Random Matrix Theory Notes

Pantelis Tassopoulos[*], Arthur Dremaux[†]

December 2024

Introduction

The aim of these notes is to provide a survey of some basic theory and classical results in random matrix theory, as well as to include more recent progress in the field.

In the first part, given by Pantelis Tassopoulos, we aim to discuss some elementary results from spectral theory that will be nonetheless essential in setting up the probabilistic framework for random matrices. Then, will outline some global properties of the spectrum of Wigner matrices and give an outline of the key ideas in the proof of Wigner's semi-circle law. We then consider issues of fluctuations around this semi-circle law and results when moment conditions are relaxed, highlighting the lack of universality therein.

Arthur's part is mainly just from Wendelin Werner, Alice Guionnet, and Edouard Maurel-Segala's notes.

Throughout, we will give pointers to the relevant literature and encourage the reader to initiate their own investigations in this rich and rewarding field.

[*]DPMMS, Cambridge University.
[†]DPMMS, Cambridge University.

1 Some Spectral Theory

We start with a general result from the theory of finite dimensional Hilbert spaces, the Spectral Theorem.

Theorem 1.1. *(Spectral Theorem) Let V be a finite-dimensional real or complex Hilbert space of dimension $n \in \mathbb{N}$ and let $T : V \to V$ be a self-adjoint linear operator. Then, there exists an orthonormal basis $(v_i)_{i=1}^n$ of V and real scalars $(\lambda_i)_{i=1}^n \in \mathbb{R}$ such that $Tv_i = \lambda_i v_i$ for all $1 \leqslant i \leqslant n$.*

Proof. Omitted, but can be found in any standard reference on the subject. $\square$

In this exposition, V will be $\mathbb{C}^n(\mathbb{R}^n)$, $n \in \mathbb{N}$ throughout these notes, where as usual we identify a self-adjoint linear operator $T : \mathbb{C}^n \to \mathbb{C}^n(\mathbb{R}^n \to \mathbb{R}^n)$ with a Hermitian (resp. symmetric) matrix.

We make an important definition that is guaranteed to us by the spectral theorem 1.1, namely that of an *eigenvalue functional.*

Definition 1.2. *(Eigenvalue Functionals) Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ be its respective eigenvalues. Then for $1 \leqslant i \leqslant n$ we define the i^{th} eigenvalue functional to be that map $A \mapsto \lambda_i(A)$ for A Hermitian.*

We have a simple characterisation of the eigenvalue functionals as a min-max of convex functions.

Theorem 1.3. *(Courant-Fischer) Let A be an $n \times n$ real symmetric or Hermitian matrix. Then we have the following characterisation of the eigenvalue functionals*

$$\begin{aligned} \lambda_i(A) &= \sup_{dim(V)=i} \inf_{v \in V:\|v\|=1} v^* A v \\ &= \inf_{dim(V)=n-i+1} \sup_{v \in V:\|v\|=1} v^* A v \end{aligned} \tag{1}$$

where V denotes any subspace of $\mathbb{R}^n$ or $\mathbb{C}^n$ respectively.[a]

[a]The $(\cdot)^*$ denotes either complex conjugation and transpose in the complex case, and just taking the transpose in of a vector in the real case.

Proof. By the spectral theorem 1.1, there exists an orthonormal basis $(e_i)_{i=1}^n$ of eigenvectors of A. Fix $1 \leqslant i \leqslant n$. Then, we easily obtain the inequality

$$\lambda_i(A) \leqslant \sup_{\dim(V)=i} \inf_{v \in V:\|v\|=1} v^* A v \tag{2}$$

by setting $V = \operatorname{span}\{e_1, \cdots, e_i\}$. To obtain the reverse inequality, observe that for any i-dimensional subspace V of $\mathbb{R}^n$ or $\mathbb{C}^n$, there exists a vector $v \in V, \|v\| = 1$ such that $v^* A v \leqslant \lambda_i(A)$. Indeed, let $W = \operatorname{span}\{e_i, \cdots, e_n\}$ which has codimension $i - 1$ and so $\dim(V \cap W) \geqslant 1$ and so we obtain

$$v^* A v \leqslant \sum_{k=1}^{n} \lambda_i |v_i|^2 \leqslant \lambda_i(A), \tag{3}$$

with v as above. For the second equality, use the equality just proved with $-A$ in place of A and note $\lambda_i(A) = -\lambda_{n-i+1}(-A)$ to conclude. $\square$

This characterisation of the spectrum allows us to make a statement about the *regularity* of the eigenvalue functionals for real symmetric or Hermitian matrices, namely, that they are Lipschitz continuous with respect to the operator norm (equivalently the Frobenius norm). This clarifies any measurability issues when considering random entries and showcases the stability of the spectrum of such matrices.

Corollary 1.4. *(Stability of spectrum) Let $n \in \mathbb{N}$, $A, B \in M_{n\times n}(\mathbb{C}\, or \mathbb{R})$ be Hermitian or real symmetric matrices respectively, then have for all $1 \leqslant i \leqslant n$*

$$|\lambda_i(A+B) - \lambda_i(A)| \leqslant \|B\|_{op} \leqslant \|B\|_F \tag{4}$$

where $\|\cdot\|_{op}$ and $\|\cdot\|_F$ denote the operator and Frobenius norms respectively.

Proof. For $v \in \mathbb{C}^n (\text{or}\, \mathbb{R}^n)$, $\|v\| = 1$, have the following

$$|v^* Bv| \leqslant \|B\|_{\text{op}} = \max\{|\lambda_1(A)|, |\lambda_n(A)|\}. \tag{5}$$

and

$$v^* Av - \|B\|_{\text{op}} \leqslant v^*(A+B)v \leqslant v^* Av + \|B\|_{\text{op}} \tag{6}$$

and conclude by invoking the theorem 1.3. □

Furthermore, we make the observation that in a sense, the 'typical' behaviour of real symmetric or Hermitian matrices is that they have *simple* spectra, that is that their eigenvalues are distinct. To be more precise it is easy to see using the eigenvalue decomposition (theorem 1.1) that the collection of matrices with simple spectra is an open, dense subset of the space of real symmetric or Hermitian matrices. This will have important ramifications for computing densities against the Lebesgue measure, of the laws of the GUE and GOE ensembles, which will be defined later on.

In fact one can obtain more regularity for the eigenvalue functionals when the spectrum is simple and in fact show the following.

Proposition 1.5. *Fix $n \in \mathbb{N}$ and let A be a real symmetric or Hermitian matrix in $M_{n\times n}(\mathbb{R}\, or \mathbb{C})$ with simple spectrum. Then the eigenvalue functionals $\lambda_i(A)$ for $1 \leqslant i \leqslant n$ depend smoothly with respect to A.*

Proof. To observe this, define the smooth function

$$\begin{gathered} F : M_{n\times n} \times \mathbb{R} \to \mathbb{R} \\ (A, \lambda) \mapsto F(A, \lambda) := \det(A - \lambda I) \end{gathered} \tag{7}$$

and observe that in a neighbourhood of a point $(A, \lambda_i(A)), 1 \leqslant i \leqslant n$ where A has simple spectrum, $F(A, \lambda_i(A)) = 0$ and $D_\lambda F \neq 0$ (roots of the characteristic polynomial are simple). Thus, one applies the Implicit Function Theorem to conclude. □

One can prove a similar result and show that one can make a locally smooth choice of eigenvectors, since they are determined up to a sign or complex phase once normalised, in a neighbourhood of a real symmetric or Hermitian matrix with simple spectrum.

