5

Joint Distribution of Eigenvalues

In the previous chapters, we have studied the moments, the Stieltjes transform and the eigenvalue density of two classical ensembles (Wigner and Wishart). These quantities in fact relate to *single eigenvalue* properties of these ensembles. By this we mean that the Stieltjes transform and the eigenvalue density are completely determined by the univariate law of eigenvalues but they do not tell us anything about the *correlations* between different eigenvalues.

In this chapter we will extend these results in two directions. First we will consider a larger class of rotationally invariant (or orthogonal) ensembles that contains Wigner and Wishart. Second we will study the joint law of all eigenvalues. In these models, the eigenvalues turn out to be strongly correlated and can be thought of as "particles" interacting through pairwise repulsion.

5.1 From Matrix Elements to Eigenvalues

5.1.1 Matrix Potential

Consider real symmetric random matrices $\mathbf{M}$ whose elements are distributed as the exponential of the trace of a certain matrix function $V(\mathbf{M})$, often called a potential by analogy with statistical physics:[1]

$$P(\mathbf{M}) = Z_N^{-1} \exp\left\{-\frac{N}{2} \operatorname{Tr} V(\mathbf{M})\right\}, \tag{5.1}$$

where Z_N is a normalization constant. These matrix ensembles are called orthogonal ensembles for they are rotationally invariant, i.e. invariant under orthogonal transformations.[2] For the Wigner ensemble, for example, we have (see Chapter 2)

[1] $V(\mathbf{M})$ is a matrix function best defined in the eigenbasis of $\mathbf{M}$ through a transformation of all its eigenvalues through a function of a scalar, $V(x)$, see Section 1.2.6.

[2] The results of this chapter extend to Hermitian ($\beta = 2$) or quarternion-Hermitian ($\beta = 4$) matrices with the simple introduction of a factor β in the probability distribution:

$$P(\mathbf{M}) \propto \exp\left\{-\frac{\beta N}{2} \operatorname{Tr} V(\mathbf{M})\right\}, \tag{5.2}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

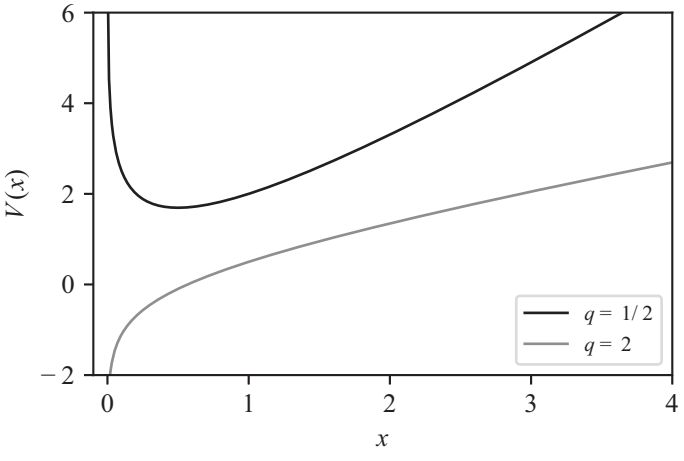

Figure 5.1 The Wishart matrix potential (Eq. (5.4)) for $q = 1/2$ and $q = 2$. The integration over positive semi-definite matrices imposes that the eigenvalues must be greater than or equal to zero. For $q < 1$ the potential naturally ensures that the eigenvalues are greater than zero and the constraint will not be explicitly needed in the computation. For $q \geq 1$, the constraint is needed to obtain a sensible result.

$$V(x) = \frac{x^2}{2\sigma^2}, \tag{5.3}$$

whereas the Wishart ensemble (at large N) is characterized by (see Chapter 4)

$$V(x) = \frac{x + (q-1)\log x}{q} \tag{5.4}$$

(see Fig. 5.1). We can also consider other matrix potentials, e.g.

$$V(x) = \frac{x^2}{2} + \frac{gx^4}{4}. \tag{5.5}$$

Note that $\mathrm{Tr}\, V(\mathbf{M})$ depends only on the eigenvalues of $\mathbf{M}$. We would like thus to write down the joint distribution of these eigenvalues alone. The key is to find the Jacobian of the change of variables from the entries of $\mathbf{M}$ to the eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$.

5.1.2 Matrix Jacobian

Before computing the Jacobian of the transformation from matrix elements to eigenvalues and eigenvectors (or orthogonal matrices), let us count the number of variables in both parameterizations. Suppose $\mathbf{M}$ can be diagonalized as

$$\mathbf{M} = \mathbf{O}\Lambda\mathbf{O}^T. \tag{5.6}$$

this factor will match the factor of β from the Vandermonde determinant. These two other ensembles are called unitary ensembles and symplectic ensembles, respectively. Collectively they are called the beta ensembles.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

The symmetric matrix $\mathbf{M}$ has $N(N+1)/2$ independent variables, and Λ has N independent variables as a diagonal matrix. To find the number of independent variables in $\mathbf{O}$ we first realize that $\mathbf{O}\mathbf{O}^T = \mathbf{1}$ is an equation between two symmetric matrices and thus imposes $N(N+1)/2$ constraints out of N^2 potential values for the elements of $\mathbf{O}$, therefore $\mathbf{O}$ has $N(N-1)/2$ independent variables. In total, we thus have $N(N+1)/2 = N + N(N-1)/2$.

The change of variables from $\mathbf{M}$ to $(\Lambda, \mathbf{O})$ will introduce a factor $|\det(\Delta)|$, where

$$\Delta := \Delta(\mathbf{M}) = \left[\frac{\partial \mathbf{M}}{\partial \Lambda}, \frac{\partial \mathbf{M}}{\partial \mathbf{O}}\right] \tag{5.7}$$

is the Jacobian matrix of dimension $N(N+1)/2 \times N(N+1)/2$.

First, let us establish the scaling properties of the Jacobian. We assume that the matrix elements of $\mathbf{M}$ have some dimension $[\mathbf{M}] = d$ (say centimeters). Using dimensional analysis, we thus have

$$[\mathcal{D}\mathbf{M}] = d^{N(N+1)/2}, \quad [\mathcal{D}\Lambda] = d^N, \quad [\mathcal{D}\mathbf{O}] = d^0, \tag{5.8}$$

since rotations are dimensionless. Hence we must have

$$[|\det(\Delta)|] \sim d^{N(N-1)/2}, \tag{5.9}$$

which has the dimension of an eigenvalue raised to the power $N(N-1)/2$, the number of distinct off-diagonal elements in $\mathbf{M}$.

We now compute this Jacobian exactly. First, notice that the Jacobian relates the volume "around" $(\Lambda, \mathbf{O})$ when Λ and $\mathbf{O}$ change by infinitesimal amounts, to the volume "around" $\mathbf{M}$ when its elements change by infinitesimal amounts. We note that volumes are invariant under rotations, so in order to compute the infinitesimal volume we can choose the rotation matrix $\mathbf{O}$ to be the identity matrix, which amounts to saying that we work in the basis where $\mathbf{M}$ is diagonal. Another way to see this is to note that the orthogonal transformation

$$\mathbf{M} \to \mathbf{U}^T\mathbf{M}\mathbf{U}; \qquad \mathbf{U}^T\mathbf{U} = \mathbf{1} \tag{5.10}$$

has a Jacobian equal to 1, see Section 1.2.7. One can always choose $\mathbf{U}$ such that $\mathbf{M}$ is diagonal.

5.1.3 Infinitesimal Rotations

For rotations $\mathbf{O}$ near the identity, we set

$$\mathbf{O} = \mathbf{1} + \epsilon\, \delta\mathbf{O}, \tag{5.11}$$

where ϵ is a small number and $\delta\mathbf{O}$ is some matrix. From the identity

$$\mathbf{1} = \mathbf{O}\mathbf{O}^T = \mathbf{1} + \epsilon(\delta\mathbf{O} + \delta\mathbf{O}^T) + \epsilon^2 \delta\mathbf{O}\delta\mathbf{O}^T, \tag{5.12}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

we get $\delta\mathbf{O} = -\delta\mathbf{O}^T$ by comparing terms of the first order in ϵ, i.e. $\delta\mathbf{O}$ is skew-symmetric.[3] A convenient basis to write such infinitesimal rotations is

$$\epsilon\,\delta\mathbf{O} = \sum_{1\le k<l\le N} \theta_{kl}\mathbf{A}^{(kl)}, \tag{5.13}$$

where $\mathbf{A}^{(kl)}$ are the elementary skew-symmetric matrices such that $\mathbf{A}^{(kl)}$ has only two non-zero elements: $[\mathbf{A}^{(kl)}]_{kl} = 1$ and $[\mathbf{A}^{(kl)}]_{lk} = -1$:

$$\mathbf{A}^{(kl)} = \begin{pmatrix} 0 & \dots & & \dots & 0 \\ \vdots & \ddots & & & \vdots \\ & & 1 & & \\ & -1 & & & \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & & \dots & 0 \end{pmatrix}. \tag{5.14}$$

An infinitesimal rotation is therefore fully described by $N(N-1)/2$ generalized "angles" θ_{kl}.

