

# Random Matrix Theory in a nutshell

## Part II: Random Matrices

Manuela Girotti

based on M. Girotti's PhD thesis,  
A. Kuijlaars' and M. Bertola's lectures from Les Houches Winter School 2012,  
and B. Eynard's notes from IPhT Saclay 2015

### Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>A little bit of history</b>	<b>2</b>
<b>3</b>	<b>Unitary matrices</b>	<b>3</b>
3.1	Eigenvalues distribution . . . . .	4
3.2	Dyson's Theorem . . . . .	6
<b>4</b>	<b>Asymptotics and universality</b>	<b>7</b>
4.1	Macroscopic behaviour . . . . .	7
4.1.1	Wigner's semicircle law . . . . .	8
4.2	Microscopic behaviour . . . . .	10
4.2.1	Bulk universality . . . . .	10
4.2.2	Soft-edge universality . . . . .	11
<b>5</b>	<b>A zoo of random matrix models</b>	<b>12</b>
<b>6</b>	<b>Bonus: applications</b>	<b>14</b>

### 1 Introduction

A random matrix is a matrix whose elements are randomly distributed. A random matrix model is characterized by a matrix ensemble  $\mathcal{E}$  and a probability measure  $d\mu(M)$  for  $M \in \mathcal{E}$  (the *random matrix law*), thus the matrix itself is a random variable.

Let  $\mathfrak{M}$  be a space of matrices of given size: e.g.

- Hermitian matrices ( $M = M^\dagger$ ) of size  $n \times n$ :  $\mathfrak{M} = \{ M \in \text{Mat}_n(\mathbb{C}) \mid M_{ij} = M_{ji}^* \}$  (**Unitary ensemble**)

- Symmetric matrices ( $M = M^T$ ) of size  $n \times n$ :  $\mathfrak{M} = \{ M \in \text{Mat}_n(\mathbb{R}) \mid M_{ij} = M_{ji} \}$  (**Orthogonal ensemble**)
- Symplectic matrices:  $M^T J = J M^T$ , with  $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \otimes \mathbf{1}_n$  of size  $2n \times 2n$  (**Symplectic ensemble**)
- Rectangular matrices of size  $n \times K$
- $\mathfrak{M} = \text{Mat}_n(\mathbb{C})$
- etc.

**Remark 1.** *The first three examples are extensively studied and their names refer to the compact group that leaves the measure invariant.*

A simple way to define a probability measure on these ensembles relies on the remark that each of these spaces is a vector space and thus carries a natural flat Lebesgue measure (invariant by translations) which we shall denote by  $dM$ .

Then, starting from  $dM$ , we can equip each of these spaces with a probability measure of the form

$$d\mu(M) = F(M)dM$$

where  $F : \mathfrak{M} \rightarrow \mathbb{R}_+$  is some suitable ( $L^1(dM)$ ) function of total integral 1 (this is a measure which is *absolutely continuous* with respect to Lebesgue measure).

The main objective in Random Matrix Theory typically is to study the statistical properties of the spectra (for square matrices ensembles) or singular values (for rectangular ensembles). In order to do so, we need to develop an understanding of the joint probability distribution functions (jpdf) of the eigen-/singular-values.

Additional interest is the study of the properties of the said statistics when the size of the matrix ensemble tends to infinity (under suitable assumption on the probability measure).

Random Matrices are one of those transversal theories who may appear in different fields of Mathematics and Physics, providing unexpected links between for example Probability, Number Theory and Integrable Systems.

Among the vast literature on Random Matrix Theory, we can mention the book by Mehta [3] and the book by Anderson, Guionnet and Zeitouni [1].

## 2 A little bit of history

The first appearance of the concept of a random matrix dates back to the Fifties and it is due to the physicist E.P. Wigner. In the field of Nuclear Physics, Wigner wished to describe the general properties of the energy levels of highly excited states of heavy nuclei, as measured in nuclear reactions. In particular, he wanted to study the spacings between those energy levels.

Such a complex nuclear system is usually represented by a Hermitian operator  $\mathcal{H}$ , called the Hamiltonian, defined on an infinite-dimensional Hilbert space and governed by physical laws. However, except for very specific and simple cases,  $\mathcal{H}$  is unknown or very hard to compute.

On the other hand, the real quantities of interest are the eigenvalues of  $\mathcal{H}$ , which represent the energy levels:

$$\mathcal{H}v = \lambda v \quad (1)$$

where  $v$  is the eigenfunction associated to the eigenvalue  $\lambda$ .

Wigner argued that one should regard a specific Hamiltonian  $\mathcal{H}$  as behaving like a large-dimension matrix with random entries. Such a matrix is thought as a member of a large class of Hamiltonians, all of which would have similar general properties as the specific Hamiltonian  $\mathcal{H}$  in question ([6]). As a consequence, the eigenvalues of  $\mathcal{H}$  could then be approximated by the eigenvalues of a large random matrix and the spacings between energy levels of heavy nuclei could be modelled by the spacings between successive eigenvalues of a random  $n \times n$ -matrix as  $n \rightarrow +\infty$ .

