INTRODUCTION TO RANDOM MATRICES

FRANÇOIS CHAPON

University of Havana, December 2012

1. Introduction

Random matrices appear in a wide range of mathematics and theoretical physics, such as combinatorics, operator algebras, noncommutative probability, quantum field theory, statistical mechanics... They also have many interesting applications in sciences and engineering, like for instance in telecommunications processing. They were first introduced in multivariate statistics in the thirties by Wishart [19] and in theoretical physics in the fifties by Wigner in his fundamental article [18]. This lectures give a brief introduction to one aspect of random matrix theory, the asymptotic behavior of the eigenvalues of random Hermitian matrices as the dimension goes to infinity. 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. On the other hand, the microscopic regime can also be investigated, as in the central limit theorem we make a renormalization of certain probabilistic quantities to obtain nondegenerate limits, such as $N\mu_{X_N}(B_N)$ where now the size of B_N goes to zero. This regime is more delicate to study, and we will only consider here the case of matrices with Gaussian coefficients, known as the Gaussian Unitary Ensemble (GUE), where the distribution of the eigenvalues is explicitly known.

Different techniques are commonly used in random matrix theory, and we will present the usual ones: the moment method, the Stieltjes transform, and the orthogonal polynomials approach. This lecture will try to be self-contained and only standard notions of probability theory will be needed. To go further, a good reference is the book by Anderson, Guionnet, and Zeitouni [1], and indeed this lecture will be mostly inspired by it. The lectures notes [6] and [7] are also a good introduction to random matrix theory, the latter presenting the deep connection with systems of particles.

The aim of this lectures is to give an introduction to the most standard results in random matrix theory, but also to present different techniques commonly used in the field. They are 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. We will also mention without proof that the extremal eigenvalues converge to the edge of the support of the semicircular distribution. 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, using an orthogonal polynomials approach, we will provide a much deeper analysis of the Gaussian Unitary Ensemble, such as the density of eigenvalues, correlation functions, and links with the so-called determinantal processes. Using Laplace method, we will see the local asymptotics of the eigenvalues of the Gaussian Unitary Ensemble. Chapter 5 briefely presents some applications of random matrix theory to telecommunications processing. At last, Chapter 6 is an appendix where we recall some complex analysis tools, and where proofs of some technical results used in the notes are postponed.

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 $1 \leq i,j \leq N$.
- We often drop the dependance in the dimension in matrix notation for readability.
- The cardinal of a set A is denoted either $\#A$ or $|A|$.

2. Global behavior. The Wigner theorem

Random matrix theory has been widely developed since Wigner's work in the fifties [18]. 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

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

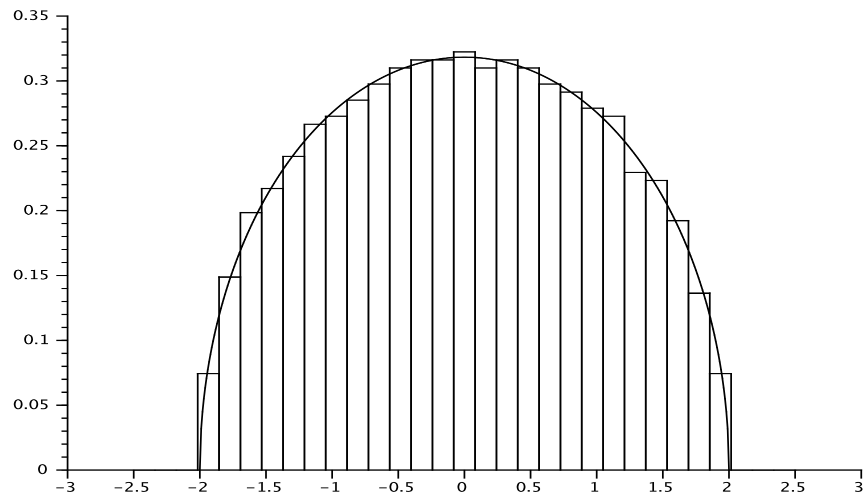

FIGURE 1. Histogram of the eigenvalues of a 1000×1000 GUE matrix and the semicircular distribution.

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

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

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

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

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

and one can prove that the determinant of the change of variables $X \mapsto UXU^*$ is equal to 1.

Figure 1 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 $n \geq 1$,*

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

As the next lemma shows, 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

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

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 $+1$ or -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.*

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 Figures 2 and 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 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 come 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. $\square$

Lemma 2.11. *Let D_n be the number of Dyck paths with $2n$ steps. Then we have $D_n = C_n$ where the C_n's are the Catalan numbers.*

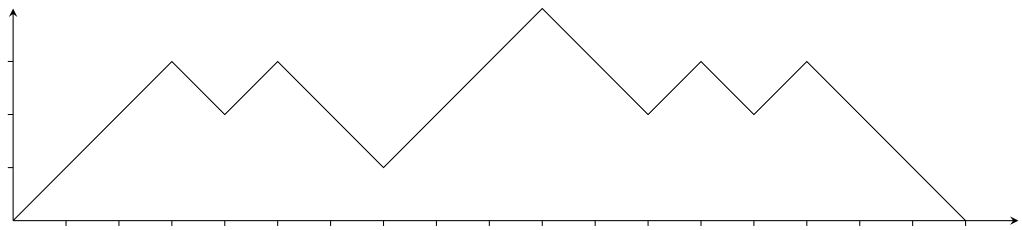

FIGURE 2. A Dyck path with 18 steps.

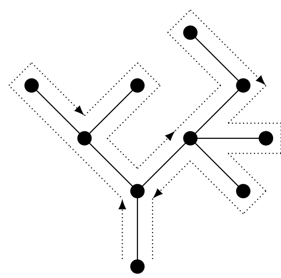

FIGURE 3. The rooted oriented tree corresponding to the Dyck path of Figure 2. The dashed line corresponds to the walk that surrounds the tree.

Proof. We prove that the D_n's satisfy the relations $D_0 = 1$ and

$$D_n = \sum_{l=1}^{n} D_{l-1} D_{n-l},$$

and show that this relation characterizes the Catalan numbers.

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

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

By shifting $2l$ to 0, we have that $\#\{\text{Dyck paths from } (2l,0) \text{ to } (2n,0)\} = D_{n-l}$. 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 $+1$, we get that

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

Hence,

$$D_n = \sum_{l=1}^{n} D_{l-1} D_{n-l}.$$

We next show that this relation characterizes the Catalan numbers. Define the series

$$S(z) = \sum_{k \geq 0} D_k z^k,$$

which is absolutely convergent for $|z| \leq 1/4$ since $D_k \leq 2^k$. Then the recurrence relation above gives

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

Hence we get that $S(z) = \frac{1-\sqrt{1-4z}}{2z}$, the minus branch being determined by the fact that $S(0) = 1$. Now the usual development in Taylor series

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

yields that $S(z) = \sum_{k\geq 0} \frac{(2k)!}{k!(k+1)!}z^k$, thus $D_k = C_k$. □

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

Lemma 2.12. *Let $G = (V, E)$ a connected graph. Then,*

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

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

Proof. We prove the lemma by induction over $|V|$. This is obviously true for $|V| = 1$. Now suppose $|V| = n$. Take a vertex v in V, and let $e_1, \ldots, e_l$ be the edges containing v, for some $l \geq 1$. Split the graph G into the graph with edges $\{e_1, \ldots, e_l\}$ and $G_1, \ldots, G_r$ connected graph with $r \leq l$. Let $G_i = (V_i, E_i)$, for $i = 1, \ldots, r$. By induction hypothesis, we have $|V_i| \leq |E_i| + 1$. Hence, since

$$\begin{aligned} |V| - 1 &= \sum_{i=1}^{r} |V_i| \\ |E| - l &= \sum_{i=1}^{r} |E_i|, \end{aligned}$$

we have

$$|V| = \sum_{i=1}^{r} |V_i| + 1 \leq \sum_{i=1}^{r} |E_i| + r + 1 = |E| - l + r + 1 \leq |E| + 1.$$

If $|V| = |E| + 1$, we claim that G is a tree, that is G does not have any loops. For this equality to hold, we must have equality in all the previous decomposition. But if there is a loop in G, we can find a vertex v with $r < l$. □

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 2.13. *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,

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

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,

$$\begin{aligned} \mathbb{E}\Big(\frac{1}{N} \operatorname{Tr}(H^k)\Big) &= \frac{1}{N} \mathbb{E}(\sum_{i_1, \ldots, i_k = 1}^{N} H_{i_1 i_2} H_{i_2 i_3} \cdots H_{i_k i_1}) \\ &= \frac{1}{N^{k/2+1}} \sum_{i_1, \ldots, i_k = 1}^{N} \mathbb{E}(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}). \end{aligned} \tag{1}$$

Let $I = (i_1, \ldots, i_k)$, and put $P(I) = \mathbb{E}(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1})$. Then, since by assumption $\sup_N \sup_{i,j} \mathbb{E}(|X_{ij}|^k) < \infty$, we have by Hölder's inequality that

$$P(I) \leq a_k,$$

where a_k is a constant independent of N. But 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.

