Preliminary version, June 10, 2020

Random Matrix Models: an introduction (Lecture notes, April-May 2017)

Filippo Colomo

1. Introduction

1.1. Prerequisites.

- Path-Integral formulation of Quantum Field Theory and Statistical Mechanics
- Saddle-Point Method
- Basics on Orthogonal Polynomials

1.2. Bibliography.

- There are many good textbooks on Random Matrix Models. My preferred one is Mehta's seminal book, in his first version [1]. Relatively short and completely self-contained, it contains all fundamentals. The third edition of the book [2] contains much more materials, and many updates to keep track of some recent developments, but beginners can easily get lost.
- A nice and complete collection of recent reviews on various aspects and applications of Random Matrices is *The Oxford Handbook of Random Matrix Theory* [3].
- The mathematically-oriented reader could browse also [4] and [5].
- The present lectures refers mainly to the most recent edition of Mehta's book, with some occasional change of notations (especially of normalizations) when more convenient for our purposes. These lectures are to be complemented with an accurate reading of the nice reminiscences talk by Dyson [6], and of seminal papers [7–10].

1.3. Why Random Matrices? Random Matrices (RM) were originally introduced by Wishart in 1928 in the domain of multivariate statistics, and more specifically in the study of covariance matrices. They were independently reintroduced in physics by Wigner in 1951, to describe the level statistics of nuclear spectra - the frontier of physics at the time. In the course of the last sixty years, they have found applications in the most disparate sectors of physics - ranging from quantum chaos to mesoscopic systems, from Quantum ChromoDynamics to 2D quantum gravity, from 2D statistical mechanics on random lattices to string theory - and mathematics - from the distributions of zeroes of Riemann ζ-function to the enumeration of graphs and knots, from integrable hierarchies of partial differential equations to growth and tiling problems. This ubiquity of random matrices

We are grateful to Claudio Bonati, Paolo Braccia and Laura Gentini for several relevant comments and corrections.

is to be ascribed to the fact that some of their features are **universal**, in the sense that they do not depend on the details of the particular random matrix model considered, but only on the symmetry class of the model.

1.4. Wigner proposal. Consider an heavy nucleus, with many protons and neutrons in strong interaction. The standard models of nuclear physics used in the fifties gave reasonable results for the ground state and the first few excited levels. Actually, the question in this case reduces to the choice of a few good quantum numbers, on the basis of the known symmetries of the model. However, in addressing for instance the resonances observed in the capture of slow neutrons by heavy nuclei, one is concerned with highly excited levels (with principal quantum number $n \approx 10^5$).

The traditional approach to find these highly excited states would be that of devising, on the basis of symmetries and of experimental scattering data, some suitable potential for the nucleons, and hence some Hamiltonian. In principle, from the diagonalization of such Hamiltonian, one can extract all the desired knownledge on the corresponding physical system. However, we here face two major difficulties: first, we are not able to write down the actual nuclear potential, and second, even if we were, it would be far too complicated to solve the corresponding Schrodinger equation, even in the crudest approximations.

In view of this situation, Wigner proposed to give up any pretension of exact knowledge of the spectrum of such complicate systems, and rather to address only statistical properties such as, e.g., the statistical distribution of spacings between levels. In a nutshell: let us assume we are unable to access the explicit form of the nuclear Hamiltonian, and thus let us choose it at random in a given statistical ensemble (dictated by the symmetries of the problem, such as time-reversal or rotational invariance), and study the statistical properties of its spectrum. We could view this as 'statistical mechanics of Hamiltonians'. Not only we renounce the knowledge of the state of the system, but of the system itself! (exception made for its symmetries).

1.5. The Gaussian Orthogonal Ensemble. Following Wigner, let us consider a generic Hamitonian H, with time-reversal and rotational invariance. It can be shown that in this case the Hamiltonian is not only Hermitean, but real-symmetric. In principle nuclear spectra are unbounded and with infinitely many levels, but if we are interested only in a finite portion of the discrete spectrum, considering an Hamiltonian over a finite Hilbert space is a reasonable assumption which simplifies significantly the problem. We shall thus define an ensemble of $N \times N$ real-symmetric matrices (and eventually, investigate the large N limit). The fact that H is real-symmetric fixes the group of invariance G of our model, $G = O(N)$, that is the group of $N \times N$ orthogonal matrices.

We are interested into the statistical properties of H, in particular of its spectrum. We thus introduce a partition function, in the sense of statistical mechanics of Hamiltonians belonging to some ensemble,

$$Z_N = \int \mathcal{D}H\mathcal{P}(H).$$

Here the integration is over all real-symmetric matrices. These have $N(N+1)/2$ independent matrix elements, that we choose as $H_{jj} = a_j$, $j = 1, \dots, N$, and

$H_{jk} = b_{jk}$, $j < k = 1, \ldots, N$. The integration measure over real-symmetric matrices is then

$$\mathcal{D}H = \prod_{j=1}^{N} \mathrm{d}a_j \prod_{\substack{j,k=1 \\ j<k}}^{N} \mathrm{d}b_{jk}$$

Concerning the probability measure $\mathcal{P}(H)$, as physicists, the most natural choice is the Gaussian distribution. To make the matrix as random as possible we further assume that the independent matrix elements are independently distributed, each one with Gaussian probability density:

$$\mathcal{P}(a_j) = \mathrm{e}^{-\frac{1}{2}a_j^2}, \qquad \mathcal{P}(b_{jk}) = \mathrm{e}^{-b_{jk}^2}, \tag{1.1}$$

orthog_integ_meas

i.e., $\mathcal{P}(H)$ is factorized in terms of the probability densities of the independent elements of H. Note that the variances appearing in (1.1) differ by a factor 2, which is necessary in order to have a $O(N)$ invariant probability density. Indeed, we can equivalently write

$$\mathcal{P}(H) = \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}[H^2]}, \tag{1.2}$$

proba_density

that is clearly invariant under the $O(N)$ symmetry group, due to the cyclic invariance of the trace. Thus, by construction, the whole integrand in Z_N is $O(N)$ invariant and defines the so-called Gaussian Orthogonal Ensemble (GOE). As we shall see later on, other ensembles can be considered as well. These differ in the invariance group G, here chosen as $O(N)$, and in the choice of the G-invariant probability density, which in the present case is specified by equation (1.2).

1.6. From matrix elements to eigenvalues. Our aim now is to give a more explicit form to the integrand of Z_N, and extract informations on the statistics of the eigenvalues of this ensemble of real-symmetric random matrices.

It will be shown in the course of these lectures that the matrix integration element can be factorized into a 'gauge part', corresponding to the invariance group, and a 'physical part', corresponding to the eigenvalues. In particular, in the present case, this can be expressed as follows: since any real-symmetric matrix H can be diagonalized by means of an orthogonal transformation, we can always write $H = O^t \Lambda O$, where O is a suitable orthogonal matrix and Λ is the diagonal matrix built from the eigenvalues of H, namely $\Lambda_{jk} = \lambda_j \delta_{jk}$. Correspondingly, the integration measure may be rewritten as

$$\begin{aligned} \mathcal{D}H &= \mathcal{D}O\, \mathcal{D}\Lambda\, |J| \\ &= \mathcal{D}O \prod_{i=j}^{N} \mathrm{d}\lambda_j \prod_{\substack{j,k=1 \\ j<k}}^{N} |\lambda_k - \lambda_j| \end{aligned} \tag{1.3}$$

jacobian

where the Jacobian J associated with the change of variable is simply given by the double product in (1.3). This will be explained in detail in Section 3

Due to the invariance of the integrand in Z_N under orthogonal transformations, it is clear that the integration over the unitary group factors out, giving simply an overall constant, and we obtain:

$$Z_N \propto \int_{-\infty}^{+\infty} \mathrm{d}^N\lambda \prod_{\substack{j,k=1 \\ j<k}}^{N} |\lambda_k - \lambda_j| \mathrm{e}^{-\frac{1}{2}\sum_{j=1}^{N} \lambda_j^2}.$$

The integrand in Z_N is the joint probability density for the eigenvalues of our random matrix. The main problem is know to evaluate the above integral for large N, and to extract from it information on the statistics of the eigenvalues of our model of random matrices. This will be the main object of these lectures.

1.7. Wigner surmise. Before going ahead, let us consider for a moment the case $N = 2$. This is very far from the physically interesting case of large N. However, it is a very simple case, where we shall be able to work out some exact results with elementary techniques. Moreover, and surprisingly enough, the obtained results will be physically sound and in very good agreement with nuclear data.

In the case $N = 2$ the evaluation of the Jacobian coming from the change of variables from real-symmetric to diagonal matrices is straightforward. Indeed, if we parameterize the real-symmetric, orthogonal, and diagonal matrices as

$$H = \begin{pmatrix} a_1 & b \\ b & a_2 \end{pmatrix}, \qquad O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix},$$

respectively, the relation $H = O^t \Lambda O$ becomes

$$\begin{pmatrix} a_1 & b \\ b & a_2 \end{pmatrix} = \begin{pmatrix} \lambda\cos^2\theta + \mu\sin^2\theta & (\mu-\lambda)\sin\theta\cos\theta \\ (\mu-\lambda)\sin\theta\cos\theta & \lambda\sin^2\theta + \mu\cos^2\theta \end{pmatrix},$$

and it is readily seen that

$$\mathcal{D}H = \mathrm{d}a_1\,\mathrm{d}a_2\,\mathrm{d}b = |\mu-\lambda|\mathrm{d}\lambda\,\mathrm{d}\mu\,\mathrm{d}\theta.$$

We thus have the following probability density for the eigenvalues

$$\rho(\lambda,\mu) \propto |\mu-\lambda|\mathrm{e}^{-\frac{1}{2}(\lambda^2+\mu^2)},$$

modulo a suitable mormalization factor.

Let us now compute the probability density for the spacing s between two consecutive levels, assuming we have normalized to 1 the average spacing (this procedure is referred to as *unfolding*, and will be explained further in the course of these lectures). In the present case we have just two levels; assuming $\mu > \lambda$, we simply have $s \propto (\mu - \lambda)$, where the unknown proportionality constant will be fixed below. With the change of variables $(\lambda,\mu) \to (s,t) \propto (\mu-\lambda, \mu+\lambda)$, we readily get the probability density

$$P(s,t) \propto s\,\mathrm{e}^{-c_1(s^2+t^2)}, \qquad\qquad s > 0$$

and thus

$$P(s) = \int_{-\infty}^{+\infty} P(s,t)\,\mathrm{d}t = c_2 s\mathrm{e}^{-c_1 s^2},$$

where the two constants are fixed by the two conditions

$$\int_0^{+\infty} P(s)\,\mathrm{d}s = 1, \qquad\qquad \int_0^{+\infty} sP(s)\,\mathrm{d}s = 1,$$

coming from normalization and unfolding, respectively, with the final result:

$$P(s) = \frac{\pi}{2}\,s\,\mathrm{e}^{-\frac{\pi}{4}s^2}.$$

This expression is known as Wigner surmise.

Wigner's result is in extremely good agreement with the experimental data for the level spacing of nuclear spectra. More precisely, if we collect the sequence of

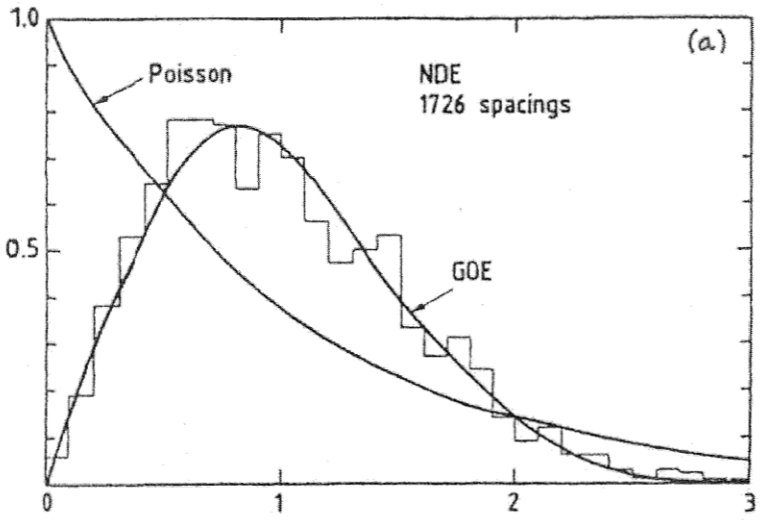


FIGURE 1. Level spacing histogram for a large set of nuclear levels (1756 resonance levels belonging to 30 sequences of 27 different nuclei). The data chosen in each sequence is believed to be complete (no missing levels) and pure (same angular momentum and parity). The solid curves correspond to the Poisson distribution (no correlation at all) and to Wigner surmise for the Gaussian Orthogonal Ensemble, respectively. (From O. Bohigas, R.U. Haq and A. Pandey, in *Nuclear data for Science and Technology* (1983) 809, ed. K.H. Böckhoff, see also [2].)

nuclear

levels with the same spin and parity for a given nucleus with time-reversal and rotational invariance, compute the corresponding level spacings, repeat this operation for many different nuclei with the same symmetries, and finally plot an histogram of all the collected level spacings, its shape is very well described by Wigner surmise, see Fig. 1 for an example.

Note that the correct large N evaluation of the spacing distribution worked out by Gaudin in 1961, gives of course a different expression. However, Wigner surmise is very close to it, differing from it at most at 1% in the most relevant interval $0 \lesssim s \lesssim 2$.

The idea of Wigner that the $N = 2$ case should capture the level spacing distribution of nuclear spectra was based on two assumptions, that in modern language we would phrase as follows: i) the two-point correlation function is linear in the eigenvalue separation; ii) higher order correlation functions are negligeably small. As we shall see in the course of these lectures, both arguments are innacurate. It is surprising that the two errors compensate so nearly each other.

Note that the evaluation of the large N limit exact result is quite involved, and came ten years after Wigner surmise. One could thus speculate that if this almost exact compensation between two unaccurate assumptions had not occured, Wigner surmise would have not been in such good agreement with experimental data, and the story of Random Matrix Models could have been quite different.

1.8. Level spacing in a random spectrum. The approach proposed by Wigner can be summarized as considering a statistical ensemble of Hamiltonian, and then extracting the resulting statistical features for their eigenvalues. But one could wonder, what about considering from the very beginning a random spectrum, rather than the spectrum following from a random Hamiltonian? What would be the resulting level spacing distribution?

Let us suppose we have a spectrum made of a large number N of energy levels that are independent identically distributed random variables, with constant distribution p over the interval $(0, N-1)$ (i.e., the average level spacing is 1 by construction). Note that normalization of probability fixes $p = 1/N$. The probability of having k levels in a given interval ΔE is given by the binomial distribution: $\binom{N}{k}(p\Delta E)^k(1-p\Delta E)^{N-k}$. Furthermore, we are interested into the case of large N and small ΔE. In this limiting situation the probability of having k levels in a given interval $(E, E+\mathrm{d}E)$ is proportional to $(\mathrm{d}E)^k/k!$. Normalizing, we actually recognize the Poisson distribution, $(\mathrm{d}E)^k\mathrm{e}^{-\mathrm{d}E}/k!$.

The probability that the spacing between two given consecutive levels lies in the interval $(s, s+\mathrm{d}s)$ can now be rephrased as follows: given the occurence of an energy level at E, what is the probability of having no level in the interval $(E, E+s)$ and one level in the interval $(E+s, E+s+\mathrm{d}s)$? Let us split the interval $(E, E+s)$ into a large number l of small parts $\mathrm{d}E$. Because the level are independent, the probability of having no level in the interval $(E, E+s)$ is the product of the probabilities of having no level in any of the first l intervals $\mathrm{d}E$; multiplying by the probability of having exactly one level in the interval $\mathrm{d}s$, we get:

$$p(s)\mathrm{d}s = \left[\mathrm{e}^{-\mathrm{d}E}\right]^l \left[(\mathrm{d}s)\mathrm{e}^{-\mathrm{d}s}\right] \simeq \mathrm{e}^{-s}\mathrm{d}s.$$

Summarizing, in a random spectrum the level spacing distribution is exponential, and maximum at $s = 0$.

1.9. Level repulsion. According to Wigner surmise, the level spacing distribution vanishes at $s = 0$, i.e., the levels tend to stay apart, to repel each other, preventing their spacing to vanish. This is a completely different behaviour with respect to the one of a truely random spectrum, discussed in the previous subsection, where the probability of coinciding eigenvalues is not only different form zero, but in fact maximal.

We anticipate here that the level repulsion observed in Wigner surmise is in fact typical of all random matrix model spectra. Thus, the spectrum of a random matrix is actually far from being random, and we shall see in the course of these lectures that in fact it exhibits very strong and typical correlations.

1.10. Some applications of Random Matrices. We have seen above a very beautiful application of Random Matrix Models to the description of essential features of nuclear spectra. Let us recall briefly some other applications in physics:

- **Quantum chaos.** Consider the quantum version of a classically chaotic system with few degrees of freedom (well known examples are: quantum billiards, i.e. systems made of a particle free to move inside some not too regular two-dimensional domain; Rydberg levels of Hydrogen atoms in a strong magnetic field, two coupled quartic oscillators with potential $x^4 + y^4 + \lambda^2 y^2$, ...). For still unclear reasons, the level spacing distribution and

of the energy spectrum of such system is described by Random Matrices with extraordinary good accuracy.

- **Mesoscopic systems I.** Electrons in metallic nanograins (i.e. metallic grains made of 10^3 atoms) at low temperatures can be modelled as free particles confined to a three-dimensional region. Since there is in fact no control on the shape of each nanograin, the energy levels look random. Their statistical properties are very well described by Random Matrix Models.
- **Mesoscopic systems II.** Consider the measurement of the conductance in mesoscopic metallic samples. The measured electronical conductance varies from sample to sample, mainly due to inhomogeneous scattering sites (impurities). However, the fluctuations of the conductance measured over a large number of samples appear to be universal, $\Delta G = G_0 = 2e^2/h$. This remarkable effect can be interpreted in terms of Random Matrix Models, where now the random matrix correspond to the unknown Scattering matrix S, rather than to the unkown Hamiltonian, of a given matellic sample. These considerations applies more generally to the statistical properties of transmission through a random medium.
- **Glassy systems.**
- **'t Hooft large N limit of $U(N)$ gauge theories.**
- **Discretized surfaces and two-dimensional quantum gravity.**
- **Discretized surfaces and string theory.**
- **Chiral effective actions and the spectrum of Dirac operator.**
- `[...]`

in mathematics:

- **Number theory:** the zeroes of Riemann ζ-function, see Fig. 2.
- **Combinatorics:** Enumeration of particular types of maps, foldings, colourings, knots, links, ...
- **Random Permutations**
- **Random matrices and random processes:** random growth models, tilings, partitions, spatial phase separation, ...
- **Integrable ierarchies** of Partial Differential Equations.
- `[...]`

and also in

- **Biophysics:** RNA folding, ...
- **Econophysics:** Return statistics and portfolio theory, ...
- **Complex Networks.**
- `[...]`

and even in everyday life:

- **Spatial distribution of trees in Scandinavian forests.**
- **Time distribution of buses in Cuernavaca bus system.**
- **Distance between parked cars in the street.**
- `[...]`

We refer to [3] for further examples and details on the many applications of Random Matrices.

1.11. Universality. This list, while still far from exhaustive, ranges from applied physics and complex systems to string theory, from number theory to algebraic

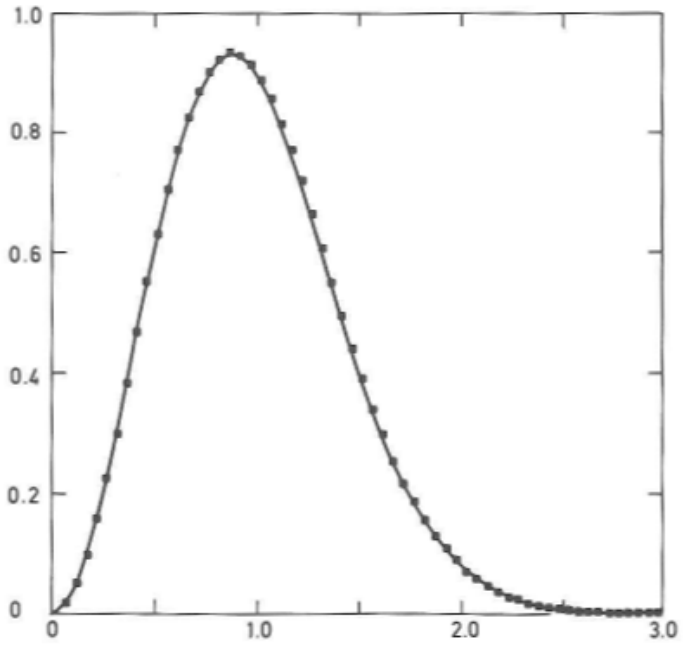


FIGURE 2. Plot of the density of normalized spacings for the 79 millions zeroes around the 10^{20}th zero of Riemann Zeta function. The solid curve is the spacing probability for the Gaussian Unitary Ensemble. (From Odlyzko (1989), see also [2].)

riemann

geometry. Furthermore, when we claim that some system is well described by a given random matrix model, we are not referring only to the fact that they have identical spacing distributions: we are making a much stronger statement, namely that of the agreement between any *local* statistical quantity (essentially, any correlation function) we are able to evaluate both in the physical system and in the corresponding random matrix model.

