Determinantal point processes and random matrix theory in a nutshell

- part II -

Manuela Girotti

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

Contents

| 1 | A little bit of history | | | | | | 1 | |
|---|-------------------------|--------|--------------------------------|---|--|--|---|---------------|
| 2 | | | Ensembles nvalues distribution | • | | | | 2 3 |
| 3 | Universality | | | | | | 5 | |
| | 3.1 | Macro | roscopic behaviour | | | | | 5 |
| | | 3.1.1 | Wigner's semicircle law | | | | | 7 |
| | 3.2 | Micro | roscopic behaviour | | | | | 8 |
| | | 3.2.1 | Bulk universality | | | | | 8 |
| | | 3.2.2 | Soft-edge universality | | | | | 9 |
| 4 | Oth | er ran | ndom matrix models | | | | | 10 |

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 $M \in \mathcal{E}$, called the random matrix law.

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.

Some relevant references about Random Matrices are the book by Mehta [3] and the book by Anderson, Guionnet and Zeitouni [1].

1 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:

$$\mathcal{H}v = \lambda v \tag{1}$$

where v is the eigenfunction associated to the eigenvalue λ .

Wigner argued that one should regard a specific Hamiltonian \mathcal{H} as behaving like a largedimension 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 ([5]). 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 $n \to +\infty$.

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 ([4]), 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).

2 Unitary Ensembles

Consider the space of $n \times n$ complex Hermitian matrices

$$\mathfrak{H}_n = \{ M \in \operatorname{Mat}_n(\mathbb{C}) \mid M = M^{\dagger} \}.$$
 (2)

This is a n^2 -dimensional 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. We now want to equip this space with a probability measure.

Definition 1. A **Unitary Ensemble** is the probability measure

$$d\mu_n(M) = \frac{1}{Z_n} e^{-\operatorname{Tr} V(M)} dM \tag{3}$$

on the space \mathfrak{H}_n , where $V: \mathbb{R} \to \mathbb{R}$ is a given function, called the potential, with suitable growth condition at $\pm \infty$ to guarantee that the probability measure above is well defined and dM represents the flat Lebesgue measure on \mathfrak{H}_n , i.e.

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

$$(4)$$

Remark 2. A sufficient (and standard) condition for the probability (3) to be well-defined is that

$$\lim_{|x| \to +\infty} \frac{V(x)}{\ln(1+x^2)} = +\infty,\tag{5}$$

which is certainly satisfied if V is a polynomial of even degree, with positive leading coefficient.

The name "Unitary Ensemble" comes from the fact that the probability distribution (3) is invariant under conjugation with a unitary matrix

$$M \mapsto UMU^{-1}, \quad U \in U(n).$$
 (6)

Paradigma. The Gaussian Unitary Ensemble (GUE) is the ensemble on Hermitian matrices \mathfrak{H}_n equipped with the probability measure

$$d\mu(M) = \frac{1}{Z_n} e^{-\frac{1}{2} \operatorname{Tr} M^2} dM.$$
 (7)

Since

$$\operatorname{Tr} M^{2} = \sum_{i=1}^{n} (M^{2})_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} M_{ij} M_{ji} = \sum_{i=1}^{n} M_{ii}^{2} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |M_{ij}|^{2} =$$

$$= \sum_{i=1}^{n} M_{ii}^{2} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left[(\Re M_{ij})^{2} + (\Im M_{ij})^{2} \right], \tag{8}$$

the probability measure (9) factorizes as a product of Gaussians

$$d\mu(M) = \frac{1}{Z_n} \prod_{i=1}^n e^{-\frac{1}{2}M_{ii}^2} dM_{ii} \prod_{i=1}^{n-1} \prod_{j=i+1}^n \left(e^{-(\Re M_{ij})^2} d\Re M_{ij} \right) \left(e^{-(\Im M_{ij})^2} d\Im M_{ij} \right). \tag{9}$$

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:

$$\Re M_{ij}, \ \Im M_{ij} \sim \mathcal{N}\left(0, \frac{1}{2}\right) \qquad M_{ii} \sim \mathcal{N}\left(0, 1\right).$$
 (10)

2.1 Eigenvalues distribution

In random matrix theory, one is interested in the distribution of the eigenvalues of the (random) matrix M. If M belongs to a unitary ensemble, the eigenvalues will be real random variables.