5.1.4 Vandermonde Determinant

Now, in the neighborhood of $(\Lambda, \mathbf{O} = \mathbf{1})$, the matrix $\mathbf{M} + \delta\mathbf{M}$ can be parameterized as

$$\mathbf{M} + \delta\mathbf{M} \approx \left(\mathbf{1} + \sum_{k,l} \theta_{kl}\mathbf{A}^{(kl)}\right)(\Lambda + \delta\Lambda)\left(\mathbf{1} - \sum_{k,l} \theta_{kl}\mathbf{A}^{(kl)}\right). \tag{5.15}$$

So to first order in $\delta\Lambda$ and θ_{kl},

$$\delta\mathbf{M} \approx \delta\Lambda + \sum_{k,l} \theta_{kl}\left[\mathbf{A}^{(kl)}\Lambda - \Lambda\mathbf{A}^{(kl)}\right]. \tag{5.16}$$

Using this local parameterization, we can compute the Jacobian matrix and find its determinant. For the diagonal contribution, we have

$$\frac{\partial M_{ij}}{\partial \Lambda_{nn}} = \delta_{in}\delta_{jn}, \tag{5.17}$$

i.e. perturbing a given eigenvalue only changes the corresponding diagonal element with slope 1.

For the rotation contribution, one has, for $k < l$ and $i < j$,

[3] The reader familiar with the analysis of compact Lie groups will recognize the statement that skew-symmetric matrices form the Lie algebra of $O(N)$.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

$$\frac{\partial M_{ij}}{\partial \theta_{kl}} = \left(\mathbf{A}^{(kl)}\Lambda - \Lambda\mathbf{A}^{(kl)}\right)_{ij} = \begin{cases} \lambda_l - \lambda_k, & \text{if } i = k, j = l, \\ 0, & \text{otherwise.} \end{cases} \tag{5.18}$$

i.e. an infinitesimal rotation in the direction kl modifies only one distinct off-diagonal element ($\mathbf{M}_{kl} \equiv \mathbf{M}_{lk}$) with slope $\lambda_l - \lambda_k$. In particular, if two eigenvalues are the same ($\lambda_k = \lambda_l$) a rotation of the eigenvectors in that subspace has no effect on the matrix $\mathbf{M}$. This is expected since eigenvectors in a degenerate subspace are only defined up to a rotation within that subspace.

Finally, the $N(N+1)/2 \times N(N+1)/2$ determinant has its first N diagonal elements equal to unity, and the next $N(N-1)/2$ are equal to all possible pair differences $\lambda_i - \lambda_j$. Hence,

$$\Delta(\mathbf{M}) = \det \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \lambda_2 - \lambda_1 & & & & \\ & & & & \lambda_3 - \lambda_1 & & & \\ & & & & & \ddots & & \\ & & & & & & \lambda_N - \lambda_{N-1} \end{pmatrix} = \prod_{k<\ell} (\lambda_\ell - \lambda_k). \tag{5.19}$$

The absolute value of Δ is then given by

$$|\Delta(\mathbf{M})| = \prod_{k<\ell} |\lambda_\ell - \lambda_k|. \tag{5.20}$$

We can check that this result has the expected dimension $d^{N(N-1)/2}$, since the product contains exactly $N(N-1)/2$ terms. The determinant $\Delta(\mathbf{M})$ is called the Vandermonde determinant as it is equal to the determinant of the following $N \times N$ Vandermonde matrix:

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 & \dots & \lambda_N \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 & \dots & \lambda_N^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \lambda_3^{N-1} & \dots & \lambda_N^{N-1} \end{pmatrix}. \tag{5.21}$$

Since the above Jacobian has no dependence on the matrix $\mathbf{O}$, we can integrate out the rotation part of (5.1) to get the joint distribution of eigenvalues:

$$P(\{\lambda_i\}) \propto \prod_{k<l} |\lambda_k - \lambda_l| \exp\left\{-\frac{N}{2}\sum_{i=1}^{N} V(\lambda_i)\right\}. \tag{5.22}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

A key feature of the above probability density is that the eigenvalues are not independent, since the term $\prod_{k<l} |\lambda_k - \lambda_l|$ indicates that the probability density vanishes when two eigenvalues tend towards one another. This can be interpreted as some effective "repulsion" between eigenvalues, as we will expand on now using an analogy with Coulomb gases.

Exercise 5.1.1 Vandermonde determinant for 2×2 matrices

In this exercise we will explicitly compute the Vandermonde determinant for 2×2 matrices. We define $\mathbf{O}$ and Λ as

$$\mathbf{O} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \text{ and } \Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \tag{5.23}$$

Then any 2×2 symmetric matrix can be written as $\mathbf{M} = \mathbf{O}\Lambda\mathbf{O}^T$.

(a) Write explicitly $\mathbf{M}_{11}$, $\mathbf{M}_{12}$ and $\mathbf{M}_{22}$ as a function of λ_1, λ_2 and θ.

(b) Compute the 3×3 matrix Δ of partial derivatives of $\mathbf{M}_{11}$, $\mathbf{M}_{12}$ and $\mathbf{M}_{22}$ with respect to λ_1, λ_2 and θ.

(c) In the special cases where θ equals 0, $\pi/4$ and $\pi/2$ show that $|\det \Delta| = |\lambda_1 - \lambda_2|$. If you have the courage show that $|\det \Delta| = |\lambda_1 - \lambda_2|$ for all θ.

Exercise 5.1.2 Wigner surmise

Wigner was interested in the distribution of energy level spacings in heavy nuclei, which he modeled as the eigenvalues of a real symmetric random matrix (time reversal symmetry imposes that the Hamiltonian be real). Let $x = |\lambda_{k+1} - \lambda_k|$ for k in the bulk. In principle we can obtain the probability density of x by using Eq. (5.22) and integrating out all other variables. In practice it is very difficult to go much beyond $N = 2$. Since the $N = 2$ result (properly normalized) has the correct small x and large x behavior, Wigner surmised that it must be a good approximation at any N.

(a) For an $N = 2$ GOE matrix, i.e. $V(\lambda) = \lambda^2/2\sigma^2$, write the unnormalized law of λ_1 and λ_2, its two eigenvalues.

(b) Change variables to $\lambda_\pm = \lambda_2 \pm \lambda_1$, integrate out λ_+ and write the unnormalized law of $x = |\lambda_-|$.

(c) Normalize your law and choose σ such that $\mathbb{E}[x] = 1$; you should find

$$P(x) = \frac{\pi}{2} x \exp\left[-\frac{\pi}{4}x^2\right]. \tag{5.24}$$

(d) Using Eq. (5.26) redo the computation for GUE ($\beta = 2$). You should find

$$P(x) = \frac{32}{\pi^2} x^2 \exp\left[-\frac{4}{\pi}x^2\right]. \tag{5.25}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

5.2 Coulomb Gas and Maximum Likelihood Configurations

5.2.1 A Coulomb Gas Analogy

The orthogonal ensemble defined in the previous section can be generalized to complex or quaternion Hermitian matrices. The corresponding joint distribution of eigenvalues is simply obtained by adding a factor of β (equal to 1, 2 or 4) to both the potential and the Vandermonde determinant:

$$P(\{\lambda_i\}) = Z_N^{-1} \left[\prod_{k<l} |\lambda_k - \lambda_l|^\beta \right] \exp\left\{ -\frac{\beta}{2} \left[\sum_{i=1}^N N V(\lambda_i) \right] \right\}$$
$$= Z_N^{-1} \exp\left\{ -\frac{\beta}{2} \left[\sum_{i=1}^N N V(\lambda_i) - \sum_{\substack{i,\, j=1 \\ j\neq i}}^N \log|\lambda_i - \lambda_j| \right] \right\}. \tag{5.26}$$

This joint law is exactly the Gibbs–Boltzmann factor ($\mathrm{e}^{-E/T}$) for a gas of N particles moving on a one-dimensional line, at temperature $T = 2/\beta$, whose potential energy is given by $NV(x)$ and that interact with each other via a pairwise repulsive force generated by the potential $V_{\mathrm{R}}(x, y) = -\log(|x - y|)$. Formally, the repulsive term happens to be the Coulomb potential in two dimensions for particles that all have the same sign. In a truly one-dimensional problem, the Coulomb potential would read $V_{1\mathrm{d}}(x, y) = -|x - y|$, but with a slight abuse of language one speaks about the eigenvalues of a random matrix as a *Coulomb gas* (in one dimension).