It turns out that the ensemble of the random eigenvalues is a determinantal point process. Therefore, studying the spacings or gaps between eigenvalues means studying the gap probabilities of the determinantal system. Furthermore, the distribution of the largest eigenvalue obeys a different law on its own and is governed by the so called “Tracy-Widom” distribution ([5]), which can still be considered as a gap probability on an interval of the type  $[s, +\infty)$ ,  $s \in \mathbb{R}$  (the eigenvalues, or in general the points of a DPP, are always confined in finite positions on the real line).

### 3 Unitary matrices

We will focus from now on on the case of Hermitian matrices (Unitary ensemble). This is a vector space with the real diagonal entries  $\{M_{ii}\}_{i=1}^n$  and the real and imaginary part of the upper diagonal elements  $\{\Re M_{ij}, \Im M_{ij}\}_{i < j}$  as independent coordinates:

$$M_{ij} = \Re M_{ij} + i\Im M_{ij}, \quad \text{with} \quad \Re M_{ij} = \Re M_{ji}, \Im M_{ij} = -\Im M_{ji}, \quad n = 1, \dots, n.$$

Its dimension is equal to

$$\dim \mathfrak{M} = \frac{n(n+1)}{2} + \frac{n(n-1)}{2} = n^2$$

and the corresponding Lebesgue measure reads

$$dM = \prod_{i=1}^n dM_{ii} \prod_{i=1}^{n-1} \prod_{j=i+1}^n d\Re M_{ij} d\Im M_{ij}. \quad (2)$$

We also recall the following properties of Hermitian matrices:

**Theorem 2 (Spectral Theorem).** *Any Hermitian matrix can be diagonalized by a Unitary matrix*

$$U \in \mathcal{U}(n) = \{ U \in \text{GL}_n(\mathbb{C}) \mid U^\dagger U = U U^\dagger = \mathbf{1}_n \}$$

*and its eigenvalues are real:*

$$M = U^\dagger X U, \quad X = \text{diag} \{x_1, \dots, x_n\}, \quad x_j \in \mathbb{R}. \quad (3)$$

**Remark 3.** *The diagonalization is not unique even if  $X$  has distinct eigenvalues, because of the ordering of the eigenvalues, so in general there are  $n!$  distinct diagonalizations.*

Additionally, the Lebesgue measure (2) is invariant under conjugation with a unitary matrix

$$dM = d(UMU^\dagger), \quad U \in \mathcal{U}(n) \quad (4)$$

(more generally, for all ensembles of square matrices the Lebesgue measure is invariant under conjugation:  $dM = d(CMC^{-1})$ ).

In view of these properties, we can perform a strategic change of variables

$$\begin{aligned} M &\mapsto (X, U) \\ \{M_{ii}, i = 1, \dots, n; \Re M_{ij}, \Im M_{ij}, i < j\} &\mapsto \{x_1, \dots, x_n; u_{ij}\}, \end{aligned} \quad (5)$$

where  $u_{ij}$  are the parameters that parametrize the unitary group. Under such transformation, the Lebesgue measure reads (thanks to the Weyl integration formula)

$$dM = c_n \Delta(X)^2 dX dU \quad (6)$$

where  $c_n = \frac{\pi^{n(n-1/2)}}{\prod_{j=1}^n j!}$ ,

$$\Delta(X) = \prod_{1 \leq i < j \leq n} (x_i - x_j) = \det \left[ x_a^{b-1} \right]_{1 \leq a, b \leq n} \quad (7)$$

is the *Vandermonde determinant* and  $dU$  is the Haar measure on  $\mathcal{U}(n)$ .

**Remark 4.** Similarly, for the other two main cases

$$\begin{aligned} \text{Orthogonal} \quad dM &\sim |\Delta(X)| dX dU \\ \text{Symplectic} \quad dM &\sim \Delta(X)^4 dX dU \end{aligned}$$

where  $dU$  is the Haar measure in the respective compact group ( $\mathcal{O}(n)$  or  $\mathcal{Sp}(2n)$ ). Since the exponent of the Vandermonde determinant  $\Delta(X)$  is  $\beta = 1, 2, 4$  (Orthogonal, Unitary, Symplectic ensembles), they are also universally known as the  $\beta = 1, 2, 4$  ensembles.

### 3.1 Eigenvalues distribution

We now want to equip the space  $\mathfrak{M}$  with a probability measure. Consider again a measure  $d\mu(M)$  that is absolutely continuous with respect to Lebesgue:

$$d\mu(M) = F(M) dM \quad (8)$$

with  $F \in L^1(\mathfrak{M}, dM)$  and  $\int F(M) dM = 1$ . Thus, under the “change of variable” performed before,

$$d\mu(\vec{x}, U) = c_n F(U^\dagger X U) \Delta(X)^2 dx_1 \dots dx_n dU$$

If we are interested only on the eigenvalues one can study the reduced measure

$$\begin{aligned} d\mu(\vec{x}) &= \Delta(X)^2 dx_1 \dots dx_n \times \left( \int_{\mathcal{U}(n)} c_n F(U^\dagger X U) dU \right) \\ &= \Delta(X)^2 \tilde{F}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned} \quad (9)$$

where  $\tilde{F}$  needs to be a symmetric function of the  $n$  arguments.