Proposition 1.6. *(Smooth dependence of eigenvectors) Fix $n \in \mathbb{N}$ and let A be a real symmetric or Hermitian matrix in $M_{n\times n}(\mathbb{R}\, or \mathbb{C})$ with simple spectrum. Then one can make a smooth selection of eigenvectors $B \mapsto u_i(B)$ for $1 \leqslant i \leqslant n$ satisfying*

$$\begin{cases} Bu_i(B) = \lambda_i(B) u_i(B) \\ u_i^*(B) u_i(B) = 1 \end{cases} \tag{8}$$

for B in a neighbourhood of A, where the $\lambda_i(A)$ are the eigenvalue functionals.

Proof. Let $A \in M_{n\times n}$ be such that its spectrum $\lambda_1(A) > \cdots > \lambda_n(A)$ is simple. Then, observe that the matrix $A - \lambda_i(A)I$ has a one-dimensional kernel. Thus, by an elementary fact from linear algebra, there is an $n-1 \times n-1$ dimensional invertible minor of $A - \lambda_i(A)I$. Since the determinant $B \mapsto \det(B - \lambda_i(B)I)$ is a smooth function of B in a neighbourhood of A, we get that in a neighbourhood of A, the same minor is invertible. This is illustrated in the figure below, the linear system 8 is represented and the invertible minor is the matrix excluding the highlighted rows and columns in yellow (the choice of rows and columns is uniform in a neighbourhood).

$$\begin{pmatrix} \times & \times & \circ & \times \\ \times & \times & \circ & \times \\ \vdots & \vdots & \vdots & \vdots \\ \circ & \circ & \circ & \circ \\ \times & \times & \circ & \times \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ 1 \\ u_n \end{pmatrix} = 0. \tag{9}$$

One can then decompose this linear system into the matrix minor and the highlighted row which by the dimensionality constraint must be expressible as a linear combination of the remaining rows. Thus, this system is invertible, assuming one fixes the coordinate corresponding to the highlighted row (to one for instance) and so one can recover smoothly chosen eigenvectors in a neighbourhood of A and to conclude one normalises them, which is again a smooth proceduce. □

The above enable us to obtain time-evolution equations for the eigenvalue functionals on a smooth one-parameter family of real symmetric or Hermitian matrices A_t. Assume that A_0 has a simple spectrum, then given local smooth choices of eigenvectors $(u_i)_{i=1}^n(t)$ and eigenvalues in a neighbourhood of A_0, hence in a neighbourhood of 0 in t, we can differentiate in time the evolution equations

$$\begin{cases} A_t u_i = \lambda_i(A_t) u_i \\ u_i^* u_i = 1 \end{cases} \tag{10}$$

and obtain *Hadamard's second variation formula*

$$\frac{\mathrm{d}^2\lambda_i}{\mathrm{d}t^2} = u_i^* \ddot{A}_t u_i + 2\sum_{j\neq i} \frac{|u_j^* \dot{A}_t u_i|^2}{\lambda_i - \lambda_j}, \quad 1 \leqslant i \leqslant n \tag{11}$$

where the overhead dots indicate taking entry-wise derivatives twice (in the t-dependent entries of A). This and higher order formulas appear in [TV15], where the authors prove a result that establishes universality results for non-Hermitian matrices assuming independence of the entries satisfying exponential decay, and a moment matching condition (to fourth order). This equation can be interpreted as saying, in analogy with physics, being a second order time evolution equation, that there is a 'repulsive' force between eigenvalues, which becomes arbitrarily large whenever the eigenvalues are sufficiently close. This equation is also in direct analogy with the stochastic differential equation obtained by the spectra of real symmetric and Hermitian matrix-valued Brownian motions, namely Dyson Brownian motion which will be discussed in the final section.

We illustrate however, with an elementary example that the local smoothness of eigenvalues in general cannot be made global, especially when the spectrum ceases to be simple.

Examples 1.7. *Non-global smoothness of eigenvalue functionals Consider the one parameter family of real symmetric matrices*

$$M_t = \begin{bmatrix} t & 0 \\ 0 & -t \end{bmatrix}, \quad t \in (-\epsilon, \epsilon) \tag{12}$$

for any $\epsilon > 0$. Then clearly $\lambda_1(M_t) = |t|$ and $\lambda_2(M_t) = -|t|$ for all t and so the eigenvalue functionals are not smooth at zero, where the spectrum becomes simple.

2 Wigner matrices and the semi-circle law

We now start with the central objects of study in random matrix theory, namely Wigner matrices which we define now. We will state everything in terms of Hermitian matrices but keep in mind that the results clearly also apply to real symmetric ones.

Definition 2.1. *(Wigner matrix) Let* $(Z_{i,j})_{1\leqslant i<j}$, $(Y_i)_{1\leqslant i}$ *be mutually independent infinite families of centred, iid random variables (only between families, not necessarily across them). Define the family of Hermitian matrices*

$$X_{n;i,j} = \overline{X}_{n;j,i} = \begin{cases} Z_{i,j}, & \text{if} 1 \leqslant i < j \leqslant n \\ Y_i, & \text{if} 1 \leqslant i = j \leqslant n \end{cases} \tag{13}$$

and call them the <u>*Wigner ensemble*</u> *corresponding to the* Z, Y *(or in an abuse of notation a Wigner matrix).*

Remark. In a sense this family while very general may seem a bit contrived. For instance the distribution of eigenvalues is not in general invariant under a change of basis. This invariance does hold true in the important special case when $\mathbb{E}[Y_1^2] = 2$ and $\mathbb{E}[Z_{1,2}^2] = 1$, and the Z, Y are gaussian, namely for the GOE and GUE ensembles, which we will explore later on. We consider the initial enlarged class of Wigner matrices because they still exhibit remarkable universal properties.

Before embarking on these universality results we make a small digression and pick up on an observation from the previous section, namely that the 'typical' behaviour of Hermitian matrices is to have a simple spectrum (being an open and dense in the set of Hermitian matrices endowed with the usual norm topology). For random matrices, this question is not so clear a priori due to possible degeneracies in the laws of the entries which could lead to a simple spectrum with high probability. However, assuming no 'asymptotic degeneration', a recent result in [TV14], the authors show that for a general class of random matrices the spectrum becomes asymptotically simple, remarkably relaxing independence between the diagonal entries, summarised in the following result.

Theorem 2.2. *([TV14, Theorem 5]) Fix* $A, \mu > 0$ *and let* $n \in \mathbb{N}$ *be sufficiently large (*$\geqslant n_0(A, \mu)$*). Then, suppose that for all* n*, have an independent family of* $\xi_{n;i,j}$, $1 \leqslant i < j \leqslant n$ *(real or complex valued) such that the 'asymptotic non-degeneration' condition is satisfied*

$$\sup_n \sup_x \mathbb{P}(\xi_{n;i,j} = x) \leqslant 1 - \mu, \quad 1 \leqslant i < j \leqslant n \tag{14}$$

and furthermore let $\xi_{n;i,i}$ *be real random variables independent of the* $\xi_{n;i,j}$ $1 \leqslant i < j \leqslant n$ *set* $\xi_{n;i,j} = \overline{\xi}_{n;j,i}$ *for* $1 \leqslant i, j \leqslant n$ *then the families* $(\xi_{n;i,j})_{1\leqslant i,j\leqslant n}$ *have simple spectrum with probability at least* $1 - 1/n^A$.