Even though we are interested in one particular value of β (namely $\beta = 1$), we can build an intuition by considering this system at various temperatures. At very low temperature (i.e. $\beta \to \infty$), the N particles all want to minimize their potential energy and sit at the minimum of $NV(x)$, but if they try to do so they will have to pay a high price in interaction energy as this energy increases as the particles get close to one another. The particles will have to spread themselves around the minimum of $NV(x)$ to minimize the sum of the potential and interaction energy and find the configuration corresponding to "mechanical equilibrium", i.e. such that the total force on each particle is zero. At non-zero temperature (finite β) the particles will fluctuate around this equilibrium solution. Since the repulsion energy diverges as any two eigenvalues get infinitely close, the particles will always avoid each other. Figure 5.2 shows a typical configuration of particles/eigenvalues for $N = 20$ at $\beta = 1$ in a quadratic potential (GOE matrix).

In the next section, we will study this equilibrium solution, which is exact at low temperature $\beta \to \infty$ or when $N \to \infty$, and is the maximum likelihood solution at finite β and finite N.

5.2.2 Maximum Likelihood Configuration and Stieltjes Transform

In the previous section, we saw that in the Coulomb gas analogy, $\beta \to \infty$ corresponds to the zero temperature limit, and that in this limit the eigenvalues *freeze* to the minimum of

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

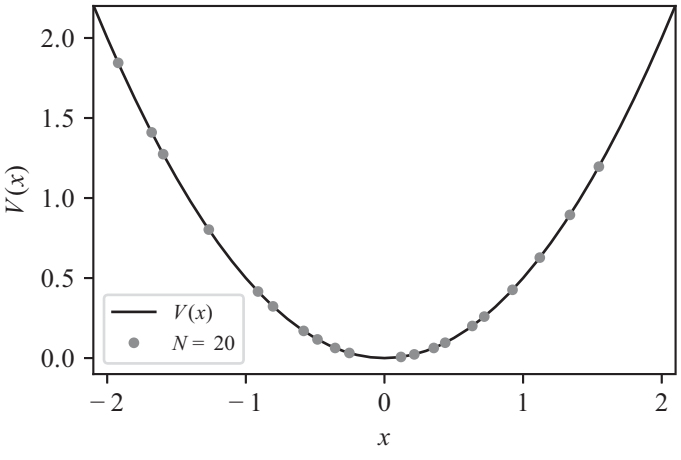

Figure 5.2 Representation of a typical $N = 20$ GOE matrix as a Coulomb gas. The full curve represents the potential $V(x) = x^2/2$ and the 20 dots, the positions of the eigenvalues of a typical configuration. In this analogy, the eigenvalues feel a potential $NV(x)$ and a repulsive pairwise interaction $V(x, y) = -\log(|x - y|)$. They fluctuate according to the Boltzmann weight $\mathrm{e}^{-\beta E/2}$ with $\beta = 1$ in the present case.

the energy (potential plus interaction). We will argue that this freezing (or *concentration of the equilibrium measure*) also happens when $N \to \infty$ for fixed β.

Let us study the minimum energy configuration. We can rewrite Eq. (5.26) as

$$P(\{\lambda_i\}) \propto \mathrm{e}^{\frac{1}{2}\beta N \mathcal{L}(\{\lambda_i\})}, \quad \mathcal{L}(\{\lambda_i\}) = -\sum_{i=1}^{N} V(\lambda_i) + \frac{1}{N} \sum_{\substack{i,\, j=1 \\ j \neq i}}^{N} \log|\lambda_i - \lambda_j|, \tag{5.27}$$

where $\mathcal{L}$ stands for "log-likelihood". For finite N and finite β, we can still consider the solution that maximizes $\mathcal{L}(\{\lambda_i\})$. This is the maximum likelihood solution, i.e. the configuration of $\{\lambda_i\}$ that has maximum probability. The maximum of $\mathcal{L}$ is determined by the equations

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \Rightarrow V'(\lambda_i) = \frac{2}{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{\lambda_i - \lambda_j}. \tag{5.28}$$

These are N coupled equations of N variables which can get very tedious to solve even for moderate values of N. In Exercise 5.2.1 we will find the solution for $N = 3$ in the Wigner case. The solution of these equations is the set of equilibrium positions of all the eigenvalues, i.e. the set of eigenvalues that maximizes the joint probability. To characterize this solution (which will allow us to obtain the density of eigenvalues), we will compute the Stieltjes transform of the $\{\lambda_i\}$ satisfying Eq. (5.28). The trick is to make algebraic manipulations to both sides of the equation to make the Stieltjes transform explicitly appear.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

In a first step we multiply both sides of Eq. (5.28) by $1/(z-\lambda_i)$ where z is a complex variable not equal to any eigenvalues, and then we sum over the index i. This gives

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\frac{V'(\lambda_i)}{z-\lambda_i} &= \frac{2}{N^2}\sum_{\substack{i,j=1\\ j\neq i}}^{N}\frac{1}{(\lambda_i-\lambda_j)(z-\lambda_i)} \\ &= \frac{1}{N^2}\left[\sum_{\substack{i,j=1\\ j\neq i}}^{N}\frac{1}{(\lambda_i-\lambda_j)(z-\lambda_i)}+\sum_{\substack{i,j=1\\ j\neq i}}^{N}\frac{1}{(\lambda_j-\lambda_i)(z-\lambda_j)}\right] \\ &= \frac{1}{N^2}\sum_{\substack{i,j=1\\ j\neq i}}^{N}\frac{1}{(z-\lambda_i)(z-\lambda_j)} = g_N^2(z)-\frac{1}{N^2}\sum_{i=1}^{N}\frac{1}{(z-\lambda_i)^2} \\ &= g_N^2(z)+\frac{g_N'(z)}{N}, \end{aligned} \tag{5.29}$$

where $g_N(z)$ is the Stieltjes transform at finite N:

$$g_N(z) := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_i}. \tag{5.30}$$

We still need to handle the left hand side of the above equation. First we add and subtract $V'(z)$ on the numerator, yielding

$$\frac{1}{N}\sum_{i=1}^{N}\frac{V'(\lambda_i)}{z-\lambda_i} = V'(z)g_N(z)-\frac{1}{N}\sum_{i=1}^{N}\frac{V'(z)-V'(\lambda_i)}{z-\lambda_i} = V'(z)g_N(z)-\Pi_N(z), \tag{5.31}$$

where we have defined a new function $\Pi_N(z)$ as

$$\Pi_N(z) := \frac{1}{N}\sum_{i=1}^{N}\frac{V'(z)-V'(\lambda_i)}{z-\lambda_i}. \tag{5.32}$$

This does not look very useful as the equation for $g_N(z)$ will depend on some unknown function $\Pi_N(z)$ that depends on the eigenvalues whose statistics we are trying to determine. The key realization is that if $V'(z)$ is a polynomial of degree k then $\Pi_N(z)$ is also a polynomial and it has degree $k-1$. Indeed, for each i in the sum, $V'(z)-V'(\lambda_i)$ is a degree k polynomial having $z=\lambda_i$ as a zero, so $(V'(z)-V'(\lambda_i))/(z-\lambda_i)$ is a polynomial of degree $k-1$. $\Pi_N(z)$ is the sum of such polynomials so is itself a polynomial of degree $k-1$.

In fact, the argument is easy to generalize to the Laurent polynomials, i.e. such that $z^k V'(z)$ is a polynomial for some $k \in \mathbb{N}$. For example, in the Wishart case we have a Laurent polynomial

$$V'(z) = \frac{1}{q}\left(1+\frac{q-1}{z}\right). \tag{5.33}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

Nevertheless, from now on we make the assumption that $V'(z)$ is a polynomial. We will later discuss how to relax this assumption.

Thus we get from Eq. (5.29) that

$$V'(z)g_N(z) - \Pi_N(z) = g_N^2(z) + \frac{g_N'(z)}{N} \tag{5.34}$$

for some polynomial Π_N of degree $\deg(V'(z)) - 1$, which needs to be determined self-consistently using Eq. (5.32). For a given $V'(z)$, the coefficients of Π_N are related to the moments of the $\{\lambda_i\}$, which themselves can be obtained from expanding $g_N(z)$ around infinity. In some cases, Eq. (5.34) can be solved exactly at finite N, for example in the case where $V(z) = z^2/2$, in which case the solution can be expressed in terms of Hermite polynomials – see Chapter 6. In the present chapter we will study this equation in the large N limit, which will allow us to derive a general formula for the limiting density of eigenvalues. Note that since Eq. (5.34) does not depend on the value of β, the corresponding eigenvalue density will also be independent of β.

Exercise 5.2.1 Maximum likelihood for 3×3 Wigner matrices

In this exercise we will write explicitly the three eigenvalues of the maximum likelihood configuration of a 3×3 GOE matrix. The potential for this ensemble is $V(x) = x^2/2$.

(a) Let $\lambda_1, \lambda_2, \lambda_3$ be the three maximum likelihood eigenvalues of the 3×3 GOE ensemble in decreasing order. By symmetry we expect $\lambda_3 = -\lambda_1$. What do you expect for λ_2?

(b) Consider Eq. (5.28). Assuming ($\lambda_3 = -\lambda_1$), check that your guess for λ_2 is indeed a solution. Now write the equation for λ_1 and solve it.