The reason at the basis of this wide range of application of random matrix models is not completely understood yet. Note however that despite the infinite variety of Random Matrix Models one could define, just by varying the form of the G-invariant probability density $\mathcal{P}(H)$, we have in fact a very small number of *universality classes*, i.e., of possible behaviours of local statistical quantities, depending essentially only on the choice of the invariance group G.

Recall the 'law of large numbers': the average of a large number of independent random variables (each with arbitrary distribution, the only requirement beeing the one of finite variance) tends to a random variable with Gaussian distribution. In the case of random matrix, we have a similar situation: once the invariance group G is fixed, the possible behaviour of correlation functions in the large N limit is determined (almost) independently of the probability density $\mathcal{P}(H)$. Thus we can speculate that a kind of 'law of large numbers' is at work.

Within such a framework, if the association of a random Hamiltonian (within a given symmetry class) with a given complicate physical system is justified, and if

this association can be done for a great number of other physical systems (with the same symmetry), then it is no surprise that the eigenvalues of all such systems have the same statistical behaviour. Why such a behaviour should be observed also, say, for the zeroes of the Riemann ζ-function, is however still completely unclear.

In the course of these lectures we shall describe some of the most relevant universality classes of Random Matrices, and characterize the corresponding statistical behaviour of their eigenvalues. Note that in principle one could investigate the universal properties of eigenvectors as well (indeed this has been done, especially in the study of quantum chaotic systems, where the eigenvectors correspond to the wave functions of the system), but these will not be considered here.

2. Probability density for matrix elements

2.1. Preliminaries. We shall here define the most common and relevant ensembles or Random Matrices. In general these are not necessarily related to any Hamiltonian, so we shall in general denote a Random Matrix by M. A given ensemble of matrices M is characterized by an invariance group G, in turn determining the integration measure $\mathcal{D}M$ and by a probability law $\mathcal{P}(M)$. We shall now discuss these three aspect in turn.

In the simplest and most obvious interpretation Random Matrices are to be thought as Hamiltonians for some complicate physical system, that is as Hermitean matrices. Our physical system can have some fundamental symmetry, and correspondingly, some integral of motion, such as total spin, or parity, and some exact quantum number. If the Hilbert space basis functions are chosen to be eigenfunctions of these conserved quantities, all the Hamiltonian matrices of the ensemble will be block diagonal, each block corresponding to a set of definite values for these quantum numbers.

Hereafter we shall assume that such a basis has already been chosen, and restrict our attention to one of these blocks. We assume for the present time that the considered block is an Hermiten matrix of size $N \times N$, with N a large but fixed positive integer. We shall eventually investigate the behaviour of our Random Matrices in the large N limit.

2.2. Time-reversal invariance. As it will become apparent below, time reversal plays a fundamental role in determining the invariance group G. It is well known that the time-reversal transformation on a quantum system is realized in terms of some anti-unitary[1] operator T. Because operating twice with T should leave our system unchanged, and its physical states altered at most by a phase, we must have $T^2 = \alpha \mathbf{1}$, $|\alpha| = 1$. Actually, more can be shown, namely that the antiunitarity of T implies in fact $\alpha = \pm 1$.

These two alternatives correspond to integer or half-integer total angular momentum of the system. Since orbital angular momentum is always integer, these two alternatives in fact correspond to even or odd number of half-integer spin components.

Another aspects that plays a role in determining the invariance group G is that of invariance of the system under space rotation. Indeed, it appears that in

[1]Recall that, given some (complex) Hilbert space H, an operator $U : \mathrm{H} \to \mathrm{H}$ is *anti-unitary* if $\langle Ux, Uy\rangle = \overline{\langle x, y\rangle}$. Recall that time-reversal, that changes the sign of p, but not that of q, may be realized only in terms of some anti-unitary operator, to leave invariant the commutator $[q, p] = \mathrm{i}\hbar$.

presence of time-reversal invariance, with $\alpha = 1$, the Hamiltonian is real-symmetric. If $\alpha = -1$, the Hamiltonian is real-symmetric only if the system is also invariant under space rotations. A systematic and exhaustive discussion can be found in the original paper by Dyson [11].

2.3. The invariance group G. As sketched above, when interpreting our random matrix as a random Hamiltonian, the invariance group G is determined by: *i)* invariance under time reversal; *i)* invariance under space rotations; *i)* integer or half-integer total angular momentum. There are three possible cases:

- *Time-reversal and integer spin, or time-reversal and rotational invariance:* It can be seen that in these two situations, the Hamiltonian is real-symmetric, $H = H^t$, and the invariance group G is the group of $N \times N$ orthogonal matrices $O(N)$.
- *No time-reversal invariance:* this corresponds to the most generic situation of an Hermitean Hamiltonian, $H = H^\dagger$, and the invariance group G is the group of $N \times N$ unitary matrices, $U(N)$. Tipically, time reversal is broken by the presence of an external magnetic field. In this case, the system is (obviously) invariant under T^2, rather than T.
- *Time-reversal invariance, with odd spin, and no rotational invariance:* In this case, $\alpha = -1$, corresponding to a systems with an odd number of half-integer spin components. The absence of rotational invariance does not allow for a symmetric Hamiltonian. It appears that in this case the Hamiltonian has to be quaternionic-real, $H = H^R$. This means that each element is quaternionic, $H_{jk} = h^a_{jk}\sigma^a$, where h^a_{jk}, $a = 0, 1, 2, 3$, are real numbers, while σ^0 is the 2×2 identity, and the σ^a, $a = 1, 2, 3$ are Pauli matrices. One can define the conjugate $\bar{H}_{jk} = h^0_{jk}\sigma^0 - \sum_{a=1}^3 h^a_{jk}\sigma^a$. The corresponding invariance group G is the unitary simplectic group $USp(N)$, (or quaternionic unitary group), namely the group of unitary matrices W that are quaternionic self dual, i.e., such that $\overline{W}_{jk} = W_{kj}$.

This defines three different ensembles, according to their invariance group, namely the Orthogonal, Unitary and Symplectic ensemble, respectively. These are known as the Wigner-Dyson ensembles,

Actually the above discussion can be somewhat generalized, giving rise to a few additional invariance group (such as the chiral analogous of the three groups above, relevant for the study of chiral effective actions for QCD), for a total of 10 different cases, for generic rank N, in agreement with Cartan's classification of symmetric spaces - that is, Lie groups with constant curvature.

2.4. The integration measure $\mathcal{D}M$. We need now to define an integration measure for each of the three ensembles of matrices discussed in the previous subsection; we require it to be invariant under the corresponding invariance group G. We thus need to define integration over the group of Symmetric, Hermitean, and Symplectic matrices. This can be expressed as the integration over the maximal set of independent matrix elements, in each case:

$$\mathcal{D}M = \prod_j^N \mathrm{d}M_{jj} \prod_{\substack{j,k=1 \\ j<k}}^N \mathrm{d}M_{jk}, \qquad M = M^t, \tag{2.1}$$

Omeas

$$\mathcal{D}M = \prod_{j}^{N} \mathrm{d}M_{jj} \prod_{\substack{j,k=1 \\ j<k}}^{N} \mathrm{d}(\mathrm{Re}M_{jk})\mathrm{d}(\mathrm{Im}M_{jk}), \qquad M = M^\dagger, \tag{2.2}$$ Umeas

$$\mathcal{D}M = \prod_{j}^{N} \mathrm{d}M_{jj} \prod_{\substack{j,k=1 \\ j<k}}^{N} \prod_{a=0}^{3} \mathrm{d}M_{jk}^a, \qquad M = M^R. \tag{2.3}$$ Smeas

For reasons which will become clear below, it is convenient to assign an index $\beta = 1, 2, 4$ to each of these ensembles. Note that β counts the number of real parameters in the off-diagonal entries.

2.5. The probability law $\mathcal{P}(M)$. We now tackle the question of the choice of the probability law $\mathcal{P}(M)$. One possible requirement is that the probability law should be invariant under the invariance group G. Another possible requirement is that the matrix should be 'as random as possible', namely that the algebraically independent matrix elements should be statistically independent, that is independently distributed, with a probability law $\mathcal{P}(M)$ factorized in terms of the distribution densities of the algebraically independent matrix elements.

2.6. Gaussian ensembles. It appears that the only possible way to satisfy these two requirements is to choose a Gaussian distribution for each of the algebraically independent matrix elements. For example in the case of Hermitean matrices:

$$\mathcal{P}(H) = \prod_{j=1}^{N} \mathrm{e}^{-\frac{1}{2}H_{jj}^2} \prod_{\substack{j,k=1 \\ j<k}}^{N} \mathrm{e}^{-\mathrm{Re}(H_{jk})^2} \mathrm{e}^{-\mathrm{Im}(H_{jk})^2} = \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}\left[H^2\right]}$$

This was indeed the original choice of Wigner, see (1.1) and (1.2).

In correspondence to this choice of Gaussian probability law, we have three fundamental ensembles, known as Gaussian Wigner-Dyson ensembles: the Gaussian Orthogonal Ensemble (GOE), Gaussian Unitary Ensemble (GUE), and Gaussian Simplectic Ensemble (GSE), together with their chiral and other extensions, for a total of ten Gaussian ensembles.

2.7. Wigner matrices. If the requirement of statistical independence is considered more fundamental that the one of invariance under the group G, then we must have a factorized probability law, where the upper diagonal matrix elements H_{jk}, $j < k$, are independent identically distributed centered random variables, while the diagonal matrix elements H_{jj} are also independent identically distributed centered random variables, with possibly a different distribution. The lower diagonal element H_{jk}, $j > k$ are then obtained by conjugation. Matrices built in this way are referred to as 'Wigner matrices'.

2.8. Invariant matrix ensembles. In many situation it is preferable to have a G-invariant probability law, even at the price of giving up statistical independence.

To build such a G-invariant probability law, let us start by recalling that, given an $N \times N$ matrix, such as those discussed above, one can build a complete set of N invariants of the matrix under the action of G, namely the traces of the matrix and of its powers,

$$T_k := \mathrm{Tr}\left[M^k\right], \qquad k = 1, \ldots, N.$$

Traces of higher powers of M are expressible in terms of $T_1, \dots, T_N$. We can thus choose as probability law any positive and normalizable function of the quantities T_k, that we can write, in full generality,

$$\mathcal{P}(M) = \mathrm{e}^{-W(T_1,\dots,T_N)} = \mathrm{e}^{-\mathrm{Tr}[V(M)]}$$

where the function $V(x)$ is know as the 'potential' of the Random Matrix Model; it is subject to the the condition $V(x) = O(|x|^\alpha)$, $\alpha > 0$, as $|x| \to \infty$, in order to make the probability law normalizable. Typically, the potential $V(x)$ will be polynomial and bounded from below. It can also be non-polynomial, e.g., $V(x)) = \log(\cosh x)$; in such cases, the function of a matrix is defined in term of its power expansion. Ensembles with such probability laws are called 'invariant matrix ensembles'.

2.9. Circular ensembles. The above discussion was carried on in the situation were the random matrix is somehow related to the Hamiltonian of the considered physical system. However, it may happen that the quantity we are interested here is related to some different object; for instance, in the investigation of the transport properties in mesoscopic conducting systems, the appropriate object is rather the scattering S-matrix. In general, this is a unitary matrix (where unitarity follows from some conservation law such as particle number or charge).

The group of unitary matrices is compact, and the probability law can thus be chosen to be constant, $\mathcal{P}(M) = 1$. The corresponding eigenvalues have modulus one and live on the unit circle of the complex plane. The analysis carried out above for Hermitean matrices can be extended to the case of unitary matrices, leading to the so-called circular ensembles. In particular, one obtains three more Wigner-Dyson ensembles: the Circular Orthogonal Ensemble (COE), Circular Unitary Ensemble (CUE), Circular Simpectic Ensemble (CSE), together with their chiral and other extensions, for a total of ten circular ensembles.

sec_jacobian

3. Probability density for the eigenvalues

We have thus determined the possible probability density functions for Random Matrix Models. Our aim is now to evaluate the probability density functions for the eigenvalues. We shall first treat the integration element $\mathcal{D}M$ for the various kind of matrices, and next discuss the probability law $\mathcal{P}(M)$.

3.1. Orthogonal matrices. In the case of the Orthogonal Ensemble, the integration element $\mathcal{D}M$ over real-symmetric matrices may be expressed as an integration over the $N(N+1)/2$ independent real entries, see (2.1). Since M is real-symmetric, it can be diagonalized by means of an orthogonal transformation, that is $M = O^t \Lambda O$, where O is a suitable orthogonal matrix, $O^t = O^{-1}$, and Λ is the corresponding diagonalized matrix, $\Lambda_{jk} = \lambda_j \delta_{jk}$. Note that such decomposition is not unique, but allows for eigenvalues 'reshuffling', i.e. a mere premutations of the eigenvectors. The variables related to Λ may be thought about as 'physical' ones, while those related to O as 'gauge' ones. Note also that O has $N(N-1)/2$ independent entries, while Λ has N entries, so that the total number of independent entries matches that of the matrix M.

We now perform the change of variable

$$\mathcal{D}M = |J|\, \mathcal{D}O\, \mathcal{D}\Lambda$$

where J is the corresponding Jacobian,

$$J = \det \left| \frac{\partial(M_{jk})}{\partial(\lambda_l, O_{\alpha\beta})} \right|,$$

with $j \leq k$, $\alpha < \beta$, and $j, k, l, \alpha, \beta = 1, \ldots, N$. A generic entry of M is expressed in terms of O and Λ as follows:

$$M_{jk} = \sum_{\alpha=1}^{N} O^t_{j\alpha} \lambda_\alpha O_{\alpha k} = \sum_{\alpha=1}^{N} \lambda_\alpha O_{\alpha j} O_{\alpha k}.$$

Thus the first N columns of J do not depend on λ's, while the remaining $N(N-1)/2$ columns are linear in λ's. It follows that J is a polynomial in variables $\lambda_1, \ldots, \lambda_N$, of total degree $N(N-1)/2$. Also, whenever two eigenvalues coincide, the $O(N)$ diagonalization transformation is not completely determined, with a residual $O(2)$ invariance; in such case the Jacobian must obviously vanish. This uniquely fixes the form of the Jacobian to

$$J = \text{const.} \prod_{\substack{j,k=1 \\ j<k}}^{N} (\lambda_k - \lambda_j) \tag{3.1}$$

orthogJacobian

The constant can be evaluated, and is exactly one (modulo a sign, under reshuffling of the eigenvalues). However, as it will be apparent below, this is irrelevant for our purposes.

3.2. Vandermonde determinant. Let us define the determinant

$$\Delta(\lambda_1, \ldots, \lambda_N) := \det[\lambda_j^{k-1}]_{j,k=1}^{N}.$$

This is known as Vandermonde determinant and is ubiquitous in Random Matrix Models. It immediately follows from its definition that: i) it is a polynomial of degree $N-1$ in each of its N variables; ii) it vanishes as soon as any two λ's coincide; iii) the coefficient of the term $\lambda_2 \lambda_3^2 \ldots \lambda_N^{N-1}$ is 1. This implies:

$$\Delta(\lambda_1, \ldots, \lambda_N) := \prod_{\substack{j,k=1 \\ j<k}}^{N} (\lambda_k - \lambda_j)$$

which coincides with the expression of the Jacobian (3.1).

3.3. Unitary matrices. We now turn to the case of the Unitary Ensemble. For the sake of illustration we shall resort to a slightly different method. The integration element $\mathcal{D}M$ over Hermitean matrices may be expressed as an integration over the N^2 independent real entries, see (2.2). Since M is Hermitean, it can be diagonalized by means of a unitary transformation, that is $M = U^\dagger \Lambda U$, where U is a suitable unitary matrix, $U^\dagger = U^{-1}$, and Λ is the corresponding diagonalized matrix, $\Lambda_{jk} = \lambda_j \delta_{jk}$. Note that such decomposition is not unique, but allows for eigenvectors 'reshuffling', as in the Orthogonal case, and also for independent multiplication of each of the N eigenvector by a phase. Thus only $N(N-1)$ parameters of the N^2 independent real components of a unitary matrix are indeed relevant for the diagonalization process.

Again, we perform the change of variables from M to Λ and U, and we have to compute the corresponding Jacobian. Since by definition the measure $\mathcal{D}M$ is invariant under the $U(N)$, the Jacobian can only depends on the eigenvalues, that

is on Λ, and not on U. It is thus sufficient to compute J in the vicinity of the identity, $U = \mathbf{1} + \mathrm{i}A$, with A Hermitean. The differential of $M = (\mathbf{1} - \mathrm{i}A^\dagger)\Lambda(\mathbf{1} + \mathrm{i}A)$ is

$$\mathrm{d}M = -\mathrm{i}\mathrm{d}A\Lambda + \mathrm{d}\Lambda + \mathrm{i}\Lambda\mathrm{d}A,$$

or, in components

$$\begin{aligned}\mathrm{d}M_{jk} =& \mathrm{i}\mathrm{d}A_{jk}(\lambda_j - \lambda_k) + \mathrm{d}\lambda_j\delta_{jk} \\ =& (\lambda_j - \lambda_k)\mathrm{d}U_{jk} + \delta_{jk}\mathrm{d}\lambda_j\end{aligned}$$

We see that the change of variable is diagonal and the Jacobian is thus simply the product

$$|J| = \prod_{\substack{j,k=1 \\ j<k}}^{N} (\lambda_k - \lambda_j)^2,$$

that is, exactly the square of the Jacobian of the Orthogonal case.

3.4. Other cases. In the case of the Simplectic Ensemble, the evaluation of the Jacobian is technically more involved, and we give here only the result:

$$|J| = \prod_{\substack{j,k=1 \\ j<k}}^{N} (\lambda_k - \lambda_j)^4.$$

We thus see that four the three considered ensembles, the Jacobian is simply given by $|\Delta(\lambda)|^\beta$, with $\beta = 1, 2, 4$, respectively.

In the case of circular ensembles, the eigenvalues do not live anymore on the real axis but rather on the unit circle in the complex plane. Using the parameterization $\lambda_j = \mathrm{e}^{\mathrm{i}\theta_j}$, $j = 1, \ldots, N$, the Jacobian in the three Wigner-Dyson circular ensembles is given by

$$J = \prod_{\substack{j,k=1 \\ j<k}}^{N} \left|\mathrm{e}^{\mathrm{i}\theta_k} - \mathrm{e}^{\mathrm{i}\theta_j}\right|^\beta \propto \prod_{\substack{j,k=1 \\ j<k}}^{N} \left|\sin\left(\frac{\theta_k - \theta_j}{2}\right)\right|^\beta$$

with $\beta = 1, 2, 4$, respectively.

3.5. Integration over the invariance group G. In view of the discussion above, in general the probability measure for the matrix elements can be rewritten as

$$\mathcal{D}M\mathcal{P}(M) = \mathcal{D}G\,\mathrm{d}^N\lambda\,|J(\lambda_1, \ldots, \lambda_N)|\,\mathrm{e}^{-\sum_{j=1}^N V(\lambda_j)}$$

where in the last term we have reexpressed the trace of the matrix potential in terms of the eigenvalues; $\mathcal{D}G$ is the Haar measure on the compact group of invariance G. It is evident from inspection, and it actually follows directly from our construction, that the above expression does not depend on the elements of the invariance group G. Integration over $\mathcal{D}G$ can thus be done, giving just an overall finite constant.

3.6. Joint probability density for the eigenvalues. For the sake of simplicity, from now on we restrict to the case of invariant matrix ensembles. From the discussion above we immediately derive the joint probability density $P(\lambda_1, \dots, \lambda_N)$ of observing the N eigenvalues at positions $(\lambda_1, \dots, \lambda_N) \in \mathbb{R}^N$:

$$P(\lambda_1, \dots, \lambda_N) = \text{const.}\ |\Delta(\lambda_1, \dots, \lambda_N)|^\beta \, \mathrm{e}^{-\sum_{j=1}^N V(\lambda_j)}.$$

where $\beta = 1, 2, 4$ according to the symmetry class of the model, and we have left undetermined a trivial normalization constant. We emphasize that the probability is totally symmetric in the λ_j's.

Note that here, as in many other expressions in the present lectures, we are leaving an overall constant undetermined. This constant may be computed but it is irrelevant for most considerations, as it is absorbed into the normalization of the probability density, and does not appear in the computation of 'expectation values'.

In fact, once we define the partition function

$$Z_N^{(\beta)} = \int_{-\infty}^{+\infty} \mathrm{d}^N \lambda \prod_{\substack{j,k=1 \\ j<k}}^N |\lambda_k - \lambda_j|^\beta \, \mathrm{e}^{-\sum_{j=1}^N V(\lambda_j)},$$

we may introduce a normalized probability density for the eigenvalues

$$\rho_N^{(\beta)}(\lambda_1, \dots, \lambda_N) := \frac{1}{Z_N^{(\beta)}} \, |\Delta(\lambda_1, \dots, \lambda_N)|^\beta \, \mathrm{e}^{-\sum_{j=1}^N V(\lambda_j)}. \tag{3.2}$$

prob_eigen

which will be the fundamental object of our future investigations. We can further define the n-eigenvalue density, which is obtained by integrating ρ_N over $N - n$ variables:

$$\rho_n^{(\beta)}(\lambda_1, \dots, \lambda_n) := \int \rho_N^{(\beta)}(\lambda_1, \dots, \lambda_N) \, \mathrm{d}\lambda_{n+1} \dots \mathrm{d}\lambda_N. \tag{3.3}$$

eigen_density

and is normalized to 1 by construction. In particular the standard eigenvalue density is $\rho^{(\beta)}(\lambda) \equiv \rho_1^{(\beta)}(\lambda)$.