To I we associate the connected graph $G(I) = (V(I), E(I))$, where the vertices $V(I) = \{i_1, \ldots, i_k\}$ and the edges are given by $E(I) = \{(i_1, i_2), \ldots, (i_k, i_{k+1})\}$, with $i_{k+1} = i_1$. Note that $G(I)$ may contain cycles. The skeleton of $G(I)$ is the connected graph given by $\widetilde{G}(I) = (\widetilde{V}(I), \widetilde{E}(I))$, where $\widetilde{V}(I)$ is the set of distinct points of $V(I)$, and $\widetilde{E}(I)$ the corresponding undirected edges without multiplicities, that is $\widetilde{G}(I)$ is the graph $G(I)$ where multiplicities and orientation have been erased.

Let I such that $P(I) > 0$, then each undirected edge appears at least twice, hence $|\widetilde{E}(I)| \leq \lfloor k/2 \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of x, and by lemma 2.12, we have $|\widetilde{V}(I)| \leq \lfloor k/2 \rfloor + 1$. 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

$$\mathbb{E}\Big(\frac{1}{N} \operatorname{Tr}(H^k)\Big) \leq a_k N^{\lfloor k/2 \rfloor - k/2}.$$

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 (1) are those for which $|\widetilde{V}(I)|$ is exactly equal to $\frac{k}{2}+1$, Lemma 2.12 implies that $\widetilde{G}(I)$ is a tree and $|\widetilde{E}(I)| = \frac{k}{2}$. Then, since $|E(I)| = k$, we have that each undirected edge appears exactly twice in $E(I)$, indeed once in each orientation, since we can explore $G(I)$ by the path $i_1 \to i_2 \to \cdots \to i_k \to i_1$. Thus, $G(I)$ appears as a fat tree, and $\widetilde{G}(I)$ is an oriented rooted tree, the root is given by the directed edge (i_1, i_2) and the order of the indices induces a cyclic order on the fat tree that uniquely prescribed an orientation. Hence, for these indices I, we have

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

Since there is $N(N-1)\cdots(N-k/2)$ choices for the distinct $k/2+1$ vertices for the same geometry of the rooted oriented tree, we get that

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

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

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

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 gives the result. We have,

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

where as before $I = \{i_1,\dots,i_k\}$, $I' = \{i'_1,\dots,i'_k\}$, 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 also denote $G(I,I') = (V(I,I'), E(I,I'))$ the graph with vertices $V(I,I') = \{i_1,\dots,i_k,i'_1,\dots,i'_k\}$, and corresponding edges. To give a non zero contribution, the graph must be connected, otherwise $E(I)\cap E(I') = \emptyset$ so $P(I,I') = P(I)P(I')$ by independence. Moreover, as before, each edge must appears at least twice and thus $|E(I,I')| \le k$, so $|V(I,I')| \le k+1$ by Lemma 2.12 and the same hold for $\widetilde{G}(I,I')$ the skeleton of $G(I,I')$. 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.

To improve to that this bound is actually of order N^{-2}, we show that the case where $|\widetilde{V}(I,I')| = |\widetilde{E}(I,I')|+1$ cannot occur. In this case, by Lemma 2.12, $\widetilde{G}(I,I')$ is a tree, and $|\widetilde{E}(I,I')| = k$ implies that each edge appears exactly twice. But $\widetilde{G}(I,I')\cap G(I)$ and $\widetilde{G}(I,I')\cap G(I')$ must share one edge, since otherwise $P(I,I') = P(I)P(I')$. This is a contradiction. Indeed, $\widetilde{G}(I,I')$ is explored by the path $i_1 \to i_2 \to \cdots \to i_k \to i_1$, so either each visited edge appears twice, which is impossible if $\widetilde{G}(I,I')\cap G(I)$ and $\widetilde{G}(I,I')\cap G(I')$ share one edge, or this path make a loop, which is also impossible since $\widetilde{G}(I,I')$ is a tree. Therefore, for all contributing indices we have $|\widetilde{V}(I,I')| \leq k$, 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

$$\mathbb{P}\Big(\big|\frac{1}{N}\operatorname{Tr}(H_N^k) - \mathbb{E}\big(\frac{1}{N}\operatorname{Tr}(H_N^k)\big)\big| > \varepsilon\Big) \leq \frac{\mathrm{cste}}{\varepsilon^2 N^2},$$

so Borel-Cantelli implies that

$$\big|\frac{1}{N}\operatorname{Tr}(H_N^k) - \mathbb{E}\big(\frac{1}{N}\operatorname{Tr}(H_N^k)\big)\big| \underset{N\to\infty}{\longrightarrow} 0, \quad \text{almost surely},$$

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

Theorem 2.14 (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 infinty, 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 2.13 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,

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

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 2.13, 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 2.13, we have that

$$\limsup_N \Big|\int_{|x|>B} f(x)\mu_{H_N}(dx) - \int_{|x|>B} P(x)\mu_{H_N}(dx)\Big| \leq KB^{-p-2q}(2\sigma)^{2(p+q)}.$$

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

$$\limsup_N \Big|\int_{|x|>B} f(x)\mu_{H_N}(dx) - \int_{|x|>B} P(x)\mu_{H_N}(dx)\Big| = 0.$$

Finally, since δ is arbitrary, we have that

$$\limsup_N \Big|\int_{\mathbb{R}} f(x)\mu_{H_N}(dx) - \int_{\mathbb{R}} f(x)\mu_{sc,\sigma^2}(dx)\Big| = 0,$$

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 [9] for a detailed proof, and only present below the definition of noncrossing partitions.

Definition 2.15. *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 4 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.

Proposition 2.16. *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

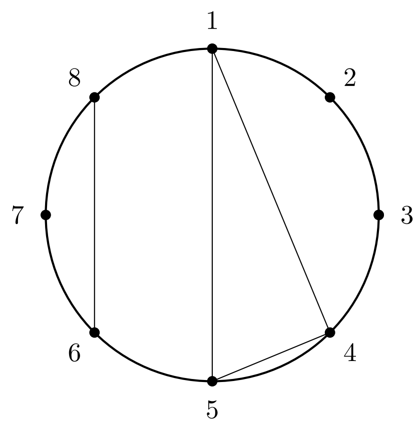

FIGURE 4. The noncrossing partition $\{1,4,5\} \cup \{2\} \cup \{3\} \cup \{6,8\} \cup \{7\}$.

is noncrossing, it induces a noncrossing partition of the set $\{1,\ldots,j-1\}$, and a noncrossing partition of the set $\{j+1,\ldots,n\}$. 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 2.11, the Catalan numbers. $\square$

2.4. **Extremal eigenvalues.** In view of Wigner theorem, one can suspect the following proposition.

Proposition 2.17 ([2]). *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 f_ε be a continuous bounded non-negative function supported on $[2-\varepsilon, 2]$, positive on $[2-\varepsilon', 2]$ with $0 < \varepsilon' < \varepsilon$. Using Wigner theorem, and the fact that $\int f_\varepsilon d\mu_{sc} > 0$, one can see that

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

Unfortunately, the corresponding upper bound on $\limsup_N \lambda_{\max}$ does not follow directly from Wigner theorem, and requires sharp combinatorial techniques. We refer to [2] for the proof of the necessary and sufficient condition for the extremal eigenvalues to converge to the edge of the support of the semicircular distribution, *i.e.* coefficients of Wigner matrices must have finite moment of order 4.

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 [12]. 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 $z \in \mathbb{C} \setminus supp(m)$).

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. □

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 measure. 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 $\int f d\mu_n \to \int f d\mu$.

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. □

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 2.11. 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

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

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

(i) $||G_M(z)|| \leq \frac{1}{|\Im z|}$,

(ii) $|G_M(i,j)(z)| \leq \frac{1}{|\Im z|}$, *for all $i, j = 1, \ldots, N$,*

(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 $\left|\frac{1}{x-z}\right| \leq \frac{1}{|\Im z|}$.

(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, 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 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* $A \in \mathcal{H}_N$,

$$\mathbb{E}(\mathrm{d}\Phi(H_N)\cdot A) = N\mathbb{E}(\Phi(H_N)\operatorname{Tr}(H_N A)).$$

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

$$\begin{aligned} \frac{d}{d\varepsilon}\big|_{\varepsilon=0} I = &\int_{\mathcal{H}_N} \mathrm{d}\Phi(M)\exp\Big(-\frac{N}{2}\operatorname{Tr}(M^2)\Big)dM \\ &+ \int_{\mathcal{H}_N} \Phi(M)\exp\Big(-\frac{N}{2}\operatorname{Tr}(M^2)\Big)(-N\operatorname{Tr}(MA))dM, \end{aligned}$$

which yields the result. $\square$

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

$$-\mathbb{E}((GAG)_{ij}) = N\mathbb{E}(G_{ij}\operatorname{Tr}(H_N A))$$

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

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

that is

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

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 Lipschitz way on many independent random variables (but not too much on any of them) is concentrated around its mean, and therefore is essentially constant. We refer to the book by Ledoux [11] for a complete treatment of the concentration of measure phenomenon.