(c) Using your solution and the definition (5.30), show that the Stieltjes transform of the maximum likelihood configuration is given by

$$g_3(z) = \frac{z^2 - \frac{1}{3}}{z^3 - z}. \tag{5.35}$$

(d) In the simple case $V(x) = x^2/2$, the zero-degree polynomial $\Pi_N(z)$ is just a constant (independent of N) that can be evaluated from the definition (5.32). What is this constant?

(e) Verify that your $g_3(z)$ satisfies Eq. (5.34) with $N = 3$.

5.2.3 The Large N Limit

In the large N limit, $g_N(z)$ is self-averaging so computing $g_N(z)$ for the most likely configuration is the same as computing the average $\mathfrak{g}(z)$. As $N \to \infty$, Eq. (5.34) becomes

$$V'(z)\mathfrak{g}(z) - \Pi(z) = \mathfrak{g}^2(z). \tag{5.36}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

Each value of N gives a different degree-$(k-1)$ polynomial $\Pi_N(z)$. From the definition (5.32), we can show that the coefficients of $\Pi_N(z)$ are related to the moments of the maximum likelihood configuration of size N. In the large N limit these moments converge so the sequence $\Pi_N(z)$ converges to a well-defined polynomial of degree $(k-1)$ which we call $\Pi(z)$.

Since Eq. (5.36) is quadratic in $\mathfrak{g}(z)$, its solution is given by

$$\mathfrak{g}(z) = \frac{1}{2}\left(V'(z) \pm \sqrt{V'(z)^2 - 4\Pi(z)}\right), \tag{5.37}$$

where we have to choose the branch where $\mathfrak{g}(z)$ goes to zero for large $|z|$.

The eigenvalues of $\mathbf{M}$ will be located where $\mathfrak{g}(z)$ has an imaginary part for z very close to the real axis. The first term $V'(z)$ is a real polynomial and is always real for real z. The expression $V'(z)^2 - 4\Pi(z)$ is also a real polynomial so $\mathfrak{g}(z)$ cannot be complex on the real axis unless $V'(z)^2 - 4\Pi(z) < 0$. In this case $\sqrt{V'(z)^2 - 4\Pi(z)}$ is purely imaginary. We conclude that, when x is such that $\rho(x) \neq 0$,

$$\mathrm{Re}(\mathfrak{g}(x)) := \fint \frac{\rho(\lambda)\mathrm{d}\lambda}{x-\lambda} = \frac{V'(x)}{2}, \tag{5.38}$$

where $\fint$ denotes the principal part of the integral. $\mathrm{Re}(\mathfrak{g}(x))/\pi$ is also called the Hilbert transform of $\rho(\lambda)$.[4]

We have thus shown that the Hilbert transform of the density of eigenvalues is (within its support) equal to $\pi/2$ times the derivative of the potential. We thus realize that the potential outside the support of the eigenvalue has no effect on the distribution of eigenvalues. This is natural in the Coulomb gas analogy. At equilibrium, the particles do not feel the potential away from where they are. One consequence is that we can consider potentials that are not confining at infinity as long as there is a confinement region and that all eigenvalues are within that region. For example, we will consider the quartic potential $V(x) = x^2/2 + \gamma x^4/4$. For small negative γ the region around $x = 0$ is convex. If all eigenvalues are found to be contained in that region, we can modify at will the potential away from it so that $V(x) \to +\infty$ for $|x| \to \infty$ and keep Eq. (5.1) normalizable, as a probability density should be.

Suppose now that we have a potential that is not a polynomial. In a finite region we can approximate it arbitrarily well by a polynomial of sufficiently high degree. If we choose the region of approximation such that for every successive approximation all eigenvalues lie in that region, we can take the limits of these approximations and find that Eq. (5.38) holds even if $V'(x)$ is not a polynomial.

We can also ask the reverse question. Given a density $\rho(x)$, does there exist a model from the orthogonal ensemble (or other β-ensemble) that has $\rho(\lambda)$ as its eigenvalue density? If the Hilbert transform of $\rho(x)$ is well defined, then the answer is yes and Eq. (5.38) gives the

[4] This is a slight abuse of terms, however, that can lead to paradoxes, if one extends Eq. (5.38) to the region where $\rho(x) = 0$. In this case, the right hand side of the equation is *not* equal to $V'(x)/2$. See also Appendix A.2.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

corresponding potential. Note that the potential is only defined up to an additive constant (it can be absorbed in the normalization of Eq. (5.1)) so knowing its derivative is enough to compute $V(x)$. Note also that we only know the value of $V(x)$ on the support of $\rho(x)$; outside this support we can arbitrarily choose $V(x)$ provided it is convex and goes to infinity as $|x| \to \infty$.

Exercise 5.2.2 Matrix potential for the uniform density

In Exercise 2.3.3, we saw that the Stieltjes transform for a uniform density of eigenvalues between 0 and 1 is given by

$$\mathfrak{g}(z) = \log\left(\frac{z}{z-1}\right). \tag{5.39}$$

(a) By computing $\mathrm{Re}(\mathfrak{g}(x))$ for x between 0 and 1, find $V'(x)$ using Eq. (5.38).

(b) Compute the Hilbert transform of the uniform density to recover your answer in (a).

(c) From your answer in (a) and (b), show that the matrix potential is given by

$$V(x) = 2[(1-x)\log(1-x) + x\log(x)] + C \quad \text{for } 0 < x < 1, \tag{5.40}$$

where C is an arbitrary constant. Note that for $x < 0$ and $x > 1$ the potential should be completed by a convex function that goes to infinity as $|x| \to \infty$.

5.3 Applications: Wigner, Wishart and the One-Cut Assumption

5.3.1 Back to Wigner and Wishart

Now we apply the results of the previous section to the Gaussian orthogonal case where $V(z) = z^2/2\sigma^2$. In this simple case, $\Pi(z)$ can be computed from its definition without knowing the eigenvalues, since

$$V'(z) = \frac{z}{\sigma^2} \Rightarrow \Pi(z) = \frac{1}{\sigma^2}. \tag{5.41}$$

Then (5.37) gives

$$\mathfrak{g}(z) = \frac{z - \sqrt[\oplus]{z^2 - 4\sigma^2}}{2\sigma^2}, \tag{5.42}$$

which recovers, independently of the value of β, the Wigner result Eq. (2.38), albeit within a completely different framework (the notation $\sqrt[\oplus]{\cdot}$ was introduced in Section 4.2.3). In particular, the cavity method does not assume that the matrix ensemble is rotationally invariant, as we do here.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

In the Wishart case, we only consider the case $q < 1$, otherwise ($q \geq 1$) the potential is not confining and we need to impose the positive semi-definiteness of the matrix to avoid eigenvalues running to minus infinity. We have

$$V'(z) = \frac{1}{q}\left(1 + \frac{q-1}{z}\right). \tag{5.43}$$

In this case $zV'(z)$ is of degree one, so $z\Pi(z)$ is a polynomial of degree zero:

$$\Pi(z) = \frac{c}{z} \tag{5.44}$$

for some constant c. Thus (5.37) then gives

$$\mathfrak{g}(z) = \frac{z + q - 1 - \overset{\oplus}{\sqrt{(z+q-1)^2 - 4cq^2 z}}}{2qz}. \tag{5.45}$$

As $z \to +\infty$, this expression becomes

$$\mathfrak{g}(z) = \frac{cq}{z} + O(1/z^2). \tag{5.46}$$

Imposing $\mathfrak{g}(z) \sim z^{-1}$ gives $c = q^{-1}$. After some manipulations, we recover Eq. (4.40).

5.3.2 General Convex Potentials and the One-Cut Assumption

For more general polynomial potentials, finding an explicit solution for the limiting Stieltjes transform is a little bit more involved. We recall Eq. (5.37):

$$\mathfrak{g}(z) = \frac{1}{2}\left(V'(z) \pm \sqrt{V'(z)^2 - 4\Pi(z)}\right). \tag{5.47}$$

For a particular polynomial $V'(z)$, $\Pi(z)$ is a polynomial that depends on the moments of the matrix $\mathbf{M}$. The expansion of $\mathfrak{g}(z)$ near $z \to \infty$ will give a set of self-consistent equations for the coefficients of $\Pi(z)$.

The problem simplifies greatly if the support of density of eigenvalues is compact, i.e. if the density $\rho(\lambda)$ is non-zero for all λ's between two edges λ_- and λ_+. We expect this to be true if the potential $V(x)$ is convex. Indeed, by the Coulomb gas analogy we could place all eigenvalues near the minimum of $V(x)$ and let them find their equilibrium configuration by repelling each other. For a convex potential it is natural to assume that the equilibrium configuration would not have any gaps. This assumption is equivalent to assuming that the limiting Stieltjes transform has a single branch cut (from λ_- and λ_+), hence the name one-cut assumption.