3.7. Electrostatic analogy. Consider the partition function associated with the probability density $P(\lambda_1, \dots, \lambda_N)$. By exponentiating the Vandermonde determinant, and suitably redefining the potential $V \to \beta N V$, we may write

$$Z_N^{(\beta)} = \int_{-\infty}^{+\infty} \mathrm{d}^N \lambda \exp\left[-\beta \left(-\frac{1}{2} \sum_{\substack{j,k=1 \\ j \neq k}}^N \log|\lambda_k - \lambda_j| + N \sum_{j=1}^N V(\lambda_j) \right) \right]. \tag{3.4}$$

coulomb

We immediately recognize the statistical mechanics partition function of a gas of N identical charged particles at temperature $1/\beta$, constrained to the real line, subject to electrostatic repulsion in the form of a logarithmic repulsive mutual interaction term, and confined by an external potential $V(x)$. This is a useful analogy we shall often exploit to get insight over the eigenvalue probability density, especially in the large N limit. In view of this analogy, the system of eigenvalues of random matrix models, is sometimes called *Log-gas* or *Coulomb gas*.

Note that this analogy is a particular instance of the well-known relation between D-dimensional quantum mechanics and $(D+1)$-dimensional statistical mechanics: indeed we have rephrased a zero-dimensional quantum field theory of $N \times N$ matrices (that has reduced after gauging away the diagonalization, to the

zero-dimensional quantum field theory of N eigenvalues) as the partition function of a one-dimensional system of N classical particles.

3.8. Qualitative discussion of the Coulomb gas. An obvious fundamental quantity in the Coulomb gas model is the density of eigenvalues $\rho_{sp}(\lambda)$, that describes in the large N limit, the *equilibrium configuration*, i.e., the configuration of charges that minimizes the free energy. The subscript *sp* stands for 'saddle-point' for reasons that will become evident later on.

Note that in equation (3.4) we have rescaled the coefficients in the potential by a factor N, so that in the large N limit the potential term and the mutual repulsion terms are both of order N^2. This is of importance for the two terms to compete and produce a non-trivial result; indeed, if the repulsive term should become irrelevant in the large N limit, we would get a trivial configuration with all eigenvalues accumulating at the minimum of the potential; viceversa, should the potential term become irrelevant, the eigenvalues would spread away at infinity.

Let us now briefly discuss the qualitative behaviour of $\rho_{sp}(\lambda)$. With the rescaling of the coefficients of the potential, as discussed above, we have genuine competition between the two terms, with the confining potential favouring the accumulation of eigenvalues towards the absolute minimum of the potential, and the repulsive interaction preventing their collapse to a single point. As a result, we expect the eigenvalues, that are real, to lie within a finite interval of the real line, with a finite width ditribution peaked in correpondence of the minimum of the potential. If the potential has several relative minima, with sufficiently hight potential barriers between them, we expect an eigenvalue spectrum with *forbidden bands* or *gaps*, and allowed bands. In this case the support of the density is expected to be a collection of disjoint intervals on the real line.

It is evident from this discussion that the eigenvalue density $\rho_{sp}(\lambda)$ cannot be universal, but rather depends on the particular form of the potential V.

4. Quantities of interest

We shall now define various quantities describing the statistical properties of the eigenvalue λ_j.

4.1. Expectation values. The most obvious statistical quantity is the expectation value of any generic function F of the eigenvalues:

$$\langle F\rangle := \int \rho_N(\lambda_1,\dots,\lambda_N)F(\lambda_1,\dots,\lambda_N)\mathrm{d}^N\lambda, \tag{4.1}$$ <F>

where the joint eigenvalue probability density ρ_N has been defined in (3.2). The expectation value is well defined for a very wide class of function F, modulo mild assumptions to guarantee the convergence of the integrand.

4.2. The eigenvalue probability density. Let us consider the observable

$$O(\mu) := \frac{1}{N}\sum_{j=1}^{N}\delta(\mu-\lambda_j), \tag{4.2}$$ observable

depending on the eigenvalues λ_j and on the additional parameter μ. Note that we have associated a factor $1/N$ with the trace appearing in (4.2), that is a sum over N terms; in this way , the observable remains finite in the large N limit.

The expectation value of the observable $O(\mu)$is the so-called *eigenvalues probability density*, i.e., the probability density of finding an eigenvalue in the vicinity of μ

$$\begin{aligned}\rho(\mu) :=& \langle O(\mu)\rangle \\ =& \int O(\mu)\rho_N(\lambda_1,\dots,\lambda_N)\mathrm{d}^N\lambda\end{aligned}$$

The eigenvalue density $\rho(\mu)$ is normalized to one, $\int \rho(\mu)\mathrm{d}\mu = \langle 1\rangle = 1$, by construction. In fact this definition coincides with the one of (3.3), as it is easily seen by using the fact that the joint density ρ_N is a totally symmetric function of its variables.

Another widely used alternative quantity, especially in the mathematically oriented literature, is the *eigenvalue density*, defined as $\sigma(\mu) := N\rho(\mu)$. This is normalized to N, and is related to the number of eigenvalues - rather than the probability of finding one of them - in the interval $\mathrm{d}\mu$.

4.3. The Green function. A useful observable is the trace of the resolvent operator $R(\mu) = (\mu\mathbf{1} - M)^{-1}$. Its expectation value, generically called *Green function* of the operator M, reads:

$$G(\mu) := \left\langle \frac{1}{N}\mathrm{Tr}\left[R(\mu)\right]\right\rangle = \int \frac{1}{\mu-\lambda}\rho(\lambda)\mathrm{d}\lambda\ ; \tag{4.3}$$ Green

The last integral expression is also known as the *Stieltjes transform* of the eigenvalue density.

The Green function encodes all the statistical properties of the spectrum, and will play a crucial row in what follows. In particular it may also also be viewed as the generating function of the *moments* of the eigenvalue probability density:

$$G(\mu) := \sum_{l=0}^{\infty}\left\langle \frac{1}{N}\mathrm{Tr}\left[M^l\right]\right\rangle \frac{1}{\mu^{l+1}}. \tag{4.4}$$ moments

and viceversa,

$$\left\langle \frac{1}{N}\mathrm{Tr}\left[M^l\right]\right\rangle = \frac{1}{2\mathrm{i}\pi}\oint_C G(z)z^l\mathrm{d}z, \tag{4.5}$$ moments2

where C is some simple positive contour enclosing all the poles singularities of the Green function.

If the eigenvalues are supported on the real axis, the discontinuity of the Green function across the real axis determines the eigenvalue probability density.

$$G(\mu+\mathrm{i}0) - G(\mu-\mathrm{i}0) = -2\mathrm{i}\pi\left\langle \frac{1}{N}\sum_{k=1}^{N}\delta(\mu-\lambda_l)\right\rangle = -2\mathrm{i}\pi\rho(\mu). \tag{4.6}$$ cut

Here we have used the Cauchy identity $1/(z\pm\mathrm{i}0) = \mathrm{PV}(1/z) \mp \mathrm{i}\pi\delta(z)$, where PV denotes the principal value.

4.4. Correlation functions. The n-point correlation function is defined as

$$\begin{aligned}R_n(\lambda_1,\dots,\lambda_n) :=& N^n\left\langle O(\lambda_1)\dots O(\lambda_n)\right\rangle \\ =& \frac{N!}{(N-n)!}\int \rho_N(\lambda_1,\dots,\lambda_N)\mathrm{d}\lambda_{n+1}\dots\mathrm{d}\lambda_N \\ =& \frac{N!}{(N-n)!}\rho_n(\lambda_1,\dots,\lambda_n)\end{aligned} \tag{4.7}$$ R_n

where we have assumed that $\lambda_1 \neq \cdots \neq \lambda_n$ to eliminate 'contact terms'. This essentially gives the probability of finding n level, regardless of order, in the infinitesimal neighborhoods of the points $\lambda_1, \ldots, \lambda_n$, the position of the remaining levels being unobserved. Note however that it is not a probability density in the strict sense since its total integral equals $N!/[(N-n)!]$, instead of 1. It is rather an n-eigenvalue density, in the sense explained above. For reasons that will become apparent later, the n-point correlation function is a somewhat more convenient object that the n-eigenvalue probability density.

4.5. Cluster functions. Just as in Quantum Field Theory (or Statistical Mechanics), for practical purposes it is usually more convenient to deal with the connected correlation functions, usually called 'cluster functions' in the context of Random Matrix Models, or also 'cumulants'; the n-point cluster is defined as

$$T_n(\lambda_1, \ldots, \lambda_n) := \sum_P (-1)^{n-m}(m-1)! \prod_{j=1}^{m} R_{|G_j|}(x_k, \text{with } k \in G_j)$$

where P is any partition of the set of indices $\{1, \ldots, n\}$ into m parts G_j. For example, for $n = 1$, $T_1(\lambda) = R_1(\lambda)$, for $n = 2$, $T_2(\lambda_1, \lambda_2) = R_1(\lambda_1)R_1(\lambda_2) - R_2(\lambda_1, \lambda_2)$, etc. Just as in Quantum Field Theory, if we know the generating function for the correlation functions R_n, taking its logarithm we immediately obtain the generating function for the cluster functions T_n.

Contrarely to what we are used in Quantum Field Theory or in Statistical Mechanics, the above n-point correlation function are not translation invariant. This is not a surprise, since the confining potential $V(\lambda)$ breaks translation symmetry. However, we will see below that translation invariance, i.e., dependence of n-point functions on differences $|\lambda_k - \lambda_j|$, can be recovered if we restrict to *small* intervals, i.e., intervals that are small enough that the eigenvalue density can be approximated as constant, and at the same time, large enough to contain a sufficiently high number n of levels, so that the notion of continuous density is still meaningful.

In other word translation invariance can be recovered if we restrict to interval containing n eigenvalues, with $1 \ll n \ll N$. In this regime, the cluster functions has the nice additional property of vanishing for large eigenvalue separations. Moreover, it is in this regime that universal features of Random Matrix Models can be observed. A preliminary necessary step is the one of 'unfolding' the spectrum, a procedure addressed in the next section.

5. Unfolding the spectrum

5.1. Mean spacing. As mentioned earlier, the eigenvalue density is not universal, as it depends on the precise form of the potential $V(\lambda)$. Anyway, this is not a big obstruction, since each physical system has its own level spectrum anyway. In particular, for each physical or Random Matrix Model spectrum, we can define a *mean spacing* between levels.

Let us consider a portion of the spectrum $\Delta\lambda$ that is small enough to consider $\sigma(\lambda)$ constant over it, but at the same time large enough to contain a sufficiently high number n of levels, so that the notion of continuous density is still meaningful; in other word we require the number n of levels within the interval $\Delta\lambda$ to satisfy $1 \ll n \ll N$. One can define the average over this interval of the distance between

two consecutive levels, that is the mean spacing

$$\Delta = \frac{1}{\sigma(\lambda)}.$$

It is not constant along the spectrum, but rather it depens on λ.

To compare two spectra coming from the most disparate physical systems and Random Matrix Models, and to look for universal properties, one need first to *rectify*, or *unfold* them, that is to perform a local rescaling to make them have the same average level spacing, say $\tilde{\Delta} = 1$. Here ans below the $\tilde{}$ refers to unfolded quantities.

5.2. The counting function. Let us introduce the counting function,

$$\begin{aligned}\mathcal{N}(\mu) &:= \left\langle \sum_{j=1}^{N} \theta(\mu - \lambda_j) \right\rangle \\ &= \int_{-\infty}^{\mu} \sigma(\lambda) \mathrm{d}\lambda\end{aligned}$$

where θ is the Heaviside step function. Clearly $\mathcal{N}(\mu)$ increases by 1 each time μ meets an eigenvalue λ_j.

Introducing the variable $x_j := \mathcal{N}(\mu_j)$, it is clear that the corresponding 'spectrum' has now constant mean spacing, $\tilde{\Delta} = 1$. It also follows directly from our definition that $\mathrm{d}x/\mathrm{d}\lambda = \sigma(\lambda)$.

It is easy to see that the density of the x_j's is constant, and in fact trivially 1. Indeed, this directly follows from the fact that $\tilde{\sigma}(x)\mathrm{d}x = \sigma(\lambda)\mathrm{d}\lambda$ must hold, by construction. The x_j are called the unfolded variables.

5.3. Unfolding. We are know able to unfold the spectrum:

- we first compute the average density $\sigma(\lambda)$, and its primitive $\mathcal{N}(\lambda)$;
- next we make a change of variables from the λ_j's to the unfolded variables $x_j := \mathcal{N}(\lambda_j)$;
- then we compute all statistical properties by averaging over the x_j's rather than the λ_j's.

As a result, as anytime we make a change of variable in a density, we have for the unfolded eigenvalue probability densities

$$\tilde{\rho}_n(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n = \rho_n(\lambda_1, \ldots, \lambda_n)\, \mathrm{d}\lambda_1 \ldots \mathrm{d}\lambda_n$$

which immediately implies

$$\tilde{\rho}_n(x_1, \ldots, x_n) = \frac{\rho_n(\lambda_1, \ldots, \lambda_n)}{\sigma(\lambda_1) \ldots \sigma(\lambda_n)}, \tag{5.1}$$

unfolded

and similarly for the n-point correlation and cluster functions. In particular we recover that $\tilde{\sigma}(x) = 1$, and $\tilde{\rho}(x) = 1/N$.

5.4. Level spacing distribution. A fundamental quantity, especially useful to address the short range correlations in the unfolded spectrum, is the so-called level spacing distribution, that we have already mentioned in the Introduction together with some of its universal features.

Given two consecutive levels x_j and x_{j+1} of the unfolded spectrum, the level spacing distribution gives the probability $p(s)\mathrm{d}s$ that their spacing $x_{j+1} - x_j$ lies in the interval $[s, s + \mathrm{d}s]$.

Let us start with considering the (unfolded) emptiness formation probability, that is the probability of having *no* eigenvalue in the interval $I = [y_1, y_2]$,

$$E(y_1, y_2) := \left\langle \prod_{j=1}^{N} [1 - \chi_I(x_j)] \right\rangle = \int_{\vec{x} \notin I^N} \tilde{\rho}_N(x_1, \dots, x_N) \mathrm{d}^N x.$$

Consider now the quantity

$$\frac{\mathrm{d}^2 E(y_1, y_2)}{\mathrm{d}y_1 \mathrm{d}y_2} = - \sum_{\substack{j,k=1,\dots,N \\ j \neq k}}^{N} \left\langle \delta(y_1 - x_j)\delta(y_2 - x_k) \prod_{\substack{l=1 \\ l \neq j,k}}^{N} [1 - \chi_I(x)] \right\rangle.$$

To perform the calculation, recall that $\chi_I(x) = \theta(x - y_1)\theta(y_2 - x)$. The RHS above is obviously proportional to the probability of having one eigenvalue at y_1, another at y_2, and none in between. Recalling that the x_j's are unordered and indistinguishable, we may write the (unfolded) level spacing distribution for $s = |y_2 - y_1|$ as

$$p(s) = -\frac{1}{N(N-1)} \frac{\mathrm{d}^2 E(y_1, y_2)}{\mathrm{d}y_1 \mathrm{d}y_2}.$$

Note that the very definition of a function $p(s)$ somehow assumes translational invariance of the spectrum, which is, striclty speaking, not true. However, taking N sufficiently large, and restricting to a portion of the spectrum which is sufficiently 'far from the boundaries', in the sense that $x_j \in [a, N - a]$, with $1 \ll a \ll N$, the assumption of translational invariance becomes a reasonable one.

The level spacing distribution, just as the emptiness formation probability, being in fact an N-point correlation function, is actually difficult to compute in general.

5.5. Unfolded correlation functions and universality. It is only after unfolding that statistical properties of Random Matrix Models can exhibit any kind of universal behaviour. The unfolded eigenvalue density is trivially universal, $\tilde{\sigma}(\lambda) = 1$, by construction.

Consider now the 2-point correlator $R_2(\lambda_1, \lambda_2)$ in the large N limit, in the bulk of the spectrum, and for eigenvalue separation $|\lambda_2 - \lambda_1| = O(1/N)$, i.e., of the same order of the mean level spacing. After unfolding, this amounts to consider the 2-point correlator $\tilde{R}_2(x_1, x_2)$, for $|x_2 - x_1| = O(1)$. As we shall see in the course of these lectures, it exhibits the following universal behaviour

$$\tilde{R}_2(x) = 1 - \left(\frac{\sin \pi x}{\pi x} \right)^2, \qquad x = O(N^0), \tag{univ_2p}$$

where $x := x_2 - x_1$.

Such universal behaviour holds in the $\beta = 2$ case and will be derived later on. Slightly different but qualitatively analogous behaviours are observed in the $\beta = 1$ and $\beta = 4$ cases. Universal behaviours are observed for other correlation functions as well, such as the level spacing distribution, or the largest eigenvalue distribution, and also in many other occurrence, such as at the edge of the spectrum, or in the vicinity of a critical point of the potential.

6. The saddle-point approximation

From now on we specialize to the case of unitary matrices, $\beta = 2$, within invariant ensembles. We are interested in the evaluation of the partition function Z_N and of the correlation functions $\langle F(M)\rangle$ for some Random Matrix Model, with a given potential $V(M)$.

Several methods have been devised to tackle the problem, for instance the Saddle-Point approximation, the Orthogonal Polynomials approach, and the Loop Equation method. We shall here focus on the first method, that has the great advantage of simplicity, although it does not allow for a sistematic treatment of corrections to the leading order approximation.

6.1. Preliminaries. For the application of the method it is essential that the potential contains a single trace (e.g., a potential of the form $\mathrm{Tr}[V_1(M)]\mathrm{Tr}[V_2(M)]$, although defining an invariant ensemble, is not suitable). It is also convenient to normalize the potential in such a way that both the (logarithm of the) Vandermonde and the potential terms are $O(N^2)$. We are thus considering an integral of the form

$$Z_N = \int \mathrm{e}^{-N^2 S(\lambda)} \mathrm{d}^N \lambda \tag{6.1}$$ Z_N

where $S(\lambda)$ is the Coulomb gas action reads

$$S(\lambda) = \frac{1}{N}\sum_{j=1}^{N} V(\lambda_j) - \frac{1}{N^2}\sum_{\substack{j,k=1\\ j\neq k}} \log|\lambda_k - \lambda_j|, \tag{6.2}$$ action

see (3.4). Note that with respect to our original definitions, see Section 2, we have rescaled the potential according to $V \to \beta N V/2 = NV$

We emphasize that, contrarely to the other situations where the saddle-point method is usually applied, in the present case the large parameter N appears not only in the exponent of the integrand, but also as the dimension of the integration space. Consequently, the saddle-point condition is not a single equation but rather a set of N coupled nonlinear equations. We will thus need some additional tool, with respect to the standard saddle-point method, to solve these N coupled equations.

6.2. The saddle-point equation. The saddle-point equations are readily calculated:

$$\frac{\partial S}{\partial \lambda_j} = 0 = V'(\lambda_j) - \frac{2}{N}\sum_{\substack{k=1\\ k\neq j}}^{N} \frac{1}{\lambda_j - \lambda_k}, \qquad j = 1, \ldots, N. \tag{6.3}$$ SPE

In the Coulomb gas analogy, these are just the requirement that the total 'force' on the jth particle is zero. Thus, in this picture, the solutions of the saddle-point equations are simply the positions of the N charged particles in the equilibrium configuration. The N charges are constrained to the real axis, and thus the solutions of the saddle-point equations will be real as well.

Denoting by $\tilde{\lambda}_j$ the actual solutions of the saddle-point equations (6.3), we immediately have the evaluation $Z_N \sim \mathrm{e}^{-N^2 S(\tilde{\lambda})}$, and more generally, for any correlation function of the form (4.1), $\langle F\rangle \sim F(\tilde{\lambda})$. In particular it follows that correlation functions of products of invariant functions factorize, $\langle FG\rangle \sim F(\tilde{\lambda})G(\tilde{\lambda}) \sim \langle F\rangle\langle G\rangle$,

and cluster functions vanishes (at leading order in $1/N$, which is the only one we are able to compute in this random matrix version of the saddle-point approximation).

6.3. The resolvent approach. As already mentioned in the qualitative discussion of the Coulomb gas, the solutions of the saddle-point equations are best described by the the large N eigenvalue density $\rho_{sp}(\lambda)$. In turn, this is completely determined by the form taken in the large N limit by the Green function $G(\mu)$, defined in (4.3). We thus need to evaluate the Green function in the saddle point approximation. In view of the previous discussion, this is simply:

$$G_{sp}(z) = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{z-\tilde{\lambda}_j} =: W(z).$$

Note that, due to the normalization condition, $\int \rho_{sp}(z)\mathrm{d}z = 1$, we have the large $|z|$ behaviour

$$W(z) = \frac{1}{z} + O\left(\frac{1}{z^2}\right), \qquad |z|\to\infty, \tag{6.4}$$

asymptotic

see (4.4).