Remark. This result applies to a class of random matrices that contains the Wigner matrices, particularly the adjacency matrices of a large family of random graphs, namely the Erdős-Renyi random graphs $G(n, p)$ with n vertices and success probability $p \in (0, 1)$, solving a long standing conjecture due to Babai, namely that the $G(n, 1/2)$ has an asymptotically simple spectrum.

2.1 Global properties of the spectrum of Wigner matrices

Having defined Wigner matrix ensembles, we now proceed with making a couple of preliminary definitions that will allow us to state the classical result due to Wigner, namely the semi-circle law.

Definition 2.3. *(Spectral measure) Fix* $n \in \mathbb{N}$ *and lest* $(X_n)_n$ *be a Wigner ensemble as in 2.1, define its corresponding family of* <u>*spectral measures*</u> *to be the atomic measures*

$$\mu_{X_n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(X_n)}, \quad n \in \mathbb{N} \tag{15}$$

where the λ_i *are the eigenvalue functionals.*

Definition 2.4. *(Semi-circle law) Call the Borel measure*

$$\sigma(A) = \frac{1}{2\pi}\int_A \mathbf{1}_{[-2,2]}(x)\sqrt{4-x^2}\,\mathrm{d}x, \quad A \in \mathcal{B}(\mathbb{R}), \tag{16}$$

where $\mathcal{B}(\mathbb{R})$ *denotes the Borel sets, the* <u>*semi-circle law*</u>.

Remark. Observe that the density is the semi-circle with radius two normalised by area.

Theorem 2.5. *(Wigner) Let* $(X_n)_n$ *be a Wigner ensemble with entries having finite second moments, satisfying the normalisation condition* $\mathbb{E}[|Z_{1,2}|^2] = 1$ *(with* $Z_{1,2}$ *as in definition 2.1). Then we have for all functions* $f \in C_b(\mathbb{R})$[a] *and* $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \int_{\mathbb{R}} f\,\mathrm{d}\mu_{\frac{X_n}{\sqrt{n}}} - \int_{\mathbb{R}} f\,\mathrm{d}\sigma \right| \geqslant \epsilon \right) = 0. \tag{17}$$

[a] That is, continuous and bounded.

Remark. Note that this convergence in probability for a fixed test function can be promoted to almost sure convergence provided the entries are almost surely uniformly bounded.

We will also henceforth make the abbreviation

$$\langle f, \mu \rangle := \int_{\mathbb{R}} f\,\mathrm{d}\mu, \quad f \in C_b(\mathbb{R}),\ \mu \in \mathcal{M}(\mathbb{R}) \tag{18}$$

where $\mathcal{M}(\mathbb{R})$ denotes the space of Borel measures on $\mathbb{R}$ with finite total variation (particular probability measures, which is all that is going to concern us in these notes).

To prove Theorem 2.5, we will first state two supporting lemmas that once combined will make the proof a fairly straightforward consequence thereof.

Lemma 2.6. *Let* $(X_n)_n$ *be a Wigner ensemble with uniformly bounded entries, satisfying the normalisation condition* $\mathbb{E}[|Z_{1,2}|^2] = 1$ *(with* $Z_{1,2}$ *as in definition 2.1). Then we have for any* $k \in \mathbb{N}$

$$\mathbb{E}\left[\langle x^k, \mu_{\frac{X_n}{\sqrt{n}}}\rangle\right] \xrightarrow{n\to\infty} \langle x^k, \sigma\rangle. \tag{19}$$

Note the convergence is deterministic.

Lemma 2.7. *Let* $(X_n)_n$ *be a Wigner ensemble as in lemma 2.6. Then, we have for all* $\epsilon > 0$ *and* $k \geqslant 1$

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \langle x^k, \mu_{\frac{X_n}{\sqrt{n}}}\rangle - \mathbb{E}\left[\langle x^k, \mu_{\frac{X_n}{\sqrt{n}}}\rangle\right] \right| \right) = 0. \tag{20}$$

Assuming for a moment that lemmas 2.6, 2.7 are true, we are led to the following proof of Wigner's Theorem.

Proof. (Wigner, sketch) Let $(X_n)_n$ be a Wigner ensemble with uniformly bounded entries, satisfying the normalisation condition $\mathbb{E}[|Z_{1,2}|^2] = 1$ (with $Z_{1,2}$ as in definition 2.1). An elementary calculation gives that the moments of the semi-circle law σ are

$$\langle x^k, \sigma \rangle = \begin{cases} C_{\frac{k}{2}}, & k \,\text{even} \\ 0, & k \,\text{odd} \end{cases} \tag{21}$$

where $C_n = \binom{2n}{n}$, $n \geqslant 1$ are the Catalan numbers and one can estimate them by $C_n \leqslant 4^n$, $n \in \mathbb{N}$. We now claim that for all $\epsilon > 0$ and $k \geqslant 1$

$$\limsup_{n\to\infty} \mathbb{P}(\langle |x|^k \cdot \mathbf{1}_{|x|>5}, \mu_{\frac{X_n}{\sqrt{n}}} \rangle > \epsilon) = 0. \tag{22}$$

Indeed, notice that by Markov's inequality and lemma 2.6

$$\begin{aligned} &\limsup_{n\to\infty} \mathbb{P}(\langle |x|^k \cdot \mathbf{1}_{|x|>5}, \mu_{\frac{X_n}{\sqrt{n}}} \rangle > \epsilon) \\ &\leqslant \limsup_{n\to\infty} \frac{\mathbb{E}\left[(\langle |x|^{2k} \cdot \mathbf{1}_{|x|>5}, \mu_{\frac{X_n}{\sqrt{n}}} \rangle > \epsilon\right]}{\epsilon \cdot 5^k} \leqslant \frac{\langle |x|^{2k}, \sigma \rangle}{\epsilon \cdot 5^k} \leqslant \frac{4^k}{\epsilon \cdot 5^k},\ k \geqslant 1. \end{aligned} \tag{23}$$

Since $\limsup_{n\to\infty} \mathbb{P}(\langle |x|^k \cdot \mathbf{1}_{|x|>5}, \mu_{\frac{X_n}{\sqrt{n}}} \rangle > \epsilon)$ is increasing in k and the upper bound decreasing geometrically to zero, it can only be that equation 22 holds. This essentially is a localisation result showing that any contribution from a test function outside the compact interval $[-5, 5]$ is negligible in the limit, in probability. One can without loss of generality consider test functions with support in $[-5, 5]$ and essentially reduce the proof to showing convergence in probability against polynomials supported on $[-5, 5]$, using Weierstrass' Approximation Theorem from Analysis. But this is nothing more that combining the results from lemmas 2.6 and 2.7 and invoking linearity.