So, for a convex polynomial potential $V(x)$, we expect that there exists a well-defined equilibrium density $\rho(\lambda)$ that is non-zero if and only if $\lambda_- < \lambda < \lambda_+$ and that $\mathfrak{g}(z)$ satisfies

$$\mathfrak{g}(z) = \int_{\lambda_-}^{\lambda_+} \frac{\rho(\lambda)}{z - \lambda} \mathrm{d}\lambda. \tag{5.48}$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

From this equation we notice three important properties of $\mathfrak{g}(z)$:

- The function $\mathfrak{g}(z)$ is potentially singular at λ_- and λ_+.
- Near the real axis ($\mathrm{Im}\, z = 0^+$) $g(z)$ has an imaginary part if $z \in (\lambda_-, \lambda_+)$ and is real otherwise.
- The function $\mathfrak{g}(z)$ is analytic everywhere else.

If we go back to Eq. (5.37), we notice that any non-analytic behavior must come from the square-root. On the real axis, the only way $g(z)$ can have an imaginary part is if $D(z) := V'(z)^2 - 4\Pi(z) < 0$ for some values of z. So $D(z)$ (a polynomial of degree $2k$) must change sign at some values λ_- and λ_+, hence these must be zeros of the polynomial. On the real axis, the other possible zeros $D(z)$ can only be of even multiplicity (otherwise $D(z)$ would change sign). Elsewhere in the complex plane, zeros should also be of even multiplicity, otherwise $\sqrt{D(z)}$ would be singular at those zeros. In other words $D(z)$ must be of the form

$$D(z) = (z - \lambda_-)(z - \lambda_+)Q^2(z), \tag{5.49}$$

for some polynomial $Q(z)$ of degree $k-1$ where k is the degree of $V'(z)$. We can therefore write $\mathfrak{g}(z)$ as

$$\mathfrak{g}(z) = \frac{V'(z) \pm Q(z)\sqrt{(z-\lambda_-)(z-\lambda_+)}}{2}, \tag{5.50}$$

where again $Q(z)$ is a polynomial with real coefficients of degree one less than $V'(z)$. The condition that

$$\mathfrak{g}(z) \to \frac{1}{z} \text{ when } |z| \to \infty \tag{5.51}$$

is now sufficient to compute $Q(z)$ and also $\lambda_\pm$ for a given potential $V(z)$. Indeed, expanding Eq. (5.50) near $z \to \infty$, the coefficients of $Q(z)$ and the values $\lambda_\pm$ must be such as to cancel the $k+1$ polynomial coefficients of $V'(z)$ and also ensure that the $1/z$ term has unit coefficient. This gives $k+2$ equations to determine the k coefficients of $Q(z)$ and the two edges $\lambda_\pm$, see next section for an illustration.

Once the polynomial $Q(x)$ is determined, we can read off the eigenvalue density:

$$\rho(\lambda) = \frac{Q(\lambda)\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi} \quad \text{for} \quad \lambda_- \le \lambda \le \lambda_+. \tag{5.52}$$

We see that generically the eigenvalue density behaves as $\rho(\lambda_\pm \mp \delta) \propto \sqrt{\delta}$ near both edges of the spectrum. If by chance (or by construction) one of the edges is a zero of $Q(z)$, then the behavior changes to δ^θ near that edge, with $\theta = n + \frac{1}{2}$ and n the multiplicity of root of $Q(z)$. A potential with generic $\sqrt{\delta}$ behavior at the edge of the density is called non-critical. Other non-generic cases are called critical. In Section 14.1 we will see how the $\sqrt{\delta}$ edge singularity is smoothed over a region of width $N^{-2/3}$ for finite N. In the critical case, the smoothed region is of width $N^{-2/(3+2n)}$.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

5.3.3 $M^2 + M^4$ Potential

One of the original motivations of Brézin, Itzykson, Parisi and Zuber to study the ensemble defined by Eq. (5.1) was to count the so-called planar diagrams in some field theories. To do so they considered the potential

$$V(x) = \frac{x^2}{2} + \frac{\gamma x^4}{4}. \tag{5.53}$$

We will not discuss how one can count planar diagrams from such a potential, but use this example to illustrate the general recipe given in the main text to compute the Stieltjes transform and the density of eigenvalues. Interestingly, for a certain value of γ, the edge singularity is $\delta^{3/2}$ instead of $\sqrt{\delta}$.

Since the potential is symmetric around zero we expect $\lambda_+ = -\lambda_- =: 2a$. We introduce this extra factor of 2, so that if $\gamma = 0$, we obtain the semi-circular law with $a = 1$. Since $V'(z) = z + \gamma z^3$ is a degree three polynomial, we write

$$Q(z) = a_0 + a_1 z + \gamma z^2, \tag{5.54}$$

where the coefficient of z^2 was chosen to cancel the γz^3 term at infinity. Expanding Eq. (5.50) near $z \to \infty$ and imposing $g(z) = 1/z + O(1/z^2)$ we get

$$\begin{aligned} a_1 &= 0, \\ 1 - a_0 + 2\gamma a^2 &= 0, \\ 2a^4\gamma + 2a^2 a_0 &= 2. \end{aligned} \tag{5.55}$$

Solving for a_0, we find

$$\mathfrak{g}(z) = \frac{z + \gamma z^3 - (1 + 2\gamma a^2 + \gamma z^2) \overset{\oplus}{\sqrt{z^2 - 4a^2}}}{2}, \tag{5.56}$$

where a is a solution of

$$3\gamma a^4 + a^2 - 1 = 0 \quad \Rightarrow \quad a^2 = \frac{\sqrt{1 + 12\gamma} - 1}{6\gamma}. \tag{5.57}$$

The density of eigenvalues for the potential (5.53) reads

$$\rho(\lambda) = \frac{(1 + 2\gamma a^2 + \gamma\lambda^2)\sqrt{4a^2 - \lambda^2}}{2\pi} \quad \text{for} \quad \gamma > -\frac{1}{12}, \tag{5.58}$$

with a defined as above. For positive values of γ, the potential is confining (it is convex and grows faster than a logarithm for $z \to \pm\infty$). In that case the equation for a always has a solution, so the Stieltjes transform and the density of eigenvalues are well defined; see Figure 5.3. For small negative values of γ, the problem still makes sense. The potential is convex near zero and the eigenvalues will stay near zero as long as the repulsion does not push them too far in the non-convex region.

There is a critical value of γ at $\gamma_c = -1/12$, which corresponds to $a = \sqrt{2}$. At this critical point, $\mathfrak{g}_c(z)$ and the density are given by

$$\mathfrak{g}_c(z) = \frac{z^3}{24}\left[\left(1 - \frac{8}{z^2}\right)^{\frac{3}{2}} - 1 + \frac{12}{z^2}\right] \text{ and } \rho_c(\lambda) = \frac{\left(8 - \lambda^2\right)^{3/2}}{24\pi}. \tag{5.59}$$

At this point the density of eigenvalues at the upper edge ($\lambda_+ = 2\sqrt{2}$) behaves as $\rho(\lambda) \sim (2\sqrt{2} - \lambda)_+^\theta$ with $\theta = 3/2$ and similarly at the lower edge ($\lambda_- = -2\sqrt{2}$). For values of

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

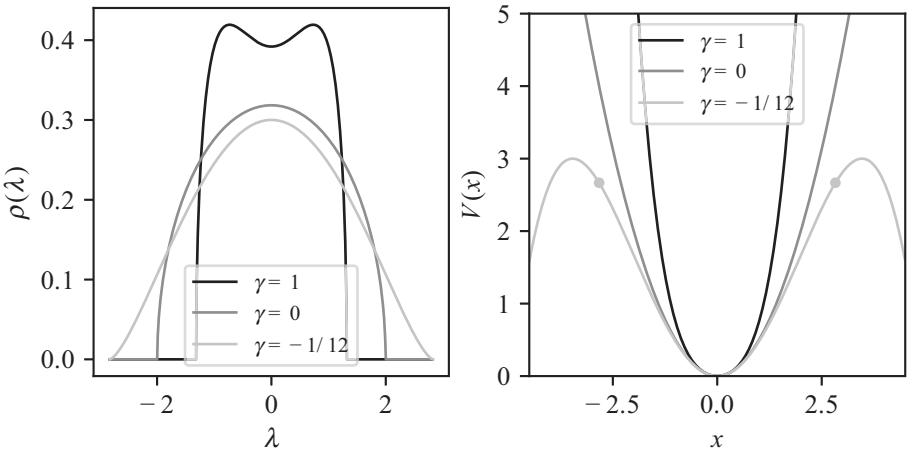

Figure 5.3 (left) Density of eigenvalues for the potential $V(x) = \frac{1}{2}x^2 + \frac{\gamma}{4}x^4$ for three values of γ. For $\gamma = 1$, even if the minimum of the potential is at $\lambda = 0$, the density develops a double hump due to the repulsion of the eigenvalues. $\gamma = 0$ corresponds to the Wigner case (semi-circle law). Finally $\gamma = -1/12$ is the critical value of γ. At this point the density is given by Eq. (5.59). For smaller values of γ the density does not exist. (right) Shape of the potential for the same three values of γ. The dots on the bottom curve indicate the edges of the critical spectrum.