Let us now evaluate the square of the function $W(z)$:

$$\begin{aligned}
W^2(z) =& \frac{1}{N^2}\sum_{j,k=1}^{N}\frac{1}{z-\tilde{\lambda}_j}\frac{1}{z-\tilde{\lambda}_k}\\
=& \frac{1}{N^2}\sum_{j=1}^{N}\frac{1}{(z-\tilde{\lambda}_j)^2} + \frac{1}{N^2}\sum_{\substack{j,k=1\\ j\neq k}}^{N}\frac{1}{z-\tilde{\lambda}_j}\frac{1}{z-\tilde{\lambda}_k}\\
=& -\frac{1}{N}W'(z) + \frac{1}{N^2}\sum_{\substack{j,k=1\\ j\neq k}}^{N}\left(\frac{1}{z-\tilde{\lambda}_j}-\frac{1}{z-\tilde{\lambda}_k}\right)\frac{1}{\tilde{\lambda}_j-\tilde{\lambda}_k}\\
=& -\frac{1}{N}W'(z) + \frac{2}{N^2}\sum_{\substack{j,k=1\\ j\neq k}}^{N}\frac{1}{z-\tilde{\lambda}_j}\frac{1}{\tilde{\lambda}_j-\tilde{\lambda}_k}\\
=& -\frac{1}{N}W'(z) + \frac{1}{N}\sum_{j=1}^{N}\frac{V'(\tilde{\lambda}_j)}{z-\tilde{\lambda}_j}
\end{aligned}$$

where in the last equality we have used the saddle-point equations (6.3). Adding and substracting $V'(z)$ in the numerator of the last term, we have

$$W^2(z) = -\frac{1}{N}W'(z) + W(z)V'(z) - \frac{1}{N}\sum_{j=1}^{N}\frac{V'(z)-V'(\tilde{\lambda}_j)}{z-\tilde{\lambda}_j}.$$

Observe that: *i)* if $V(z)$ is a polynomial of degree $d\geq 2$ (which is the situation we shall mainly consider in these lectures), then the last term is a polynomial of degree $d-2$ in z, that we shall denote $P(z)$, and whose explicit form is of course still unknwon at this level; *ii)* the obtained equation is a non-linear differential equation, which can be difficult to solve. However, in the saddle-point approach, at leading

order in $1/N$, the term containing $W'(z)$ is subleading and can be neglected. In the saddle-point approximation, we are thus left with a quadratic equation for $W(z)$,

$$W^2(z) - W(z)V'(z) + P(z) = 0,$$

whose solution is

$$W(z) = \frac{1}{2}\left(V'(z) - \sqrt{V'^2(z) - 4P(z)}\right), \tag{6.5}$$

Wsol

where the choice of the minus sign is dictated by the asymptotic behaviour (6.4).

6.4. Poles and cuts. We observe that for finite N the saddle-point resolvent $W(z)$ is by construction a rational function of z, while in the large N limit it turns into an algebraic function. In other words, it looks like the N poles of the resolvent $G(z)$, that is, the N solutions of the saddle-point equations (6.3), coalesce to form cuts, as $N \to \infty$. We shall see that this is indeed the case.

We can imagine that in the large N limit the $\tilde{\lambda}_j$ form a continuous distribution with density $\rho_{sp}(\tilde{\lambda})$ on a real support S, made of just one interval or several disjoint intervals, see the qualitative discussion of the Coulomb gas. We may write:

$$W(z) = \int_S \frac{1}{z-\mu}\rho_{sp}(\mu)\mathrm{d}\mu \qquad z \in \mathbb{C} \setminus S.$$

resolvent2

Thus $W(z)$ is an analytic function all over the complex plane of z, with one or more cuts in correspondence of the real support S of the density $\rho_sp(z)$.

In this language, the saddle-point equation (6.3) may be written

$$2\mathrm{PV}\int_S \frac{1}{z-\mu}\rho_{sp}(\mu)\mathrm{d}\mu = V'(z) \tag{6.6}$$

SPE2

Moreover, it follows immediately from Cauchy identity, see (4.6), that the density is determined by the discontinuity across the cut,

$$W(z+\mathrm{i}0) - W(z-\mathrm{i}0) = -2\mathrm{i}\pi\rho_{sp}(z). \tag{6.7}$$

density

and that $W(z)$ must satisfy the following equation

$$W(z+\mathrm{i}0) + W(z-\mathrm{i}0) = 2\,\mathrm{PV}\int_S \frac{1}{z-\mu}\rho_{sp}(\mu)\mathrm{d}\mu, \qquad z \in S.$$

This last equation, together with (6.6), allows to rewrite the saddle-point equation as

$$W(z+\mathrm{i}0) + W(z-\mathrm{i}0) = V'(z), \qquad z \in S. \tag{6.8}$$

SPElinear

which is *linear*, to be compared with the quadratic equation derived above.

Finally, note that in the present 'complex plane' reformulation, with $\rho(x)$ given by (6.7), the normalization condition $\int_S \rho(x)\mathrm{d}x = 1$ can be rewritten as

$$\frac{1}{2\mathrm{i}\pi}\oint_{C_S} W(z)\mathrm{d}z = 1, \tag{6.9}$$

normalization

where C_S is a simple positive contour enclosing the support S. This expression can be viewed as the saddle-point limit of (4.5), in the case $l = 0$; also, by deforming C_S to a very large contour, consistency with (6.4) is easily verified.

In conclusion, the problem of evaluating the density eigenvalues in the saddle-point approximation has been reduced to that of solving the functional equation (6.8), supplemented by the normalization condition (6.9). This determines the resolvent, and, via (6.7), the eigenvalue density.

6.5. One-cut assumption. If $V(z)$ has only one minimum, according to the Coulomb gas picture, the eigenvalues, i.e. the roots of the saddle-point equations, accumulate towards this minimum, and it is reasonable to expect that the support of $\rho_{sp}(z)$ is made of just one interval on the real axis, $S=[a,b]$. It immediately follows that the solution of (6.5) must have the form

$$W(z)=\frac{1}{2}\left[V'(z)-Q(z)\sqrt{(z-a)(z-b)}\right], \qquad z\in\mathbb{C}\setminus[a,b], \tag{6.10}$$ Wsol_onecut

where $Q(z)$ is a polynomial of degree $d-2$. Note that to completely determine $W(z)$, we need to find the coefficients of $Q(z)$, and the two end-points of the density support S, for a total of $d+1$ parameters. Observing that expression (6.10) has the large $|z|$ behaviour $W(z)=O(z^{d-1})$, it is easily verified that requiring the asymptotic behaviour (6.4) provides exactly $d+1$ conditions, allowing for the complete determination of the unknown parameters and thus of $W(z)$.

Note that, under our assumption of a polynomial potential, from the form of the resolvent $W(z)$, it is clear that the density will be of the form $\rho(z)=\frac{1}{2\pi}Q(z)\sqrt{(z-a)(b-z)}$, see (6.7). Thus it immediately follows that, independently from the particular form of the potential, in the vicinity of, say, the upper edge of the spectrum, $z\sim b$, we always have the behaviour $\rho(z)\sim\sqrt{b-z}$. This is a first example of universal behaviour, that holds as long as the polynomial $Q(z)$ does not vanish at the edge of the spectrum.

Note that the solution (6.10) makes sense only as long as $Q(z)$ is positive over the interval $[a,b]$, since otherwise the corresponding density would turn negative. If $Q(z)$ appears to be negative for some $z\in[a,b]$, this means that the one-cut assumption is not valid.

Finally, it is easily verified that the solution (6.10) for $W(z)$ indeed solve the linear version (6.8) of the saddle-point equation.

6.6. Example: the quadratic potential and Wigner semicircular law. Let us consider the case $V(\lambda)=g\lambda^2/2$, that is, at $g=1$, the Gaussian Unitary Ensemble. The polynomial $Q(z)$ in (6.10) reduces to a constant, and we have

$$W(z)=\frac{1}{2}\left[gz-C\sqrt{(z-a)(z-b)}\right], \qquad z\in\mathbb{C}\setminus[a,b],$$ Wsol_GUE

where the three currently unknown constants a, b, C, will be determined by the normalization condition (6.4). Indeed, recalling that $\sqrt{1+x}\sim 1+x/2-x^2/8$, $x\sim 0$, we have

$$\begin{aligned} W(z)&=\frac{1}{2}\left(gz-Cz\sqrt{(1-\frac{a+b}{z}+\frac{ab}{z^2}}\right),\\ &\sim\frac{g-C}{2}z+C\frac{a+b}{4}+C\frac{(b-a)^2}{16}\frac{1}{z}, \qquad |z|\to\infty \end{aligned}$$

This should match (6.4), thus implying $C=g$, $a=-2/\sqrt{g}$, and $b=2/\sqrt{g}$, with

$$W(z)=\frac{g}{2}\left(z-\sqrt{z^2-4/g}\right). \tag{6.11}$$ W_quadratic

Using (6.7), the corresponding density is readily calculated

$$\rho(z)=\frac{g}{2\pi}\sqrt{\frac{4}{g}-z^2} \qquad z\in\left[-\frac{2}{\sqrt{g}},\frac{2}{\sqrt{g}}\right]. \tag{6.12}$$ semicir

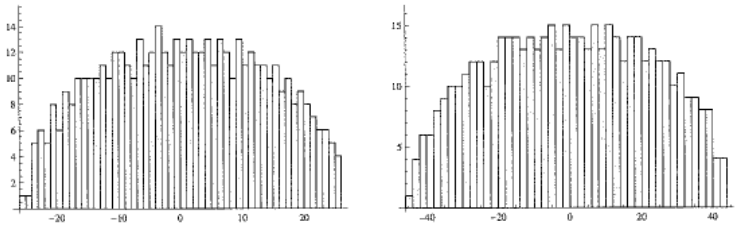


FIGURE 3. Histogram of the eigenvalues of a typical real symmetric Wigner matrix of size 500×500, with i.i.d elements with probability density as in (6.13), left, or (6.14), right. wigner

This is the celebrated Wigner's semicircle law.

Remark: note that in the literature one often find the alternative expression

$$\sigma(\mu) = \frac{1}{\pi}\sqrt{2N - \mu^2},$$

the difference being due to *(i)* the rescaling of potential, corresponding to the choice $g = 2/N$, and *(ii)* the definition of density $\sigma(\mu) = N\rho(\mu)$.

Remark: We have already commented that, generally speaking, in the case of invariant ensembles, the eigenvalue density is not universal, depending on the form of the potential. Indeed this is manifest from expression (6.5) for the resolvent. However, if one considers instead Wigner matrices, that is matrix ensembles where the probability law $\mathcal{P}(M)$ is factorized, although not invariant under the matrix symmetry group, one observes that Wigner semicircle law is observed for a wide variety of factorized probability laws. In particular, focussing for definiteness on the case of real symmetric matrices, $\beta = 1$, and chosing for example one of the following probability densities for the matrix elements M_{ij}, $i \le j$:

$$\mathcal{P}(M_{ij}) = \begin{cases} \frac{1}{2}, & M_{ij} \in [-1,1], \\ 0, & M_{ij} \notin [-1,1], \end{cases} \qquad (6.13) \quad \text{wigner\_dist1}$$

$$\mathcal{P}(M_{ij}) = 1/2[\delta(M_{ij}+1) + \delta(M_{ij}-1)], \qquad (6.14) \quad \text{wigner\_dist2}$$

numerical experiments for large values of N clearly show the emergence of Wigner semicircle law for the eigenvalue density, see Fig. 3.

sec.quartic

6.7. Example: the (single well) quartic potential. Let us consider now the quartic potential $V(\lambda) = \frac{1}{2}\lambda^2 + \frac{1}{4}g\lambda^4$. From (6.10), we readily write for the resolvent the expression

$$W(z) = \frac{1}{2}\left(z + gz^3 - M(z)\sqrt{(z-a)(z-b)}\right), \qquad z \in \mathbb{C} \setminus [a,b], \qquad (6.15) \quad \text{W\_quartic1}$$

where $M(z)$ is a polynomial of degree 2. We have thus five unknown constants, the three coefficients c_0, c_1 and c_2 of $M(z)$, and the two end-points a, b, to be determined by means of the five conditions coming from the requirement that the large $|z|$ asymptotic behaviour of (6.15) matches (6.4). The calculation is left as an exercise. [Hint: to simplify calculations, recall that, the potential $V(\lambda)$ being even, the eigenvalue density must be an even function of z, and so must be the singular part of $W(z)$; this immediately implies $c_1 = 0$ and $b = -a$].

The result is

$$W(z) = \frac{1}{2}\left[z + gz^3 - (1 + \frac{g}{2}a^2 + gz^2)\sqrt{z^2 - a^2}\right], \qquad z \in \mathbb{C} \setminus [-a, a], \tag{6.16}$$

W_quartic2

where a^2 is the solution of

$$3ga^4 + 4a^2 - 16 = 0, \tag{6.17}$$

W_quartic_a1

that is

$$a^2 = \frac{2}{3g}\left(-1 + \sqrt{1 + 12g}\right), \tag{6.18}$$

W_quartic_a2

the choice of the plus sign being dictated by $a^2 \geq 0$. The corresponding eigenvalue probability density is readily evaluated,

$$\rho(z) = \frac{1}{2\pi}(1 + \frac{g}{2}a^2 + gz^2)\sqrt{a^2 - z^2} \qquad z \in [-a, a],$$

and all invariant quantities such as the free energy, the moments, etc., can be calculated.

Remark that at $g = -1/12 =: g_c$, the end-point a^2 has a square-root singularity. Correspondingly, all physical quantity, including the free energy, become singular. To understand the physical meaning of this singular behaviour, let us plot the potential $V(\lambda)$ for various values of g. For positive g, we have a regular potential well, and correspondingly a well-defined matrix integral. When g is negative, our matrix integral is ill-defined, and in fact this is manifest in the usual perturbative approach, where one has a vanishing convergence radius.

What is surprising is that in the present situation, at negative but sufficiently small values of g, $g > g_c$, even if the original matrix integral is ill-defined, our saddle-point solution is still meaningful, and corresponds to the fact that a finite-depth well is still present in the vicinity of the origin. Since we are in a 'classical' limit, tunelling of eigenvalues is not allowed, and a stable and sensible solution still exists. This holds as far as the finite-depth well at the origin is sufficiently deep to accommodate the eigenvalues, that is for $g > g_c$.

[... analogy with QM in semiclassical framework, ...]

[... interchange of large N and perturbative expansions is a nontrivial issue... For further details, see [8]...]

Consider the behaviour of the density $\rho(z)$ at the end-point $z \sim a$. For all $g > g_c$, we have $\rho(z) \sim \sqrt{a - z}$, in agreement with universality of edge behaviour. But for $g = g_c$, $a^2 = 8$, and we have instead

$$\rho(z) = \frac{1}{24\pi}(8 - z^2)^{3/2},$$

which implies the behaviour $\rho(z) \sim (a - z)^{3/2}$, $z \sim a = \sqrt{8}$. This is due to the fact that at $g = g_c$, the polynomial $Q(z)$ vanishes.

Note that this milder behaviour of $\rho(z)$ at the edge of the spectrum, see Fig. ..., can be interpreted euristically as the fact that we are at a limiting situation, beyond which the eigenvalue repulsion will start to overcome the confining potential, and escape outside the finite-depth potential well. [Note that, denoting by $\pm b$ the positions of the two maxima of the potential, we have, for any $g < 0$, $a < b = 1/\sqrt{-g}$].

The edge behaviour $\rho(z) \sim (a - z)^{3/2}$ is in fact observed in a wide variety of potentials, upon fine tuning of some appropriate parameter, and is referred to as a

'critical' universal behaviour. It is clear that polynomial potentials of higher order, upon suitable tuning of their parameters, can create double or higher order zeroes, and correspondingly, universal edge behaviours of the form $\rho(z) \sim (a-z)^{m+1/2}$, where $m+1$ is the order of the critical point, associated with a potential V of degree $2m+2$. While the exponent of the edge behaviour, directly related to the exponent of the singular behaviour of the free energy at $g \sim g_c$, $F^{(0)}(g) \sim (g-g_c)^{m+3/2}$, is universal, the actual value of g_c is not, as usual in critical phenomena.

It appears that the for any given value m, the corresponding matrix model describes the coupling to gravity of the minimal model $(2m+1, 2)$ of Conformal Field Theory[2], with central charge $c = 1 - 3(2m-1)/(2m+1)$. In particular, the critical point of the quartic model considered above describes the $(3, 2)$ minimal model of Conformal Field Theory, with central charge $c = 0$, known as the 'pure gravity' critical point.

sec.free_energy_i

6.8. Evaluation of the free energy (i). As an instructive example, let us evaluate the saddle-point (or 'genus zero', in jargon, see Section 7.7) free energy, defined as

$$F^{(0)}(g) := E^{(0)}(g) - E^{(0)}(0), \qquad E^{(0)}(g) := -\lim_{N\to\infty} \frac{1}{N^2} \log Z_N(g)$$

We thus need to evaluate

$$E^{(0)}(g) = \int_S \rho(\lambda) V(\lambda) \mathrm{d}\lambda - \int_S \int_S \rho(\lambda)\rho(\mu) \log|\lambda-\mu| \mathrm{d}\lambda \mathrm{d}\mu. \tag{6.19}$$

last

Integrating (6.6) with respect to z, between 0 and λ, we get

$$\frac{1}{2} V(\lambda) = \int_S \rho(\mu) \left[\log|\lambda-\mu| - \log|\mu|\right] \mathrm{d}\mu,$$

that we now plug into (6.19), obtaining

$$E^{(0)}(g) = \int_S \rho(\lambda) \left[\frac{1}{2} V(\lambda) - \log|\lambda|\right] \mathrm{d}\lambda.$$

In other words, in the saddle-point approximation, the free energy is completely determined by the eigenvalue density.

In the considered case of the quartic potential, we thus get

$$\begin{aligned} E^{(0)}(g) &= \int_0^a \left[\frac{1}{2\pi}(1 + \frac{g}{2}a^2 + gz^2)\sqrt{a^2-z^2}\right] \left(\frac{\lambda^2}{2} + \frac{g}{4}\lambda^4 - 2\log\lambda\right) \mathrm{d}\lambda \\ &= \frac{1}{384}(-a^4 + 40a^2 + 144) - \frac{1}{2}\ln\frac{a^2}{4}, \end{aligned}$$

where we have re-expressed g in terms of a, see (6.17). Substracting the $g = 0$ (or $a = 2$) contribution, we finally get for the free energy

$$F^{(0)}(g) = \frac{1}{384}(a^2-4)(36-a^2) - \frac{1}{2}\ln\frac{a^2}{4}, \tag{6.20}$$

free_quartic

where a is given by (6.18).

[2]We briefly recall that a particular family of Conformal Field Theories is constituted by the so-called minimal models, labelled by two coprime integers, p and q. The (p, q) minimal model has central charge $c = 1 - 6(p-q)^2/(pq)$.

sec.free_energy_ii

6.9. Evaluation of the free energy (ii). We present here an alternative method for the evaluation of the free energy in the saddle-point approximation. Clearly,

$$\frac{\partial}{\partial g}E^{(0)}(g) = -\lim_{N\to\infty}\frac{1}{N^2}\frac{\partial}{\partial g}\log Z_N(g) = \frac{1}{4}\langle\lambda^4\rangle_{sp};$$

recalling the relation between the moments and the large z expansion of the Green function (or of the resolvent, in the saddle-point approximation), see (4.4) and (4.5), we thus have

$$\frac{\partial}{\partial g}E^{(0)}(g) = \frac{a^4}{256}(12-a^2),$$

where the RHS is just (modulo a factor $1/4$) the coefficient of z^{-5} in the large z expansion of $W(z)$ as given by (6.16). Multiplying both sides by $\partial_{a^2}g$ we readily get

$$\partial_{a^2}E^{(0)}(g) = \frac{4}{3}\frac{a^2-8}{a^6}\frac{a^4}{256}(12-a^2) = -\frac{a^2}{192}+\frac{5}{48}-\frac{1}{2a^2},$$

and, integrating with respect to a^2,

$$F^{(0)}(g) = -\frac{a^4}{384}+\frac{5a^2}{48}-\frac{1}{2}\log a^2 + \text{const.};$$

fixing the integration constant through the condition $F^{(0)}(0)=0$, we easily recover (6.20).

The convenience in using one or the other of the two methods presented may depend on the situation.

6.10. Discussion. Let us comment on the expression (6.20) for the free energy of the model with potential $V(\lambda) = \frac{1}{2}\lambda^2+\frac{1}{4}g\lambda^4$. As a function of a^2, the free energy is analitic, except for the branch point of the logarithmic, $a^2=0$, corresponding to $g=\infty$. Recalling (6.18), we thus see that, as a function of g, the free energy is analytic for $g > -\frac{1}{12}$. In particular, it is analytic in the neighbourhood of $g=0$, and for (not too) negative values of g. The physical picture behind this result has already been discussed above.

We want here to investigate the behaviour of the free energy in a neighbourhood of its singularity at $g_c = -\frac{1}{12}$. Setting $\epsilon = 12(g-g_c)$, we thus write, from (6.18),

$$\begin{aligned} a^2 &= 8\frac{1-\sqrt{\epsilon}}{1-\epsilon} \\ &= 8[1-\epsilon^{1/2}+\epsilon-\epsilon^{3/2}+\epsilon^4+O(\epsilon^{5/2})] \end{aligned}$$

and replacing into (6.20), we get

$$F^{(0)}(g) = \text{const.} + \frac{1}{12}\epsilon - \frac{1}{8}\epsilon^2 + \frac{4}{15}\epsilon^{5/2} + O(\epsilon^3).$$

Thus, in the vicinity of g_c, the free energy of the model develops a singular behaviour

$$F^{(0)}(g) \sim (g-g_c)^{5/2},$$

in agreement with the discussion at the end of Section 6.7 (recall that we are considering here the case $m=1$). The same result will be recovered by means of a different technique later on.