According to the spectral theorem, any Hermitian matrix M can be written as $M = U\Lambda U^{-1}$, where $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ is the matrix of eigenvalues and $U \in U(n)$ is the matrix of corresponding eigenvectors. Therefore, we can perform a strategic change of variables

$$M \mapsto (\Lambda, U)$$

$$\{M_{ii}, i = 1, \dots, n; \Re M_{ij}, \Im M_{ij}, i < j\} \mapsto \{\lambda_1, \dots, \lambda_n; u_{ij}\},$$
(11)

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

$$dM = c_n \, \Delta(\lambda)^2 \, d\lambda_1 \dots d\lambda_n \, dU \tag{12}$$

where

$$c_n = \frac{\pi^{n(n-1/2)}}{\prod_{j=1}^n j!}, \qquad \Delta(\lambda) = \prod_{i < j} (\lambda_i - \lambda_j)$$
(13)

and dU is the Haar measure on U(n).

Furthermore, since $\operatorname{Tr} V(M) = \sum_{j} V(\lambda_{j})$ depends only on the eigenvalues of M, we can conclude that the probability measure on the space of matrices (3) induces a joint probability density on the eigenvalues given by

$$d\mu(\lambda_1, \dots, \lambda_n) = \frac{1}{\widetilde{Z}_n} \Delta(\lambda)^2 \prod_{j=1}^n e^{-V(\lambda_j)} d\lambda_1 \dots d\lambda_n$$
(14)

with \widetilde{Z}_n a normalization constant. By writing the Vandermonde determinant as a determinant and by making some column ad row transformations we obtain that

$$\frac{1}{\widetilde{Z}_n} \Delta(\lambda)^2 \prod_{j=1}^n e^{-V(\lambda_j)} = \frac{1}{n!} \det \left[K_n(x_i, x_j) \right]_{i,j=1}^n$$
 (15)

with

$$K_n(x,y) = e^{-\frac{V(x)+V(y)}{2}} \sum_{j=0}^{n-1} \phi_j(x)\phi_j(y),$$
(16)

 $\{\phi_i\}$ being the set of orthonormal polynomials with respect to the weight $w(x) = e^{-V(x)}$.

Therefore, the set of random eigenvalues of a (unitary) matrix ensemble is a determinantal point process.

Paradigma (part II). In the case of GUE matrices, the polynomials $\phi_j(x)$ are the Hermite polynomials.

Furthermore, the following result holds: if we consider n non-intersecting Brownian paths $X_1(t), \ldots, X_n(t)$, all starting at x = 0 and finishing at x = 0 after time T > 0. Their transition probability is a Gaussian

$$p_t(x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$$
(17)

and thanks to Karlin-McGregor theorem, we have that the joint probability distribution of the paths at any time 0 < t < T is equal to

$$\frac{1}{Z_n} \det \left[F_{j-1}(x_i) e^{-\frac{(x_i - a)^2}{2t}} \right]_{i,j=1}^n \det \left[G_{j-1}(x_i) e^{-\frac{(x_j - b)^2}{2(T-t)}} \right]_{i,j=1}^n$$
(18)

 $(F_{j-1} \text{ and } G_{j-1} \text{ are some polynomials of degree } j-1)$. In conclusion, it is easy to see that the positions of the paths are distributed as an ensemble of GUE-eigenvalues (see Figure 1).

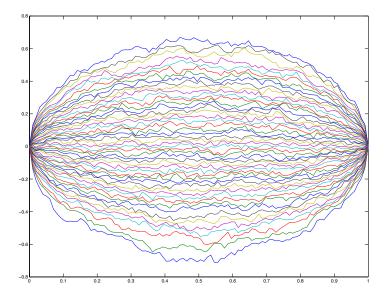


Figure 1: Numerical simulation of 50 non-intersecting Brownian paths in the confluent case with one starting and one ending point.

3 Universality

A property is "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).

Universality will arise in limits where the matrix size n is large, i.e. when the number of eigenvalues grows.

3.1 Macroscopic behaviour

We are interested in the distribution of the eigenvalues as the dimension of the matrix grows. Consider a matrix ensemble and denote the (ordered) eigenvalues by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The empirical spectral distribution of the eigenvalues is defined by

$$d\mu_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}(x) dx$$
 (19)

where δ_{λ} is the Dirac delta function centered at λ .

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 hystograms of the eigenvalues: see Figure 2.

Indeed, in many cases of interest the eigenvalue density has a finite limit as $n \to +\infty$, called the equilibrium density $\rho(x)$.