γ more negative than γ_c, there are no real solutions for a and Eq. (5.56) ceases to make sense. In the Coulomb gas analogy, the eigenvalues push each other up to a critical point after which they run off to infinity. There is no simple argument that gives the location of the critical point (except for doing the above computation). It is given by a delicate balance between the repulsion of the eigenvalues and the confining potential. In particular it is not given by the point $V'(2a) = 0$ as one might naively expect. Note that at the critical point $V''(2a) = -1$, so we are already outside the convex region.

5.4 Fluctuations Around the Most Likely Configuration

5.4.1 Fluctuations of All Eigenvalues

The most likely positions of the N eigenvalues are determined by the equations (5.28) that we derived in Section 5.2.2. These equations balance a Coulomb (repulsive) term and a confining potential. On distances d of order $1/N$, the Coulomb force on each "charge" is of order $d^{-1} \sim N$, whereas the confining force V' is of order unity. Therefore, the Coulomb force is dominant at small scales and the most likely positions must be locally equidistant. This is expected to be true within small enough intervals

$$\mathcal{I} := [\lambda - L/2N, \lambda + L/2N], \tag{5.60}$$

where L is sufficiently small, such that the average density $\rho(\lambda)$ does not vary too much, i.e. $\rho'(\lambda)L/N \ll \rho(\lambda)$.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

Of course, some fluctuations around this crystalline order must be expected, and the aim of this section is to introduce a simple method to characterize these fluctuations. Before doing so, let us discuss how the number of eigenvalues $n(L)$ in interval $\mathcal{I}$ behaves. For a perfect crystalline arrangement, there are no fluctuations at all and one has $n(L) = \rho(\lambda)L + O(1)$, where the $O(1)$ term comes from "rounding errors" at the edge of the interval.

For a Poisson process, where points are placed independently at random with some density $N\rho(\lambda)$, then it is well known that

$$n(L) = \bar{n} + \xi\sqrt{\bar{n}}, \quad \text{with} \quad \bar{n} := \rho(\lambda)L, \tag{5.61}$$

where ξ is a Gaussian random variable $\mathcal{N}(0,1)$. For the eigenvalues of a large symmetric random matrix, on the other hand, the exact result is given by

$$n(L) = \bar{n} + \sqrt{\Delta}\xi, \qquad \Delta := \frac{2}{\pi^2}\left[\log(\bar{n}) + C\right] + O(\bar{n}^{-1}), \tag{5.62}$$

where C is a numerical constant.[5] This result means that the typical fluctuations of the number of eigenvalues is of order $\sqrt{\log \bar{n}}$ for large L, much smaller than the Poisson result $\sqrt{\bar{n}}$. In fact, the growth of Δ with L is so slow that one can think of the arrangement of eigenvalues as "quasi-crystalline", even after factoring in fluctuations.

Let us see how this $\sqrt{\log \bar{n}}$ behavior of the fluctuations can be captured by computing the Hessian matrix $\mathbf{H}$ of the log-likelihood $\mathcal{L}$ defined in Eq. (5.27). One finds

$$\mathbf{H}_{ij} := -\frac{\partial^2 \mathcal{L}}{\partial\lambda_i \partial\lambda_j} = \begin{cases} V''(\lambda_i) + \frac{1}{N}\sum_{k\neq i} \frac{2}{(\lambda_i-\lambda_k)^2} & (i=j), \\ -\frac{2}{N}\frac{1}{(\lambda_i-\lambda_j)^2} & (i\neq j). \end{cases}$$

This Hessian matrix should be evaluated at the maximum of $\mathcal{L}$, i.e. for the most likely configurations of the λ_i. If we call ϵ_i/N the *deviation* of λ_i away from its most likely value, and assume all ϵ to be small enough, their joint distribution can be approximated by the following multivariate Gaussian distribution:

$$P(\{\epsilon_i\}) \propto \exp\left[-\frac{\beta}{4N}\sum_{i,j}\epsilon_i \mathbf{H}_{ij}\epsilon_j\right], \tag{5.63}$$

from which one can obtain the covariance of the deviations as

$$\mathbb{E}[\epsilon_i\epsilon_j] = \frac{2N}{\beta}[\mathbf{H}^{-1}]_{ij}. \tag{5.64}$$

Since the most likely positions of the λ_i are locally equidistant, one can approximate $\lambda_i - \lambda_j$ as $(i-j)/N\rho$ in the above expression. This is justified when the contribution of far-away eigenvalues to $\mathbf{H}^{-1}$ is negligible in the region of interest, which will indeed be the case because the off-diagonal elements of $\mathbf{H}$ decay fast enough (i.e. as $(i-j)^{-2}$).

[5] For $\beta = 1$, one finds $C = \log(2\pi) + \gamma + 1 + \frac{\pi^2}{8}$, where γ is Euler's constant, see Mehta [2004].

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

For simplicity, let us consider the case where $V(x) = x^2/2$, in which case $V''(x) = 1$. The matrix $\mathbf{H}$ is then a Toeplitz matrix, up to unimportant boundary terms. It can be diagonalized using plane waves (see Appendix A.3), with eigenvalues given by[6]

$$\mu_q = 1 + 4N\rho^2 \sum_{\ell=1}^{N-1} \frac{1 - \cos\frac{2\pi q\ell}{N}}{\ell^2}, \qquad q = 0, 1, \ldots, N-1, \tag{5.65}$$

where ρ is the local average eigenvalue density. In the large N limit, the (convergent) sum can be replaced by an integral:

$$\mu_q = 1 + 4N\rho^2 \times \frac{2\pi|q|}{N} \int_0^\infty \mathrm{d}u \, \frac{1 - \cos u}{u^2} = 1 + 4\pi^2\rho^2|q|. \tag{5.66}$$

The eigenvalues of $\mathbf{H}^{-1}$ are then given by $1/\mu_q$ and the covariance of the deviations for $i - j = n$ is obtained from Eq. (5.64) as an inverse Fourier transform. After transforming $q \to uN/2\pi$ this reads

$$\mathbb{E}[\epsilon_i \epsilon_j] = \frac{2N}{\beta} \frac{1}{N} \int_{-\pi}^{\pi} \frac{\mathrm{d}u}{2\pi} \frac{\mathrm{e}^{-\mathrm{i}un}}{N^{-1} + 2\pi\rho^2|u|}. \tag{5.67}$$

Now, the fluctuating distance d_{ij} between eigenvalues i and $j = i + n$ is, by definition of the ϵ,

$$d_{ij} = \frac{n}{N\rho} + \frac{\epsilon_i - \epsilon_j}{N}. \tag{5.68}$$

Its variance is obtained, for $N \to \infty$, from

$$\mathbb{E}[(\epsilon_i - \epsilon_j)^2] \approx \frac{2}{\beta\pi^2(\rho N)^2} \int_0^\pi \mathrm{d}u \, \frac{1 - \cos un}{u} \approx_{n \ll 1} \frac{2}{\beta\pi^2\rho^2} \log n. \tag{5.69}$$

The variables ϵ are thus long-ranged, log-correlated Gaussian variables. Interestingly, there has been a flurry of activity concerning this problem in recent years (see references at the end of this chapter).

Finally, the fluctuating local density of eigenvalues can be computed from the number of eigenvalues within a distance d_{ij}, i.e.

$$\frac{1}{N} \frac{n}{d_{ij}} \approx \rho + \rho^2 \frac{\epsilon_i - \epsilon_j}{n} \qquad (n \gg 1). \tag{5.70}$$

Its variance is thus given by

$$\mathbb{V}[\rho] = \frac{2\rho^2}{\beta\pi^2 n^2} \log n. \tag{5.71}$$

Hence, the fluctuation of the number of eigenvalues in a fixed interval of size L/N and containing on average $\bar{n} = \rho L$ eigenvalues is

$$\mathbb{V}[\rho L] = L^2 \mathbb{V}[\rho] = \frac{2}{\beta\pi^2} \log \bar{n}. \tag{5.72}$$

This argument recovers the leading term of the exact result for all values of β (compare with Eq. (5.62) for $\beta = 1$).