Finally, we observe that the critical exponent $\frac{5}{2}$ gives rise to a divergence of the free-energy only at the level of its third derivative.

6.11. Example: the two-well quartic potential.

6.11.1. *The one-cut case.* Let us consider again the quartic potential, with parameters such that it is even, confining, and with two minima, that is $V(x) = \frac{1}{4}\lambda^4 - \frac{1}{2}\alpha\lambda^2$, with $\alpha > 0$. Let us consider the case of small α, for which we expect the density to be supported on a single interval $[-a, a]$. We have

$$W(z) = \frac{1}{2}\left[z^3 - \alpha z - M(z)\sqrt{z^2 - a^2}\right],$$

where $M(z)$ is a polynomial of degree two, with leading coefficient equal to one, to comply with the requested large $|z|$ behaviour, and vanishing odd-order coefficient due to the definite parity of the potential: $M(z) = z^2 + c$. Recalling that $\sqrt{1+x} \sim 1 + x/2 - x^2/8$, $x \sim 0$, and requiring condition (6.4), we get:

$$\frac{a^2}{2} = \alpha + c, \qquad a^4 + 4ca^2 - 16 = 0,$$

implying

$$a^2 = \frac{2}{3}\left(\alpha + \sqrt{\alpha^2 + 12}\right),$$

where the plus sign in front of the square root follows from the condition that $a^2 > 0$. Correspondingly, the density reads:

$$\rho(x) = -\frac{1}{2\pi\mathrm{i}}\left[W(z + \mathrm{i}0) - W(z + \mathrm{i}0)\right] = \frac{1}{2\pi}\left(x^2 + \frac{a^2}{2} - \alpha\right)\sqrt{a^2 - x^2}.$$

The condition that the density should be positive all over the interval $[-a, a]$ implies $\frac{1}{2}a^2 - \alpha \geq 0$, that is $\alpha \leq 2$.

subsec.quartic_twocut

6.11.2. *The two-cut case.* It is clear that for $\alpha = 2$, the density vanishes at the origin, and we expect that further increasing α, the support of the density should split into two distinct interval, $S = [-b, -a] \cup [a, b]$. We thus expect the resolvent to have two cuts, on these two intervals. We are led to look for a resolvent of the form

$$W(z) = \frac{1}{2}\left[z^3 - \alpha z - z\sqrt{(z^2 - a^2)(z^2 - b^2)}\right].$$

Requiring the asymptotic behaviour (6.4) we determine the end-points of the support of the density,

$$a = \sqrt{\alpha - 2}, \qquad b = \sqrt{\alpha + 2}.$$

Note that for $\alpha = 2$, $a = 0$, meaning that the two cuts are merging. When $\alpha < 2$, a becomes imaginary, in correspondence to the fact that the two wells are not deep enough to keep the eigenvalues separated, and one should turn back to the one-cut problem.

6.11.3. *The free energy.* The (first derivative of the) free energy of the model is easily computed by resorting to the method of Section (6.9); we have:

$$\partial_\alpha E^{(0)}(\alpha) = -\frac{1}{2}\langle\lambda^2\rangle_{sp}$$

We thus need to evaluate the coefficient of z^{-3} in the large z expansion of $W(z)$, obtaining

$$\partial_\alpha E_1^{(0)}(\alpha) = -\frac{1}{216}(\alpha + \sqrt{12+\alpha^2})(\alpha^2 + 24 + \alpha\sqrt{12+\alpha^2}) \qquad \alpha < 2,$$
$$\partial_\alpha E_2^{(0)}(\alpha) = -\frac{\alpha}{2} \qquad \alpha > 2, \tag{6.21}$$

deriv_FE

respectively, in the two phases. Note that the derivative of the free energy is continuous at the transition,

$$\partial_\alpha E_1^{(0)}(\alpha)\Big|_{\alpha=2^-} = \partial_\alpha E_2^{(0)}(\alpha)\Big|_{\alpha=2^+} = -1.$$

It is just a matter of calculation to verify that the same holds true for the second derivative as well. However, it is evident from expressions (6.21) that the third derivative vanishes for all values of the coupling $\alpha > 2$ while it is non vanishing for all values $\alpha < 2$. In other words, the splitting of the density support from one to two intervals gives rise to a third-order phase transition[3].

This phenomenon was first observed by Gross and Witten [9], and independently by Wadia [12], in a slightly different model, describing the large N limit of the two-dimensional $U(N)$ lattice gauge theory. This is a first example of a number of third-order transitions appearing in random matrix models.

6.12. The multicut case. As seen in the qualitative discussion of the electrostatic model, and more in detail in last example, in the case of a potential with several wells, the support of the density may be disconnected, consisting in the union of a finite numbers n of real intervals, $S = [a_1, b_1] \cup \cdots \cup [a_n, b_n]$. Correspondingly, the solution (6.5) for the resolvent in the saddle-point approximation can be written

$$W(z) = \frac{1}{2}\left[V'(z) - Q(z)\sqrt{\sigma(z)}\right] \qquad z \in \mathbb{C} \setminus S,$$

Wsol_multicut

where we have introduced the polynomial $\sigma(z) := (z-a_1)\dots(z-b_n)$, of degree $2n$. To compensate the large z behaviour of $V'(z)$, the polynomial $Q(z)$ must now be of degree $d-1-n$. In all we thus have $d+n$ unknown parameters, while the normalization condition (6.4) provides us with only $d+1$ conditions. Therefore in the multicut case, the saddle-point equation is not sufficient to determine $W(z)$ completely, and we need $n-1$ additional conditions.

Turning once again to the electrostatic analogy, the saddle-point equations essentially correspond to the requirement that the force on all particles inside a given well (or interval) vanishes; in terms of the free energy, this amounts to the request that the free energy inside each well is constant. However this does not prevent that such constant varies from one well to the other. Global equilibrium corresponds to the fact that this constant is the same in all n wells. This requests provided exactly the $n-1$ missing conditions.

[3]Actually, to complete the argument, one should also verify that the free energy is indeed continuous at the transition. This would required to evaluate the integration constant indipendently in both phases. This can actually be done, with some effort, for instance by resorting to the method of Section 6.8.

The above condition can be rephrased as the request that the total energy for moving a single eigenvalue from one well to another vanishes. The force acting on a single eigenvalue lying outside S is

$$f(z) \;=\; -V'(z) + \frac{2}{N}\sum_{j=1}^{N}\frac{1}{z-\lambda_j} \quad \overrightarrow{{}_{N\to\infty}} \quad -V'(z) + 2W(z), \qquad z \in \mathbb{C}\setminus S.$$

The energy to transport the eigenvalue from the jth to the $(j+1)$th well is simply

$$\begin{aligned}\Delta E_j &= \int_{b_j}^{a_{j+1}} f(z)\mathrm{d}z \\ &= \int_{b_j}^{a_{j+1}} \left[2W(z) - V'(z))\right]\mathrm{d}z \\ &= -\int_{b_j}^{a_{j+1}} Q(z)\sqrt{\sigma(z)}\,\mathrm{d}z.\end{aligned}$$

We thus have $n-1$ additional conditions,

$$\int_{b_j}^{a_{j+1}} Q(z)\sqrt{\sigma(z)}\,\mathrm{d}z = 0, \qquad j = 1,\dots,n-1. \tag{6.22}$$ additional_cond

which in principle allows to completely determine $W(z)$. The main problem is that actually solving these conditions may be technically involved. Indeed, in the two-cut case, the above conditions may be written in terms of Elliptic Integrals, while in the n-cut case, HyperElliptic integrals appears.

Note however that, in the two-cut case, in presence of an even potential, one may get rid of such technical complication: indeed, due to the parity symmetry of the problem, $Q(z)$ is an odd polynomial, and the condition (6.22) is automatically fulfilled; at the same time, the symmetry of the problem implies $a_1 = -b_2$ and $b_1 = -a_2$, so that the number of parameters to be determined is in actually smaller. Indeed this is the trick we used in the evaluation of the resovent in Section 6.11.2.

To conclude, note that

$$\begin{aligned}\phi_j :=& \frac{1}{4\pi\mathrm{i}}\int_{a_j}^{b_j} Q(z)\,\left[\sqrt{\sigma(z+\mathrm{i}0)} - \sqrt{\sigma(z-\mathrm{i}0)}\right]\,\mathrm{d}z \\ =& \frac{1}{2\pi\mathrm{i}}\oint_{C_{[a_j,b_j]}} W(z)\mathrm{d}z, \qquad\qquad j = 1,\dots,n,\end{aligned}$$

with $C_{[a_j,b_j]}$ a closed contour enclosing the interval $[a,b]$, is the fraction of eigenvalues lying in this interval. It is easily verified that $\sum_{j=1}^{n}\phi_j = 1$, as it should.

6.13. General solution of the saddle-point equation. The expression of the saddle-point equation (6.3) for the resolvent provided above holds only for a potential $V(x)$ of polynomial form. Under this hypothesis, we have obtained the general solution (6.5), which under the one-cut assumption becomes of the form (6.10).

We shall now discuss the linear form of the saddle-point equation, (6.8), for generic potential. The derivation is more formal, but applies to a much wider range of situations. For reasons which will become apparent below, we rewrite (6.8) as

$$W(x+\mathrm{i}0) + W(x-\mathrm{i}0) = U(x), \qquad x \in S, \tag{6.23}$$ SPElinear2

where S is the (still unknown) support of the density, consisting in general in the union of n disjoint intervals on the real axis, $S = \cup_{j=1}^{n}[a_{2j-1}, a_{2j}]$, with $a_1 < \cdots < a_{2n}$. We introduce the function

$$\sigma(z) = \prod_{j=1}^{2n}(z - a_j), \qquad z \in \mathbb{C},$$

sigma

and consider its square root, choosing the determination for which $\sqrt{\sigma(z)} > 0$ for $z > a_{2n}$. Correspondingly,

$$\sqrt{\sigma(x \pm \mathrm{i}0)} = (\pm \mathrm{i})^{2n-k}\sqrt{|\sigma(x)|}, \qquad x \in [a_k, a_{k+1}], \qquad k = 0, \dots, 2n,$$

where $a_0 = -\infty$, $a_{2n+1} = +\infty$.

Then the general solution of (6.23) is

$$W(z) = \frac{\sqrt{\sigma(z)}}{4\pi \mathrm{i}} \int_{\Gamma_S} \frac{1}{z-u}\frac{U(u)}{\sqrt{\sigma(u)}}\mathrm{d}u, \tag{6.24}$$

spesol

where Γ_S is a simple positive oriented contour enclosing the support S and no other singularity of the integrand. To verify that (6.24) is indeed the solution of (6.23), it is sufficient to rewrite the contour integral as an integral over S of the discontinuity of the integrand across its cuts,

$$W(z) = \frac{\sqrt{\sigma(z)}}{2\pi} \sum_{j=1}^{n}(-1)^{n-j} \int_{a_{2j-1}}^{a_{2j}} \frac{1}{z-u}\frac{U(u)}{\sqrt{|\sigma(u)|}}\mathrm{d}u, \qquad z \in \mathbb{C} \setminus S, \tag{6.25}$$

proofspesol

and than use the Cauchy identity, $\frac{1}{x \pm \mathrm{i}0} = PV\left(\frac{1}{x}\right) \mp \pi \mathrm{i}\delta(x)$.

Deforming the contour Γ_S, expression (6.24) for the resolvent may be rewritten

$$W(z) = \frac{1}{2}\left[U(z) - M(z)\sqrt{\sigma(z)}\right], \tag{6.26}$$

spesol2

with

$$M(z) := \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_U} \frac{1}{z-u}\frac{U(u)}{\sqrt{\sigma(u)}}\mathrm{d}u \tag{6.27}$$

spesol3

where Γ_U is a simple counterclockwise contour enclosing all and only the singularities of integrand that are induced by $U(u)$, including possibly the point at infinity.

Note that the solution (6.24) contains the $2n$ unknown parameters, $a_1, \dots, a_{2n}$. In the one-cut case, these are completely determined by the normalization condition (6.4). However, as seen in the previous Section, in the n-cut case the normalization condition is not sufficient, and we have to resort to the $n-1$ additional conditions:

$$\int_{a_{2j}}^{a_{2j+1}} M(x)\sqrt{\sigma(x)}\mathrm{d}x = 0, \qquad j = 1, \dots, n-1, \tag{6.28}$$

multicutcond

ensuring equilibrium not only within each potential wall, but also among different ones.

Remark: in some cases, it may be more convenient to evaluate directly (6.25), instead of using (6.26) and (6.27).

6.14. Some examples of one-cut problems. Let $S = [a, b]$, and $\sigma(z) = (z-a)(z-b)$. Let us consider first the case of a polynomial potential of degree d; we have

$$M(z) = \operatorname*{Res}_{u=\infty} \frac{1}{z-u} \frac{U(u)}{\sqrt{\sigma(u)}}.$$

onecutpoly

which is clearly a polynomial of degree $d-2$ in z, in agreement with the solution (6.10) previously obtained by other means. Note that the cut of the resolvent is just given by that of $\sqrt{\sigma(z)}$. Recalling (6.7), we have

$$\rho(z) = \frac{1}{2\pi} M(z) \sqrt{|\sigma(z)|}. \tag{6.29}$$

rhoresult

Note that in more general situations (non-polynomial potentials, saturated intervals, etc.) $U(u)$ is not anymore a polynomial and $M(z)$ may develop branch points. As a result, the expression for the density is not anymore given simply by (6.29).

As an example, `[...to be checked!...]` let us consider the potential $V(u) = (u-u_0)\log(u-u_0) - (u-u_0)$, $u > u_0$ (that is, we assume an 'hard wall' at $u = u_0$). We also assume that the support of the eigenvalue density consists of one interval $S = [a, b]$, with $b > a > u_0$. We thus have $U(u) = \log(u-u_0)$, and, from (6.27), for $z \in \mathbb{C} \setminus \{[-\infty, u_0] \cup [a, b]\}$,

onecutlog

$$\begin{aligned} M(z) &= \int_{-\infty}^{u_o} \frac{1}{z-x} \frac{1}{\sqrt{|\sigma(x)|}} \mathrm{d}x \\ &= \frac{1}{\sqrt{\sigma(z)}} \log \frac{(\sqrt{(z-a)(b-u_0)} - \sqrt{(z-b)(a-u_0)})(\sqrt{z-a} + \sqrt{z-b})}{(\sqrt{(z-a)(b-u_0-)} + \sqrt{(z-b)(a-u_0)})(\sqrt{z-a} - \sqrt{z-b})}, \end{aligned}$$

see `[Gr-Ryz: 2.252.2]` and `[Gr-Ryz: 2.261]`. In this case, the term $\sqrt{\sigma(z)}$ appearing in (6.26) is canceled by the denominator of $M(z)$, and the corresponding cut does not appear any more in the resolvent. Note however the additional cut induced by the term $U(z)$. The density is determined as usual by the discontinuity of $W(z) = \frac{1}{2}[\log(z-u_0) - M(z)\sqrt{\sigma(z)}]$ across the real axis. Explicit calculation shows that $\rho(x) \equiv 0$ outside the interval $x \in [a, b]$, and `[...to be checked!...]`

$$\rho(x) = \frac{2}{\pi} \arctan \frac{\sqrt{a(b-x)}}{\sqrt{b(a-x)}}, \qquad x \in [a, b];$$

thus, after all, the support of (the non trivial part of) the eigenvalue density still coincides with the initially imposed interval $S = [a, b]$.

6.15. The discrete log-gas. Let us consider a variant of our log-gas, no more directly related to any random matrix model. Such variant appears in many problems of statistical mechanics (growth process, tilings, dimer models, ...), combinatorics (asymptotics of Young diagrams, plane partitions, ...), and high energy physics (two-dimensional Yang-Mills theory, string theory, ...). The partition function of such models has a general expression of the form

$$Z_N \propto \sum_{m_1 \in \mathbb{Z}} \cdots \sum_{m_N \in \mathbb{Z}} \prod_{1 \le j < k \le N} (m_k - m_j)^2 \prod_{j=1}^{N} \mu_N(m_j). \tag{6.30}$$

RMM1

After rescaling $m_j \to N x_j$, sums can be reinterpreted as Riemann sums, and in the large N limit replaced by integrals, so that, as $N \to \infty$, we obtain again an expression of the form (6.1), (6.2), where now $V(x) = -\lim_{N\to\infty} \frac{1}{N} \log \mu_N(Nx)$.

The usual random matrix saddle-point analysis can be applied, provided that one imposes a suitable additional constraint to keep track of the discreteness of the m_j's, see [10]. In (6.30), all m_j's must be distinct, otherwise the Vandermonde determinant vanishes; therefore $|m_k - m_j| \geq 1$, for all $j \neq k$. In terms of the density $\rho(x)$ of the rescaled variables x_j, the constraint simply reads [10]:

$$\rho(x) \leq 1. \tag{6.31}$$ leqone

If $\rho(x) = 1$ (or $0 < \rho(x) < 1$, or $\rho(x) = 0$) $\forall x \in I$, then I is called a saturated interval (or a band, or a void, respectively).

In the case of a discrete log-gas, the saddle-point equation is still of the form (6.23). In absence of saturated intervals, $U(x)$ is obviously still equal to the derivative $V'(x)$ of the potential of the model. However in presence of a saturated interval $[\alpha, \beta]$, relation (6.7) implies the presence of a term of the form $\log[(z-\alpha)/(z-\beta)]$ in the resolvent. It is convenient to remove the logarithmic branch cut by introducing an auxiliary resolvent,

$$\tilde{W}(z) := W(z) - \log \frac{z-\alpha}{z-\beta}$$

still satisfying (6.23), now with $U(z)$ replaced by $\tilde{U}(z) = U(z) - 2\log \frac{z-\alpha}{z-\beta}$, while the support S is restricted to $\tilde{S} = S \setminus [\alpha, \beta]$. The solution is thus

$$W(z) = \log \frac{z-\alpha}{z-\beta} + \frac{\sqrt{\tilde{\sigma}(z)}}{4\pi \mathrm{i}} \int_{\Gamma_{\tilde{S}}} \frac{1}{z-u} \frac{\tilde{U}(u)}{\sqrt{\tilde{\sigma}(u)}} \mathrm{d}u, \tag{6.32}$$ spesoltilde

where $\tilde{\sigma}(z)$ is such that its square root has cuts in correspondence of $\tilde{S}$. Note that, even in the case of compact support of the density, the presence of saturation intervals may lead to a multicut situation.

6.16. Douglas-Kazakov phase transition. Let us consider in particular a quadratic potential, $V(x) = \frac{1}{2}\alpha x^2$, that is $U_0(x) = \alpha x$, with $\alpha > 0$. Clearly, for sufficiently small α, the eigenvalues will tend to spread on a wider interval, without ever saturating the constraint (6.31). The saddle-point solution for the resolvent will be exactly the one already computed above for the harmonic potential, and the density will be again given by Wigner's semicircle.

However, if we increase α, at some stage we expect that at the center of the support, near the origin, the density $\rho(x)$ becomes larger than one. Indeed, recalling that Wigner's semicircle density reads $\rho(x) = \frac{\alpha}{2\pi}\sqrt{4/\alpha - x^2}$, we easily find that this happens for $\alpha = \alpha_c = \pi^2$. For $\alpha > \alpha_c$, our solution is not valid anymore, and we have a new scenario, with the density saturating the constraint (6.31) on some central interval $[-a, a]$, with two unsaturated intervals $[-b, -a]$ and $[a, b]$ (note that we have exploited the parity of the potential). We thus have $\tilde{S} = [-b, -a] \cup [a, b]$, and $\tilde{\sigma}(z) = (z^2 - a^2)(z^2 - b^2)$. We can again resort to (6.24), or directly to (6.26) and (6.27).

Note that the considered potential is increasing mildly enough as large $|z|$, so that it does not give rise to a residue at infinity, when calculating $M(z)$. We also assume the presence of a saturation interval $[-a, a]$. The auxiliary resolvent $\tilde{W}(z) := W(z) - \log \frac{z+a}{z-a}$, is given by (6.32), with $\tilde{U}(z) = \alpha z - 2\log \frac{z+a}{z-a}$ and

$\tilde{S} = [-b, -a] \cup [a, b]$. We get[4]

$$\tilde{W}(z) = \frac{1}{2}\alpha z - \log\frac{z+a}{z-a} - \frac{1}{2}M(z)\sqrt{\tilde{\sigma}(z)},$$

with, from (6.27),

$$M(z) = -\frac{1}{\pi \mathrm{i}}\oint_{\Gamma_U}\frac{1}{z-u}\frac{1}{\sqrt{\tilde{\sigma}(u)}}\log\frac{u+a}{u-a} = 2\int_{-a}^{a}\frac{1}{z-u}\frac{1}{\sqrt{|\tilde{\sigma}(u)|}},$$

leading to

$$W(z) = \frac{1}{2}\alpha z - \sqrt{\sigma(z)}\int_{-a}^{a}\frac{1}{z-u}\frac{1}{\sqrt{|\tilde{\sigma}(u)|}}. \tag{6.33}$$ WDK

Requiring the correct large z behaviour gives the two conditions:

$$\int_{-a}^{a}\frac{1}{\sqrt{(a^2-u^2)(b^2-u^2)}}\mathrm{d}u - \frac{\alpha}{2} = 0,$$
$$\frac{a^2+b^2}{2}\int_{-a}^{a}\frac{1}{\sqrt{(a^2-u^2)(b^2-u^2)}}\mathrm{d}u - \int_{-a}^{a}\frac{u^2}{\sqrt{(a^2-u^2)(b^2-u^2)}}\mathrm{d}u = 1.$$

Resorting to [Gr-Ryz: 3.152.7] and [Gr-Ryz: 3.153.5], these may be rewritten as

$$\frac{2}{b}K(a/b) - \frac{\alpha}{2} = 0,$$
$$\frac{a^2+b^2}{b}K(a/b) - 2b\,[K(a/b) - E(a/b)] = 1$$

where $E(k)$ and $K(k)$ are the complete elliptic integrals of the first and second kind, respectively, and the elliptic modulus has been set to $k = a/b$. This reproduces the result of [10]. Note that here again, the parity of the potential guarantees the equality of the filling fractions, and the problem is determined without resorting to condition (6.28)

The free energy of the model may be evaluated, for instance by resorting to the method of Section 6.9. We have

$$\partial_\alpha E^{(0)}(\alpha) = \frac{1}{2}\langle\lambda^2\rangle_{sp}.$$

We thus need just to evaluate the coefficient of z^{-3} in the large z expansion of $W(z)$, given by (6.11), with $g \to \alpha$, when $\alpha < \alpha_c$, and by (6.33) when $\alpha > \alpha_c$. It appears that, as a function of α, the free energy is continuous at $\alpha_c = \pi^2$, but it exhibits a discontinuity in its third derivative. This third-order phase transition was first observed by Douglas and Kazakov in the context of the large N limit of continuous Yang-Mills theory in two-dimensions [10].

