det(K^{edge}(y_i, y_j))_{i,j \leq k} \quad (4.15)$$

where

$$K^{edge}(x, y) = \frac{Ai(x)Ai'(y) - Ai'(x)Ai(y)}{(x - y)} \quad (4.16)$$

Sketch of Proof of (4.13): From (4.9), (4.6), we may find the limit of

$$\frac{1}{Nf_{sc}(u)} K_N(u + \frac{s}{Nf_{sc}(u)}, u + \frac{t}{Nf_{sc}(u)}).$$

We express the kernel K_N given by (4.7) thanks to Cristoffel-Darboux formula (see Appendix)

$$\begin{aligned} K_N(X, Y) &= \frac{k_{N-1} q_N(X)q_{N-1}(Y) - q_N(Y)q_{N-1}(X)}{k_N X - Y} \exp(-\frac{N}{4}(X^2 + Y^2)) \\ K_N(X, Y) &= \frac{1}{2^N(N-1)!\sqrt{\pi}} \frac{h_N(\sqrt{N/2} X)h_{N-1}(\sqrt{N/2} Y) - h_N(\sqrt{N/2} Y)h_{N-1}(\sqrt{N/2} X)}{X - Y} \\ &\quad \times \exp(-\frac{N}{4}(X^2 + Y^2)) \end{aligned}$$

with k_N the highest coefficient in q_N . Then, set $X = u + \frac{s}{Nf_{sc}(u)}$, $Y = u + \frac{t}{Nf_{sc}(u)}$, $u = 2 \cos(\Phi)$. Then, $f_{sc}(u) = \frac{\sin(\Phi)}{\pi}$ and

$$x = \sqrt{N/2}X = \sqrt{2N}(\cos(\Phi) + \frac{\pi s}{2N \sin(\Phi)}).$$

In order to use Plancherel-Rotach formulas, we express x as

$$x = \sqrt{2N+1} \cos(\Phi_N).$$

A development gives

$$\Phi_N = \Phi + \frac{a}{2N} + O(\frac{1}{N^2})$$

with $a = \frac{1}{2 \tan(\Phi)} - \frac{\pi s}{\sin^2(\Phi)}$. Then,

$$\sin(2\Phi_N) - 2\Phi_N = (\sin(2\Phi) - 2\Phi) + \frac{a}{N}(\cos(2\Phi) - 1) + O(\frac{1}{N^2})$$

and

$$(\sin(\Phi_N))^{-1/2} = (\sin(\Phi))^{-1/2}(1 + O(1/N)).$$

Formula (4.12) gives:

$$\exp(-x^2/2)h_N(x) = b_N(\sin(\Phi))^{-1/2}\left\{\sin\left[\left(\frac{N}{2} + \frac{1}{4}\right)(\sin(2\Phi) - 2\Phi) + \frac{a}{2}(\cos(2\Phi) - 1) + \frac{3\pi}{4}\right] + O(1/N)\right\}$$

We make the same transformations for $e^{-x^2/2}h_{N-1}(x)$, $e^{-y^2/2}h_N(y)$, $e^{-y^2/2}h_{N-1}(y)$ giving ϕ'_N , Ψ_N and Ψ'_N associated respectively to:

$$a' = -\frac{1}{2 \tan(\Phi)} - \frac{\pi s}{\sin^2(\Phi)}, \quad b = \frac{1}{2 \tan(\Phi)} - \frac{\pi t}{\sin^2(\Phi)}, \quad b' = -\frac{1}{2 \tan(\Phi)} - \frac{\pi t}{\sin^2(\Phi)}.$$

Then, we replace in the product $h_N(x)h_{N-1}(y)$ the product of two sinus by a trigonometric formula and then in the difference, we obtain a linear combination of cosinus, The difference of two of them cancels using that $a' + b = a + b'$. Then, we use again a trigonometric formula. After some computations, the kernel K^{bulk} appears. The Airy kernel appears, using the second formula of Plancherel-Rotach.