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

$$\gamma_{n,\sigma^2}(|f - \int f d\gamma_{n,\sigma^2}| \geq \delta) \leq 2\exp\Big(- \frac{\kappa\delta^2}{c^2\sigma^2} \Big).$$

We postpone the proof to Appendix, section 6.2. Note that the above inequality is dimension free.

For a function $F\colon \mathbb{R} \to \mathbb{R}$, we define its extension to $\mathcal{H}_N$, still denoted F, by $F(M) = U\mathit{Diag}(F(\lambda_1), \dots, F(\lambda_N))U^*$, if $M = U\mathit{Diag}(\lambda_1, \dots, \lambda_N)U^*$. 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 $||M||_2 = \sqrt{\operatorname{Tr}(M^2)}$. In particular, the function $M \mapsto \frac{1}{N}\operatorname{Tr}(F(M))$ is $\frac{c}{\sqrt{N}}$-Lipschitz.*

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

$$A = U\mathit{Diag}(\lambda_1(A), \dots, \lambda_N(A))U^*$$
$$B = V\mathit{Diag}(\lambda_1(B), \dots, \lambda_N(B))V^*,$$

with U, V unitary matrices. Then, we have

$$||A - B||_2^2 = \operatorname{Tr}\big((A - B)^2\big) = \operatorname{Tr}(A^2) + \operatorname{Tr}(B^2) - 2\operatorname{Tr}(AB),$$

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

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

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 $f_N(z) = \mathbb{E}(g_N(z))$. Since $\mathbb{E}(g_N(z)^2) + z\mathbb{E}(g_N(z)) + 1 = 0$ by proposition 3.8, we have

$$\mathbb{V}\mathrm{ar}(g_N(z)) = -f_N(z)^2 - zf_N(z) - 1,$$

hence by the above estimate on the variance of $g_N(z)$, we get

$$|f_N(z)^2 + zf_N(z) + 1| \leq \frac{K}{N^2|\Im z|^2}.$$

Furthermore, we have $|f_N(z)| \leq \frac{1}{|\Im z|}$, thus the sequence $(f_N(z))_{N\geq 1}$ is uniformly bounded and analytic on compact sets of $\mathbb{C}^{(2)} = \{z \in \mathbb{C} \mid |\Im z| > 2\}$. Hence by the classical Montel's theorem, see Theorem 6.1 in the Appendix, there exists a subsequence $(f_{N_k}(z))_{k\geq 1}$ which converges uniformly on compact sets to some analytic function f. Passing to the limit, one can easily see that f satisfy properties (i)-(iii) of Proposition 3.2, as well as the equation

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

Hence, f is the Stieltjes transform of a probability measure, and by the above equation we get $f(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4})$, the sign of the square root being determined by condition (ii) of Proposition 3.2. Hence f is uniquely determined, that is does not depend on the choice of the subsequence of $(f_N(z))_{N\geq 1}$, and is equal to the Stieltjes transform of the semicircular distribution μ_{sc} as we have seen in Remark 3.4. Thus, $(f_N(z))_{N\geq 1}$ converges uniformly on compact sets of $\mathbb{C}^{(2)}$ to f. Now, using Markov inequality and Proposition 3.11, we have

$$\mathbb{P}(|g_N(z) - f_N(z)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{V}\mathrm{ar}(g_N(z)) \leq \frac{K}{N^2 \varepsilon |\Im z|^2}.$$

Hence, Borel-Cantelli lemma implies that for all $z \in \mathbb{C}^{(2)}$,

$$f_N(z) - g_N(z) \to_{N\to\infty} 0 \quad a.s.,$$

so $g_N(z) \to f(z)$ a.s., as N goes to infinity.

Let $z_0 \in \mathbb{C}^{(2)}$. There exists a measurable set Ω_0 such that $\mathbb{P}(\Omega_0) = 1$, such that $g_N(z_0)$ converges to $f(z_0)$ on Ω_0. We continue the same procedure for points $z_1, z_2, \ldots$ such that $(z_j)_{j\geq 0}$ has an accumulation point in $\mathbb{C}^{(2)}$, that is there exists $(\Omega_j)_{j\geq 0}$ such that $\mathbb{P}(\Omega_j) = 1$ for all $j \geq 0$, and $g_N(z_j)$ converges to $f(z_j)$ on Ω_j. Hence, Vitali's theorem (see Appendix Theorem 6.2) asserts that g_N converges to f uniformly on compacts of $\mathbb{C}^{(2)}$ on $\widetilde{\Omega} = \cap_{j\geq 0}\Omega_j$ which has probability one, and this ends the proof of the theorem using Proposition 3.3. $\square$

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 [10]). 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]).

4. Local behavior

We start this section by briefly recalling the notion of orthogonal polynomials.

4.1. **Orthogonal polynomials.**

Definition 4.1. *Given a weight $w\colon \mathbb{R} \to \mathbb{R}_+$, that is a nonnegative function such that $\int_{\mathbb{R}} |x|^k w(x)dx < \infty$, for all $k \geq 0$, the orthogonal polynomials $(q_k(x))_{k\geq 0}$ with respect to w are defined by*

(i) *$q_k(x)$ is a polynomial of degree k, $q_k(x) = u_k x^k + \cdots$, with leading term $u_k > 0$,*
(ii) *they satisfy the orthonormality condition,*

$$\langle q_k, q_l\rangle := \int_{\mathbb{R}} q_k(x)q_l(x)w(x)dx = \delta_{k,l}.$$

Note that the orthogonalization procedure of Gramm-Schmidt enables to construct such a sequence of polynomials.

Every sequence $(q_k(x))_{k\geq 0}$ of orthogonal polynomials satisfies a three terms recurrence relation of the form

$$q_n(x) = \alpha_n x q_{n-1}(x) + \beta_n q_{n-1}(x) + \gamma_n q_{n-2}(x). \tag{3}$$

Indeed, since $q_n(x) = u_n x^n + \cdots$, it follows that

$$\frac{q_n(x)}{u_n} - \frac{xq_{n-1}(x)}{u_{n-1}}$$

is a polynomial of degree $n-1$, thus

$$\frac{q_n(x)}{u_n} = \frac{xq_{n-1}(x)}{u_{n-1}} + \sum_{k=0}^{n-1} a_k q_k(x), \quad \text{where } a_k = \left\langle \frac{q_n}{u_n} - \frac{Xq_{n-1}}{u_{n-1}}, q_k\right\rangle,$$

where $X\colon x \mapsto x$. Using $\langle Xq, p\rangle = \langle q, Xp\rangle$, we get

$$a_k = \frac{1}{u_n}\langle q_n, q_k\rangle - \frac{1}{u_{n-1}}\langle q_{n-1}, Xq_k\rangle = 0, \quad \text{for } k = 0, \ldots, n-3.$$

Moreover,

$$a_{n-2} = -\frac{1}{u_{n-1}}\langle q_{n-1}, Xq_{n-2}\rangle = -\frac{u_{n-2}}{u_{n-1}^2},$$

since we can write $xq_{n-2}(x) = \frac{u_{n-2}}{u_{n-1}}q_{n-1}(x) + \text{polynomial of degree } n-2$. Therefore, putting $\alpha_n = u_n/u_{n-1}$, $\beta_n = a_{n-1}u_n$, and $\gamma_n = -u_n u_{n-2}/u_{n-1}^2$, gives the three terms relation

$$q_n(x) = \alpha_n x q_{n-1}(x) + \beta_n q_{n-1}(x) + \gamma_n q_{n-2}(x).$$

This recurrence relation is useful to prove the following formula.

Proposition 4.2 (Christoffel-Darboux formula)**.** *Let $(q_k(x))_{k\geq 0}$ be a sequence of orthogonal polynomials. Then we have,*

$$\sum_{k=0}^{n-1} q_k(x)q_k(y) = \begin{cases} \dfrac{u_{n-1}}{u_n}\dfrac{q_n(x)q_{n-1}(y)-q_{n-1}(x)q_n(y)}{x-y}, & \text{if } x \neq y, \\ \frac{u_{n-1}}{u_n}(q_n'(x)q_{n-1}(x)-q_{n-1}'(x)q_n(x)), & \text{if } x = y. \end{cases}$$

Proof. We just have to prove the $x \neq y$ case, since the formula for $x = y$ is obtained by taking the limit $y \to x$. By the three terms recurrence relation (3), we have for $k \geq 1$,

$$\begin{aligned} q_{k+1}(x)q_k(y) - q_k(x)q_{k+1}(y) &= \big((\alpha_{k+1}x+\beta_{k+1})q_k(x)+\gamma_{k+1}q_{k-1}(x)\big)q_k(y) \\ &\quad - q_k(x)\big((\alpha_{k+1}y+\beta_{k+1})q_k(y)+\gamma_{k+1}q_{k-1}(y)\big) \\ &= \alpha_{k+1}q_k(x)q_k(y)(x-y)+\gamma_{k+1}(q_{k-1}(x)q_k(y)-q_k(x)q_{k-1}(y)). \end{aligned}$$