7. Random matrices and discretized surfaces

7.1. Preliminaries. As explained above, random matrix models were originally introduced as random Hamiltonians. In this section we shall discuss a completely different interpretation, which emerges when regarding the partition function Z_N as that of a zero-dimensional quantum field theory. In the calculation of the matrix integral in terms of Feynman diagrams, the large N expansion leads to an ordering of these diagrams in terms of their topological character. Since

[4]We outline the presence of a recurrent bad in paper [10]. In Eqs. (30) and (31), the sign of the log should be +; in Eq. (34) the sign of the h^{-1} should be +; and maybe others.

Feynman diagrams are in bijections with triangulated surfaces, Z_N can be reinterpreted as describing a theory of random triangulated surfaces, ordered according to their topological properties. Such interpretation has found fruitful applications in the statistical mechanics of lattice models on random graphs, in two-dimensional quantum gravity, and in string theory.

Consider our usual matrix integral, in the $\beta = 2$ case, that is M is an $N \times N$ Hermitean matrix,

$$Z_N = \int \mathcal{D}M \mathrm{e}^{-N\mathrm{Tr}[V(M)]},$$

with

$$\begin{aligned} V(M) &= \frac{g}{2}M^2 + \frac{g_3}{3}M^3 + \cdots + \frac{g_n}{n}M^n \\ &= \frac{g}{2}M^2 + \delta V. \end{aligned}$$

We now regard Z_N as the path-integral partition function of a zero-dimensional quantum field theory. Being unable to compute the integral exactly, we resort to the usual perturbative approach, viewing δV as a perturbation of the Gaussian integral (i.e., free theory), and performing a power series expansion of $\exp\{-N\mathrm{Tr}[V(M)]\}$:

$$\begin{aligned} Z_N(g_3, \ldots, g_n) &= \int \mathcal{D}M \mathrm{e}^{-N\mathrm{Tr}[M^2]} \left[1 - N\mathrm{Tr}[\delta V] + \frac{N^2}{2!}(\mathrm{Tr}[\delta V])^2 + \ldots\right] \quad (7.1) \\ &= Z_N(0, \ldots, 0) \left[1 - N \left\langle \mathrm{Tr}[\delta V] \right\rangle_0 + \frac{N^2}{2!} \left\langle (\mathrm{Tr}[\delta V])^2 \right\rangle_0 + \ldots\right] \end{aligned}$$

Z_pert

where we have used the obvious notation $\langle \ldots \rangle_0$ for the Gaussian average.

7.2. Wick's theorem and Feynman diagrams. Before proceeding, it is useful to recall the evaluation of Gaussian averages. Consider a generic N dimensional Gaussian measure

$$\frac{1}{\sqrt{(2\pi)^N \det A}} \mathrm{e}^{-\frac{1}{2}x_\alpha A_{\alpha\beta} x_\beta} \mathrm{d}^N x,$$

that we have normalized to one for convenience (note that repeated indices are suumed over). According to a well-know theorem, a generic Gaussian average can be expressed as sum of products of second moments:

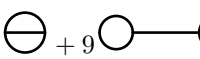
$$\begin{aligned} \langle x_{\alpha_1} \ldots x_{\alpha_{2r}} \rangle &= \frac{1}{\sqrt{(2\pi)^N \det A}} \int x_{\alpha_1} \ldots x_{\alpha_{2r}} \mathrm{e}^{-\frac{1}{2}x_\alpha A_{\alpha\beta} x_\beta} \mathrm{d}^N x \\ &= \sum_{p \in P_{2r}} A^{-1}_{\alpha_{p(1)}\alpha_{p(2)}} \ldots A^{-1}_{\alpha_{p(2r-1)}\alpha_{p(2r)}}, \end{aligned}$$

where P_{2r} denotes the set of all $(2r-1)!!$ distinct pairings of numbers $1, \ldots, 2r$. Note that with each pairing between x_α and x_β is associated a matrix element $A^{-1}_{\alpha\beta} = \langle x_\alpha x_\beta \rangle$, that is, a 'propagator'. In the context of perturbative quantum field theory this is essentially Wick's theorem. For instance, for $r = 2$, one has $\langle x_1 x_2 x_3 x_4 \rangle = \langle x_1 x_2 \rangle \langle x_3 x_4 \rangle + \langle x_1 x_3 \rangle \langle x_2 x_4 \rangle + \langle x_1 x_4 \rangle \langle x_2 x_3 \rangle$. Note further that in the simple case $A_{\alpha\beta} = g\delta_{\alpha\beta}$, the propagator reduces to $(1/g)\delta_{\alpha\beta}$.

For illustration, consider the 'vacuum' perturbative expansion for the case of the cubic theory, $\left\langle \exp[(g_3/3)\sum_\alpha x_\alpha^3] \right\rangle$. The first order vanishes (odd number of

fields); at second order, we have:

$$\frac{1}{2!}\left(\frac{g_3}{3}\right)^2 \sum_{\alpha,\beta=1}^{N} \langle x_\alpha^3 x_\beta^3 \rangle = \frac{1}{2!}\left(\frac{g_3}{3}\right)^2 \frac{1}{(g)^3} \left[6 \; + 9 \; \right] \tag{7.2}$$



QFTcubic

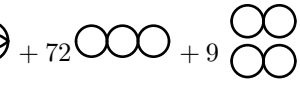
Similarly, for the quartic theory, $\left\langle \exp[(g_4/4)\sum_\alpha x_\alpha^4] \right\rangle$, at first order we have

$$\left(\frac{g_4}{4}\right) \sum_{\alpha=1}^{N} \langle x_\alpha^4 \rangle = \left(\frac{g_4}{4}\right) \frac{1}{(g)^2} \left[3 \; \right], \tag{7.3}$$

QFTquartic1

while at second order,

$$\frac{1}{2!}\left(\frac{g_4}{4}\right)^2 \sum_{\alpha,\beta=1}^{N} \langle x_\alpha^4 x_\beta^4 \rangle = \frac{1}{2!}\left(\frac{g_4}{4}\right)^2 \frac{1}{(g)^4} \left[24 \; + 72 \; + 9 \; \right].$$



QFTquartic2

The Feynman rules for this simple theory can be summarized as follows: at some given order, say n,

- draw all possible vacuum Feyman diagrams of order n, that is all possible graphs with n_3 cubic vertices, ..., n_k k-vertices, $n_3 + \cdots + n_k = n$ (and no vertices of order one or two);
- with each line associate a factor $1/g$;
- with each k-vertex associate a factor g_k/k;
- each diagram has a given symmetry factor Ω, associated with the number of different pairings giving rise to it.

7.3. Feynman rules for Hermitean matrices. We now treat the case of Hermitean matrices fields, where some additional features arise, in relation to the facts that *i)* there are two indices; *ii)* the field is complex. Coming back to the expression (7.1) for Z_N, we have for the propagator the expression

$$\begin{aligned} \langle M_{ij} M_{kl} \rangle_0 &:= \int \mathcal{D}M M_{ij} M_{kl} \, \mathrm{e}^{-\frac{1}{2} N g M_{\alpha\beta} M^\dagger_{\beta\alpha}} \\ &= \frac{1}{Ng} \delta_{il} \delta_{jk} \end{aligned}$$

To keep track of the two indices of the matrix, it is convenient to represent the propagator graphically as a double line, see Fig. 4. Since M is Hermitean, $M_{ij} = M^*_{ji} \neq M_{ji}$, and indices can not be permuted, we have to keep track of their position (as first or second indices). Graphically, this can be done by always orienting the lines of the propagator, for instance, from first indices to second ones. The meaning of the arrows is the usual one for complex fields, being associated with a conserved charge, or 'colour', in the language of Yang-Mills gauge field.

In the case of symmetric matrices (real entries, neutral fields, corresponding to the case $\beta = 1$), we would have instead a double line propagator with no arrows, in relation to the fact that the indices can be interchanged.

Turning now to the vertices, and considering for definiteness to a quartic interaction $-N(g_4/4)\mathrm{Tr}[M^4]$, we can write it as:

$$-N(g_4/4) M_{ij} M_{kl} M_{mn} M_{pq} \delta_{jk} \delta_{lm} \delta_{np} \delta_{qi},$$

where sum over reapeated indices is understood. We thus associate to each quartic vertex a terms $-N(g_4/4)\delta_{jk}\delta_{lm}\delta_{np}\delta_{qi}$, that we may represent graphically as in Fig. 4. Again, the index structure is here represented by oriented lines. More generally,

each monome $-N(g_k/k)\mathrm{Tr}M^k$ gives rise to a diagram with k external legs, one for each matrix element in the trace.

Finally, we have to connect the lines of the vertices with propagators, respecting the direction of arrows. Connecting lines means setting their indices equal (this is imposed by the Kronecker's δ's in the propagators), and sum over them. Since we are considering only 'vacuum' diagrams, there are no external lines, and all lines are closed. Each closed thin line correspond to a sum of the form

$$\sum_{\alpha_1,\dots\alpha_n=1}^{N}\delta_{\alpha_1\alpha_2}\delta_{\alpha_2\alpha_3}\dots\delta_{\alpha_n\alpha_1}=\sum_{\alpha_1=1}^{N}\delta_{\alpha_1\alpha_1},$$

that obviously sums up to N.

Summarizing, for matrix models, we have the following Feyman rules: at some given perturbative order v, draw all possbile diagrams made of v vertices. Then for each diagram:

- assign a weight $-Ng_k/k$ to each vertex of type k
- assign a weight $1/(Ng)$ to each propagator
- assign a term N to each thin closed line.

We now have all necessary tools to compute quantities such has

$$Z_N=\sum_{\{v_k\}}\left\langle\prod_k\frac{1}{v_k!}\left(-N\frac{g_k}{k}\mathrm{Tr}\left[M^k\right]\right)^{v_k}\right\rangle_0$$

Each term in the sum is expanded according to Wick theorem, and corresponds to all fat diagrams with v_k k-vertices.

Each of such diagrams contribute with:

$$\frac{1}{\Omega}\prod_k(-Ng_k)^{v_k}\left(\frac{1}{Ng}\right)^p N^l\sim N^{v-p+l},$$

where Ω is a combinatorial factor, including the various $v_k!$, k^{v_k}, and the number of different ways of connecting vertex legs to obtain the given diagram; $v=\sum_k v_k$, $p=(\sum_k kv_k)/2$ and l are the total numbers of vertices, of fat propagator, and of thin closed lines, respectively.

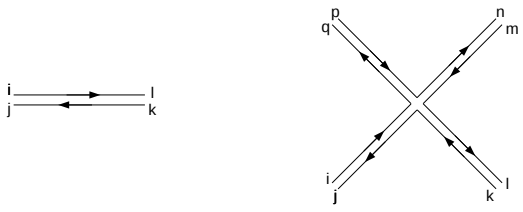


FIGURE 4. Graphical representation for the propagator (left), and for the quartic vertex (right), of Hermitean matrix fields. fatgraphs

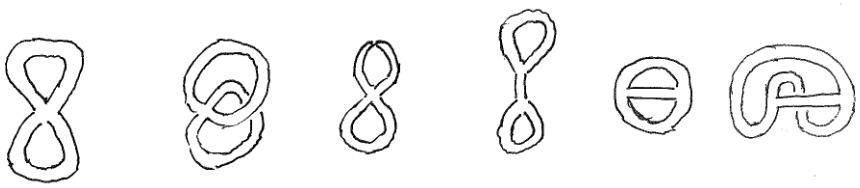


FIGURE 5. The diagrams $d_1, \ldots, d_6$ considered in the text. fatgraphs_2

7.4. Examples. Consider the first perturbative order, in the case of a matrix model with quartic interaction; we have

$$\begin{aligned} -N\frac{g_4}{4}\left\langle \mathrm{Tr}\left[M^4\right]\right\rangle_0 &= -\frac{Ng_4}{4}\frac{1}{(Ng)^2}\left[2d_1+d_2\right] \\ &\sim \frac{1}{N}\left[2N^3+N\right] \\ &\sim 2N^2+N^0, \end{aligned}$$

where d_1 and d_2 are the diagrams drawn in Fig. 5. It is easily verified that indeed the quantity $v-p+l$ takes the values 2 and 0, respectively. Note also that in the case of scalar/vector field theory, diagrams d_1 and d_2 coincide, with a total coefficient of 2+1=3, see (7.3).

When considering symmetric (rather than Hermitean) matrices one more diagram d_3, see Fig. 5, would contribute, which is however incompatible with the orientation of arrows in the case of Hermitean matrices. Note further that this diagrams has two thin closed lines, thus giving a factor N^2, and a combinatorial factor equal to two. In the case of symmetric matrix we would thus have an additional term equal to $2N$ contributing to the last line of the above calculation.

Consider now the case of a matrix model with cubic interaction. Clearly, there are no vacuum diagrams at first perturbative order. At second order we have:

$$\begin{aligned} \frac{1}{2!}\left(\frac{-Ng_3}{3}\right)^2\left\langle \mathrm{Tr}\left[M^3\right]\mathrm{Tr}\left[M^3\right]\right\rangle_0 &= \frac{1}{2!}\left(\frac{-Ng_3}{3}\right)^2\frac{1}{(Ng)^3}\left[9d_4+3d_5+3d_6\right] \\ &\sim \frac{1}{N}\left[9N^3+3N^3+3N\right] \\ &\sim 12N^2+3N^0, \end{aligned}$$

where d_4, d_5, and d_6 are drawn in Fig. 5. Again, it is easily verified that indeed the quantity $v-p+l$ takes the values 2, 2, and 0, respectively. Note that the coefficient 9, 3, 3 should be compared to the coefficients 9, 6 that one would have in the scalar or vector fields, were the diagrams d_5, and d_6 coincide, see (7.2).

7.5. Large N expansion. From the examples provided above, it is clear that in the large N limit the perturbative series is reordered in powers of N^{v-p+l}, with leading contribution coming from all diagrams with $v-p+l=2$, that turns out to be 'planar' (in a sense that will be made more precise in next Section) while other diagrams are arranged according to some topological features, still to be understood, and labeled by the value of $v-p+l$.

Moreover we note that oriented diagrams come with even powers of N while unoriented ones comes with odd powers (and can appear only in symmetric rather than Hermitean matrix models).

We conclude by observing that if we modify some diagrams without altering its topology, the corresponding power of N is unaltered. Consider for definiteness a cubic matrix model: altering a generic diagrams by simply adding somewhere a 'fat' propagators, as follows,

it is easily seen that the number of vertices is increased by 2, the number of propagator by 3, and that of closed thin lines (loops) by 1. Thus the quantity $v - p + l$ is indeed unchanged, and is a genuine topological invariant.

[...]

7.6. Duality between fat diagrams and surfaces. Fat diagrams has a natural two-dimensional structure: propagators are ribbons, and we can try to draw diagrams on two-dimensional surfaces. We have already observed that some diagrams are planar (in the sense they can be drawn on a plane), and some are not, in the sense that they need at least a torus to be drawn on, see for instance d_2 and d_6 in Fig. 5. As for diagram d_3 in Fig. 5, it can not be drawn even on a torus, actually it is non-orientable (just like the famous Moebius strip).

Actually the correspondence between fat diagrams and surface can be much more precise, indeed we shall now associate to each fat diagram its dual graph, which can in turn be viewed as a polygonal surface: associate to each k-valent vertex of the fat diagram a $k - gone$, and identify pairs of edges of such k-gones whenever the corresponding legs of the k-valent vertices of the original fat diagram were connected by a propagator. Note that according to this correspondence, each closed thin line of the original fat diagram is actually mapped to a vertex of the polygonal surface

Summarizing, we have the following dictionary:

fat diagram	polygonal surface
k-vertex	k-gone
propagator	edge
loop	vertex
# vertices v	# faces F
# propagator p	# edges E
# loops l	# vertices V
Ω (symmetry factor)	# of automorphisms

In all, it appears that we can write the partition function Z_N as a sum over all possible k-gonal surfaces, each with weight $\frac{1}{\Omega}\left(\frac{1}{g}\right)^E (-g_k)^F N^{V-E+F}$ This is however not completely true, because the duality is exact only for connected diagrams. But we know that to get rid of disconnected ones, we just need to take the logarithm.

We thus have:

$$\log Z_N = \sum_{\substack{\text{connected}\\ \text{vacuum diagrams}}} \frac{1}{\Omega}\left(\frac{1}{g}\right)^E (-g_k)^F N^{v-p+l}$$

$$= \sum_{\substack{\text{closed}\\ k-\text{gonal surfaces}}} \frac{1}{\Omega}\left(\frac{1}{g}\right)^E (-g_k)^F N^{V-E+F}$$

This can be interpreted as the grand-canonical partition function for a set of random polygonal surfaces, with fugacity $-g_k$ for k-gones, and binding energy $\ln g$ between them. The power of N will be clarified in next Section, but let us first make a few commnens.

As a matter of principle fugacity should be positive, implying negative couplings $g_k < 0$, and correspondingly, for our random matrix model, a potential that is unbounded from below, and an ill-defined matrix integral. But we have already seen how this problem can be cured in the large N limit: the positive g_k solution is well-defined, and can be analytically continued to negative values of g_k (although not too far from the origin), as long as the potential maintain a sufficiently deep local minimum. This defines a critical value for g_k, see below for details.

Finally, note that, at large N, discretized surfaces can be viewed as a good approximation to continuous ones. Thus in the large N limit we are actually summing over two-dimensional surfaces, that is making a path integral for two-dimensional euclidean gravity. The factor $1/\Omega$ becomes in this limit the suface diffeomorphism group.

topology

7.7. The Euler characteristic. [...]

7.8. Topological expansion. [...]

7.9. Multimatrix models. Let us consider the following two-matrix model:

$$Z_N = \int \mathcal{D}M_1 \mathcal{D}M_2 \mathrm{e}^{-N\mathrm{Tr}\left[\frac{1}{2}M_1^2+\frac{1}{2}M_2^2+g_1M_1^3+g_2M_2^3-cM_1M_2\right]} \tag{7.4}$$

partition_ising

with M_1, M_2 Hermitean $N \times N$ matrices. We have two different three-leg vertices and, inverting the quadratic form

$$\begin{pmatrix} 1 & -c \\ -c & 1 \end{pmatrix}^{-1} = \frac{1}{1-c^2}\begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix},$$

three different propagators, connecting two vertices of type 1, or two of type 2, or finally two of different type, see Fig. 6. Thus the dual discretized surfaces associated to the Feynman diagrams are made of white (g_1) and black (g_2) triangles, that we can reinterpret as spins, $\sigma = \pm 1$, respectively.

.

Each Feynman diagram has associated a weight

$$\frac{1}{\Omega} g_1^{n_+} g_2^{n_-} c^{p_{+-}} \frac{1}{(1-c^2)^p}$$

where we have ignored the factor N^χ, and:

- n_σ is the number of triangles with spin σ;

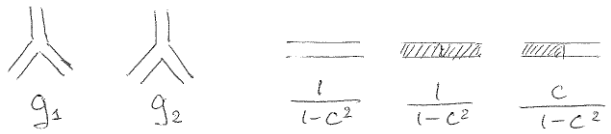


FIGURE 6. The two vertices and the three propagators associated to the random matrix model in (7.4)

ising

- p_{+-} is the total number of pairs of adjacent spins with opposite orientation;
- p is the total number of pairs of adjacent spins.