To relax the boundedness assumption and prove the result in the generality of the statement of the theorem, one needs to perform a truncation argument, wherein one considers the entries

$$\hat{X}_{n;i,j} = \frac{1}{\sigma_{i,j}(C)}(X_{n;i,j} \cdot \mathbf{1}_{|X_{n;i,j}| \leqslant C} - \mathbb{E}\left[X_{n;i,j} \cdot \mathbf{1}_{|X_{n;i,j}| \leqslant C}\right]),\ n \geqslant 1 \tag{24}$$

where

$$\sigma^2_{i,j}(C) = \begin{cases} \mathrm{Var}(X_{n;i,j}\mathbf{1}_{|X_{n;i,j}| \leqslant C}), & 1 \leqslant \quad i \neq j \leqslant n \\ 1, & 1 \leqslant i = j \leqslant n \end{cases} \tag{25}$$

and $C > 0$ is a sufficiently large cut-off constant such that $\sigma^2_{i,j}(C) > 0$, which can be chosen uniformly in $n, i, j \in \mathbb{N}$. It is not hard to show that

$$\hat{X}_{n;i,j} \overset{C\to\infty}{\longrightarrow} X_{n;i,j},\ n \in \mathbb{N},\ \text{in}\ L^2(\mathbb{P}) \tag{26}$$

uniformly in $n, i, j \in \mathbb{N}$. Now observe that we have shown Wigner's theorem applies to the matrix ensemble $(\hat{X}_n)_n$ and can conclude be showing weak convergence with respect to bounded Lipschitz functions which are dense in $C_b(\mathbb{R})$ with the local uniform topology. $\square$

Proof. (Lemma 2.6) First notice that we have the identity for $k \geqslant 1$

$$\langle x^k, \mu_n \rangle = \int_{\mathbb{R}} x^k \mu_n(\mathrm{d}x) = \frac{1}{n^{\frac{k}{2}+1}} \mathrm{Tr} X_n^k \tag{27}$$

and taking expectations yields

$$\mathbb{E}\langle x^k, \mu_n \rangle = \frac{1}{n^{\frac{k}{2}+1}} \sum_{i_1,\cdots,i_k=1} \mathbb{E} X_{i_1,i_2} \cdot \cdots \cdot X_{i_{k-1},i_k} \cdot X_{i_k,i_1}. \tag{28}$$

Observe that every term in the above expectation corresponds to a path of length k on the set of vertices $\{i_1, \cdots, i_k\}$ with k edges $i_j i_{j \mathrm{mod} k+1}$, $1 \leqslant j \leqslant k$. Since the X_i, j are centred and independent, we only get contributions from paths where every edge is traversed at least twice (possibly in reverse). Hence, there can be at most $k/2$ unique edges and at most $k/2 + 1$ unique vertices in $\{i_1, \cdots, i_k\}$.

Figure 1: Illustration of type sequences induced from their corresponding paths when $k = 4$.

Define the weight t of a sequence $\underline{i} = i_1, i_2, \cdots, i_k$ to be the number of distinct indices. For $1 \leqslant t \leqslant \frac{k}{2} + 1$, define the set

$$\pi_t = \{1 \leqslant i_1, \cdots, i_k \leqslant n : |i_1, \cdots, i_k| = t\}. \tag{29}$$

For $1 \leqslant t \leqslant \frac{k}{2}+1$, define the equivalence relation $\sim$ on π_t, where we declare $\underline{i} \sim \underline{i}'$ for $\underline{i}, \underline{i}' \in \pi_t$ if and only if there exists a bijection π on $\{1, \cdots, n\}$ such that $i_j \mapsto_\pi i'_j$, $1 \leqslant j \leqslant k$. The number of distinct equivalence classes in π_t only depends on k (and not n) since one can always pick a representative in $\{1, \cdots, k\}$ for every $\underline{i} \in \pi_t$. Then using that there exists some deterministic $C > 0$ such that for all i, j $|X_{i,j}| \leqslant C$ almost surely, we can conclude that for $t < \frac{k}{2} + 1$

$$\begin{aligned} &\frac{1}{n^{\frac{k}{2}}} \sum_{t=1}^{\frac{k}{2}} \sum_{\underline{i} \in \pi_t} |\mathbb{E} X_{i_1,i_2} \cdot \cdots \cdot X_{i_{k-1},i_k} \cdot X_{i_k,i_1}| \\ &\leqslant C^k \frac{1}{n^{\frac{k}{2}+1}} \sum_{t=1}^{\frac{k}{2}} \cdot \#\{[\underline{i}]_\sim : \underline{i} \in \pi_t\} \leqslant O\left(n^{t-\frac{k}{2}+1}\right) \to 0 \text{ as } n \to \infty \end{aligned} \tag{30}$$

Where we used that $\#\{[\underline{i}]_\sim : \underline{i} \in \pi_t\} = \binom{n}{t} \leqslant n^t$. Thus, we see that if k is odd, then $\mathbb{E}\langle x^k, \mu_n \rangle \to 0 = \langle x^k, \sigma \rangle$ as $n \to \infty$. When k is even, asymptotically the only contribution comes from $t = \frac{k}{2} + 1$.

Thus, let k be even and consider $\pi_{\frac{k}{2}+1}$, that is the collection of sequences $\underline{i}$ corresponding to paths visiting exactly $\frac{k}{2}$ edge where each edge is traversed exactly twice. Define the type value of an an edge appearing in the path $\underline{i}$ to be $+1$ if it has not been traversed before and -1 otherwise. Now, using the equivalence relation $\sim$, we see that $\underline{i} \sim \underline{i}'$ if and only if they have the same type sequence.

Now, taking the cumulative sum of the type values of a path $\underline{i}$, (and linearly interpolating between integer values), we obtain a path from 0 to k starting and ending at zero. For instance, when $k = 4$, we have two possible such paths corresponding to type sequences, see figure 1 below.
Thus, we have

$$\#\{\text{equivalence classes in } \pi_{\frac{k}{2}+1}\} = \#\{\text{type sequences of length } k\} = C_{\frac{k}{2}}, \tag{31}$$

where that last equality is a known fact and can be shown by establishing a recurrence relation for $\#\{\text{type sequences of length } k\}$ and showing that it is the same as that for the Catalan numbers $C_{\frac{k}{2}}$. Piecing together the above with 28 and 30, we conclude the proof. ☐

Proof. (Lemma 2.7, [Fei12, Lemma 3.2.2] sketch) One can argue similarly as in the proof of lemma 2.6 using an albeit more involved combinatorial argument and show that

$$\operatorname{Var}\left(\frac{1}{n}\operatorname{Tr}(X_n^k)\right) = O\left(\frac{1}{n^2}\right) \to 0,\ n \to \infty \tag{32}$$

which in fact is sufficient to promote the convergence in probability to almost sure convergence, by an application of Borel-Cantelli, assuming the boundedness condition throughout. ☐

Remark. (Theorem 2.5) A shortfall of this method (i.e. the moment method) is that it is not constructive in the sense that the proof does not construct the semi-circle law explicitly as we only show convergence to the moments which do not have to a priori uniquely determine a probability distribution. That they do can be shown using a separate argument and follows from the geometric growth of the moments.

2.2 Free probability

We make another small digression into the notion of free probability and its connection to random matrices. In particular, one can interpret Wigner's theorem, theorem 2.5 as an analogue of the Central Limit Theorem for a different kind of probability, namely free probability where the notion of an underlying sample space is abstracted away and one proceeds with a formalism where the primitive objects are the random variables themselves, which can exist as elements of a more abstract space with certain closure properties and convergence in distribution is replaced a notion extending that of convergence of moments. To begin establishing this connection, we make the following definition, namely that of a non-commutative probability space.