Hence, dividing by $(x-y)\alpha_{k+1}$, we get, noticing that $\frac{1}{\alpha_{k+1}} = \frac{u_k}{u_{k+1}}$ and $\frac{\gamma_{k+1}}{\alpha_{k+1}} = -\frac{u_{k-1}}{u_k}$,

$$q_k(x)q_k(y) = S_{k+1}(x,y) - S_k(x,y),$$

where

$$S_k(x,y) = \frac{u_{k-1}}{u_k}\frac{q_k(x)q_{k-1}(y)-q_{k-1}(x)q_k(y)}{x-y}.$$

The proposition follows by taking the telescopic sum

$$\sum_{k=0}^{n-1} q_k(x)q_k(y) = q_0(x)q_0(y) + S_n(x,y) - S_1(x,y),$$

and noticing that $q_0(x) = u_0$ and $S_1(x,y) = u_0u_0 = q_0(x)q_0(y)$. □

The Hermite polynomials $(H_n)_{n\geq 0}$ are defined by

$$H_n(x) = (-1)^n e^{x^2}\frac{d^n}{dx^n}e^{-x^2}.$$

They are orthogonal for the weight $w(x) = e^{-x^2}$, and satisfy the orthogonality relation

$$\int_{\mathbb{R}} H_k(x)H_l(x)e^{-x^2}dx = \sqrt{\pi}2^k k!\delta_{k,l},$$

and the three terms recurrence relation

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1(x)},$$

and also $H_n'(x) = 2nH_{n-1}(x)$. The coefficient of the leading term of $H_n(x)$ is 2^n. They are well known, and a standard reference on orhtogonal polynomials is the book of Szegő [14].

4.2. **Eigenvalues' distribution.**

Proposition 4.3. *Let $X_N \in \mathcal{H}_N$ be a random matrix distributed according to the GUE, and denote $\lambda_1, \dots, \lambda_N$ its eigenvalues. The joint distribution of $(\lambda_1, \dots, \lambda_N)$ has density with respect to Lebesgue measure given by*

$$p_N(x_1, \dots, x_N) = \frac{1}{Z_N} |\Delta(x)|^2 \exp\left(- \frac{1}{2} \sum_{i=1}^N x_i^2 \right)$$

where $\Delta(x) = \prod_{1 \le i < j \le N} (x_j - x_i)$ is the Vandermonde determinant, and Z_N is a normalization constant equal to $Z_N = \pi^{N/2} 2^{-N(N-1)/2} \prod_{j=2}^n j!$.

We refer to [1] for a proof of this proposition. Its relies on an integration formula due to Weyl. Heuristically, the term $\exp(-\frac{1}{2N} \sum_{i=1}^N x_i^2)$ comes from the term $\exp(-\frac{1}{2} \operatorname{Tr}(M^2))$ in the GUE distribution, and the Vandermonde Δ comes from the Jacobian of the map $M \mapsto U Diag(\lambda_1, \dots, \lambda_N) U^*$.

Note that Δ is actually a determinant, and indeed one has

$$\Delta(x) = \det(x_i^{j-1})_{1 \le i,j \le N}.$$

This claim is proved by induction on N. This is clearly true for $N = 2$. By multilinearity of the determinant, making the operation $C_i \to C_i - x_1 C_{i-1}$, starting from column C_N until column C_2, we get

$$\begin{aligned}
\begin{vmatrix} 1 & x_1 & \cdots & x_1^{N-1} \\ 1 & x_2 & \cdots & x_2^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^{N-1} \end{vmatrix} &= \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 1 & x_2 - x_1 & \cdots & x_2^{N-2}(x_2 - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N - x_1 & \cdots & x_N^{N-2}(x_N - x_1) \end{vmatrix} \\
&= \begin{vmatrix} x_2 - x_1 & \cdots & x_2^{N-2}(x_2 - x_1) \\ \vdots & \ddots & \vdots \\ x_N - x_1 & \cdots & x_N^{N-2}(x_N - x_1) \end{vmatrix} \\
&= (x_N - x_1) \cdots (x_2 - x_1) \begin{vmatrix} 1 & x_2 & \cdots & x_2^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \cdots & x_N^{N-2} \end{vmatrix},
\end{aligned}$$

hence the claim follows by induction hyptohesis.

4.3. **Correlation functions of the GUE eigenvalues.**

Definition 4.4. *The n-point correlation functions of the GUE eigenvalues $\lambda_1, \dots, \lambda_N$ are defined by*

$$\rho_N^{(n)}(x_1, \dots, x_n) = \frac{N!}{(N-n)!} \int_{\mathbb{R}^{N-n}} p_N(x_1, \dots, x_N) dx_{n+1} \cdots x_N,$$

for $n = 1, \dots, N$, where p_N is the density distribution of $(\lambda_1, \dots, \lambda_N)$.

The n-point correlation functions are, up to a constant, the marginal distributions of p_N. Heuristically, they are the probability of finding an eigenvalue at each of the position $x_1, \dots, x_n$, but note that it is not fixed which eigenvalue is at which position. Note also that we can define in the same way correlation functions of N particles evolving according to some symmetric density distribution p_N.

The next goal is to compute the correlation functions of the GUE eigenvalues. Before doing this, let us see what kind of probabilistic quantities correlation functions enable to compute.

Let f be a Borel function on $\mathbb{R}$. Then,

$$\begin{aligned}\mathbb{E}\Big(\prod_{i=1}^{N}(1+f(\lambda_i))\Big) &= \mathbb{E}\Bigg(\sum_{k=0}^{N}\sum_{1\le i_1<\cdots<i_k\le N} f(\lambda_{i_1})\cdots f(\lambda_{i_k})\Bigg)\\ &= \sum_{k=0}^{N}\frac{1}{k!}\sum_{\substack{i_1,\dots,i_k\\ \text{all different}}}\mathbb{E}(f(\lambda_{i_1})\cdots f(\lambda_{i_k}))\\ &= \sum_{k=0}^{N}\frac{1}{k!}\frac{N!}{(N-k)!}\mathbb{E}(f(\lambda_1)\cdots f(\lambda_k))\\ &= \sum_{k=0}^{N}\frac{1}{k!}\int_{\mathbb{R}^k} f(x_1)\cdots f(x_k)\rho_N^{(k)}(x_1,\dots,x_k)dx_1\cdots dx_k.\end{aligned}$$

Now, let $f=-\mathbb{1}_B$ where B is a Borel set of $\mathbb{R}$. Then the *gap probability* is expressed as

$$\mathbb{P}(\text{no eigenvalues lies in } B) = \sum_{k=0}^{N}\frac{(-1)^k}{k!}\int_{B^k}\rho_N^{(k)}(x_1,\dots,x_k)dx_1\cdots dx_k.$$

In particular, if $B=]s,+\infty[$,

$$\mathbb{P}(\lambda_{\max}\le s) = \sum_{k=0}^{N}\frac{(-1)^k}{k!}\int_{]s,+\infty[^k}\rho_N^{(k)}(x_1,\dots,x_k)dx_1\cdots dx_k.$$

Also, *the density of states* is defined as

$$\mathbb{E}\Bigg(\frac{1}{N}\sum_{k=1}^{N}f(\lambda_i)\Bigg) = \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^N} f(x_i)p_N(x_1,\dots,x_N)dx_1\cdots dx_N = \int_{\mathbb{R}} f(x)\frac{1}{N}\rho_N^{(1)}(x)dx,$$

that is, the measure $\frac{1}{N}\rho_N^{(1)}(x)dx$ represents the expectation of the spectral distribution of X_N.

The next proposition expresses the n-point correlation functions as a determinant of some kernel, giving a structure of a so-called determinantal process to the GUE eigenvalues.

Proposition 4.5. *The n-point correlation functions of the GUE eigenvalues are given by*

$$\rho_N^{(n)}(x_1,\dots,x_n) = \det\big(K_N(x_i,x_j)\big)_{1\le i,j\le n},$$

where the kernel K_N is given by

$$K_N(x,y) = \sqrt{w(x)}\sqrt{w(y)}\sum_{k=0}^{N-1} q_k(x)q_k(y),$$

where w is the weight $w(x)=\exp(-x^2/2)$ and $(q_k(x))_{k\le 0}$ are orthonormal polynomials with respect to the weight w, given by

$$q_k(x) = \frac{1}{\sqrt{2^k k!}}\frac{1}{(2\pi)^{1/4}}H_k\Big(\frac{x}{\sqrt{2}}\Big),$$

where $(H_k)_{k\geq 0}$ are the Hermite polynomials.