Corollary 4.3 (*Fluctuations of λ_{max}*)

The fluctuations of the largest eigenvalue of the GUE around 2 are given by:

$$\mathbb{P}(N^{2/3}(\lambda_{max} - 2) \leq x) = F_2(x)$$

where F_2 is called the Tracy-Widom distribution and is given by

$$F_2(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{]x, \infty[^k} \det(K^{edge}(y_i, y_j))_{1 \leq i, j \leq k} d^k y.$$

F_2 can be written $F_2(x) = \det(I - \mathbf{K})_{L^2(x, \infty)}$ where \mathbf{K} is the integral operator on L^2 with kernel $K^{edge}(x, y)$ and the det is the Fredholm determinant.

4.5 Comments

1. The computation of the correlation functions which have a determinantal form is specific to the unitary case and do not hold for the GOE case.
2. We refer to [D], [Me] for others computations involving correlation functions such as the spacing distribution.

3. The Tracy-Widom distribution can also be expressed as

$$F_2(x) = \exp\left(-\int_x^\infty (y-x)q^2(y)dy\right)$$

where $q''(x) = xq(x) + q^3(x)$ with $q(x) = Ai(x)(1 + o(1))$ as $x \rightarrow \infty$. The function q is called the solution of Painlevé II equation (see [TW]).

4. One of the important ideas of the theory is that of **universality**. This idea, in fact a conjecture, is that the statistical properties of the eigenvalues in the local scale do not depend asymptotically on the ensemble, that is the sine kernel (4.14) is "universal" and appears in other models of Hermitian random matrices.

This has been shown for

- a class of Hermitian Wigner matrices: Soshnikov [S] (for the edge), Johansson [J]. See also the recent preprints on ArXiv (Erdos, Ramirez, Schlein, Yau and Tao, Vu).
- unitary invariant ensemble of the form

$$P_N(dM) = C_N \exp(-N \operatorname{Tr}(V(M)))dM$$

for a weight V satisfying some assumptions. See [DKMVZ], [PS]. Note that the GUE corresponds to the quadratic weight $V(x) = \frac{1}{2}x^2$. For example, for the Wishart ensemble (associated to the Laguerre polynomials), we have the same asymptotic kernel as in the GUE, while the density of state is not universal (semicircular for GUE and Marchenko-Pastur distribution for Wishart). The main difficulty for general V is to derive the asymptotics of orthogonal polynomials. This can be done using Riemann-Hilbert techniques (see [D]).

5 Appendix

5.1 Orthogonal polynomials (see [D], [S])

Let $w(x)$ a positive function on \mathbb{R} such that $\int_{\mathbb{R}} |x|^m w(x) dx < \infty$ for all $m \geq 0$. On the space of real polynomials $P[X]$, we consider the scalar product

$$(P|Q) = \int_{\mathbb{R}} P(x)Q(x)w(x)dx.$$

Then the orthogonalisation procedure of Schmidt enables to construct of sequence of orthogonal polynomials (p_l) : p_l is of degree l and

$$\int_{\mathbb{R}} p_m(x)p_n(x)w(x)dx = 0 \text{ if } m \neq n.$$

We denote by a_l the coefficient of x^l in $p_l(x)$ and $d_l = \int_{\mathbb{R}} p_l(x)^2w(x)dx$.

Example: If $w(x) = \exp(-x^2)$, the Hermite polynomials h_l are orthogonal with $a_l = 2^l$ and $d_l = 2^l l! \sqrt{\pi}$.

Christoffel-Darboux formula: We consider a family of orthonormal polynomials (p_l) ($d_l = 1$) for the weight w . We denote by K_n the kernel defined by:

$$K_n(x, y) = \sum_{l=0}^{n-1} p_l(x)p_l(y).$$

K_n is the kernel associated to the orthogonal projection in the space of polynomials of degree less than $n - 1$. This kernel has a simple expression based upon a three terms recurrence relation between the (p_l) :

$$xp_n(x) = \alpha_n p_{n+1}(x) + \beta_n p_n(x) + \alpha_{n-1} p_{n-1}(x)$$

for some coefficients $\alpha_n = \frac{a_n}{a_{n+1}}$ and β_n (depending on a_n and the coefficient b_n of x^{n-1} in p_n).

From this relation, one obtains:

$$K_n(x, y) = \alpha_{n-1} \frac{p_n(x)p_{n-1}(y) - p_n(y)p_{n-1}(x)}{x - y} \quad (5.1)$$

For the orthonormal polynomials q_l defined in (4.8),

$$a_l (= a_{N,l}) = \left(\frac{N}{2\pi}\right)^{1/4} \left(\sqrt{\frac{N}{l!}}\right)^l$$

and $\alpha_{N-1} = 1$.