Simple considerations leads to:

$$2p = 3(n_+ + n_-), \qquad \sum_{\langle ij \rangle} \sigma_i \sigma_j = p - 2p_{+-}, \qquad \sum_i \sigma_i = n_+ - n_-.$$

Setting now $n := n_+ + n_-$, we may rewrite the weight associated to each diagram as:

$$\begin{aligned} \frac{1}{\Omega} (g_1 g_2)^{p/3} \left(\frac{g_1}{g_2}\right)^{\frac{1}{2}\sum_i \sigma_i} c^{\frac{1}{2}p - \frac{1}{2}\sum_{\langle ij \rangle} \sigma_i \sigma_j} \frac{1}{(1-c^2)^{\frac{3}{2}n}} \\ = \frac{1}{\Omega} \left(\frac{\sqrt[3]{g_1 g_2}\sqrt{c}}{1-c^2}\right)^{\frac{3}{2}n} \left(\frac{g_1}{g_2}\right)^{\frac{1}{2}\sum_i \sigma_i} c^{-\frac{1}{2}\sum_{\langle ij \rangle} \sigma_i \sigma_j} \\ = \frac{1}{\Omega} \mathrm{e}^{\Lambda n} \mathrm{e}^{-J\sum_{\langle ij \rangle} \sigma_i \sigma_j} \mathrm{e}^{-h\frac{1}{2}\sum_i \sigma_i} \end{aligned}$$

where in the last line we have introduces the constants

$$\Lambda = \left(\frac{\sqrt[3]{g_1 g_2}\sqrt{c}}{1-c^2}\right)^{\frac{3}{2}}, \qquad J = \ln c, \qquad h = \log\sqrt{\frac{g_2}{g_1}}.$$

The constant J and h are immediately interpreted as the coupling constant and the external magnetic field for an Ising model on a random triangular lattice. The quantity Λ is the chemical potential associated to the addition of a triangular plaquette to the random triangular lattice, that is, in gravitation theory, the 'cosmological constant' coupled to the total area of our two-dimensional discretized surface. We have already seen in previous Sections that the random matrix partition function has a natural interpretation as a sum of all possible discretized surfaces. Our two-matrix model has thus a natural interpretation as an Ising model coupled to two-dimensional euclidean gravity. It has a critical point, in a sense discussed below, at some value $J = J_c$, $h = 0$, where it is described by the $(4, 3)$ minimal model of Conformal Field Theory, coupled to gravity.

Playing with the order of the interaction (3 in the above example), and with the number of different matrices in the model (2 in the above example) one can actually describe all minimal models of Conformal Field Theory, coupled to gravity.

[...]

8. The orthogonal polynomials method

8.1. Preliminaries. Saddle-point method is very elegant and powerful: its direct relation with the classical electrostatic analogy is quite helpful in guiding our intuition, and the technique allows for evaluation of many relevant quantities. However the approach has its limitations, the main one being the impossibility to compute the subleading corrections to the large N leading term.

We now turn to a more powerful approach, based on orthogonal polynomials, providing a systematic procedure to evaluate subleading corrections, and to obtain non trivial results on the behaviour of correlation functions.

Again we shall limit ourselves to Hermitean ensembles, $\beta = 2$, for the sake of simplicity, but the approach can be applied to other ensembles as well, at the price of some additional technicalities, that we shall not treat here.

To start with, let us briefly recall a well known fact. Consider a real nonnegative weight function $\mu(x)$, $x \in \mathbb{R}$, vanishing faster than any polynomial as $|x| \to \infty$, so that all its moments are finite,

$$c_n := \int_{-\infty}^{+\infty} x^n \mu(x) \mathrm{d}x, \qquad n = 0, 1, \dots .$$

Given such a function $\mu(x)$, it is alway possible, by means of the so-called Gram procedure, to build the associated system of real orthogonal polynomials, that is, $\{P_n(x)\}_{n=0}^{\infty}$, with $\deg P_n(x) = n$, such that

$$\int_{-\infty}^{+\infty} P_n(x) P_m(x) \mu(x) \mathrm{d}x = h_n \delta_{nm}. \tag{8.1}$$

orthog

Note that at this stage the squared norms h_n are still to be determined. The $P_n(x)$'s are instead completely determined, modulo a normalization. Once this is fixed, the h_n can be determined as well. For instance, one can choose the $P_n(x)$ to be orthonormal, that is $h_n = 1$, $n = 0, 1, \dots$. In the present context it is convenient to choose the $P_n(x)$ to be *monic*, that is, with leading coefficient equal to one, $P_n(x) = x^n + \dots$; the corresponding values for the squared norms h_n can be determined consistently.

Such a set of orthogonal polynomials is known to form a complete basis for $L^2(\mathbb{R})$, the Hilbert space of square integrable function, and to satisfy a three-term recurrence relation. Note that, for simplicity, we are considering polynomials defined on the whole real axis, but all this can be extended with minor modifications to the case of polynomials and weight functions defined on any interval $I \subseteq \mathbb{R}$, or on some contour in the complex plane, or even on a discrete set of points.

8.2. The main idea. Let us consider our usual Hermitean matrix model integral,

$$Z_N = \int_{-\infty}^{+\infty} \prod_{\substack{j,k=1 \\ j<k}}^{N} |\lambda_k - \lambda_j|^2 \, \mathrm{e}^{-N \sum_{j=1}^{N} V(\lambda_j)} \mathrm{d}^N \lambda.$$

We recall that the double product can be written as a Vandermonde determinant,

$$\Delta(\lambda_1, \dots, \lambda_N) := \det \left[\lambda_j^{k-1} \right]_{j,k=1}^{N} = \prod_{1 \le j < k \le N} (\lambda_k - \lambda_j).$$

Note that we can always add to a given column any linear combination of other columns without changing the determinant. This means that we can always write

$$\Delta(\lambda_1,\dots,\lambda_N)=\det\left[P_{k-1}(\lambda_j)\right]_{j,k=1}^{N}, \tag{8.2}$$ defdet0

where the $P_n(\lambda)$'s, $n=0,1,\dots$ are an totally arbitrary family of monic polynomials. Note that at this stage we do not even require them to be orthogonal.

From the definition of the determinant, we have

$$\Delta(\lambda_1,\dots,\lambda_N)=\sum_{\sigma\in S_N}(-1)^{[\sigma]}\prod_{j=1}^{N}P_{\sigma_j-1}(\lambda_j), \tag{8.3}$$ defdet

where the sum is taken over the permutations $\sigma:\{1,\dots,s\}\mapsto\{\sigma_1,\dots,\sigma_s\}$, i.e. over the elements of the symmetric group S_N, with $[\sigma]$ denoting the parity of σ. We may thus write

$$Z_N=\sum_{\sigma,\tau\in S_N}(-1)^{[\sigma]}(-1)^{[\tau]}\prod_{j=1}^{N}\int_{-\infty}^{+\infty}P_{\sigma_j-1}(\lambda_j)P_{\tau_j-1}(\lambda_j)\mathrm{e}^{-NV(\lambda_j)}\mathrm{d}\lambda_j.$$ ZNdoubledet

Let us now choose the $P_n(x)$'s to be a complete set of monic orthogonal polynomials associated to the measure $\exp[-NV(x)]$. Then,

$$Z_N=\sum_{\sigma,\tau\in S_N}(-1)^{[\sigma]}(-1)^{[\tau]}\delta_{\sigma_j\tau_j}\cdot\prod_{j=1}^{N}h_{\sigma_j-1}=N!\prod_{j=0}^{N-1}h_j. \tag{8.4}$$ ZNpoly

Thus the evaluation of the matrix integral is reduced to the standard problem of constructing the complete set of orthogonal polynomials associated to a given measure.

8.3. Slater determinant and free-fermion picture. In such orthogonal polynomials framework, Z_N is essentially the square norm of the ground state wave function of a sytem of N non-interacting fermions confined in some potential well. Indeed, let us introduce the normalized one-particle wave function

$$\psi_n(\lambda)=\frac{1}{\sqrt{h_n}}P_n(\lambda)\mathrm{e}^{-\frac{N}{2}V(\lambda)},\qquad n=0,1,\dots, \tag{8.5}$$ ff

corresponding to the nth discrete level for a single particle in some potential well. Note that the potential well is not that of our original matrix model, but rather the one whose ground state wave function is $\propto\exp[-NV(\lambda)/2]$. Since we are considering bound states, we can always choose our wave functions $\psi_n(\lambda)$ to be real valued.

Consider now a system of N non-interacting fermions in the same potential well. The N-particle ground state is constructed by placing the N fermions in the N lowest levels, one fermion per level. The N fermions constitute a Fermi sea, and the energy of the state $N-1$ is the Fermi level. The corresponding wavefunction must be totally antisymmetric in the N particle coordinates, and this is realized with the so-called Slater determinant:

$$\Psi_N(\lambda_1,\dots,\lambda_N)=\det\left[\psi_{j-1}(\lambda_k)\right]_{j,k=1}^{N},$$

whose squared norm is simply $N!$.

Thus, recalling (8.2), it is easily seen that we may write

$$Z_N = \prod_{j=0}^{N-1} h_j \int_{-\infty}^{\infty} |\Psi_N(\lambda_1, \dots, \lambda_N)|^2 \, \mathrm{d}^N\lambda.$$

This expression is somewhat trivial, in the sense that it adds nothing to what we already know from expression (8.4) for the partition function. However this interpretation gives useful insight when turning to correlation functions. In particular, the presence of a Fermi sea implies that in the large N limit, most properties of our matrix models will be controlled essentially by the behaviour in the vicinity of the Fermi level, that is by the properties of those P_n's with $n \approx N$.

We emphasize that, at variance with the electrostatic picture, here Z_N is not anymore viewed as the partition function of a classical system, but rather as the (square of the) ground state N-body wave function for a system of non-interacting fermions, subjected only to the exclusion principle (implemented by the Vandermonde). The Coulomb repulsion of the electrostatic picture is here interpreted as a purely quantum effect, the exchange interaction.

8.4. The N-point correlation function. The orthogonal polynomial framework is very efficient to treat correlation functions. To fix notations, from now on $P_n(x)$, $n = 0, 1, \dots$, are monic orthogonal polynomial associated to the weight function $\mu(x) = \exp[-NV(x)]$, with square norm h_n, see (8.1), while $\psi_n(x)$, $n = 0, 1, \dots$, are the corresponding normalized one-particle wave function, (8.5).

In this framework, the N-point correlation function (4.7) can be written as

$$\begin{aligned} R_N(\lambda_1, \dots, \lambda_N) &= N! \rho_N(\lambda_1, \dots, \lambda_N) \\ &= |\Psi_N(\lambda_1, \dots, \lambda_N)|^2 \\ &= \det{}^2 \left[\psi_{j-1}(\lambda_k)\right]_{j,k=1}^N \\ &= \det \left[\sum_{l=0}^{N-1} \psi_l(\lambda_j)\psi_l(\lambda_k)\right]_{j,k=1}^N . \end{aligned}$$

In the last line we have used the identity $\det^2 A = \det(A^{\mathrm{t}} A)$, that holds for any given real $N \times N$ matrix A, chosing for the entries $A_{lk} = \psi_{l-1}(\lambda_k)$ and $(A^{\mathrm{t}})_{jl} = \psi_{l-1}(\lambda_j)$.

We now introduce the kernel,

$$K_N(\lambda, \nu) := \sum_{l=0}^{N-1} \psi_l(\lambda)\psi_l(\nu),$$

that is an integral kernel, satisfying

$$\int_{-\infty}^{\infty} K_N(\lambda, \lambda) \mathrm{d}\lambda = N, \qquad \int_{-\infty}^{\infty} K_N(\lambda, \mu) K_N(\mu, \nu) \mathrm{d}\mu = K_N(\lambda, \nu). \tag{8.6}$$

kernel_properties

In quantum mechanics, the kernel $K_N(\lambda, \nu)$ is just the projector onto the space of the the N lowest states.

From the above definition, it immediately follows that

$$R_N(\lambda_1, \dots, \lambda_N) = \det \left[K_N(\lambda_j, \lambda_k)\right]_{j,k=1}^N .$$

8.5. The Dyson-Mehta theorem. We shall now state and prove a simple theorem, that allows to express all correlation functions in terms of the integral kernel $K_N(\lambda,\mu)$.

THEOREM 8.1. *Suppose that the integral kernel $K_N(\lambda,\mu)$ satisfies the projection properties* (8.6)*. Then*

$$\int_{-\infty}^{\infty} \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N} \mathrm{d}\lambda_N = \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N-1}, \tag{8.7}$$

th1

and

$$\int_{-\infty}^{\infty} \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N} \mathrm{d}\lambda_{n+1}\ldots\mathrm{d}\lambda_N = (N-n)!\det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{n}. \tag{8.8}$$

th2

PROOF. From the definition of the determinant (8.3), we may write

$$\det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N} = \sum_{\sigma\in S_N}(-1)^{[\sigma]}\prod_{j=1}^{N} K_N(\lambda_{\sigma_j},\lambda_j).$$

It is convenient to split the sum over permutations into N sums over the N subsets of S_N with given value $\sigma_N = l$, $l = 1,\ldots,N$. The lth sum is thus over the subset $S_N^{(l)} := \{\sigma\in S_N | \sigma_N = l\}$. It is clear that each $S_N^{(l)}$ may be identified with S_{N-1}. It is also clear that when $l = N$ the parity of the permutation is preserved, while for $l\neq N$, it is inverted.

We now integrate over λ_N. When $\sigma_N = N$, the first of the two properties (8.6) gives simply N times the desired determinant $\det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N-1}$. When $\sigma_N = l\neq N$, then, for some j, $\sigma_j = N$, and upon integration over λ_N, the second of the two properties (8.6) transform the product of the two kernels involving λ_N, say, $K_N(\lambda_l,\lambda_N)K_N(\lambda_N,\lambda_j)$ into $K_N(\lambda_l,\lambda_j)$. The sum over permutation in the subset $S_N^{(l)}$ then gives rise again to the desired determinant. Notes that we have $N-1$ contributions of this kind, all coming with a minus sign. In all we thus have

$$\int_{-\infty}^{\infty} \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N} \mathrm{d}\lambda_N = [N-(N-1)]\det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N-1},$$

and identity (8.7) is thus proven. Similarly, one can prove

$$\int_{-\infty}^{\infty} \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{n} \mathrm{d}\lambda_n = [N-(n-1)]\det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{n-1}.$$

Applying this identity recursively, one immediately proves identity (8.8) as well. □

8.6. Correlation functions. The Dyson-Mehta theorem allows to express all correlation functions in terms of the integral kernel $K_N(\lambda,\mu)$. Indeed, recalling the definition (4.7), one has

$$\begin{aligned} R_n(\lambda_1,\ldots,\lambda_n) &= \frac{1}{(N-n)!}\int R_N(\lambda_1,\ldots,\lambda_N)\mathrm{d}\lambda_{n+1}\ldots,\mathrm{d}\lambda_N \\ &= \frac{1}{(N-n)!}\int \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{N}\mathrm{d}\lambda_{n+1}\ldots,\mathrm{d}\lambda_N \\ &= \det\left[K_N(\lambda_j,\lambda_k)\right]_{j,k=1}^{n}. \end{aligned}$$

For example,

$$R_1(\lambda) = K_N(\lambda,\lambda), \tag{8.9}$$ sigma_K
$$R_2(\lambda,\mu) = K_N(\lambda,\lambda)K_N(\mu,\mu) - K_N^2(\lambda,\mu), \quad \text{2p-corr}$$
$$T_2(\lambda,\mu) = -K_N^2(\lambda,\mu), \quad \text{2p-cluster}$$

and so on. All other correlation functions, even the most sophisticate, such has the level spacing distribution, or the distribution of the rightmost eigenvalue, can similarly be expressed in terms of the kernel $K_N(\lambda,\mu)$. This is a direct consequence of the free-fermion nature of the model.

It thus follows that our task reduces to devise some method to determine at some suitable level of precision the explicit form of the orthogonal polynomials $P_n(x)$ associated to the weight function $\exp[-NV(x)]$, and to build the corresponding kernel.

8.7. Some standard facts on orthogonal polynomials. We shall here state some very standard facts from the theory of orthogonal polynomials, of great importance in the present context of random matrix models. Derivations are sketched in Appendix A.

It is well known that, given a weight function $\mu(x)$, one can evaluate (at least in principle) all its moments, and from them construct recursively the corresponding family of orthogonal polynomials $\{P_n(x)\}_{n=0}^{\infty}$, through the Gram procedure. However this approach is not very efficient in practice. For our purposes, it appears more convenient to resort to the so-called 3-terms recurrence relation.

Given a family of orthogonal polynomials, any three consecutive polynomials are related by a linear relation, known as the three-terms recurrence relation:

$$xP_n(x) = A_nP_{n+1}(x) + B_nP_n(x) + C_nP_{n-1}(x), \quad \text{RR}$$

with

$$A_n = \frac{k_n}{k_{n+1}}, \qquad B_n = \frac{k'_n}{k_n} - \frac{k'_{n+1}}{k_{n+1}}, \qquad C_n = \frac{k_{n-1}h_n}{k_nh_{n-1}}, \tag{8.10}$$ RR_coeff

where k_n, k'_n are the leading coefficients of $P_n(z) = k_nx^n + k'_nx^{n-1} + \dots$, and the h_n's are the squared norms.

If the weight function is even, $\mu(x) = \mu(-x)$, then the polynomials have definite parity, and the k'_ns vanish, implying $B_n = 0$. If we choose our polynomials to be monic, $k_n = 1$, then $A_n = 1$. All together, the three-terms recurrence relation simplifies to

$$xP_n(x) = P_{n+1}(x) + \frac{h_n}{h_{n-1}}P_{n-1}(x). \tag{8.11}$$ RR_even

From now on we specialize to the case of monic polynomials, and to a weight function of the form $\mu(x) = \exp[-NV(x)]$. We denote by $\langle f,g\rangle$ the 'scalar product' of two functions $f(x)$, $g(x)$, against the measure $\mu(x)$. We have the obvious relation

$$\langle P_l, P'_n\rangle + \langle P'_lP_n\rangle - N\langle V'P_l, P_n\rangle = 0. \tag{8.12}$$ trivial

Clearly, we can always expand P'_n in terms of lower order polynomials. Inserting such expansion in (8.12), it follows that, if the 'potential' $V(x)$ is a polynomial of

degree d, then

$$P_n'(x) = nP_{n-1}(x) + \sum_{k=n-d}^{n-2} s_{nk} P_k(x)$$

derivative

where the s_{nk}'s are $d-1$ suitable constants.

At the present stage all the coefficients h_n, C_n, s_{nk} are still to be determined. All together, the three-terms recurrence relation, the obvious 'partial integration' identity and the derivative relation give enough conditions to determine all of them. In particular, one is able to work out some recurrence relations that they should satisfy. In practice, however, such relations cannot be solved exactly. They nevertheless provide sufficient informations on the large n behaviour of the $P_n(x)$ and h_n.

To conclude, let us recall the Christoffel-Darboux identity, see appendix A, that in the case of monic polynomials read

$$\sum_{n=0}^{N} \frac{1}{h_n} P_n(x)P_n(y) = \frac{1}{h_N} \frac{P_{N+1}(x)P_n(y) - P_N(x)P_{N+1}(y)}{x-y}. \tag{8.13}$$

CD_formula

This immediately implies

$$K_N(\lambda,\mu) = \sqrt{\frac{h_N}{h_{N-1}}} \frac{\psi_N(\lambda)\psi_{N-1}(\mu) - \psi_{N-1}(\lambda)\psi_N(\mu)}{\lambda-\mu}. \tag{8.14}$$

K_darboux

It is apparent from this expression that the large N behaviour of the kernel $K_N(\lambda,\mu)$ is in fact completely determined only by the behaviour of $P_{N-1}(\lambda)$ and $P_N(\lambda)$, in agreement with the 'Fermi sea' physical picture discussed above. It is also clear that if we want to extract some information on the behaviour of the correlation functions, the knowledge of the sole h_n's, or C_n's, is not sufficient, and we need the asymptotic behaviour of the $P_n(x)$'s as well.

8.8. Example: the gaussian ensemble. Let us consider the quadratic potential $V(x) = \frac{1}{2}gx^2$. The potential being even, the recurrence relation simplifies to (8.11), and in the derivative relation the only undetermined coefficient, s_{n2}, vanishes. We thus have:

$$\begin{aligned} xP_n(x) &= P_{n+1}(x) + \frac{h_n}{h_{n-1}} P_{n-1}(x), \\ P_n'(x) &= nP_{n-1}(x). \end{aligned}$$

Plugging into

$$\langle P_l, P_n' \rangle + \langle P_l' P_n \rangle - Ng \langle xP_l, P_n \rangle = 0,$$

see (8.12), we get a nontrivial relation only for $n = l \pm 1$, namely

$$h_n = \frac{n}{Ng} h_{n-1},$$

that is immediately solved,

$$h_n = \frac{n!}{(Ng)^n} h_0, \qquad h_0 = \sqrt{\frac{2\pi}{Ng}}.$$

Consequently, the three-terms recurrence relation reads now

$$xP_n(x) = P_{n+1}(x) + \frac{n}{Ng} P_{n-1}(x),$$

which reminds that for Hermite polynomials.

Recall that the Hermite polynomials $H_n(x)$ are orthogonal with respect to the weight function e^{-x^2}, with $h_n = \sqrt{\pi} 2^n n!$. Their leading coefficient is 2^n. They satisfy the recursion relation $2xH_n = H_{n+1} + 2nH_{n-1}$. They satisfy the (confluent hypergometric) differential equation $H_n'' - 2xH_n' + 2nH_n = 0$. These two relations further imply $H_{n+1} = 2xH_n - H_n'$.

The solution of our recurrence relation is thus given by Hermite polynomials, modulo suitable rescaling of the variable x and of the normalizations h_n's:

$$P_n(x) = \left(\frac{1}{\sqrt{2Ng}}\right)^n H_n\left(\sqrt{\frac{Ng}{2}}x\right).$$

Note that this immediately implies an exact closed form expression for the kernel $K_N(x,y)$, and hence for the density and all other correlation functions.