The connection to Orthogonal Polynomials becomes possible when  $\tilde{F}$  is the product of a single function of the individual eigenvalues:

$$\tilde{F}(x_1, \dots, x_n) \sim \prod_{j=1}^n e^{-V(x_j)} \quad (10)$$

( $V(x)$  is called the *potential*).

**Remark 5.** A sufficient condition for the probability (9) to be well-defined is that

$$\lim_{|x| \rightarrow +\infty} \frac{V(x)}{\ln(1+x^2)} = +\infty. \quad (11)$$

A standard example is when  $V(x)$  is a polynomial of even degree, with positive leading coefficient (e.g.  $V(x) = x^2$ ).

In conclusion, the probability measure on the space of matrices (8) induces a joint probability density on the eigenvalues given by

$$d\mu(x_1, \dots, x_n) = \frac{1}{Z_n} \Delta(x_1, \dots, x_n)^2 \prod_{j=1}^n e^{-V(x_j)} dx_1 \dots dx_n \quad (12)$$

with  $Z_n = \int_{\mathbb{R}^n} d\mu(x_1, \dots, x_n)$  a suitable normalization constant (*partition function*).

**Paradigma.** The **Gaussian Unitary Ensemble** (GUE) is the ensemble on Hermitian matrices equipped with the probability measure

$$d\mu(M) = \frac{1}{Z_n} e^{-\frac{1}{2} \text{Tr } M^2} dM. \quad (13)$$

Since

$$\begin{aligned} \text{Tr } M^2 &= \sum_{i=1}^n (M^2)_{ii} = \sum_{i=1}^n \sum_{j=1}^n M_{ij} M_{ji} = \sum_{i=1}^n M_{ii}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n |M_{ij}|^2 = \\ &= \sum_{i=1}^n M_{ii}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[ (\Re M_{ij})^2 + (\Im M_{ij})^2 \right], \end{aligned} \quad (14)$$

the probability measure (13) factorizes as a product of Gaussians

$$d\mu(M) = \frac{1}{Z_n} \prod_{i=1}^n e^{-\frac{1}{2} M_{ii}^2} dM_{ii} \prod_{i=1}^{n-1} \prod_{j=i+1}^n \left( e^{-(\Re M_{ij})^2} d\Re M_{ij} \right) \left( e^{-(\Im M_{ij})^2} d\Im M_{ij} \right). \quad (15)$$

Therefore, in GUE all the entries  $\{\Re M_{ij}, \Im M_{ij}\}_{i < j}$  and  $\{M_{ii}\}$  are mutually independent normal random variable with zero mean and different variances for the diagonal and off-diagonal entries:

$$\Re M_{ij}, \Im M_{ij} \sim \mathcal{N}\left(0, \frac{1}{2}\right) \quad M_{ii} \sim \mathcal{N}(0, 1). \quad (16)$$

Furthermore, the induced joint probability density of the eigenvalues is

$$d\mu(x_1, \dots, x_n) = \frac{1}{Z_n} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 e^{-\frac{1}{2} \sum_{j=1}^n x_j^2} dx_1 \dots dx_n. \quad (17)$$

### 3.2 Dyson's Theorem

We start from the eigenvalue distribution. Calling  $W(X)$  the Vandermonde matrix

$$W(X) = \begin{bmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{bmatrix} \quad (18)$$

then, the Vandermonde determinant is  $\Delta(X) = \det W(X)$ :

$$\begin{aligned} d\mu(x_1, \dots, x_n) &= \frac{1}{Z_n} \Delta(x_1, \dots, x_n)^2 \prod_{i=1}^n e^{-V(x_i)} dx_1 \dots dx_n \\ &= \frac{1}{Z_n} \left( \det \left[ x_i^{j-1} \right]_{1 \leq i, j \leq n} \right)^2 \prod_{i=1}^n e^{-V(x_i)} dx_1 \dots dx_n \\ &= \frac{1}{Z_n} \prod_{i=1}^n e^{-\frac{V(x_i)}{2}} \det \left[ x_i^{j-1} \right]_{1 \leq i, j \leq n} \cdot \det \left[ x_i^{j-1} \right]_{1 \leq i, j \leq n} \prod_{i=1}^n e^{-\frac{V(x_i)}{2}} dx_1 \dots dx_n \\ &= \frac{1}{Z_n} \det \left[ e^{-\frac{V(X)}{2}} \right] \det \left[ x_i^{j-1} \right]_{1 \leq i, j \leq n} \cdot \det \left[ x_i^{j-1} \right]_{1 \leq i, j \leq n} \det \left[ e^{-\frac{V(X)}{2}} \right] dx_1 \dots dx_n \\ &= \frac{1}{Z_n} \det \left[ x_i^{j-1} e^{-\frac{V(x_i)}{2}} \right]_{1 \leq i, j \leq n} \cdot \det \left[ x_i^{j-1} e^{-\frac{V(x_i)}{2}} \right]_{1 \leq i, j \leq n} dx_1 \dots dx_n \end{aligned} \quad (19)$$

where we used  $V(X) = \text{diag}\{V(x_1), \dots, V(x_n)\}$ .