[6] In fact, the Hessian $\mathbf{H}$ can be diagonalized exactly in this case, without any approximations, see Agarwal et al. [2019] and references therein.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

5.4.2 Large Deviations of the Top Eigenvalue

Another interesting question concerns the fluctuations of the top eigenvalue of a matrix drawn from a β-ensemble. Consider Eq. (5.26) with $\lambda_{\max}$ isolated:

$$P(\lambda_{\max};\{\lambda_i\}) = P_{N-1}(\{\lambda_i\}) \exp\left\{-\frac{N\beta}{2}\left[V(\lambda_{\max}) - \frac{2}{N}\sum_{i=1}^{N-1}\log(\lambda_{\max}-\lambda_i)\right]\right\}, \quad (5.73)$$

with

$$P_{N-1}(\{\lambda_i\}) := Z_N^{-1} \exp\left\{-\frac{\beta}{2}\left[\sum_{i=1}^{N-1} NV(\lambda_i) - \sum_{\substack{i,j=1\\ j\neq i}}^{N-1}\log|\lambda_i-\lambda_j|\right]\right\}. \quad (5.74)$$

At large N, this probability is dominated by the most likely configuration, which is determined by Eq. (5.27) with $\lambda_{\max}$ removed. Clearly, the most likely positions $\{\lambda_i^*\}$ of the $N-1$ other eigenvalues are only changed by an amount $O(N^{-1})$, but since the log-likelihood is close to its maximum, the change of $\mathcal{L}$ (i.e. the quantity in the exponential) is only of order N^{-2} and we will neglect it. Then one has the following *large deviation* expression:

$$\frac{P(\lambda_{\max};\{\lambda_i^*\})}{P(\lambda_+;\{\lambda_i^*\})} := \exp\left[-\frac{N\beta}{2}\Phi(\lambda_{\max})\right], \quad (5.75)$$

with

$$\Phi(x) = V(x) - V(\lambda_+) - \frac{2}{N}\sum_{i=1}^{N-1}\log(x-\lambda_i^*) + \frac{2}{N}\sum_{i=1}^{N-1}\log(\lambda_+-\lambda_i^*), \quad (5.76)$$

where λ_+ is the top edge of the spectrum. Note that $\Phi(\lambda_+) = 0$ by construction. To deal with the large N limit of this expression, we take the derivative of $\Phi(x)$ with respect to x to find

$$\Phi'(x) = V'(x) - \frac{2}{N}\sum_{i=1}^{N-1}\frac{1}{x-\lambda_i^*} \overset{N\to\infty}{\to} V'(x) - 2\mathfrak{g}(x). \quad (5.77)$$

Hence,

$$\Phi(x) = \int_{\lambda_+}^{x}\big(V'(s) - 2\mathfrak{g}(s)\big)\,\mathrm{d}s. \quad (5.78)$$

When the potential $V'(x)$ is a polynomial of degree $k \geq 1$, we can use Eq. (5.37), yielding

$$\Phi(x) = \int_{\lambda_+}^{x}\sqrt{V'(s)^2 - 4\Pi(s)}\,\mathrm{d}s, \quad (5.79)$$

where λ_+ is the largest zero of the polynomial $V'(s)^2 - 4\Pi(s)$. Since $\Pi(s)$ is a polynomial of order $k-1$, for large s one always has

$$V'(s)^2 \gg |\Pi(s)|, \quad (5.80)$$

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

and therefore

$$\Phi(x) \approx V(x). \tag{5.81}$$

As expected, the probability to observe a top eigenvalue very far from the bulk is dominated by the potential term, and the Coulomb interaction no longer plays any role. When $x - \lambda_+$ is small but still much larger than any inverse power of N, we have

$$\sqrt{V'(s)^2 - 4\Pi(s)} \approx C(s - \lambda_+)^\theta, \tag{5.82}$$

where C is a constant and θ depends on the type of root of $V'(s)^2 - 4\Pi(s)$ at $s = \lambda_+$. For a single root, which is the typical case, $\theta = 1/2$. Performing the integral we get

$$\Phi(\lambda_{\max}) = \frac{C}{\theta + 1}(\lambda_{\max} - \lambda_+)^{\theta+1}, \qquad \lambda_{\max} - \lambda_+ \ll 1. \tag{5.83}$$

Note that the constant C can be determined from the eigenvalue density near (but slightly below) λ_+, to wit, $\pi\rho(\lambda) \approx C(\lambda_+ - \lambda)^\theta$.

For a generic edge with $\theta = 1/2$, one finds that $\Phi(\lambda_{\max}) \propto (\lambda_{\max} - \lambda_+)^{3/2}$, and thus the probability to find $\lambda_{\max}$ just above λ_+ decays as

$$P(\lambda_{\max}) \sim \exp\left[-\frac{2\beta C}{3}u^{3/2}\right]; \qquad u = N^{2/3}(\lambda_{\max} - \lambda_+), \tag{5.84}$$

where we have introduced the rescaled variable u in order to show that this probability decays on scale $N^{-2/3}$. We will further discuss this result in Section 14.1, where we will see that the $u^{3/2}$ behavior coincides with the right tail of the Tracy–Widom distribution.

For a unit Wigner, we have (see Fig. 5.4)

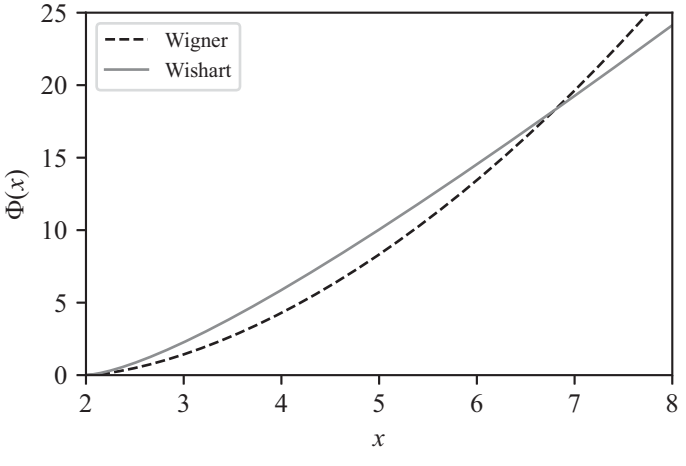

Figure 5.4 Large deviation function for the top eigenvalue $\Phi(x)$ for a unit Wigner and a Wishart matrix with $\lambda_+ = 2$ ($q = 3 - 2\sqrt{2}$). The Wishart curve was obtained by numerically integrating Eq. (5.78).

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

$$\Phi(x) = \frac{1}{2}x\sqrt{x^2-4} - 2\log\left(\frac{\sqrt{x^2-4}+x}{2}\right) \quad \text{for} \quad x > 2. \tag{5.85}$$

For a Wishart matrix with parameter $q = 1$,

$$\Phi(x) = \sqrt{(x-4)x} + 2\log\left(\frac{x-\sqrt{(x-4)x}-2}{2}\right) \quad \text{for} \quad x > 4. \tag{5.86}$$

Remember that the Stieltjes transform $\mathfrak{g}(z)$ and the density of eigenvalues $\rho(\lambda)$ are only sensitive to the potential function $V(x)$ for values in the support of ρ ($\lambda_- \leq x \leq \lambda_+$). The large deviation function $\Phi(x)$, on the other hand, depends on the value of the potential for $x > \lambda_+$. For Eq. (5.79) to hold, the same potential function must extend analytically outside the support of ρ. In Section 5.3.3, we considered a non-confining potential (when $\gamma < 0$). We argued that we could compute $\mathfrak{g}(z)$ and $\rho(\lambda)$ as if the potential were confining as long as we could replace the potential outside (λ_-, λ_+) by a convex function going to infinity. In that case, the function $\Phi(x)$ depends on the choice of regularization of the potential. Computing Eq. (5.78) with the non-confining potential would give nonsensical results.

Exercise 5.4.1 Large deviations for Wigner and Wishart

(a) Show that Eqs. (5.85) and (5.86) are indeed the large deviation function for the top eigenvalues of a unit Wigner and a Wishart $q = 1$. To do so, show that they satisfy Eq. (5.77) and that $\Phi(\lambda_+) = 0$ with $\lambda_+ = 2$ for Wigner and $\lambda_+ = 4$ for Wishart $q = 1$.

(b) Find the dominant contribution of both $\Phi(x)$ for large x. Compare to their respective $V(x)$.

(c) Expand the two expressions for $\Phi(x)$ near their respective λ_+ and show that they have the correct leading behavior. What are the exponent θ and constant C?

5.5 An Eigenvalue Density Saddle Point

As an alternative approach to the most likely configuration of eigenvalues, one often finds in the random matrix literature the following density formalism. One first introduces the density of "charges" $\omega(x)$ for a given set of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ (not necessarily the most likely configuration), as

$$\omega(x) = \frac{1}{N}\sum_{i=1}^{N}\delta(\lambda_i - x). \tag{5.87}$$

Expressed in terms of this density field, the joint distribution of eigenvalues, Eq. (5.26) can be expressed as

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

$$P(\{\omega\}) = Z^{-1} \exp\left(-\frac{\beta N^2}{2}\left[\int \mathrm{d}x\omega(x)V(x) - \oint \mathrm{d}x\mathrm{d}y\omega(x)\omega(y)\log|x-y|\right] - N\int \mathrm{d}x\omega(x)\log\omega(x)\right), \tag{5.88}$$

where the last term is an entropy term, formally corresponding to the change of variables from the $\{\lambda_i\}$ to $\omega(x)$. Since this term is of order N, compared to the two first terms that are of order N^2, it is usually neglected.[7]

One then proceeds by looking for the density field that maximizes the term in the exponential, which is obtained by taking its functional derivative with respect to all $\omega(x)$:

$$\frac{\delta}{\delta\omega(x)}\left[\int \mathrm{d}y\omega(y)V(y) - \oint \mathrm{d}y\mathrm{d}y'\omega(y)\omega(y')\log|y-y'| - \zeta\int \mathrm{d}y\omega(y)\right]_{\omega^*} = 0, \tag{5.89}$$

where ζ is a Lagrange multiplier, used to impose the normalization condition $\int \mathrm{d}x\omega(x) = 1$. This leads to

$$V(x) = 2\oint \mathrm{d}y\omega^*(y)\log|x-y| + \zeta. \tag{5.90}$$

We can now take the derivative with respect to x to get

$$V'(x) = 2\oint \mathrm{d}y\frac{\omega^*(y)}{x-y}, \tag{5.91}$$

which is nothing but the continuum limit version of Eq. (5.28), and is identical to Eq. (5.38). Although this whole procedure looks somewhat ad hoc, it can be fully justified mathematically. In the mathematical literature, it is known as the large deviation formalism.