8.9. The large N limit: calculation for the free energy. We know from (8.4) that the evaluation of the partition function is reduced to the calculation of the square norms h_n of the set of orthogonal polynomials $P_n(x)$ associated to the measure $\mathrm{e}^{-NV(x)}$. It is convenient to write

$$Z_N = N! \prod_{j=0}^{N-1} h_j = N! h_0^N \prod_{j=0}^{N-1} C_j^{N-j}.$$

In general the exact expressions for the h_n's and C_n's are not accessible, but it is often possible to evaluate them in the limit of large n. This allows to write:

$$\begin{aligned} F_N &:= \frac{1}{N^2} \log Z_N \\ &\sim \frac{1}{N^2}\left[N\log N - N + N\log h_0 + \sum_{j=1}^{N}(N-j)\log C_j\right] \\ &\sim \int_0^1 (1-\xi)\log C(\xi)\mathrm{d}\xi, \qquad N\to\infty. \end{aligned} \tag{8.15}$$

free-energ

This result holds under the assumption that the limit $\lim_{n\to\infty} C_n$ exists, which is actually granted only in the one-cut case (for example, in the two-cut case, the sequences C_{2j} and C_{2j+1} have two different limits).

8.10. Example: the quartic potential. Let us consider the quartic potential $V(\lambda) = \frac{1}{2}x^2 + \frac{1}{4}gx^4$. Taking into account that the potential, and the corresponding weigth function, are even, we have:

$$\begin{aligned} xP_n(x) &= P_{n+1}(x) + C_n P_{n-1}(x) \\ P_n'(x) &= nP_{n-1}(x) + s_n P_{n-3} \\ 0 &= \langle P_l', P_n\rangle + \langle P_l, P_n'\rangle - N\left\langle P_l, (x+gx^3)P_n\right\rangle \end{aligned} \tag{8.16}$$

quartic_pol3

Now, setting $l = n+3$ in last equation, and using the other two, we get

$$s_{n+3}h_n - Ngh_{n+3} = 0.$$

Similarly, setting $l = n+1$ in (8.16), we get

$$(n+1)h_n - Nh_{n+1} - Ng(C_n + C_{n+1} + C_{n+2})h_{n+1} = 0,$$

that, after shifting $n \to n-1$, and recalling that $C_n = h_n/h_{n-1}$, leads to the recurrence relation for the C_n's:

$$\frac{n}{N} = C_n \left[1 + g(C_{n-1} + C_n + C_{n+1})\right]. \tag{8.17}$$ C_recurr

Unfortunately now the recurrence relation is non linear, and we are unable to solve it exactly. However we can try to solve it in the limit of large n and N, with fixed ratio n/N.

In this limit, the quantity n/N can be approximated by the continuous variable $\xi \in [0,1]$, and the recurrence relation is turned into the algebraic equation

$$\xi = C(\xi)\left[1 + 3gC(\xi)\right],$$

with solution

$$C(\xi) = \frac{-1 + \sqrt{1 + 12g\xi}}{6g}, \tag{8.18}$$ Cxi

the sign of the square root being determined by the Gaussian limit, $g \to 0$, where we know that $C(\xi) = \xi$.

Note that, due to the square root, the above expression is well-defined only for $g > -1/(12\xi)$; recalling that $\xi \in [0,1]$, we recover the critical point $g_c = -1/12$, already obtained previously by using the saddle-point method.

Using expression (8.15), it is now straightforward to evaluate the free energy of the model, and recover the singular behaviour $F^{(0)} \sim (g - g_c)^{5/2}$, and all other results of [8].

Finally, note that in principle, and at variance with the resolvent approach, the present method allows to go beyond the planar limit result (8.18). For instance, taking into account $O(1/N)$ corrections, the recurrence relation (8.17) becomes

$$\xi = C(\xi)\left[1 + 3gC(\xi) + gC''(\xi)\right],$$

This can be reconducted to the famous and well-studied Painlevé equation, which would be hoverver beyond the scope of these notes.

8.11. Gaussian ensemble: the eigenfunctions. Let us come back to the Gaussian ensemble, $V(x) = \frac{1}{2}g\lambda^2$. The single particle normalized wavefunction is

$$\psi_n(\lambda) = \frac{1}{\sqrt{(2Ng)^n h_n}} \mathrm{e}^{-\frac{Ng}{4}\lambda^2} H_n\left(\sqrt{\frac{Ng}{2}}\lambda\right), \qquad h_n = \sqrt{\frac{2\pi}{Ng}}\frac{n!}{(Ng)^n}$$

This is the nth eigenfunction of the harmonic oscillator. It is known to have oscillatory (exponentially decaying) behaviour, for values of the argument $\sqrt{Ng/2}\lambda$ inside (outside) the interval $[-\sqrt{2n+1}, \sqrt{2n+1}]$. This property is easily derived by evaluating the 'classical inversion point' for the corresponding quantum oscillator, $\frac{1}{2}\omega^2 x^2 = (n + \frac{1}{2})\omega$, in our case, $\omega = 1$.

To evaluate the large N behaviour of the kernel $K_N(\lambda, \mu)$, that is the building block for all correlation functions, we need first to work out the corresponding behaviour for the wavefunction $\psi_n(\lambda)$.

It is known that for fixed x (or uniformely in any bounded interval),

$$H_n(x) = \frac{\Gamma(n+1)}{\Gamma(\frac{n}{2}+1)} \mathrm{e}^{x^2/2}\left[\cos\left(\sqrt{2n+1}x - n\frac{\pi}{2}\right) + O(n^{-1/4})\right], \qquad n \to \infty.$$

The above large n behaviour implies

$$\psi_n(\lambda) \sim \frac{1}{\sqrt{\pi}} \left(\frac{Ng}{n}\right)^{1/4} \cos\left(\sqrt{2n+1}\sqrt{\frac{Ng}{2}}\lambda - n\frac{\pi}{2}\right), \qquad n \to \infty,$$

for the wavefunction. Note however that the condition of fixed x translates into $\lambda \lesssim 1/\sqrt{N}$, that is, the above estimates holds only within a small neighbourhood of the origin. This will be nevertheless sufficient for our purposes here.

In particular, setting $n = N, N-1$, which are the only relevant cases for our purposes, see (8.14), we have:

$$\begin{aligned} \psi_N(\lambda) &\sim \frac{g^{1/4}}{\sqrt{\pi}} \cos\left(N\sqrt{g}\lambda - N\frac{\pi}{2}\right), \\ \psi_{N-1}(\lambda) &\sim -\frac{g^{1/4}}{\sqrt{\pi}} \sin\left(N\sqrt{g}\lambda - N\frac{\pi}{2}\right), \\ \psi'_N(\lambda) &\sim -\frac{Ng^{3/4}}{\sqrt{\pi}} \sin\left(N\sqrt{g}\lambda - N\frac{\pi}{2}\right), \end{aligned} \tag{8.19}$$

psi_n

from which we shall now evaluate the kernel $K_N(\lambda, \mu)$, although with λ, μ confined to a small neighbourhood of the origin.

8.12. Gaussian ensemble: the sine-kernel. To start with, we should evaluate the eigenvalue density $\sigma(\lambda)$. From (8.9) and (8.14), we have

$$\begin{aligned} \sigma(\lambda) &= K_N(\lambda, \lambda) \\ &= \frac{1}{\sqrt{g}} \left[\psi'_N(\lambda)\psi_{N-1}(\lambda) - \psi'_{N-1}(\lambda)\psi_N(\lambda)\right] \\ &\sim \frac{N\sqrt{g}}{\pi} \cos(\sqrt{g}\lambda), \qquad N \to \infty, \end{aligned}$$

that differs from the expected Wigner semicircle. Recall however that the present results holds only for $\lambda \lesssim 1/\sqrt{N} \sim 0$. Thus what we have actually computed is $\sigma(0)$, and our result is in agreement with the Wigner semicircle obtained in the saddle-point approximation, see (6.12), and recall that $\sigma(\lambda) := N\rho(\lambda)$:

$$\sigma(\lambda) = \frac{Ng}{2\pi}\sqrt{4/g - \lambda^2}, \qquad V(\lambda) = \frac{g}{2}\lambda^2.$$

In order to recover the full semicircle, one should use more accurate estimate for the large n asymptotic behaviour of the Hermite polynomials (Plancherel-Rotah formulae), and of wavefunction $\psi_n(x)$, see [13].

We now proceed to evaluate the kernel $K_N(\lambda, \mu)$. From (8.14), we have

$$K_N(\lambda, \mu) = \frac{1}{\sqrt{g}} \frac{\psi_N(\lambda)\psi_{N-1}(\mu) - \psi_{N-1}(\lambda)\psi_N(\mu)}{\lambda - \mu}$$

We know that in order to observe some universal behaviour we need to unfold the spectrum, that is, to fix some point λ_0, and, in its neighbourhood, to rescale the variable λ (and μ) to some suitable variable x such that $\mathrm{d}x/\mathrm{d}\lambda|_{\lambda_0} = \sigma(\lambda_0) =: \sigma_0$,

$$\begin{aligned} \lambda &\to \lambda_0 + \frac{x}{\sigma_0}, \\ \mu &\to \lambda_0 + \frac{y}{\sigma_0}. \end{aligned} \tag{8.20}$$

rescaling_var

Moreover, since n-point correlation functions are of degree n in the kernel, we need also to divide the kernel by σ_0, see (5.1). In all we have thus the 'unfolded kernel'

$$\tilde{K}_N(x,y) := \frac{1}{\sigma_0} K_N(\lambda_0 + \frac{x}{\sigma_0}, \lambda_0 + \frac{y}{\sigma_0}).$$

Note that our rescaling is such that for generic finite values of x and y, of order $O(N^0)$, we are testing the correlations at separation $|\lambda - \mu| = O(1/N)$.

Let us implement this procedure in the vicinity of the point $\lambda_0 = 0$, for simplicity. We thus have $\sigma_0 = \sigma(0) = N\sqrt{g}/\pi$. Note also that by virtue of the rescaling of variables (8.20), we have $\lambda, \mu = O(1/N)$, that is, much smaller than $1/\sqrt{N}$, and the expansions (8.19) can be used. We immediately get

$$\tilde{K}_N(x,y) \sim \frac{1}{\sigma_0} \frac{1}{\sqrt{g}} \frac{\sqrt{g}}{\pi} \frac{\sin[N\sqrt{g}(x-y)/\sigma_0]}{(x-y)/\sigma_0} = \frac{\sin[\pi(x-y)]}{\pi(x-y)}.$$

where the last expression is known as the sine-kernel.

We have thus shown that in the vicinity of the origin, and at small separation $x - y = O(N^0)$, the unfolded kernel of the GUE tends to the sine-kernel in the large N limit. Actually, the result is much more general. In particular it holds for arbitrary choice of λ_0 in the bulk of the spectrum, that is within the support of the eigenvalue density, $|\lambda_0| < 2/\sqrt{g}$. One just need the above mentioned Plancherel-Rotah formulae for the large n behaviour of the Hermite polynomials.

The orthogonal polynomial approach allows to perform a similar analysis for a generic polynomial potential $V(x)$. A remarkable result is that, in the bulk of the spectrum (that is, away from end-points and critical points), and for unitary ensembles, the large N limit of the unfolded kernel at small separations, $O(1)$, always gives the sine-kernel. Similar results have been derived for the other ensembles as well. See [14] for a review.

Besides being ubiquitous in random matrix models, the sine-kernel appears also in the correlation functions of the impenetrable Bose gas (or Tonk-Girardeau model), and of the XX0 Heisenberg quantum spin chain, that are free fermion models as well.

8.13. Gaussian ensemble: the Airy kernel. [...]

Appendix A. Orthogonal polynomials

appOP

A.1. Basic definitions. In this appendix we recall some elementary fact from the theory of orthogonal polynomials. For definiteness, we shall treat here the case of a weight function $\mu(x)$ defined over the real axis. We shall also restrict to a particular normalization, namely that corresponding to the case of monic polynomials. The discussion can be easily extended to other situations.

Consider a real nonnegative weight function $\mu(x)$, $x \in \mathbb{R}$, vanishing faster than any polynomial as $|x| \to \infty$, so that all its moments are finite,

$$c_n := \int_{-\infty}^{+\infty} x^n \mu(x) \mathrm{d}x, \qquad n = 0, 1, \dots.$$

In principle, it is always possible to build the corresponding set of orthogonal polynomials, that is, $\{P_n(x)\}_{n=0}^{\infty}$, such that

$$\mathrm{deg} P_n(\lambda) = n, \qquad P_n(x) = x^n + k_n' x^{n-1} + \dots,$$

and

$$\int_{-\infty}^{+\infty} P_n(x)P_m(x)\mu(x)\mathrm{d}x = h_n\delta_{nm}.$$

orthog_app

The most direct procedure could be the standard Gram-Schmidt orthogonalization method. Starting from $P_0(x) = 1$, that is fixed by the condition of being monic, we may evaluate $h_0 = c_0$. Then we have $P_1(x) = x + a$, where a is some constant. The scalar product $\langle P_1, P_0\rangle = 0$ gives $c_1 + ah_0 = 0$, determining $P_1(x)$, etc. The solution of the iteration of the above procedure can be written as a ratio of determinants of the moments c_n:

$$P_n(x) \propto \frac{\det\left[c_{j+k-2}\ \left|x^{j-1}\right.\right]_{j,k=1}^{n,n-1}}{\det\left[c_{j+k-2}\right]_{j,k=1}^{n}}.$$

The overall constant is determined by the normalization, that can be chosen at convenience. Such a set of orthogonal polynomials is known to form a complete basis for $L^2(\mathbb{R})$, the Hilbert space of square integrable function, and to satisfy a three-term recurrence relation.

A.2. Recurrence relation. Obviously, we can always expand a generic polynomial of degree n over the first $n+1$ polynomials $P_0(x),\ldots,P_n(x)$. In particular, we have

$$xP_n(x) = A_nP_{n+1}(x) + B_nP_n(x) + C_nP_{n-1}(x) + \sum_{k=0}^{n-2} r_{nk}P_k(x). \qquad \text{(A.1)}$$

xPn

Taking the scalar product with $P_l(x)$, we have

$$\langle P_l, xP_n\rangle = 0, \qquad l > n+1.$$

Note that we can also move the monomial x to the 'bra' vector, and replace the expansion (A.1) into $xP_l(x)$, instead. This gives

$$\langle xP_l, P_n\rangle = 0, \qquad n > l+1,$$

that is, $l < n-1$. Together, these two relations ensure that the above scalar product vanishes for all values of l such that $|l-n| > 1$. In particular, setting $l = 0, 1, \ldots, n-2$, we get that all coefficients r_{nl} in (A.1) vanish. Thus we are left with

$$xP_n(x) = A_nP_{n+1}(x) + B_nP_n(x) + C_nP_{n-1}(x), \qquad \text{(A.2)}$$

recurr_app

that is, any family of orthogonal polynomials satisfies a three-terms recurrence relation (it is clear from the context that $P_{-1}(x)$ should be treated as zero).

Let us know determine the form of the coefficients A_n, B_n, and C_n, appearing in the recurrence relation. First, the $P_n(x)$ being monic, $A_n = 1$. Second, in the obvious relation $\langle P_{n-1}, xP_n\rangle = \langle xP_{n-1}, P_n\rangle$, we plug in the recurrence relation (A.2), obtaining

$$C_nh_{n-1} = h_n.$$

Finally, we compare the coefficients of x^n in both sides of the recurrence relation (A.2), obtaining

$$k'_n = k'_{n+1} + B_n.$$

In all we thus have

$$xP_n(x) = P_{n+1}(x) + (k'_n - k'_{n+1})P_n(x) + \frac{h_n}{h_{n-1}}P_{n-1}(x).$$

We recall that here the square norms h_n have been chosen such that the $P_n(x)$ are monic. In general the coefficients have instead the form (8.10), that can be derived along the same lines. Note also that if the case of an even measure, $\mu(x) = \mu(-x)$, then the corresponding polynomials have definite parity, $P_n(-x) = (-1)^n P_n(x)$, and the k_n vanish. The recurrence relation reduces then to (8.11).

A.3. Derivative relation. We shall consider here a relation which is not a general property of orthogonal polynomials, but rather a specific feature of orthogonal polynomials associated to a measure of the form $\mu(x) = \exp[-NV(x)]$, with $V(x)$ some polynomial. Given such set $\{P_n(x)\}_{n=0}^{\infty}$, let us consider the derivative of $P_n(x)$. This is clearly a polynomial of degree $n-1$, and can thus be expanded over the first n polynomials of the same set:

$$P_n'(x) = nP_{n-1}(x) + \sum_{k=0}^{n-2} s_{nk} P_k(x), \tag{A.3}$$

deriv_expansion

where the coefficient of P_{n-1} is fixed by the condition of being monic, and s_{nk}, $k = 0, 1, \ldots, n-1$ are some suitable coefficients.

Consider now the relation

$$\langle P_n, P_l' \rangle + \langle P_n', P_l \rangle - N \langle V' P_n, P_l \rangle = 0, \tag{A.4}$$

trivial_app

that immedatiely follows from the definition of the scalar product, and integration by parts. Note that, $V(x)$ being a polynomial of degree d, the multiplication of $P_n(x)$ by $V'(x)$ can be expanded over a finite number of the complete $\{P_N(x)\}_{n=1}^{\infty}$, by using $d-1$ times the three-term recurrence relation. Also, we can replace derivatives of polynomials with the expansion (A.3). From the orthogonality relation it immediately follows that the first term of relation (A.4) is different from zero only for $n < l$, the second one only for $n > l$, and the last one only for $n - d < l < n + d$. Focus now on some value n such that $l \leq n - d$; only the second term in (A.4) survives, implying that

$$0 = \langle P_n', P_l \rangle = \sum_{k=0}^{n-2} s_{nk} \langle P_k, P_l \rangle = s_{nl} h_l \qquad l \leq n - d,$$

and therefore that $s_{nl} = 0$ for $l = 0, 1, \ldots, n-d-1$. We thus have

$$P_n'(x) = nP_{n-1}(x) + \sum_{k=n-d}^{n-2} s_{nk} P_k(x),$$

deriv_expansion_2

that is, the expansion involves only the d polynomials $P_{n-d}(x)$, ..., $P_{n-1}(x)$.

A.4. Christoffel-Darboux formula. From the three-term recurrence relation, it follows that

$$\begin{aligned}(x-y)P_n(x)P_n(y) = A_n \left[P_{n+1}(x)P_n(y) - P_n(x)P_{n+1}(y)\right] \\ + C_n \left[P_{n-1}(x)P_n(y) - P_n(x)P_{n-1}(y)\right].\end{aligned}$$

Recalling the expression for the coefficients of the recursion relation, and dividing both sides by h_n, we have

$$(x-y)\frac{1}{h_n}P_n(x)P_n(y) = \frac{1}{h_n}\left[P_{n+1}(x)P_n(y) - P_n(x)P_{n+1}(y)\right] - \frac{1}{h_{n-1}}\left[P_n(x)P_{n-1}(y) - P_{n-1}(x)P_n(y)\right].$$

Now, summing both sides over n, till some value N, and observing that the right-hand side is a telescopic sum, we get

$$(x-y)\sum_{n=0}^{N}\frac{1}{h_n}P_n(x)P_n(y) = \frac{1}{h_N}\left[P_{N+1}(x)P_n(y) - P_N(x)P_{N+1}(y)\right],$$

from which the monic polynomial version (8.13) of the Christoffel-Darboux formula immediately follows.

References

M-67 [1] M.L. Mehta, *Random matrices*, Academic Press, 1967.

M-04 [2] M.L. Mehta, *Random matrices*, Elsevier, 2004.

ABD-11 [3] G. Akemann, J. Baik, and P. Di Francesco, *The Oxford Handbook of Random Matrix Theory*, Oxford University Press, 2011.

AGZ-10 [4] G.W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, 2010.

F-10 [5] P.J. Forrester, *Log-Gases and Random Matrices*, Princeton University Press, 2010.

D-02 [6] F.J. Dyson, *Random Matrices, Neutron Capture Levels, Quasicrystals and Zeta-Function Zeros*, 2002. Talk given at MSRI, Berkeley, 23 September.

H-74 [7] G. Hooft, *A planar diagram theory for strong interactions*, Nuclear Physics B **72** (1974), no. 3, 461–473.

BIPZ-78 [8] E. Brézin, C. Itzykson, G. Parisi, and J.-B. Zuber, *Planar diagrams*, Communications in Mathematical Physics **59** (1978), no. 1, 35–51.

GW-80 [9] D. J. Gross and E. Witten, *Possible third-order phase transition in the large-N lattice gauge theory*, Phys. Rev. D **21** (1980), 446–453.

DK-93 [10] M.R. Douglas and V.A. Kazakov, *Large N phase transition in continuum QCD_2*, Phys. Lett. B **319** (1993), 219–230.

D-62 [11] F.D. Dyson, *Statistical theory of the energy levels of complex systems. i*, Journal of Mathematical Physics **3** (1962), 140–156.

W-80 [12] S. R. Wadia, *$N = \infty$ phase transition in a class of exactly soluble model lattice gauge theories*, Phys. Lett. B **93** (1980), 403–410.

S-39 [13] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, 1939.

D-00 [14] P. Deift, *Orthogonal Polynomials and Random Matrices: a Riemann-Hilbert approach*, American Mathematical Society, 2000.

INFN, Sezione di Firenze, Via G. Sansone 1, 50019 Sesto Fiorentino (FI), Italy
E-mail address: colomo@fi.infn.it