**Proposition 6.** *The partition function is*

$$Z_n = n! \det \mathbb{M}, \quad (20)$$

where the matrix  $\mathbb{M}$  has entries

$$\mathbb{M}_{ab} = \int_{\mathbb{R}^n} x^{a+b} e^{-V(x)} dx \quad 0 \leq a, b \leq n-1. \quad (21)$$

*Proof.* From the definition of the constant  $Z_n$  and (19), use the definition of the determinant of a matrix to obtain the right-hand-side.  $\square$

Even better,

**Proposition 7.**

$$\frac{1}{Z_n} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n e^{-V(x_i)} = \frac{1}{n!} \det \left[ K_n(x_i, x_j) \right]_{1 \leq i, j \leq n} \quad (22)$$

where

$$K(x, y) = e^{-\frac{V(x)+V(y)}{2}} \sum_{j,k=0}^{n-1} x^j [\mathbb{M}]_{jk}^{-1} y^k \quad (23)$$

*Proof.*

$$\begin{aligned}
\frac{1}{n!} \det [K_n(x_a, x_b)]_{1 \leq i, j \leq n} &= \frac{1}{n!} \det \left[ \sum_{j, k} e^{-\frac{V(x_a)}{2}} x_a^j [\mathbb{M}]_{jk}^{-1} x_b^k e^{-\frac{V(x_b)}{2}} \right] = \frac{1}{n!} \det \left[ e^{-\frac{V(X)}{2}} W \mathbb{M}^{-1} W^T e^{-\frac{V(X)}{2}} \right] \\
&= \frac{1}{n! \det \mathbb{M}} [\det W(X)]^2 e^{-\text{Tr } V(X)} = \frac{1}{Z_n} \Delta^2(X) e^{-\text{Tr } V(X)}
\end{aligned} \tag{24}$$

□

Finally,

**Proposition 8.** *The kernel  $K(x, y)$  has the following properties:*

1. (**reproducibility**)  $\int_{\mathbb{R}} K(x, z) K(z, y) dz = K(x, y)$
2. (**normalization**)  $\int_{\mathbb{R}} K(x, x) dx = n$
3. (**marginals**)  $\int_{\mathbb{R}} \det [K(x_i, x_j)]_{i, j \leq r} dx_r = (n - r - 1) \det [K(x_i, x_j)]_{i, j \leq r-1}$
4. (**marginals**)  $\int_{\mathbb{R}^{n-r}} \det [K(x_i, x_j)]_{i, j \leq n} dx_{r+1} \dots dx_n = (n - r)! \det [K(x_i, x_j)]_{i, j \leq r-1}$

**Conclusion 1:** Dyson's theorem says that the joint probability density function of the eigenvalues and all its marginals are in a determinantal form. Therefore, the set of random eigenvalues of a (unitary) matrix ensemble is a determinantal point process!

**Conclusion 2:** As already seen in the DPP part of the notes, the whole statistical information is contained in the kernel  $K(x, y)$ .

## 4 Asymptotics and universality

As it was said in the introduction, a typical question one asks when dealing with random matrices is what happens to the statistical properties of the eigenvalues when the size of the matrix ensemble tends to infinity (or equivalently, when the number of eigenvalues grows).

The goal is to find **asymptotic behaviours** or asymptotic properties of the probability distribution and its related quantities.

A property is considered **universal** if it only depends on the matrix ensemble, and not – or almost not – on the probability measure (in particular, we have independency with respect to the choice of the potential  $V(x)$ ).

This is a quite vague description of the picture and in fact there are different ways to study the asymptotics of a matrix ensemble.

### 4.1 Macroscopic behaviour

We are interested in the distribution of the eigenvalues as a whole, when the dimension of the matrix grows.

Consider a matrix ensemble and denote the (ordered) eigenvalues by  $x_1 \leq x_2 \leq \dots \leq x_n$ . The empirical spectral distribution of the eigenvalues is defined by

$$d\mu_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}(x) dx \quad (25)$$

where  $\delta_x$  is the Dirac delta function centered at  $x$ .

Numerical evaluations shows that as we increase the size of the matrix and at the same time we scale down the phase space by a suitable power of  $n$ , we can notice a limiting shape appearing from the histograms of the eigenvalues: see Figure 1.

This means that the (rescaled) eigenvalues do not escape at infinity. In many cases of interest the eigenvalue density has a finite limit as  $n \rightarrow +\infty$ , called the **equilibrium density**  $\rho(x)$ .

We saw that the ensemble of eigenvalues is a determinantal point process and in particular the 1-point correlation function (or density function) can be given in terms of the correlation kernel as

$$\rho_1(x) = K_n(x, x). \quad (26)$$

**Proposition 9.** *The equilibrium density (if it exists) can be computed as*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} K_n(x, x) = \rho(x). \quad (27)$$

#### 4.1.1 Wigner's semicircle law

One notable example is the case where we consider the GUE ensemble. It is possible to show that the second moment of the eigenvalue distribution measure of the matrix  $M$  behaves like  $n$  ( $\mathbb{E}[M^2] \sim n$ ) and therefore it is divergent. On the other hand, if we “smartly” rescale the matrices

$$\widetilde{M} = \frac{1}{\sqrt{n}} M, \quad (28)$$

then the corresponding eigenvalue distribution density has finite moments. Its limit distribution has a very peculiar shape as a semicircle.