Definition 2.8. *(Non-commutative probability space) A complex unital algebra $\mathcal{A}$ with an involution $A \to A, x \mapsto x^*$ s.t.*

- $(\lambda x + \mu y)^* = \overline{\lambda} x^* + \overline{\mu} y^*$
- $(xy)^* = y^* x^*$
- $x^{**} = x$
- $\mathbf{1} = \mathbf{1}^*$*, where* $\mathbf{1}$ *denotes the unit of* $\mathcal{A}$

for all $x, y \in A, \lambda, \mu \in \mathbb{C}$ and a $\mathbb{C}$-linear operator $\tau : \mathcal{A} \to \mathbb{C}$ such that $\tau(x^*) = \overline{\tau(x)}$, $x \in \mathcal{A}$ and $\tau(\mathbf{1}) = 1$ is called a non-commutative probability space.

Remark. The operator τ can be thought of as generalising the expectation operator and the condition $\tau(\mathbf{1}) = 1$ is supposed to reflect the fact that in Kolmogorov's formalism, the probability of the sample space (corresponding the the expectation of the constant random variable equal to unity) is one.

Examples 2.9 (Examples of Free probability spaces)**.** *The pairs $(L^{\infty-}, \mathbb{E}), (M_{n\times n}, \frac{1}{n} Tr), (L^{\infty-} \otimes M_{n\times n}, \mathbb{E}\frac{1}{n} Tr)$ where the $*$ operation is the complex conjugation and hermitian transpose respectively, are all examples of non-commutative probability spaces.*

The point is to define a notion of convergence that allows for comparison against different domains of random variables that depends on a generalisation of convergence of moments. We thus are led to make the definition of convergence in moments.

Definition 2.10. *(Free convergence) Let $(\mathcal{A}_n, \tau_n)_{n\in\mathbb{N}}$, $(\mathcal{A}_\infty, \tau_\infty)$ be non-commuting probability spaces. For each $n \in \mathbb{N}$, let X_n be a sequence of elements (random variables) in $\mathcal{A}_n$ and likewise X_∞ in $\mathcal{A}_\infty$. We say X_n converges to X_∞ in the sense of moments is for all $k \geqslant 1$,*

$$\tau_n(X_n) \overset{n\to\infty}{\longrightarrow} \tau_\infty(X_\infty). \tag{33}$$

Definition 2.11. *(Semi-circular element) Let $(\mathcal{A}, \tau)$ be a non-commutative probability space and let $X \in \mathcal{A}$ be such that $X^* = X$ and have moments $\tau(X^n) = 0$, n odd and $\tau(X^n) = C_{n/2}$, n even where $C_\cdot$ denote the Catalan numbers. Then, X is called a semi-circular element of unit variance.*

We can thus re-interpret lemma 2.6 as saying the following.

Theorem 2.12. *(Free Central Limit Theorem) Let $(X_n)_{n\in\mathbb{N}}$ be a family of Wigner matrices where the entries have al moments finite (could assume entries with sub-gaussian tails, say) and $\mathbb{E}[|Z_{1,2}|^2] = 1$. Then the rescaled matrices $M_n = X_n/\sqrt{n} \in L^{\infty-} \otimes M_{n\times n}$ converge in the sense of moments to any semi-circular element belonging to some non-commutative probability space.*

Remark. Examples of semi-circular elements are the identity function $x \mapsto x$ in $L^\infty(\mathrm{d}\sigma)$ endowed with trace operator $\tau(f) = \int_{\mathbb{R}} f \,\mathrm{d}\sigma$, the function $\theta \mapsto 2\cos(\theta)$ in the Lebesgue space $L^\infty([0,\pi], \frac{2}{\pi}\sin^2(\theta)\,\mathrm{d}\theta)$ and less trivial examples like $U + U^*$, where $U : e_n \mapsto e_{n+1}$ is the right shift operator on $\ell^2(\mathbb{N})$ on the underlying non-commutative space of all bounded operators $B(\ell^2(\mathbb{N}))$ with trace $\tau(X) = \langle e_0, Xe_0\rangle_{\ell^2(\mathbb{N})}$ where $e_0, e_1, \cdots$ is the standard basis of $\ell^2(\mathbb{N})$ and $*$ denotes the Hilbert adjoint. [Tao10] has a nice introductory exposition of free probability and goes on to discuss more important aspects of the theory that involve tools from spectral theory.

3 Limits of Universality

We saw in the previous section that the Wigner's theorem, theorem 2.5 shows that the global convergence of the spectrum of Wigner matrices is largely stable against heavy-tailed laws, only requiring the first two moments of entries being finite. However, universality has limits when considering the fluctuations of local statistics of the spectrum of Wigner matrices. In this direction, we start by providing a result regarding the fluctuations of the largest eigenvalues λ_1 of Wigner matrix ensembles which exhibits a remarkable asymmetry, in a sense which be clarified below.

A natural question one may ask having established Wigner's theorem is whether the largest eigenvalue converges in probability to the right boundary of the support of the semi-circle measure σ. One direction is a very quick corollary from Wigner's theorem which we now prove.

Corollary 3.1. *Let $(X_n)_n$ be a Wigner ensemble whose entries have finite second moments, satisfying the normalisation condition $\mathbb{E}[|Z_{1,2}|^2] = 1$ (with $Z_{1,2}$ as in definition 2.1). Then, for all $\delta > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\lambda_1\left(\frac{X_n}{\sqrt{n}}\right) < 2 - \delta\right) = 0. \tag{34}$$

Proof. Fix $\delta > 0$ and let f be a non-negative smooth, compactly supported function supported on $[2-\delta, 2]$ satisfying the condition $\langle f, \sigma\rangle = 1$. On the event where $\lambda_1(X_n/\sqrt{n}) < 2 - \delta$ $n \in \mathbb{N}$, the spectral measure $\mu_{X_n/\sqrt{n}}$ is supported on $(-\infty, 2-\delta)$ and so $\langle f, \mu_{X_n/\sqrt{n}}\rangle = 0$. Hence, by inclusion we have the upper bound

$$\mathbb{P}\left(\lambda_1\left(\frac{X_n}{\sqrt{n}}\right) < 2 - \delta\right) \leqslant \mathbb{P}\left(|\langle f, \mu_{X_n/\sqrt{n}}\rangle - \langle f, \sigma\rangle|\right) \xrightarrow{n\to\infty} 0 \tag{35}$$

by Wigner's theorem. □

One might hope to attain a complimentary estimate for the probability of $\lambda_1(X_n/\sqrt{n}) > 2 + \delta$, $\delta > 0$, but this is not true in general as $\lambda_1(X_n/\sqrt{n})$ may even fail to converge in general when the entries of the Wigner matrices have sufficiently heavy tails. For instance, in [BY88] the authors show that $\lambda_1(X_n/\sqrt{n}) \to 2$ almost surely if and only if the fourth moments are finite. However, imposing all moments being finite and satisfying certain growth conditions, we obtain the following.