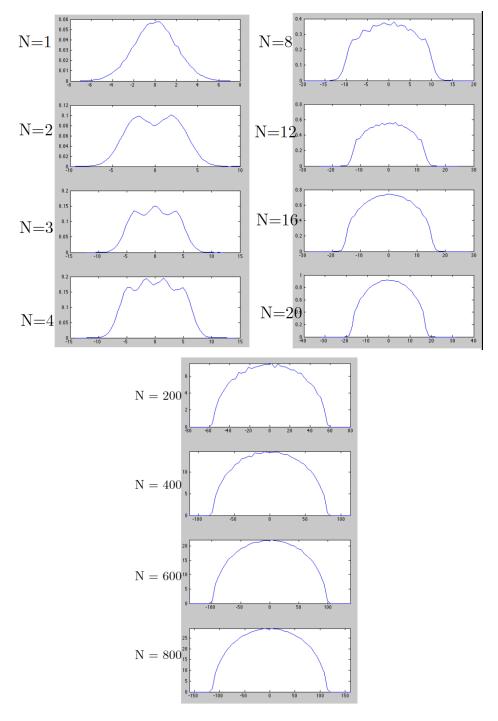


Figure 2: Histograms of the eigenvalues of GUE matrices as the size of the matrix increases (such histograms are produced with rescaling the spectrum by a factor $1/\sqrt{n}$). Numerical simulation with MATLAB (courtesy of prof. Ken McLaughlin).

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

$$\rho_1(x) = K_n(x, x). \tag{20}$$

Therefore, the equilibrium density (if it exists) can be computed as

$$\lim_{n \to +\infty} \frac{1}{n} K_n(x, x) = \bar{\rho}(x). \tag{21}$$

3.1.1 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

$$\widetilde{M} = \frac{1}{\sqrt{n}}M,\tag{22}$$

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

Theorem 3 (Wigner's semicircle law). Consider the GUE ensemble of size n with matrices $\frac{1}{\sqrt{n}}M$, then the spectrum distribution converges as $n \to +\infty$, in probability, to the following deterministic probability density

$$\bar{\rho}(x) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2} & \text{if } |x| \le 2\\ 0 & \text{otherwise.} \end{cases}$$
 (23)

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. \Box

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, ..., n. Such type of matrices are generally called Wigner matrices.

Such feature is one aspect of the **universality** theory we talked about in the introduction of this section.

Remark 4. 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.

3.2 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 $n \to +\infty$.

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**). In order to study their statistical behaviour we will make use of the results seen in the DPP theory.

We recall the following result bout transformations of DPPs.

Proposition 5. Let \mathbb{P} and \mathbb{P}_n be determinantal point processes with kernels K and K_n respectively. Let K_n converge to K

$$\lim_{n \to \infty} K_n(x, y) = K(x, y) \tag{24}$$

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

$$x \mapsto Cn^{\gamma}(x - x^*) \tag{25}$$

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:

$$\lim_{n \to \infty} \frac{1}{Cn^{\gamma}} K_n \left(x^* + \frac{x}{Cn^{\gamma}}, x^* + \frac{y}{Cn^{\gamma}} \right) = K(x, y)$$
 (26)

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

3.2.1 Bulk universality

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

In the GUE case, the kernel is defined in terms of Hermite polynomials

$$K_n(x,y) = e^{-\frac{V(x)+V(y)}{2}} \sum_{j=0}^{n-1} \phi_j(x)\phi_j(y),$$
(27)

therefore, calculating the limit of K_n (properly translated and rescaled) is equivalent of studying the asymptotic behaviour of the polynomials as $n \to +\infty$. The standard way to achieve such results is via Classical (Laplace) and Non-linear (Deift-Zhou) Steepest Descent Methods.

Proposition 6 (Bulk universality at the origin). For GUE, the local behaviour in the bulk of the spectrum is described by a DPP with correlation kernel given by

$$\lim_{n \to +\infty} \frac{\pi}{\sqrt{2n}} K_n \left(\frac{\pi x}{\sqrt{2n}}, \frac{\pi y}{\sqrt{2n}} \right) = K_{\text{bulk}}(x, y)$$
 (28)

with

$$K_{\text{bulk}}(x,y) = K_{\text{sine}}(x,y) = \frac{\sin \pi (x-y)}{\pi (x-y)}.$$
 (29)

3.2.2 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: $\bar{\rho}(x) \sim \sqrt{x-a}$. On the other hand for special choices of the potential, we can have that the vanishing of the density has a different regime $\bar{\rho}(x) \sim (x-a)^{\frac{p}{q}}$ for some positive integers p,q; in this case, the edges are called **critical**.