Proof. By multi-linearity of the determinant, and since q_k has degree k, we have

$$\Delta(x) = \det(x_i^{j-1})_{1\leq i,j\leq N} = c_N \det(q_{j-1}(x_i))_{1\leq i,j\leq N},$$

where c_N is a constant. Since $w(x) = \exp(-\frac{1}{2}x^2)$, the density of the GUE eigenvalues writes

$$\begin{aligned} p_N(x_1,\dots,x_N) &= c_N \frac{1}{Z_N}(\det(q_{j-1}(x_i))_{1\leq i,j\leq N})^2 \prod_{i=1}^N w(x_i) \\ &= \frac{c_N}{Z_N}\det\left(\sum_{k=1}^N q_{k-1}(x_i)q_{k-1}(x_j)\right)\prod_{i=1}^N w(x_i), \end{aligned}$$

using $(\det A^tA) = \det A \det {}^tA$. Hence, from the definition of the Kernel K_N, we obtain

$$p_N(x_1,\dots,x_N) = c_N \det\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq N}$$

and,

$$\rho_N^{(n)}(x_1,\dots,x_n) = c_N\frac{N!}{(N-n)!}\int_{\mathbb{R}^{N-n}}\det\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq N}dx_{n+1}\cdots dx_N.$$

We need then to integrate $N-n$ times, each step being similar. First observe that K_N satisfy the following two equalities:

$$\int_{\mathbb{R}} K_N(x,x)dx = N \tag{4}$$

$$\int_{\mathbb{R}} K_N(x,y)K_N(y,z)dy = K_N(x,z). \tag{5}$$

Indeed, since $(q_k(x))_{k\geq 0}$ are orthonormal with respect to the weight w, we have

$$\int_{\mathbb{R}} K_N(x,x)dx = \sum_{k=0}^{N-1}\int_{\mathbb{R}} w(x)q_k(x)q_k(x)dx = \sum_{k=0}^{N-1}\langle q_k,q_k\rangle = N,$$

and,

$$\int_{\mathbb{R}} K_N(x,y)K_N(y,z)dy = \sum_{k,l=0}^{N-1}\sqrt{w(x)}\sqrt{w(z)}q_k(x)q_l(z)\langle q_k,q_l\rangle = K_N(x,z).$$

Now, for $m\leq N$, we develop the determinant along the last column to get

$$\begin{aligned} \det\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq m} &= K_N(x_m,x_m)\det\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq m-1} \\ &\quad + \sum_{k=1}^{m-1}(-1)^{m-k}K(x_k,x_m)\det\begin{pmatrix}\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq m-1, i\neq k} \\ \big(K_N(x_m,x_j)\big)_{1\leq j\leq m-1}\end{pmatrix} \\ &= K_N(x_m,x_m)\det\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq m-1} \\ &\quad + \sum_{k=1}^{m-1}(-1)^{m-k}\det\begin{pmatrix}\big(K_N(x_i,x_j)\big)_{1\leq i,j\leq m-1, i\neq k} \\ \big(K(x_k,x_m)K_N(x_m,x_j)\big)_{1\leq j\leq m-1}\end{pmatrix}. \end{aligned}$$

Hence, using the two equalities (4) and (5), we have

$$
\begin{aligned}
\int_{\mathbb{R}} \det\big(K_N(x_i,x_j)\big)_{1\le i,j\le m} dx_m &= N\det\big(K_N(x_i,x_j)\big)_{1\le i,j\le m-1} \\
&\quad + \sum_{k=1}^{m-1}(-1)^{m-k}\det\begin{pmatrix}\big(K_N(x_i,x_j)\big)_{1\le i,j\le m-1, i\neq k}\\ \big(K(x_k,x_j)\big)_{1\le j\le m-1}\end{pmatrix}\\
&= N\det\big(K_N(x_i,x_j)\big)_{1\le i,j\le m-1}\\
&\quad + \sum_{k=1}^{m-1}(-1)^{m-k}\det\big(K_N(x_i,x_j)\big)_{1\le i,j\le m-1}\\
&= (N-m+1)\det\big(K_N(x_i,x_j)\big)_{1\le i,j\le m-1},
\end{aligned}
$$

since $\sum_{k=1}^{m-1}(-1)^{m-k} = -m+1$. Applying this result to $m = N, N-1, \dots, n+1$ using Fubini theorem, we obtain

$$\rho_N^{(n)}(x_1,\dots,x_n) = c_N N! \det\big(K_N(x_i,x_j)\big)_{1\le i,j\le n}.$$

It remains to determine c_N. Since c_N depends only on N, we can compute it for the $n=1$ case. By the above computation, we have $\rho_N^{(1)} = c_N N! K_N(x,x)$. Since $\int_{\mathbb{R}}\rho^{(1)}(x)dx = N\int_{\mathbb{R}^N} p_N(x)dx = N$, and $\int_{\mathbb{R}} K_N(x,x)dx = N$, we get $c_N = 1/N!$, which yields the proposition. $\square$

4.4. **Asymptotic of GUE eigenvalues, the local regime.** Let X_N be a GUE random matrix, and consider as usual the renormalized matrix $H_N = \frac{1}{\sqrt{N}}X_N$. Note that the eigenvalues of H_N are given by $\lambda_i/\sqrt{N}$, $i = 1,\dots,N$, where the λ_i's are the eigenvalues of X_N. Let us denote, for a Borel set B

$$\nu_N(B) = \#\Big\{i\in\{1,\dots,N\} \mid \frac{1}{\sqrt{N}}\lambda_i \in B\Big\} = N\mu_{H_N}(B).$$

By Wigner theorem, we have, as N goes to infinity, $\nu_N(B)\sim N\int_B f_{sc}(x)dx$, where f_{sc} is the density of the semicircular distribution. In the local regime asymptotic, we are looking at Borel sets whose size goes to zero. Hence, two cases have to be considered.

(i) *Inside the bulk:* Let $B_N = [u-\varepsilon_N, u+\varepsilon_N]$, where u is such that $f_{sc}(u) > 0$, that is $u \in]-2,2[$. Then $\nu_N(B_N)\sim N\varepsilon_N f_{sc}(u)$, so it has order of a constant for $\varepsilon_N \sim 1/N$. This suggests to study the renormalized eigenvalues inside the bulk $\mu_1,\dots,\mu_N$ defined by

$$\frac{\lambda_i}{\sqrt{N}} = u + \frac{\mu_i}{N f_{sc}(u)}, \quad i = 1,\dots,N,$$

(that is $\lambda_i/\sqrt{N} \in [u-1/N, u+1/N] \Leftrightarrow \mu_i \in [-f_{sc}(u), f_{sc}(u)]$).

(ii) *At the edge of the spectrum:* Let $u = 2$ or -2. Then $f_{sc}(u) = 0$, and

$$\nu_N([2-\varepsilon_N, 2]) \sim N\frac{1}{2\pi}\int_{2-\varepsilon_N}^{2}\sqrt{4-x^2}dx \sim N\varepsilon_N^{3/2},$$

so the renormalization at the edge of the spectrum is $\varepsilon_N \sim 1/N^{2/3}$.

4.5. **Inside the bulk.** To simplify the readability, we will only consider the local asymptotics of the GUE eigenvalues inside the bulk at $u = 0$.

Theorem 4.6. *Let $\lambda_1, \dots, \lambda_N$ be the eigenvalues of a random matrix distributed according the GUE distribution. Then for all compact $A \subset \mathbb{R}$ one has,*

$$\lim_{N\to\infty} \mathbb{P}\Big(\sqrt{N}\lambda_1 \notin A, \dots, \sqrt{N}\lambda_N \notin A\Big) = 1+\sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{A^k} \det\left(S(x_i, x_j)\right)_{1\le i,j\le k} dx_1 \cdots dx_k,$$

where S is the sine kernel *defined by*

$$S(s,t) = \frac{\sin(s-t)}{\pi(s-t)}.$$

Recall that we have seen that the gap probability is expressed as

$$\mathbb{P}(\text{no eigenvalues lies in } B) = \sum_{k=0}^{N} \frac{(-1)^k}{k!} \int_{B^k} \rho_N^{(k)}(x_1, \dots, x_k) dx_1 \cdots dx_k,$$

where $\rho_N^{(k)}$ are the n-point correlation functions of the GUE eigenvalues, and consider the rescaled eigenvalues inside the bulk defined by

$$\mu_i = \sqrt{N}\lambda_i, \quad i = 1, \dots, N.$$

Then a simple change of variables shows that the correlations functions $\rho_N^{(n,bulk)}$ of $(\mu_1, \dots, \mu_N)$ are given by

$$\rho_N^{(n,bulk)}(y_1, \dots, y_k) = \frac{1}{N^{n/2}} \rho_N^{(n)}\left(\frac{y_1}{\sqrt{N}}, \dots, \frac{y_n}{\sqrt{N}}\right).$$

Hence using the determinantal form of $\rho_N^{(n)}$ given in Proposition 4.5, we have to find the limit of

$$\frac{1}{\sqrt{N}} K_N\left(\frac{x}{\sqrt{N}}, \frac{y}{\sqrt{N}}\right).$$

The above theorem will then follow immediately using the uniform convergence of the following proposition.

Proposition 4.7. *One has*

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} K_N\left(\frac{x}{\sqrt{N}}, \frac{y}{\sqrt{N}}\right) = \frac{1}{\pi} \frac{\sin(x-y)}{x-y},$$

uniformly for (x, y) in a compact set.

The proof of this proposition is done using Laplace method described in the next subsection.