Theorem 3.2. *([Kem13, Theorem 6.2]) Let $(X_n)_n$ be a Wigner ensemble whose entries have all moments finite, satisfying the normalisation condition $\mathbb{E}[|Z_{1,2}|^2] = 1$ (with $Z_{1,2}$ as in definition 2.1) and a moment growth condition (see [Kem13, proof of Theorem 6.2]). Then, for all $\epsilon, \delta > 0$*

$$\lim_{n\to\infty} \mathbb{P}\left(n^{1/6-\epsilon}\left(\lambda_1\left(\frac{X_n}{\sqrt{n}}\right) - 2\right) > \delta\right) = 0. \tag{36}$$

Remark. This finally gives convergence in probability of the largest eigenvalue $\lambda_1(X_n/\sqrt{n})$ to two and can be interpreted as saying that there is no non-trivial structure at scales with exponent arbitrarily close to 1/6, though in general this is not tight and can be improved to $O(n^{-2/3})$ for matrices with symmetric distributions, see [SS98] where the authors show that largest eigenvalues at this scale have a non-trivial distributional limit in the family of the *Tracy-Wigom* distributions.

One can also obtain various limiting results for heavy-tailed entries which do not correspond to the Gaussian case, particularly the following.

Theorem 3.3. *([AG11, Thm 21.2.7] Heavy-tailed distributions) Let $(Z_{i,j})_{1\leqslant i<j}$, $(Y_i)_{1\leqslant i}$ be mutually independent infinite families of centred, iid random variables distributed with law P on $\mathbb{R}$ such that there is a 'slowly varying function' $L(\cdot)$ such that*

$$P(|x| \geqslant u) \leqslant \frac{L(u)}{u^\alpha}, \quad u > 0 \tag{37}$$

for some $\alpha \in (0,2)$. Then, with $\alpha_n = \inf\{u > 0 : |P(|x| \geqslant u)| \leqslant 1/n\}$, $n \in \mathbb{N}$ (which is of order $\sim n^{\frac{1}{\alpha}} \gg n^{\frac{1}{2}}$ for $\alpha < 2$), the empirical measures

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(X_n)/\alpha_n}, \quad n \in \mathbb{N} \tag{38}$$

converge for all test functions $f \in C_b(\mathbb{R})$ almost surely

$$\langle f, \mu_n \rangle \overset{n\to\infty}{\longrightarrow} \langle f, \mu_\alpha, \rangle, \tag{39}$$

where μ_α is symmetric, with unbounded support and smooth density satisfying the asymptotics $\rho_\alpha(x) \sim 1/|x|^{\alpha+1}, |x| \to \infty$.

Finally, we mention an example where local fluctuations around the semi-circle law, when interpreted as the maximal separation between the cumulative distribution functions of the laws of the empirical and limiting measures, are not in general universal and with decay that depends on the number of moments available.

Theorem 3.4. *([Bai93, Theorem 4.1] Local fluctuation cdf) With*

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(X_n)/\sqrt{n}}, \quad n \in \mathbb{N} \tag{40}$$

the spectral measures of Wigner matrix ensemble $(X_n)_n$, then if $\mathbb{E}[|Y_1|^4] + \mathbb{E}[|Z_{1,2}|^4] < \infty$

$$\sup_{x\in\mathbb{R}} |\mathbb{E}[\mu_n((-\infty, x]) - \sigma([-2, x \vee -2])]| = O(n^{-\frac{1}{4}}) \tag{41}$$

and if instead $\mathbb{E}[|Y_1|^6] + \mathbb{E}[|Z_{1,2}|^6] < \infty$, then

$$\sup_{x\in\mathbb{R}} |\mathbb{E}[\mu_n((-\infty, x]) - \sigma([-2, x \vee -2])]| = O(n^{-\frac{1}{2}}). \tag{42}$$

4 GOE and GUE

One of the initial motivating reasons that led Wigner to study random matrices is that they are randomly chosen linear operators. Viewed in this way, the chosen (orthonormal) basis of $\mathbb{R}^N$ that one chooses to represent this linear operator as a matrix should not really matter in the model. In other words, one would like to have a model of random matrices with the property that for any $N \times N$ orthogonal matrix O, the law of the random $N \times N$ matrix X is invariant under $X \mapsto OXO^{-1}$.

For general Wigner matrices, when one changes the orthonormal basis, then it looks as if some correlations between the different obtained matrix elements do appear, and this is indeed the case in the generic case — which then shows that such matrices with independent inputs are in general not invariant in distribution under change of basis.

As we shall see now, there however exists a natural and very interesting particular choice of random symmetric matrices that have both properties: (a) the matrix elements are independent, and (b) the law of the matrix is invariant under (orthonormal) change of coordinates. This is the Gaussian orthogonal ensemble that will be discussed now.

Definition 4.1 (Unnormalized Gaussian Orthogonal Ensemble (GOE))**.** *In this model, $X_N \in \mathcal{S}_N(\mathbb{R})$ is a matrix with independent entries, such that:*

- $X_{i,i} \sim \mathcal{N}(0, 2)$,
- *for* $i < j$, $X_{i,j} \sim \mathcal{N}(0, 1)$.

To see that this is what we looked for, notice how we can describe the law of the GOE. We write $x = (x_{i,j})$ and denote the eigenvalues of x by λ_1, ..., λ_N. Then

Lemma 4.2. *The probability distribution of $(X_{i,j})$ at $(x_{i,j})$ with respect to the Lebesgue measure dx is proportional to :*

$$\exp\left(-\frac{1}{4}\mathrm{Tr}(x^2)\right) = \exp\left(-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2\right)$$

Proof.

$$-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2 = -\frac{1}{4}\mathrm{Tr}(x^2) = -\frac{1}{4}\sum_{i=1}^{N}(x^2)_{i,i} = -\frac{1}{4}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{i,j}x_{j,i} = -\frac{1}{2}\sum_{i<j}x_{i,j}^2 - \frac{1}{4}\sum_{i=1}^{N}x_{i,i}^2$$

□

So the probability distribution at x only depends of the eigenvalues of x, which are of course invariant under conjugation by an orthogonal matrix. This allows us to get :

Lemma 4.3. *The GOE distribution is invariant under conjugation $X \mapsto OXO^T$ by an orthogonal matrix $O \in \mathcal{O}_N(\mathbb{R})$.*

Proof. Density proportional to $\exp(-\frac{1}{4}\mathrm{Tr}(X^2))$ which is invariant under conjugation, and the determinant of the jacobian is 1 (these are basically rotations in N dimensional space) (conjugation is a group action and compactness of $\mathcal{O}_N(\mathbb{R})$)
OR : It is a centered Gaussian vector, we can compute the covariances... □

Little detour by Haar measure : Now that we know the GOE is invariant by conjugation by an orthogonal matrix, we would like to find a probability law on $\mathcal{O}_n(\mathbb{R})$ that is in a sense "uniform over $\mathcal{O}_n(\mathbb{R})$", i.e. invariant under translation, just as the most natural probability measure (i.e., uniform measure) on the circle is defined by rotation invariance. Phrased slightly differently, the distribution of a uniform random element of $\mathcal{O}_n(\mathbb{R})$ should be a translation invariant probability measure μ on $\mathcal{O}_n(\mathbb{R})$: for any measurable subset $\mathcal{A} \subseteq \mathcal{O}_n(\mathbb{R})$ and any fixed $M \in \mathcal{O}_n(\mathbb{R})$,

$$\mu(M\mathcal{A}) = \mu(\mathcal{A}\mathcal{M}) = \mu(\mathcal{A})$$

We actually know how to do this in a very general case :

Theorem 4.4 (Haar)**.** *If G is a locally compact topological group, then there exists an unique (up to multiplicative constant) Borel measure μ such that :*

- *μ is invariant by left-translation*
- *μ is finite on compact sets*

$\mathcal{O}_n(\mathbb{R})$ is a compact group, which means the Haar measure is finite on it, so we can see it as a probability measure.
Now that we have this measure, we are finally able to describe the GOE as "$X = O\Delta O^T$", where Δ is diagonal and gives the law of the eigenvalues of the GOE, and O is independant and follows $\mathrm{Haar}(\mathcal{O}_n(\mathbb{R}))$.