**Theorem 10 (Wigner's semicircle law).** *Consider the GUE ensemble of size  $n$  with matrices  $\frac{1}{\sqrt{n}}M$ , then the spectrum distribution converges as  $n \rightarrow +\infty$ , in probability, to the following deterministic probability density*

$$\rho(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2} & \text{if } |x| \leq 2 \\ 0 & \text{otherwise.} \end{cases} \quad (29)$$

*Observation on the proof.* There are several ways to prove the theorem. Historically this was proven using the so-called “moments method”, but it can also be proved using the Stieltjes transform and other more recent methods.  $\square$

The semicircle law is characteristic for a large class of random matrices. The minimum requirements are that the matrices are hermitian (or symmetric, if we are considering matrices with real entries), with mean zero and finite variance independent entries:  $\mathbb{E}[M_{ij}] = 0$ ,  $\mathbb{E}[M_{ij}^2] < +\infty$ ,  $i = 1, \dots, n$ ,  $j = i, \dots, n$ . Such type of matrices are generally called Wigner matrices.

This is already one example of the **universality** feature mentioned at the beginning of this section.



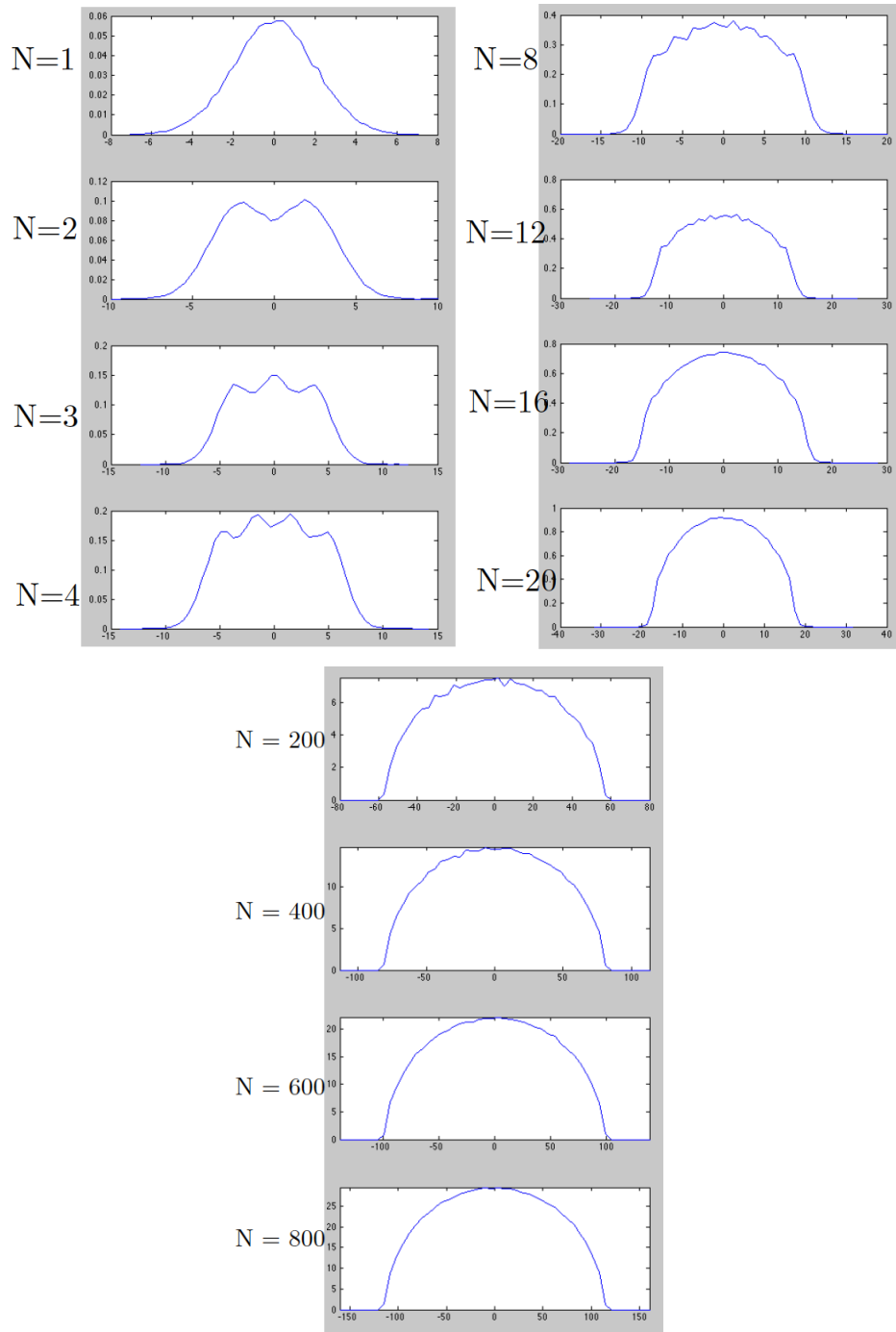


Figure 1: Histograms of the eigenvalues of GUE matrices as the size of the matrix increases (such histograms are produced with rescaling the spectrum by a factor  $1/\sqrt{n}$ ). Numerical simulation with MATLAB (courtesy of prof. Ken McLaughlin).