In the case of GUE, both edges (on $a = \pm 2\sqrt{n}$) are regular and the microscopic limit is described as

Proposition 7 (Soft-edge universality on the right endpoint).

$$\lim_{n \to +\infty} \frac{1}{2n^{\frac{1}{6}}} K_n \left(2\sqrt{n} + \frac{x}{2n^{\frac{1}{6}}}, 2\sqrt{n} + \frac{y}{2n^{\frac{1}{6}}} \right) = K_{\text{soft}}(x, y)$$
 (30)

with

$$K_{\text{soft}}(x,y) = K_{\text{Airy}}(x,y) = \frac{\text{Ai}(x)\text{Ai}'(y) - \text{Ai}'(x)\text{Ai}(y)}{x - y}.$$
(31)

Note 8. The function Ai(t) is the Airy function. It satisfies the second-order ODE

$$y'' = ty$$
, such that $\lim_{t \to +\infty} y(t) = 0$. (32)

It can be represented as a contour integral

$$\operatorname{Ai}(t) = \int_{\gamma} e^{\frac{u^3}{3} - tu} \frac{\mathrm{d}u}{2\pi i} \tag{33}$$

where the curve $\gamma \subseteq \mathbb{C}$ is an oriented contour starting at the point at ∞ with argument $-\frac{\pi}{3}$ and ending at the point at ∞ 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 in terms of the τ -function of the Painlevé II integrable system.

Theorem 9 (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

$$\det\left(\operatorname{Id}_{L^{2}(\mathbb{R})} - K_{\operatorname{Ai}}\Big|_{[s,+\infty)}\right) = \exp\left\{-\int_{s}^{\infty} (x-s)q^{2}(x)\mathrm{d}x\right\}$$
(34)

where q(x) is the Hasting-Mc Leod solution to the Painlevé II equation:

$$q''(x) = 2q^{3}(x) + sq(x)$$

$$q(x) \sim \operatorname{Ai}(x) \qquad x \to +\infty.$$
(35)

4 Other random matrix models

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

The corresponding eigenvalue distribution has the following expression

$$d\mu(\lambda) = \prod_{i < j} |\lambda_i - \lambda_j| \prod_i \lambda_i^{\alpha} \exp\left\{\sum_i \frac{\lambda_i}{2}\right\} d\lambda_1 \dots d\lambda_n.$$
 (36)

If we properly rescale the ensemble by

$$\widetilde{M} = \frac{1}{n}M\tag{37}$$

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

$$a_{-} = (1 - \sqrt{\kappa})^{2}, \ a_{+} = (1 + \sqrt{\kappa})^{2}$$
 (38)

and it is equal to

$$\bar{\rho}_{\kappa}(x) = \frac{1}{2\pi\kappa} \frac{\sqrt{(a_{+} - x)(x - a_{-})}}{x}.$$
(39)

This distribution is called Marchenko-Pastur law (see Figure 3).

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

$$\bar{\rho}_1(x) = \frac{1}{2\pi} \sqrt{\frac{4-x}{x}} \qquad x \in (0,4].$$
(40)

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 $n \to +\infty$) is described by a universal kernel called Bessel kernel

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

$$(41)$$

(Gaussian) β -ensembles.

$$d\mu(\lambda) = \prod_{i < j} |\lambda_i - \lambda_j|^{\beta} \exp\left\{-\frac{\beta}{4} \sum_i \lambda_i^2\right\} d\lambda_1 \dots d\lambda_N$$
 (42)

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.

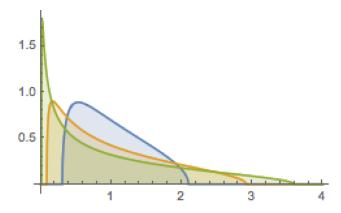


Figure 3: Marchenko-Pastur distribution for different values of κ (picture taken from WolframAlpha website).

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 $\mathcal{E} = H_N \times H_N$ with measure

$$d\mu(M_1, M_2) = e^{-\operatorname{Tr}(V_1(M_1) + V_2(M_2) - M_1 M_2)} dM_1 dM_2$$
(43)

where V_1 and V_2 are two potentials. This can be generalized to the matrix chain on $(H_N)^k$ with or without the so-called "external field", a deterministic fixed matrix which breaks the invariance under conjugation of the original model

$$d\mu = e^{\text{Tr}(M^2) - AM} dM. \tag{44}$$

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 β -ensembles. J. Math. Phys., 43:5830–5847, 2008.
- [3] M. L. Mehta. Random Matrices. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [4] C. Tracy and H. Widom. Level spacing distributions and the Airy kernel. *Comm. Math. Phys.*, 159(1):151–174, 1994.
- [5] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.*, 62:548–564, 1955.