Lemma 4.5. *There exists a probability ν on $\mathbb{R}^N$ and λ with law ν independent of $O \sim \mathrm{Haar}(\mathcal{O}_n(\mathbb{R}))$ such that $O\mathrm{diag}(\lambda)O^T$ follows the law of the GOE.*

Proof. Let X be a GOE. Almost surely, the eigenvalues are distinct. Let Λ be the diagonal matrix of the ordered spectrum, and (O_i) an orthonormal family of associated eigenvectors. Then $X = O\Lambda O^T$.

If $P \sim \mathrm{Haar}(\mathcal{O}_n(\mathbb{R}))$ independently of X, and $Q = PO$. Then $Q \sim \mathrm{Haar}(\mathcal{O}_n(\mathbb{R}))$ independently of Λ and $Q\Lambda Q^T \sim PXP^T \sim X$. So it is a GOE.

□

Theorem 4.6. *Let $\Delta(\lambda) = \prod_{i<j}(\lambda_j - \lambda_i)$ be the Vandermonde determinant of $\lambda \in \mathbb{R}^N$. Consider the random variable λ with density $\Delta(\lambda)e^{-\sum \lambda_i^2/2}d\lambda$. Let $O \sim$ Haar$(\mathcal{O}_N(\mathbb{R}))$ be a random orthogonal matrix independent of λ. Then $O\,\mathrm{diag}(\lambda)O^T$ follows the law of the GOE.*

Proof. We only have to check that the law of the spectrum of $O\,\mathrm{diag}(\lambda)O^T$ is the same as that of the GOE. Two ways to do it : either directly, using that since the spectrum is simple, the map $\Psi : (O, \lambda) \mapsto O\mathrm{diag}(\lambda)O^T$ is $N!$-to-1, it suffices to compute the determinant of the Jacobian of this map. Or we can do it directly on a diffeomorphism by looking in a neighborhood of identity at the exponential map over antisymmetric matrices.

Since $O^T = O^{-1}$ we have

$$D_{O,\Lambda}\Psi(dO, d\Lambda) = Od\Lambda O^T + dO\Lambda O^T - O\Lambda O^{-1}dOO^{-1}$$

Since everything is invariant under conjugation, it suffices to look at $O = I_N$.

$$D_{I,\Lambda}\Psi(dO, d\Lambda) = d\Lambda + dO\Lambda - \Lambda dO$$

So we have

$$\begin{aligned} D_{I,\Lambda}\Psi(0, E_{i,i}) &= E_{i,i} \\ D_{I,\Lambda}\Psi(E_{i,i}, 0) &= E_{i,i} \\ D_{I,\Lambda}\Psi(E_{i,j}, 0) &= (\lambda_i - \lambda_j)E_{i,j} \end{aligned}$$

So the determinant is $\Delta(\lambda)$. □

Remark. The $\Delta(\lambda)$ factor is due to the shape of $\mathcal{S}_N(\mathbb{R})$.

Remark. This gives us the law of the eigenvalues of the unnormalized GOE. A priori it is not easy to guess the order of magnitude of these eigenvalues because there are two competing effects: the $e^{-\sum \lambda_i^2/2}$ gaussian term prevent them from being too large, while $\Delta(\lambda)$ tend to repel the N eigenvalues from one another, which means they could become large when we make $N \to \infty$. Wigner's theorem in fact gives us that they will localize at order of magnitude $\sqrt{N}$.

Definition 4.1 (Gaussian Unitary Ensemble (GUE)). In this model, $X^N \in \mathcal{H}_N(\mathbb{C})$ is a matrix with independent entries, such that:

- $X_{i,i} \sim \mathcal{N}(0,1)$,
- for $i < j$, $\mathrm{Re}(X_{i,j})$ and $\mathrm{Im}(X_{i,j})$ are independent and follow $\mathcal{N}\left(0, \frac{1}{2}\right)$.

The normalization is chosen such that $\mathbb{E}\left[|X_{i,j}^N|^2\right] = 1$, regardless of the choice of i and j.

Remark. The law of the GUE has density $\exp\left(-\frac{1}{2}\mathrm{Tr}H^2\right)$ at $H \in \mathcal{H}_N(\mathbb{C})$.

Theorem 4.7. *For a Gaussian Unitary Ensemble (GUE), the empirical spectral measure $\hat{\mu}_N$ of the eigenvalues of X_N converges almost surely in distribution to the semicircular law as $N \to \infty$. The law of the eigenvalues has density $\Delta(\lambda)^2 e^{-\sum \lambda_i^2/2} d\lambda$.*

Proof. Similar to GOE. See [TV15] for further details.
((Stieltjes transform + concentration inequality (Talagrand Lipschitz))) □

Now that we know the law of the spectra of the GOE and GUE, it is natural to have a look at the process defined by replacing the Gaussian random variables by independent Brownian motions. We get a random matrix $X(t)$, with $X_{i,j}(t) = B_{i,j}(t)$ for $i < j$ and $X_{i,i}(t) = B_{i,i}(t)$. Then $X(t)/\sqrt{t}$ has the law of a GOE.

For any fixed t, we know the law of the eigenvalues of $X(t)$, but we want to study the coupling, i.e. the evolution of $\lambda_i(t)$ with t.

Theorem 4.8 (Dyson Brownian motion)**.** *The eigenvalues $\lambda_i(t)$ follow the law of the Dyson Brownian motion, i.e.*

$$d\lambda_i(t) = \sqrt{\frac{2}{\beta}} dB_{i,i}(t) + \sum_{k \neq i} \frac{1}{\lambda_i - \lambda_k} dt$$

with $\beta = 1$ for the GOE and $\beta = 2$ for the GUE.