Equation (5.91) is a singular integral equation for $\omega(x)$ of the so-called Tricomi type. Such equations often have explicit solutions, see Appendix A.2. In the case where $V(x) = x^2/2$, one recovers, as expected, the semi-circle law:

$$\omega^*(x) = \frac{1}{2\pi}\sqrt{4-x^2}. \tag{5.92}$$

One interesting application of the density formalism is to investigate the case of Gaussian orthogonal random matrices conditioned to have all eigenvalues strictly positive. What is the probability for this to occur spontaneously? In such a case, what is the resulting distribution of eigenvalues?

The trick is to solve Eq. (5.91) with the constraint that $\omega(x < 0) = 0$. This leads to

$$x = 2\oint_0^\infty \mathrm{d}y\frac{\omega^*(y)}{x-y}. \tag{5.93}$$

The solution to this truncated problem can also be found analytically using the general result of Appendix A.2. One finds

$$\omega^*(x) = \frac{1}{4\pi}\sqrt{\frac{\lambda_+ - x}{x}}(\lambda_+ + 2x), \quad \text{with} \quad \lambda_+ = \frac{4}{\sqrt{3}}. \tag{5.94}$$

The resulting density has a square-root divergence close to zero, which is a stigma of all the negative eigenvalues being pushed to the positive side. The right edge itself is pushed from 2 to $4/\sqrt{3}$, see Figure 5.5.

[7] Note however that when $\beta = c/N$, as considered in Allez et al. [2012], the entropy term must be retained.

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

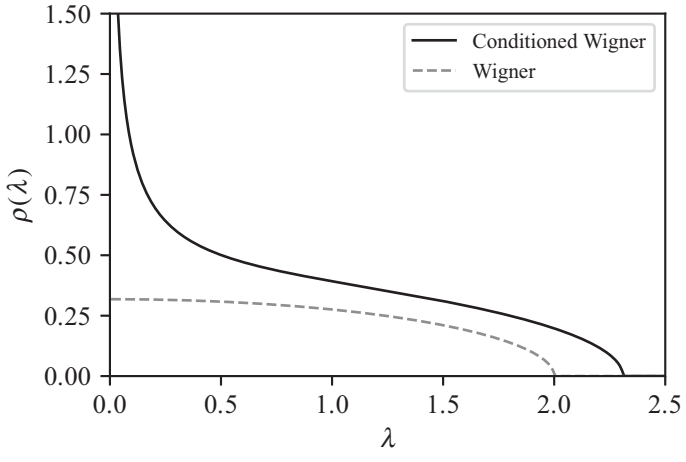

Figure 5.5 Density of eigenvalues for a Wigner matrix conditioned to be positive semi-definite. The positive part of the density of a standard Wigner is shown for comparison.

Injecting this solution back into Eq. (5.88) and comparing with the result corresponding to the standard semi-circle allows one to compute the probability for such an exceptional configuration to occur. After a few manipulations, one finds that this probability is given by $e^{-\beta C N^2}$ with $C = \log 3/4 \approx 0.2746...$. The probability that a Gaussian random matrix has by chance all its eigenvalues positive therefore decreases extremely fast with N.

Other constraints are possible as well, for example if one chooses to confine all eigenvalues in a certain interval $[\ell_-, \ell_+]$ with $\ell_- \geq \lambda_-$ or $\ell_+ \leq \lambda_+$. (In the cases where $\ell_- < \lambda_-$ and $\ell_+ > \lambda_+$ the confinement plays no role.) Let us study this problem in the case where there is no external potential at all, i.e. when $V(x) \equiv C$, where C is an arbitrary constant, but only confining walls. In this case, the general solution of Appendix A.2 immediately leads to

$$\omega^*(x) = \frac{1}{\pi \sqrt{(x - \ell_-)(\ell_+ - x)}}, \tag{5.95}$$

which has a square-root divergence at both edges. This law is called the arcsine law, which appears in different contexts, see Sections 7.2 and 15.3.1.

Note that the minimization of the quadratic form $\int \mathrm{d}x \mathrm{d}y \omega(x)\omega(y)G(|x - y|)$ for a general power-law interaction $G(u) = u^{-\gamma}$, subject to the constraint $\int \mathrm{d}x \omega(x) = 1$, has been solved in the very different, financial engineering context of optimal execution with quadratic costs. The solution in that case reads

$$\omega^*(x) = A(\gamma)\ell^{-\gamma} \sqrt{(x - \ell_-)^{\gamma-1}(\ell_+ - x)^{\gamma-1}}, \tag{5.96}$$

with $\ell := \ell_- + \ell_+$ and $A(\gamma) := \gamma\Gamma[\gamma]/\Gamma^2[(1 + \gamma)/2]$. The case $G(u) = -\log(u)$ formally corresponds to $\gamma = 0$, in which case one recovers Eq. (5.95).

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

Bibliographical Notes

- For a general introduction to the subject, see
  - M. L. Mehta. *Random Matrices*. Academic Press, San Diego, 3rd edition, 2004,
  - P. J. Forrester. *Log Gases and Random Matrices*. Princeton University Press, Princeton, NJ, 2010,

  and for more technical aspects,
  - B. Eynard, T. Kimura, and S. Ribault. Random matrices. *preprint arXiv:1510.04430*, 2006.
- One-cut solution and the $\mathbf{M}^2 + \mathbf{M}^4$ model:
  - E. Brézin, C. Itzykson, G. Parisi, and J.-B. Zuber. Planar diagrams. *Communications in Mathematical Physics*, 59(1):35–51, 1978.
- On the relation between log gases and the Calogero model, see
  - S. Agarwal, M. Kulkarni, and A. Dhar. Some connections between the classical Calogero-Moser model and the log gas. *Journal of Statistical Physics*, 176(6), 1463–1479, 2019.
- About spectral rigidity and log-fluctuations of eigenvalues, see
  - M. L. Mehta. *Random Matrices*. Academic Press, San Diego, 3rd edition, 2004,
  - M. V. Berry. Semiclassical theory of spectral rigidity. *Proceedings of the Royal Society of London, Series A*, 400(1819):229–251, 1985,

  for classical results, and
  - Y. V. Fyodorov and J. P. Keating. Freezing transitions and extreme values: Random matrix theory, and disordered landscapes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 372(2007):2012053, 2014,
  - Y. V. Fyodorov, B. A. Khoruzhenko, and N. J. Simm. Fractional Brownian motion with Hurst index $H = 0$ and the Gaussian Unitary Ensemble. *The Annals of Probability*, 44(4):2980–3031, 2016,
  - R. Chhaibi and J. Najnudel. On the circle, $gmc^{\gamma} = c\beta e_{\infty}$ for $\gamma = \sqrt{2/\beta}$, $(\gamma \leq 1)$. *preprint arXiv:1904.00578*, 2019,

  for the recent spree of activity in the field of log-correlated variables.
- On the probability of a very large eigenvalue, see Chapter 14 and
  - S. N. Majumdar and M. Vergassola. Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices. *Physical Review Letters*, 102:060601, 2009.
- On the continuous density formalism, see
  - D. S. Dean and S. N. Majumdar. Extreme value statistics of eigenvalues of Gaussian random matrices. *Physical Review E*, 77:041108, 2008

  for a physicists' introduction, and

https://doi.org/10.1017/9781108768900.007 Published online by Cambridge University Press

  - G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probability Theory and Related Fields*, 108(4):517–542, 1997,
  - G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, Cambridge, 2010

  for more mathematical discussions in the context of large deviation theory. A case where the entropy term must be retained is discussed in
  - R. Allez, J. P. Bouchaud, and A. Guionnet. Invariant beta ensembles and the Gauss-Wigner crossover. *Physical Review Letters*, 109(9):094102, 2012.
- On the relation with optimal execution problems, see
  - J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, 22(3):445–474, 2012,
  - J.-P. Bouchaud, J. Bonart, J. Donier, and M. Gould. *Trades, Quotes and Prices*. Cambridge University Press, Cambridge, 2nd edition, 2018.