5.2 Fredholm determinant

Let $K(x, y)$ a bounded measurable kernel on a space (X, μ) where μ is a finite measure on X . The Fredholm determinant of K is defined by

$$D(\lambda) = \det(I - \lambda K) := 1 + \sum_{k=1}^{\infty} \frac{(-\lambda)^k}{k!} \int_{X^k} \det(K(x_i, x_j))_{1 \leq i, j \leq k} \mu(dx_1) \dots \mu(dx_k).$$

The serie converges for all λ .

This can be extended to a space with a σ finite measure and a kernel of trace class (see [Si]).

References

- [AG] Akhiezer, N. I. and Glazman, I. M. *Theory of linear operators in Hilbert space*. Vol. 1 Dover Publications, Inc., New York, 1993.
- [BY] Bai, Z. D.; Yin, Y. Q. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* 16 (1988), no. 4, 1729–1741.
- [B] Bai, Z.D. Methodologies in spectral analysis of large-dimensional random matrices, a review. With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author. *Statist. Sinica* 9 (1999), no. 3, 611–677.
- [BS] Bai, Z.D and Silverstein, J. *Spectral Analysis of Large Dimensional Random matrices*. Mathematics Monograph Series 2, Science Press, Beijing, 2006.
- [DKMVZ] Deift, P.; Kriecherbauer, T.; McLaughlin, K. T-R; Venakides, S.; Zhou, X. Uniform asymptotics for polynomials orthogonal with respect to varying exponential weights and applications to universality questions in random matrix theory. *Comm. Pure Appl. Math.* 52 (1999), no. 11, 1335–1425.
- [D] Deift, P.A. *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*. Courant Lecture Notes in Mathematics, 3. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1999.
- [G] Guionnet, A. Large random matrices: lectures on macroscopic asymptotics. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006. *Lecture Notes in Mathematics*, 1957. Springer-Verlag, Berlin, 2009.
- [He] Henrici, P. Applied and computational complex analysis. Vol. 2. Wiley Interscience [John Wiley & Sons], New York-London-Sydney, 1977.

- [HP] Hiai, F. and Petz, D. The semicircle law, free random variables and entropy. *Mathematical Surveys and Monographs*, 77. American Mathematical Society, Providence, RI, 2000.
- [J] Johansson, K. Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices. *Comm. Math. Phys.* 215 (2001), no. 3, 683–705.
- [KKP] Khorunzhy, A., Khoruzhenko, B.A., Pastur, L.A.: On asymptotic properties of large random matrices with independent entries. *J. Math. Phys.* 37, 5033–5060 (1996).
- [L] Ledoux, M. The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 89. American Mathematical Society, Providence, RI, 2001.
- [Me] Mehta, M.L. *Random matrices*. Third edition. *Pure and Applied Mathematics (Amsterdam)*, 142. Elsevier/Academic Press, Amsterdam, 2004.
- [Mu] Muirhead, R.J. *Aspects of multivariate statistical theory*. *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1982.
- [PS] Pastur, L. and Shcherbina, M. Universality of the local eigenvalue statistics for a class of unitary invariant random matrix ensembles. *J. Statist. Phys.* 86 (1997), no. 1-2, 109–147.
- [PL] Pastur, L and Lejay, A. Matrices aléatoires: statistique asymptotique des valeurs propres. (French) [Random matrices: asymptotic eigenvalue statistics] *Séminaire de Probabilités, XXXVI*, 135–164, *Lecture Notes in Math.*, 1801, Springer, Berlin, 2003.
- [Si] Simon, B. *Trace ideals and their applications*. Second edition. *Mathematical Surveys and Monographs*, 120. American Mathematical Society, Providence, RI, 2005.
- [S] Szego, G. *Orthogonal polynomials*. Fourth edition. *American Mathematical Society, Colloquium Publications*, Vol. XXIII. American Mathematical Society, Providence, R.I., 1975.

- [TW] Tracy, C.A and Widom, H. Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* 159 (1994), no. 1, 151–174.
- [Wig] E. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* 67 1958 325–327.
- [Wis] J. Wishart The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* 20 (1928), 35-52.