4.6. **Laplace method.** Laplace method deals with the asymptotic evaluation, as $s \to \infty$, of integrals of the form

$$\int f(x)^s g(x) dx,$$

where f achieves a global maximum. To illustrate the method, suppose that f achieves a global maximum at a, $f(a) > 0$ and g is a nice function. Using Taylor's expansion, we have

$$f(x) = f(a) + \frac{1}{2} f''(a)(x-a)^2 + o((x-a)^2),$$

hence

$$\begin{aligned}\sqrt{s}f(a)^{-s}\int f(x)^s g(x)dx &\approx \sqrt{s}\int\left(1+\frac{1}{2}\frac{f''(a)}{f(a)}(x-a)^2\right)^s g(x)dx\\ &\approx \sqrt{s}\int\exp\left(\frac{s}{2}\frac{f''(a)}{f(a)}(x-a)^2\right)g(x)dx\\ &\approx \int\exp\left(\frac{1}{2}\frac{f''(a)}{f(a)}y^2\right)g\left(a+\frac{y}{\sqrt{s}}\right)dy\\ &\approx \sqrt{-\frac{2\pi f(a)}{f''a)}}g(a),\end{aligned}$$

as s goes to infinity, using Gauss integral. Intuitively, this means that as s goes to infinity, $(\frac{f(x)}{f(a)})^s$ near $x=a$ looks at the microscopic level more and more like a Gauss curve, whereas $f(x)^s$ becomes negligible elsewhere. Along the same lines, one can prove the following (see [1] for details).

Theorem 4.8 (Laplace method)**.** *Let $f:\mathbb{R}\to\mathbb{R}_+$ be a function such that, for some $a\in\mathbb{R}$ and positive constant ε', c, the following hold.*

(i) *$f(x)\le f(y)$ if either $a-\varepsilon'\le x\le y\le a$ or $a\le y\le x\le a+\varepsilon'$,*
(ii) *for all $0<\varepsilon<\varepsilon'$, $\sup_{|x-a|>\varepsilon} f(x)\le f(a)-c\varepsilon^2$,*
(iii) *f is twice continuously differentiable in $]a-2\varepsilon', a+2\varepsilon'[$,*
(iv) *$f''(a)<0$.*

Then, for any measurable bounded function g, locally Lipschitz near a, such that $\int f(x)^{s_0}|g(x)|dx<\infty$ for some $s_0>0$, we have

$$\lim_{s\to\infty}\sqrt{s}f(a)^{-s}\int f(x)^s g(x)dx=\sqrt{-\frac{2\pi f(a)}{f''(a)}}g(a),$$

the convergence being uniform over such g.

We are now ready to prove Proposition 4.7 which describes the local asymptotics of the GUE eigenvalues inside the bulk.

Proof of Proposition 4.7. Define

$$\psi_n(x)=e^{-x^2/4}q_n(x)=e^{-x^2/4}\frac{2^{-n/2}H_n\left(\frac{x}{\sqrt{2}}\right)}{(2\pi)^{1/4}\sqrt{n!}},$$

where H_n are the Hermite polynomials. Let

$$S_N(x,y)=\frac{1}{\sqrt{N}}K_N\left(\frac{x}{\sqrt{N}},\frac{y}{\sqrt{N}}\right),$$

and recall that by Proposition 4.5, the kernel K_N is given by

$$K_N(x,y)=\sum_{k=0}^{N-1}q_k(x)q_k(y)e^{-(x^2+y^2)/4}=\sum_{k=0}^{N-1}\psi_k(x)\psi_k(y).$$

Thanks to Christoffel-Darboux formula (Proposition 4.2), we can express this kernel as

$$S_N(x,y)=\frac{\psi_N\left(\frac{x}{\sqrt{N}}\right)\psi_{N-1}\left(\frac{y}{\sqrt{N}}\right)-\psi_{N-1}\left(\frac{x}{\sqrt{N}}\right)\psi_N\left(\frac{y}{\sqrt{N}}\right)}{x-y},$$

the leading coefficients of q_n being $u_n = \frac{1}{\sqrt{n!}(2\pi)^{1/4}}$. Writing

$$\frac{f(x)-f(y)}{x-y} = \int_0^1 f'(tx+(1-t)y)dt,$$

hence,

$$\begin{aligned}\frac{f(x)g(y)-f(y)(x)}{x-y} &= \frac{f(x)-f(y)}{x-y}g(y) - f(y)\frac{g(x)-g(y)}{x-y}\\ &= g(y)\int_0^1 f'(tx+(1-t)y)dt - f(y)\int_0^1 g'(tx+(1-t)y)dt,\end{aligned}$$

for differentiable functions f, g, we get

$$\begin{aligned}S_N(x,y) = \psi_{N-1}\big(\frac{y}{\sqrt{N}}\big)&\int_0^1 \psi_N'\Big(t\frac{x}{\sqrt{N}} + (1-t)\frac{y}{\sqrt{N}}\Big)dt\\ &- \psi_N\big(\frac{y}{\sqrt{N}}\big)\int_0^1 \psi_{N-1}'\Big(t\frac{x}{\sqrt{N}} + (1-t)\frac{y}{\sqrt{N}}\Big)dt.\end{aligned}$$

Recall that Hermite polynomials satisfy $H_k'(x) = 2kH_{k-1}(x)$. Hence, $\psi_k'(x) = \sqrt{k}\psi_{k-1}(x) - \frac{x}{2}\psi_k(x)$, so we have

$$\begin{aligned}S_N(x,y) = \psi_{N-1}\big(\frac{y}{\sqrt{N}}\big)&\int_0^1 \Big[\sqrt{N}\psi_{N-1}(z) - \frac{z}{2}\psi_N(z)\Big]\Big|_{z=t\frac{x}{\sqrt{N}}+(1-t)\frac{y}{\sqrt{N}}} dt\\ &- \psi_N\big(\frac{y}{\sqrt{N}}\big)\int_0^1 \Big[\sqrt{N-1}\psi_{N-2}(z) - \frac{z}{2}\psi_{N-1}(z)\Big]\Big|_{z=t\frac{x}{\sqrt{N}}+(1-t)\frac{y}{\sqrt{N}}} dt.\end{aligned}$$

Now we use the following lemma, whose proof is deferred at the end of the proof.

Lemma 4.9. *Let $\nu = N-k$, where k is independent of N. Then, one has*

$$\Big|N^{1/4}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) - \frac{1}{\sqrt{\pi}}\cos\Big(x - \frac{\pi\nu}{2}\Big)\Big| \to 0,$$

as $N \to \infty$, uniformly for x in a compact interval.

Since

$$\begin{aligned}S_N(x,y) = N^{1/4}\psi_{N-1}\big(\frac{y}{\sqrt{N}}\big)&\int_0^1 \Big[N^{1/4}\psi_{N-1}(z) - N^{-1/4}\frac{z}{2}\psi_N(z)\Big]\Big|_{z=t\frac{x}{\sqrt{N}}+(1-t)\frac{y}{\sqrt{N}}} dt\\ - N^{1/4}\psi_N\big(\frac{y}{\sqrt{N}}\big)&\int_0^1 \Big[N^{-1/4}\sqrt{N-1}\psi_{N-2}(z) - N^{-1/4}\frac{z}{2}\psi_{N-1}(z)\Big]\Big|_{z=t\frac{x}{\sqrt{N}}+(1-t)\frac{y}{\sqrt{N}}} dt,\end{aligned}$$

using the lemma, one gets

$$\begin{aligned}S_N(x,y) \underset{+\infty}{\sim} \frac{1}{\pi}\cos\Big(y - \frac{\pi(n-1)}{2}\Big)&\int_0^1 \cos\Big(tx+(1-t)y - \frac{\pi(n-1)}{2}\Big)dt\\ - \frac{1}{\pi}\cos\Big(y - \frac{\pi n}{2}\Big)&\int_0^1 \cos\Big(tx+(1-t)y - \frac{\pi(n-2)}{2}\Big)dt.\end{aligned}$$

Using the trigonometric formulas $\cos(p+\frac{\pi}{2}) = -\sin(p)$ and $\sin(p-q) = \sin p\cos q - \cos p\sin q$, one obtains

$$S_N(x,y) \underset{+\infty}{\sim} \frac{1}{\pi}\frac{\sin(x-y)}{x-y},$$

uniformly in x, y, which yields the result. $\square$

It remains to prove Lemma 4.9, which is done using Laplace method of Theorem 4.8.