**Remark 11.** *If we consider random square matrices with independent entries but without symmetry (i.e. the entries  $M_{ij}$  are independent for all  $i, j$ ) another universal pattern emerges, the so-called circular law.*

*In particular, if the entries  $M_{ij}$  have zero mean and finite variance, the empirical density of eigenvalues converges to the uniform measure on the unit disk in the complex plane.*

*If independence is dropped, one can get many different density profiles.*

## 4.2 Microscopic behaviour

Another aspect of interest about the distribution of eigenvalues is the local (infinitesimal) behaviour of the eigenvalue distribution in specific points of the spectrum in the limit as  $n \rightarrow +\infty$ .

Focusing on the GUE ensemble, the two settings that we can consider are points that lie in the interior of the spectrum (the **bulk**) or that lie on the boundary of the spectrum (the **edge**, meaning the largest or the smallest eigenvalue). In order to study their statistical behaviour we will again make use of the results seen in the DPP theory.

We recall the following result about transformations of DPPs.

**Proposition 12.** *Let  $\mathbb{P}$  and  $\mathbb{P}_n$  be determinantal point processes with kernels  $K$  and  $K_n$  respectively. Let  $K_n$  converge to  $K$*

$$\lim_{n \rightarrow \infty} K_n(x, y) = K(x, y) \quad (30)$$

*uniformly over compact subsets of  $\mathbb{R} \times \mathbb{R}$ . Then, the point processes  $\mathbb{P}_n$  converge to  $\mathbb{P}$  weakly.*

Given a fixed reference point  $x^*$  of the spectrum, center and scale the DPP of eigenvalues, i.e. apply the change of variables

$$x \mapsto Cn^\gamma(x - x^*) \quad (31)$$

to the correlation kernel, with suitable values of  $C$ ,  $\gamma > 0$ , depending on the RM model and on where we are focusing on (the edges behaviour or the bulk behaviour). We can now perform the limit:

$$\lim_{n \rightarrow \infty} \frac{1}{Cn^\gamma} K_n \left( x^* + \frac{x}{Cn^\gamma}, x^* + \frac{y}{Cn^\gamma} \right) = K(x, y) \quad (32)$$

with  $x, y$  the new local coordinates of the limit-DPP.

### 4.2.1 Bulk universality

A point  $x^*$  lies in the bulk of the spectrum if the equilibrium density doesn't vanish  $\rho(x^*) \neq 0$  (we are actually requiring that the density doesn't vanish in a whole neighbourhood of  $x^*$ ).

Pick a point  $x^*$  in the bulk of the spectrum and introduce the following change of variables:

$$x = x^* + \frac{\xi}{n\rho(x^*)} \quad y = x^* + \frac{\eta}{n\rho(x^*)} \quad (33)$$

**Theorem 13 (Bulk universality at the origin).** *For the Unitary Ensemble, the local behaviour in the bulk of the spectrum is described by a DPP with correlation kernel given by*

$$\lim_{n \rightarrow +\infty} \frac{1}{n\rho(x^*)} K_n \left( x^* + \frac{\xi}{n\rho(x^*)}, x^* + \frac{\eta}{n\rho(x^*)} \right) = K_{\text{sine}}(\xi, \eta) \quad (34)$$

with

$$K_{\text{sine}}(\xi, \eta) = \frac{\sin(\pi(\xi - \eta))}{\pi(\xi - \eta)}. \quad (35)$$

This results holds regardless on the choice of the potential  $V(x)$  (despite some reasonable properties that we require  $V(x)$  to have). Here is another universality results.

#### 4.2.2 Soft-edge universality

In the case of points  $x^*$  close to a spectral edge  $a$ , the definition of an edge microscopic limit will depend on the behaviour of the equilibrium density  $\bar{\rho}(x)$  near  $a$ .

For a generic potential  $V$ , we have **regular edges** or **soft edges** if the density vanishes as a square root:  $\rho(x) \sim \sqrt{x - a}$ . On the other hand for special choices of the potential, we can have that the vanishing of the density has a different regime  $\rho(x) \sim (x - a)^{\frac{p}{q}}$  for some positive integers  $p, q$ ; in this case, the edges are called **critical**.

In the case of UE with regular one-cut potential (e.g.  $V(x) = x^2$ ), both edges (on  $a = \pm 2\sqrt{n}$ ) are regular and the microscopic limit is described as

**Theorem 14 (Soft-edge universality).**

$$\lim_{n \rightarrow +\infty} \frac{1}{2n^{\frac{1}{6}}} K_n \left( \pm 2\sqrt{n} + \frac{\xi}{2n^{\frac{1}{6}}}, \pm 2\sqrt{n} + \frac{\eta}{2n^{\frac{1}{6}}} \right) = K_{\text{Airy}}(\xi, \eta) \quad (36)$$

with

$$K_{\text{Airy}}(\xi, \eta) = \frac{\text{Ai}(\xi)\text{Ai}'(\eta) - \text{Ai}'(\xi)\text{Ai}(\eta)}{\xi - \eta}. \quad (37)$$

**Note 15.** The function  $\text{Ai}(z)$  is the Airy function. It satisfies the second-order ODE

$$y'' = zy, \quad \text{such that} \quad \lim_{z \rightarrow +\infty} y(z) = 0. \quad (38)$$

It can be represented as a contour integral

$$\text{Ai}(z) = \int_{\gamma} e^{\frac{\zeta^3}{3} - z\zeta} \frac{d\zeta}{2\pi i} \quad (39)$$

where the curve  $\gamma \subseteq \mathbb{C}$  is an oriented contour starting at  $\infty$  with argument  $-\frac{\pi}{3}$  and ending at  $\infty$  with argument  $\frac{\pi}{3}$ .