Proof. (Sketch)(We prove it for the GOE; the proof for the GUE case is analogous) Let $O(t), \Lambda(t)$ be such that $X(t) = O(t)\Lambda(t)O(t)^T$ with $\Lambda(t)$ diagonal and independent of $O(t)$ orthogonal (see [AGZ10, Theorem 4.3.2] for details). Then for all t, $X(t)O(t) = O(t)\Lambda(t)$. If we admit they are semi-martingales, $dO = dM^O + V^O dt$ and $d\Lambda = dM^\Lambda + V^\Lambda dt$. Differentiating and taking the martingale and finite variation parts, we get (double differentials indicate covariation processes)

$$dXO + XdM^O = dM^O\Lambda + OdM^\Lambda$$

$$XV^O dt + \frac{1}{2} dXdO = V^O \Lambda dt + OV^\Lambda dt.$$

Once again, by invariance under orthogonal transformations, we can change basis and so it suffices to look at $O(t) = I_N$, so $X(t) = \Lambda$ (fixing t arbitrary), treating the stochastic and Lebesgue differentials as 'infinitesimals'. Now,

$$dX = \left[dM^O\Lambda - \Lambda dM^O\right] + dM^\Lambda \tag{43}$$

$$V^\Lambda dt = \left[\Lambda V^O - V^O\Lambda\right] dt + \frac{1}{2} dXdM^O \tag{44}$$

Let us now compute $dXdM^O$. O and Λ are independent so $dM^\Lambda dM^O = 0$ and by (43) we get for $i \neq j$ (since it is clearly 0 on the diagonal):

$$\left[dM^O\Lambda - \Lambda dM^O\right]_{i,j} = (\lambda_i - \lambda_j) dM_{i,j}^O = dX_{i,j} - dM_{i,j}^\Lambda = dX_{i,j}$$

So

$$\frac{1}{2}\left(dXdM^O\right)_{i,i} = \sum_{k\neq i}(\lambda_i-\lambda_k)dM^O_{i,k}dM^O_{k,i} = -\sum_{k\neq i}\frac{1}{\lambda_i-\lambda_k}dX_{i,k}dX_{k,i} = -\sum_{k\neq i}\frac{1}{\lambda_i-\lambda_k}dt$$

Since $dM^O\Lambda - \Lambda dM^O$ and $\Lambda V^O - V^O\Lambda$ have zero diagonal,

$$d\lambda_i(t) = (dM^\Lambda)_{i,i} + (V^\Lambda)_{i,i}dt = dX_{i,i} + \frac{1}{2}\left(dXdM^O\right)_{i,i} = \sqrt{\frac{2}{\beta}}dB_{i,i}(t) + \sum_{k\neq i}\frac{1}{\lambda_i-\lambda_k}dt.$$

One can also consult [TV15] and [AGZ10, Theorem 4.3.2] for more details. □

4.1 Dyson Brownian motion as non-intersecting Brownian motions and the Pitman transform

Finally, we mention that in the papers [O'C02], the author establishes a procedure for reconstructing the law of $n \in \mathbb{N}$ Dyson Brownian motions from n independent standard Brownian motions by performing deterministic operations on said Brownian motions, which amounts to performing a 'bubble sort', where the sorting operator on pairs of continuous functions known as the *Pitman transform*, which we briefly describe below.

With $f = (f_1, f_2)$ where $f_i : [0,\infty) \mapsto \mathbb{R}$ for $i = 1,2$. For $f \in \mathcal{C}^2_{*,*}([0,\infty))$, we define $Wf = (\mathrm{W}f_1, \mathrm{W}f_2) \in \mathcal{C}^2_{*,*}([0,\infty))$ (the space of pairs of continuous functions on the positive reals), the Pitman transform of f as follows. For $x < y \in [0,\infty)$, define the maximal gap size

$$G(f_1,f_2)(x,y) := \max\{\max_{s\in[x,y]}(f_2(s)-f_1(s)), 0\}\,.$$

Then define

$$\begin{aligned}\mathrm{W}f_1(t) &= f_1(t) + G(f_1,f_2)(0,t)\,,\\ \mathrm{W}f_2(t) &= f_2(t) - G(f_1,f_2)(0,t)\,,\end{aligned} \tag{45}$$

for all $t \in [0,\infty)$.

Suppose we replace f_1, f_2 with two independent Brownian motions B_1, B_2 starting from the origin. We now analyse the law of

$$WB_1(\cdot) = B_1(\cdot) + \left(\max_{0\leqslant s\leqslant \cdot}(B_2(s)-B_1(s))\right) \vee 0. \tag{46}$$

as a random continuous function in $\mathcal{C}_{*,*}([0,y])$. We have the following characterisation of two dimensional Dyson Brownian motion $(\lambda_1^{\mathrm{GUE}}, \lambda_2^{\mathrm{GUE}})$ (using the GUE version).

Theorem 4.9. *([O'C02, Theorem 2]) The processes $(\lambda_1^{GUE}, \lambda_2^{GUE})(\cdot) \overset{d}{=} (WB_1, WB_2)(\cdot)$, where the distributional equality is understood as that for elements of $\mathcal{C}_{*,*}([0,\infty))$.*

Furthermore, observe that by an orthogonal transformation, we have the distributional equality (as $\mathcal{C}_{*,*}([0,y])$-valued random variables)

$$WB_1(\cdot) \overset{d}{=} \frac{B_1(\cdot)-B_2(\cdot)}{\sqrt{2}} + \sqrt{2}\left(\max_{0\leqslant s\leqslant\cdot} B_2(s)\right),$$

since Brownian motion attains positive values arbitrarily close to the origin almost surely. By [Pit74], we have that

$$-B_2(\cdot) + 2\cdot\max_{0\leqslant s\leqslant\cdot}\cdot B_2(s) \overset{d}{=} \mathrm{BES}^0(3)(\cdot) \tag{47}$$

where $\mathrm{BES}^3(0)$ denotes the Bessel-3 process started from zero, that is to say to say the distribution of the radial part of a three-dimensional Brownian motion started from the origin. Now, observing that $WB_1(\cdot)$ corresponds to the top line rate of a two Dyson Brownian motion (arising from two non-colliding Brownian motions), we have the decomposition

$$\lambda_1^{\mathrm{GUE}} \stackrel{d}{=} \frac{1}{\sqrt{2}}\mathrm{BM}(0) + \frac{1}{\sqrt{2}}\mathrm{BES}^3(0), \tag{48}$$

where $\mathrm{BM}(0)$ and $\mathrm{BES}^3(0)$ denote a Brownian motion starting from the origin and an independent Bessel-3 process starting from the origin, respectively.

References

[AG11] G Ben Arous and Alice Guionnet. Wigner matrices. *The Oxford Handbook of Random Matrix Theory*, pages 433–451, 2011.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge university press, 2010.

[Bai93] Zhi Dong Bai. Convergence rate of expected spectral distributions of large random matrices. part i. wigner matrices. *The Annals of Probability*, pages 625–648, 1993.

[BY88] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

[Fei12] Adina Roxana Feier. *Methods of proof in random matrix theory*, volume 9. Citeseer, 2012.

[Kem13] Todd Kemp. Math 247a: Introduction to random matrix theory. *Lecture notes*, pages 3–4, 2013.

[O'C02] Neil O'Connell. Random matrices, non-colliding processes and queues, 2002.

[Pit74] J. W. Pitman. One-dimensional brownian motion and the three-dimensional bessel process. *Advances in Applied Probability*, 6(2):223–224, 1974.

[SS98] Ya G Sinai and Aleksandr Borisovich Soshnikov. A refinement of wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *FUNCTIONAL ANALYSIS AND ITS APPLICATIONS C/C OF FUNKTSIONAL'NYI ANALIZ I EGO PRILOZHENIIA*, 32:114–131, 1998.

[Tao10] Terrenc Tao. https://terrytao.wordpress.com/2010/02/10/245a-notes-5-free-probability/, Feb 2010.

[TV14] Terence Tao and Van Vu. Random matrices have simple spectrum, 2014.

[TV15] Terence Tao and Van Vu. Random matrices: Universality of local spectral statistics of non-hermitian matrices. *The Annals of Probability*, 43(2), March 2015.