Proof of Lemma 4.9. To use Laplace method, we use the following integral representation of Hermite polynomials,

$$H_\nu(x) = \frac{1}{2\sqrt{\pi}} e^{x^2} \int_{\mathbb{R}} (i\xi)^\nu e^{-\xi^2/4 - i\xi x} d\xi,$$

which follows from the formula for the characteristic function of a normal random variable (with variance 2)

$$e^{-x^2} = \frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{-\xi^2/4 - i\xi x} d\xi,$$

and differentiating under the integral sign using the definition of Hermite polynomials, $H_\nu(x) = e^{x^2}(-1)^\nu \frac{d^\nu}{dx^\nu} e^{x^2}$. This yields to

$$\begin{aligned}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) &= e^{-\frac{x^2}{4N}} \frac{2^{-\nu/2} H_\nu\big(\frac{x}{\sqrt{2}\sqrt{N}}\big)}{(2\pi)^{1/4}\sqrt{\nu!}} \\ &= \frac{e^{-\frac{x^2}{4N}}}{(2\pi)^{3/4}\sqrt{\nu!}} \int_{\mathbb{R}} (i\xi)^\nu e^{-\xi^2/2 - i\xi x/\sqrt{N}} d\xi.\end{aligned}$$

Making the change of variable $\zeta = \frac{\xi}{\sqrt{N}}$, we get

$$\begin{aligned}N^{1/4}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) &= \frac{N^{1/4} e^{\frac{x^2}{4N}}}{(2\pi)^{3/4}\sqrt{\nu!}} i^\nu \sqrt{N}^{\nu+1} \int_{\mathbb{R}} \zeta^\nu e^{-N\zeta^2/2 - i\zeta x} d\zeta \\ &= \frac{N^{1/4} e^{\frac{x^2}{4N}}}{(2\pi)^{3/4}\sqrt{\nu!}} i^\nu \sqrt{N}^{\nu+1} \int_{\mathbb{R}} \left(\zeta e^{-\zeta^2/2}\right)^N e^{-i\zeta x} \zeta^{\nu - N} d\zeta.\end{aligned}$$

Recall Stirling's formula

$$\nu! = \sqrt{2\pi\nu}\left(\frac{\nu}{e}\right)^\nu (1 + o(1)).$$

Therefore, the term before the integral in the last expression of $N^{1/4}\psi_\nu(\frac{x}{\sqrt{N}})$ becomes

$$\begin{aligned}\frac{N^{1/4} e^{\frac{x^2}{4N}}}{(2\pi)^{3/4}\sqrt{\nu!}} \sqrt{N}^{\nu+1} &= \frac{N^{1/4} e^{\frac{x^2}{4N}} N^{\nu/2 + 1/2}}{(2\pi)^{3/4}(2\pi)^{1/4}\nu^{1/4}(\frac{\nu}{e})^{\nu/2}}(1 + o(1)) \\ &= \frac{1}{2\pi}\left(\frac{N}{\nu}\right)^{\nu/2}\left(\frac{N}{\nu}\right)^{1/4} e^{\nu/2}\sqrt{N} e^{\frac{x^2}{4N}}(1 + o(1)) \\ &= \frac{1}{2\pi}\sqrt{N} e^{N/2}(1 + o(1)),\end{aligned}$$

where for the last equality, recalling that $N - \nu$ is a constant independent of N, we use that $\frac{N}{\nu} \sim 1$, $\left(\frac{N}{\nu}\right)^{\nu/2} \sim e^{(N-\nu)/2}$ and $e^{\frac{x^2}{4N}} \to 0$, uniformly in x since x is in bounded interval. Hence, we get, uniformly in x in a bounded interval,

$$N^{1/4}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) = \frac{\sqrt{N} e^{N/2}}{2\pi} i^\nu \int_{\mathbb{R}} \left(\zeta e^{-\zeta^2/2}\right)^N e^{-i\zeta x} \zeta^{\nu - N} d\zeta \ (1 + o(1))$$

Furthermore, since ψ_ν is real, we have

$$\begin{aligned} N^{1/4}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) &\sim \frac{\sqrt{N}e^{N/2}}{2\pi}\int_{\mathbb{R}}\left|\zeta e^{-\zeta^2/2}\right|^N \Re\left[(i\mathrm{sign}(\zeta))^\nu e^{-i\zeta x}\right]|\zeta|^{\nu-N}d\zeta \\ &= \frac{\sqrt{N}e^{N/2}}{\pi}\int_0^{+\infty}\left(\zeta e^{-\zeta^2/2}\right)^N \Re\left[i^\nu e^{-i\zeta x}\right]\zeta^{\nu-N}d\zeta \\ &= \frac{\sqrt{N}e^{N/2}}{\pi}\int_0^{+\infty}\left(\zeta e^{-\zeta^2/2}\right)^N \cos\left[\zeta x-\frac{\nu\pi}{2}\right]\zeta^{\nu-N}d\zeta. \end{aligned}$$

Now we are ready to use Laplace method. Let

$$f(\zeta)=\zeta e^{-\zeta^2/2}, \quad g(\zeta)=\cos\left(\zeta x-\frac{\nu\pi}{2}\right)\zeta^{\nu-N}.$$

Note that f admits a global maximum at 1 such that $f(1)=e^{-1/2}$, $f'(1)=0$, and $f''(1)=-2e^{-1/2}$. Recall that $N-\nu$ does not depend on N. There are four ways such that g does not depend on N according to the parity of N and $N-\nu$, each of them is treated in the same way using Laplace method, which yields to

$$N^{1/4}\psi_\nu\Big(\frac{x}{\sqrt{N}}\Big) \sim \frac{1}{\sqrt{\pi}}\cos\left(x-\frac{\pi\nu}{2}\right),$$

as $N\to\infty$, uniformly for x in a bounded interval, and proves the lemma. □

4.7. **At the edge of the spectrum.** Using for instance the *steepest descent method*, which is a general, more elaborate version of Laplace method, one can prove along the same lines than in the previous section but in a more tricky way, the following asymptotic behavior of the rescaled eigenvalues at the edge of the spectrum, see [1] for details.

Proposition 4.10. *Consider the rescaled eigenvalues at the edge of the spectrum $\gamma_1,\dots,\gamma_N$ defined by*

$$\frac{\lambda_i}{\sqrt{N}}=2+\frac{\gamma_i}{N^{2/3}}, \quad i=1,\dots,N,$$

where the λ_i's are the eigenvalues of a GUE matrix. Denote by $\rho_N^{(n,edge)}$, for $n=1,\dots,N$, the n-point correlation functions of $(\gamma_1,\dots,\gamma_N)$. Then we have,

$$\lim_{N\to\infty}\rho_N^{(n,edge)}(y_1,\dots,y_n)=\det\left(\mathrm{A}(y_i,y_j)\right)_{1\le i,j\le n},$$

where A *is the Airy kernel defined by*

$$\mathrm{A}(x,y)=\frac{Ai(x)Ai'(y)-Ai'(x)Ai(y)}{x-y},$$

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

$$y''(t)-ty(t)=0,$$

with asymptotic behavior $Ai(t)\underset{+\infty}{\sim}(2\sqrt{\pi})^{-1}t^{-1/4}e^{-\frac{2}{3}t^{3/2}}$.

Corollary 4.11 ([16])**.** *The fluctuations of the largest eigenvalues $\lambda_{\max}$ of the GUE distribution are given by*

$$\lim_{N\to\infty}\mathbb{P}(N^{1/6}(\lambda_{\max}-2\sqrt{N})\le t)=F_2(t),$$

where F_2 is the Tracy-Widom distribution given by

$$F_2(t) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k!} \int_{]t,+\infty[^k} \det\left(\mathrm{A}(y_i, y_j)\right)_{1\le i,j\le k} dy_1 \cdots dy_k.$$

The distribution F_2 has the following representation

$$F_2(t) = \exp\left(-\int_t^{+\infty} (x-t) q(x)^2 dx\right),$$

where q is the solution of the Painlevé II equation

$$q''(t) = tq(t) + 2q(t)^3,$$

with $q(t) \underset{+\infty}{\sim} Ai(t)$.

4.8. **Remarks.** An alternative proof of the convergence of the correlation functions of the GUE eigenvalues both inside the bulk and at the edge of the spectrum can be done using the well known asymptotics of Hermite polynomials given in the following proposition.

Proposition 4.12 (Plancherel-Rotach formulas [14])**.** *Let $(H_n(x))_{n\ge 0}$ denote the Hermite polynomials. Let $\varepsilon > 0$. We have,*

(i) *If $x = \sqrt{2n+1}\cos\theta$, with $\varepsilon \le \theta \le \pi - \varepsilon$,*

$$e^{-x^2/2} H_n(x) = 2^{n/2+1/4}(n!)^{1/2}(\pi n)^{-1/4}(\sin\theta)^{-1/2} \times \left(\sin\left[\left(\frac{n}{2}+\frac{1}{4}\right)(\sin 2\theta - 2\theta) + \frac{3\pi}{4}\right] + O(n^{-1})\right).$$

(ii) *If $x = \sqrt{2n+1} - 2^{-1/2}3^{-1/3}n^{-1/6}t$, with $t \in \mathbb{C}$ bounded,*

$$e^{-x^2/2} H_n(x) = 3^{1/3}\pi^{-3/4}2^{n/2+1/4}(n!)^{1/2}n^{-1/12}\left(A(t) + O(n^{-2/3})\right),$$

where $A(t) = \pi Ai(-t/3^{1/3})$, where Ai is Airy's function, defined as the solution of the differential equation

$$y''(t) - ty(t) = 0,$$

with asymptotic behavior $Ai(t) \underset{+\infty}{\sim} (2\sqrt{\pi})^{-1} t^{-1/4} e^{-\frac{2}{3}t^{3/2}}$.

In all these formulas the O-terms hold uniformly.