The corresponding gap probabilities of this (limit) DPP describe the local behaviour of the largest eigenvalue in the spectrum and it infinitesimal random oscillations are described by the celebrated Tracy–Widom distribution:

**Theorem 16 (Tracy–Widom distribution).** Consider the semi-infinite interval  $[s, +\infty)$ , then the distribution of the largest eigenvalue of the GUE ensemble obeys the following law

$$\det \left( \mathbf{1}_{L^2(\mathbb{R})} - K_{\text{Ai}} \Big|_{[s, +\infty)} \right) = \exp \left\{ - \int_s^\infty (x - s) q^2(x) dx \right\} \quad (40)$$

where  $q(x)$  is the Hasting-McLeod solution to the Painlevé II equation:

$$\begin{aligned} q''(x) &= 2q^3(x) + sq(x) \\ q(x) &\sim \text{Ai}(x) \quad x \rightarrow +\infty. \end{aligned} \tag{41}$$

**Remark 17 (food for thought).** The Tracy–Widom distribution can be seen as the  $\tau$ -function of the Painlevé II integrable system.

## 5 A zoo of random matrix models

In these notes we only focused on unitary ensembles, i.e. squared Hermitian matrices with a given (smooth) potential  $V(x)$ . On the other hand, there is a vast variety of possible matrix models and the sky is the limit.

Here we'll mention just a few.

**Wishart ensemble.** The Wishart ensemble is the ensemble of matrices of the form  $M = XX^T$ , where  $X$  is a rectangular matrix  $X \in \text{Mat}_{n \times m}(\mathbb{R})$  with i.i.d entries,  $\mathbb{E}[X_{ij}] = 0$ ,  $\mathbb{E}[X_{ij}^2] = 1$ .

The corresponding eigenvalue distribution has the following expression

$$d\mu(x) = \prod_{i < j} |x_i - x_j| \prod_i x_i^\alpha e^{\frac{1}{2} \sum_i x_i} dx_1 \dots dx_n. \tag{42}$$

If we properly rescale the ensemble by

$$\widetilde{M} = \frac{1}{n} M \tag{43}$$

and take the limit as  $n \rightarrow +\infty$ , while assuming that  $\frac{m}{n} \rightarrow \kappa$  ( $\kappa \in (0, 1]$ ), then the limit distribution of eigenvalues is defined on a bounded interval  $[a_-, a_+]$  depending on  $\kappa$

$$a_- = (1 - \sqrt{\kappa})^2, \quad a_+ = (1 + \sqrt{\kappa})^2 \tag{44}$$

and it is equal to

$$\rho_\kappa(x) = \frac{1}{2\pi\kappa} \frac{\sqrt{(a_+ - x)(x - a_-)}}{x}. \tag{45}$$

This distribution is called Marchenko–Pastur law (see Figure 2).

In particular, if  $\kappa = 1$ , then the distribution has a square-root singularity at  $x = 0$ :

$$\bar{\rho}_1(x) = \frac{1}{2\pi} \sqrt{\frac{4 - x}{x}} \quad x \in (0, 4]. \tag{46}$$

In this configuration, the point  $x = 0$  is called **hard-edge** and the local infinitesimal behaviour in a neighbourhood of  $x = 0$  (in the limit as  $n \rightarrow +\infty$ ) is described by a universal kernel called Bessel kernel

$$K_{\text{Bessel}}(x, y) = \frac{J_\alpha(\sqrt{x})\sqrt{y}J'_\alpha(\sqrt{y}) - J'_\alpha(\sqrt{x})\sqrt{x}J_\alpha(\sqrt{y})}{2(x - y)} \quad x, y \in \mathbb{R}_+, \quad \alpha > -1. \tag{47}$$

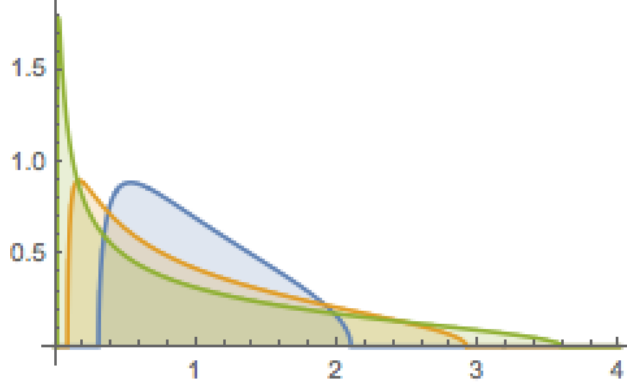


Figure 2: Marchenko-Pastur distribution for different values of  $\kappa$  (picture taken from WolframAlpha website).

**Note 18.** The function  $J_\alpha(z)$  is the Bessel function (of first kind). It satisfies the second-order ODE

$$z^2 y'' + z y' + (z^2 - \alpha^2) y = 0, \quad (48)$$

such that  $\lim_{z \rightarrow 0} y(z) < \infty$  for integer or positive  $\alpha$ . It admits a series expansion at  $x = 0$  of the form

$$J_\alpha(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \Gamma(n + \alpha + 1)} \left(\frac{z}{2}\right)^{2n+\alpha} \quad (49)$$

(alternatively, it admits a representation in terms of a contour integral).

**(Gaussian)  $\beta$ -ensembles.**

$$d\mu(x) = \prod_{i < j} |x_i - x_j|^\beta e^{-\frac{\beta}{4} \sum_i x_i^2} dx_i \dots dx_n \quad (50)$$

As seen before, for  $\beta = 1$ , the matrix ensemble is called Gaussian Orthogonal Ensemble and it is the ensemble of real symmetric matrices (its distribution is invariant under orthogonal conjugation). For  $\beta = 4$ , the ensemble is given by quaternionic Hermitian matrices (its distribution is invariant under conjugation by the symplectic group) and it is called Gaussian Symplectic Ensembles.

For general  $\beta > 1$ , it is possible to realize this distribution as the distribution of eigenvalues of certain random tri-diagonal matrices with independent entries (Dumitriu, Edelman [2]).

The adjective “Gaussian” refers to the fact that we’re still considering a quadratic potential  $V(x) = x^2$  in the definition of the probability measure.

**Multi-matrix models and external field.** The matrix models which we have considered so far could be called one-matrix models, as the corresponding integrals involved only one matrix. A natural generalization is to consider integrals over multiple matrices, and the corresponding multi-matrix models. For example, a two-matrix model can be defined from the ensemble

$$\mathcal{E} = \mathfrak{M} \times \mathfrak{M},$$

where  $\mathfrak{M}$  is the set of all  $n \times n$  Hermitian matrices, with measure (for example)

$$d\mu(M_1, M_2) = e^{-\text{Tr} [V_1(M_1) + V_2(M_2) - M_1 M_2]} dM_1 dM_2 \quad (51)$$

where  $V_1$  and  $V_2$  are two potentials. This can be generalized to the matrix chain on  $\otimes^k \mathfrak{M}$  with or without the so-called “external field”, a deterministic fixed matrix which breaks the invariance under conjugation of the original model: for example,

$$d\mu(M) = e^{\text{Tr}(M^2) - AM} dM. \quad (52)$$

## 6 Bonus: applications

We conclude these notes with a classical application of random matrices to Statistics/Data Analysis [4].

Consider  $X \in \mathbb{R}^p$  a random vector with covariance matrix

$$\Sigma = \mathbb{E} [(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

The goal of Principal Component Analysis is to learn a low dimensional representation of the vector  $X$  such that the residual variance is minimized. This is solved by finding a sequence of orthonormal eigenvectors  $\{v_k\}_{k=1, \dots, p}$  of  $\Sigma$ .

In practice we observe  $n$  i.i.d. realizations  $X^{(i)}$ ,  $i = 1, \dots, n$  and  $\Sigma$  is unknown. We estimate the  $v_k$ ’s by their empirical counterparts  $\hat{v}_k$  defined as a sequence of orthonormal eigenvectors of the empirical covariance matrix

$$S = \frac{1}{n} \sum_{i=1}^n (X^{(i)} - \bar{X}) (X^{(i)} - \bar{X})^T,$$

with eigenvalues  $\{\hat{\lambda}_k\}$ ; here  $\bar{X} \in \mathbb{R}^p$  is the mean of the  $X^{(i)}$ ’s over each entry/feature. Then,  $S$  can be viewed as a random matrix.

In general, one can test the existence of a one-dimensional signal (i.e. correlation between the features/entries of  $X$ ) over a Gaussian white noise. In the case where there is no correlation, we can reasonably argue that the  $n$  i.i.d. realizations of the random vector  $X \in \mathbb{R}^p$  are sampled from a multivariate distribution  $\mathcal{N}(\mu, \Sigma)$  and the corresponding empirical covariance matrix  $nS$  follows a Wishart distribution  $W_p(n-1, \Sigma)$ .

We test the null hypothesis  $H_0 : \{\Sigma = I_p\}$  (no correlations) against the alternative  $H_1 : \{\Sigma = \lambda \theta \theta^T + I_p\}$ , where  $\theta \in \mathbb{R}^p$ . Denote by  $\hat{\lambda}$  the largest eigenvalue of  $S$ . Under  $H_0$  the asymptotic distribution of  $\hat{\lambda}$  is given by the Tracy-Widom law  $TW_1$ . An asymptotic test of level  $1-\epsilon$ ,  $0 < \epsilon < 1$ , is to accept  $H_0$  if  $\hat{\lambda}$  is smaller or equal to the  $(1-\epsilon)$ -quantile of  $TW_1$ , and to accept it otherwise.

## References

- [1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2009.

- [2] I. Dumitriu and A. Edelman. Matrix models for  $\beta$ -ensembles. *J. Math. Phys.*, 43:5830–5847, 2008.
- [3] M. L. Mehta. *Random Matrices*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [4] D. Paul and A. Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 2013.
- [5] C. Tracy and H. Widom. Level spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.
- [6] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.*, 62:548–564, 1955.