We refer for instance to [12] for the proof of the local asymptotics of the GUE eigenvalues using Plancherel-Rotach formulas.

One of the important ideas of the theory is that of universality. This idea, which is a conjecture, is that the statistical properties of the eigenvalues in the local regime do not depend asymptotically on the ensemble, that is the sine kernel is "universal" and appears in other models of Hermitian random matrices. This has been shown for a class of Hermitian Wigner matrices and for unitary invariant ensemble of the form

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

for a general potential V satisfying some assumptions (one recovers the GUE for $V(x) = |x|^2/2$), which is largely motivated by applications in physics. The main difficulty for general V is to derive the asymptotics of orthogonal polynomials associated to V, which are not known explicitly. This can be done using Riemann-Hilbert techniques (see [5]).

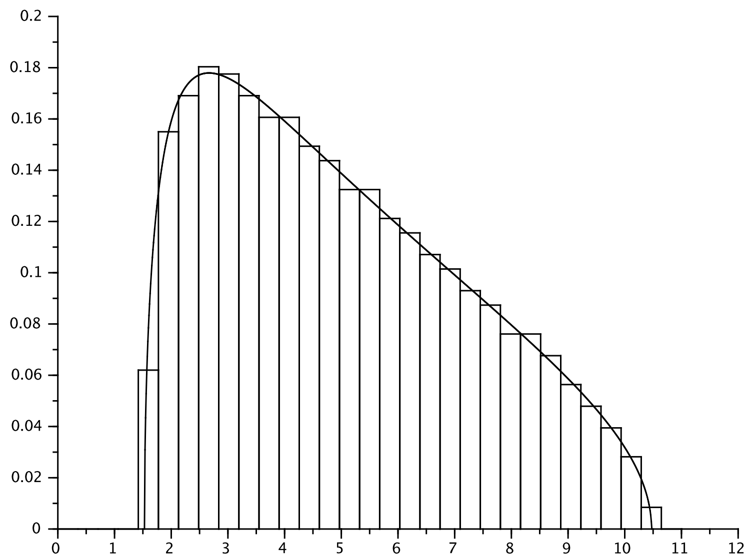

FIGURE 5. Histogram of the singular values of a 1000×5000 Gaussian random matrix and the Marchenko-Pastur distribution.

5. Some generalizations and applications

In telecommunications processing, one is naturally led to study matrix models known as the Wishart ensemble, which are Hermitian positive random matrices of the form AA^*, for A a rectangular random matrix with i.i.d. coefficients. The following theorem extend the Wigner law to the case of Wishart ensemble.

Theorem 5.1 ([3]). *Let A_N be a $N\times p$ ($N\le p$) rectangular random matrix with i.i.d. complex centered Gaussian coefficients, with variance $\mathbb{E}|A_{ij}|^2=1$. Let $W_N=\frac{1}{N}AA^*$. Suppose that as $N,p\to\infty$, we have $\frac{p}{N}\to c$, $c\in[1,+\infty[$. Then, the spectral measure μ_{W_N} of W_N converges weakly almost surely to the probability measure*

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

where $c_\pm=(1\pm\sqrt{c})^2$. The measure μ_{MP} is called the Marčenko-Pastur distribution.

Remark 5.2. One can easily see that if X is a random variable distributed according the semicircular distribution μ_{sc}, then X^2 is distributed according the Marčenko-Pastur distribution $\mu_{MP,1}$.

Figure 5 shows a simulation of the eigenvalues of a large random Gaussian matrix of the Wishart ensemble. Moreover, one has the following.

Proposition 5.3. *Let A_N be a $N \times p$ ($N \leq p$) rectangular random matrix with i.i.d. complex centered Gaussian coefficients, with variance $\mathbb{E}|A_{ij}|^2 = 1$, and $\mathbb{E}|A_{ij}|^4 < \infty$. Suppose that as $N, p \to \infty$, we have $\frac{p}{N} \to c$, $c \in [1, +\infty[$. Then almost surely, one has*

$$\lambda_{\max}(AA^*) \to (1 + \sqrt{c})^2,$$

as N goes to infinity.

Such matrix model can be seen as the sample covariance matrix of a random vector, indeed if one considers for instance a system with N receivers, the signal received at time p will be given by a matrix of the form A. We refer for instance to the book [17] or to the different works of Hachem, Loubaton, Najim... (see for instance [8] and references therein) for such development of random matrix theory for applications to telecommunications. A maybe more useful matrix model in applications to telecommunications is the so-called spiked model. Consider a system with N receivers and r sources, such as antennas. We assume that the number of sources is $r \ll N$. At time p, such a system can be modelized by the matrix

$$\Sigma_N = A_N + P_N,$$

where A is $N \times p$ random matrix with Gaussian coefficients which represents the noise of the system, and P is $N \times p$ fixed r-ranked deterministic matrix, which is the signal actually transmitted. This is a fixed ranked perturbation of a Gaussian random matrix. One can prove that such perturbation does not perturb the system in the global regime of the eigenvalues. That is,

Theorem 5.4. *With the notations above, if μ_N denotes the spectral measure of $\Sigma_N \Sigma_N^*$, we have almost surely*

$$\mu_N \to \mu_{MP},$$

weakly as N, p go to infinity such that $\frac{N}{p} \to c \in]0, 1]$, where μ_{MP} is the Marčenko-Pastur distribution.

However, $\Sigma_N \Sigma_N^*$ might have isolated eigenvalues, that is some eigenvalues will converge outside the support of the Marčenko-Pastur distribution. This is the content of the following proposition.

Proposition 5.5 ([4])**.** *Let r be independent of N. Consider the perturbed matrix model $\Sigma_N = A_N + P_N$, where A_N is a $N \times p$ random matrix with complex Gaussian coefficients, and P_N is a deterministic r-ranked matrix with eigenvalues $\pi_1^{(N)}, \ldots, \pi_r^{(N)}$ such that $\pi_i^{(N)} \to \theta_i$ as $N \to \infty$, for $i = 1, \ldots, r$. Assume that $\frac{N}{p} \to c \in]0, 1]$, as $N \to \infty$. Denote $K = \max\{1 \leq i \leq r | \theta_i > \sqrt{c}\}$ and let $\lambda_{i,N}$ be the eigenvalues of $\Sigma_N \Sigma_N^*$. Then, we have for $i = 1, \ldots, K$,*

$$\lambda_{i,N} \to \frac{(c + \theta_i)(1 + \theta_i)}{\theta_i},$$

while $\lambda_{K+1,N} \to (1 + \sqrt{c})^2$, as N goes to infinity, almost surely.

Note that $\frac{(c+\theta_i)(1+\theta_i)}{\theta_i} > (1 + \sqrt{c})^2$ when $\theta_i > \sqrt{c}$.

The study of the spikes, that is the eigenvalues outside the support of the Marčenko-Pastur distribution, is very useful in applications to telecommunications, for instance in passive signal detection, source localization...

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 [13]*):* $\mathcal{F}$ 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 $[-\pi,\pi]\backslash\{0\} \ni \theta \mapsto u = i\frac{1+e^{i\theta}}{1-e^{i\theta}} = -i\cot\frac{\theta}{2} \in \mathbb{R}$, of the restriction to $[-\pi,\pi] \setminus \{0\}$ of σ. Then we get,

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

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

6.2. **Proof of the concentration inequality for the Gaussian measure of Proposition 3.9.** We follow the argument of Maurey and Pisier as in Tao's book [15]. First recall Jensen's inequality.

Lemma 6.5. *Let X be a real random variable and φ a convex function. Then one has,*

$$\varphi(\mathbb{E}X) \leq \mathbb{E}\varphi(X).$$

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 $\int f d\gamma_n = 0$, denoting $\gamma_n = \gamma_{n,1}$. By symmetry, it then 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

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

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

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

As an exercice, one can prove that $X_\theta := Y\cos\theta + X\sin\theta$ and its derivate $X'_\theta := -Y\sin\theta + X\cos\theta$ are independent Gaussian variables with variance 1. Using again Jensen's inequality, we get

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

and Fubini's theorem gives

$$\mathbb{E}\exp\big(t(f(X) - f(Y))\big) \leq \frac{2}{\pi}\int_0^{\pi/2} \mathbb{E}\exp\Big(\frac{2t}{\pi}\nabla f(X_\theta)\cdot X'_\theta\Big)d\theta.$$

Since X_θ and X'_θ are independent, conditioning by X_θ gives that $\frac{2t}{\pi}\nabla f(X_\theta)\cdot X'_\theta$ is a Gaussian variable with variance bounded by $\frac{4t^2}{\pi^2}$ since $|\nabla f(x)| \leq 1$. Thus one obtains

$$\mathbb{E}\left(\exp\Big(\frac{2t}{\pi}\nabla f(X_\theta)\cdot X'_\theta\Big)\right) \leq \exp(Ct^2),$$

for some absolute constant C, and the proposition follows.

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] 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.

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

[4] 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.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

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

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

[15] 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/`.

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

[17] 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.

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

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

Université Paul Sabatier, Institut de Mathématiques de Toulouse, 118 route de Narbonne, F-31062 Toulouse Cedex 9

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