2

Wigner Ensemble and Semi-Circle Law

In many circumstances, the matrices that are encountered are large, and with no particular structure. Physicist Eugene Wigner postulated that one can often replace a large complex (but deterministic) matrix by a typical element of a certain ensemble of random matrices. This bold proposal was made in the context of the study of large complex atomic nuclei, where the "matrix" is the Hamiltonian of the system, which is a Hermitian matrix describing all the interactions between the neutrons and protons contained in the nucleus. At the time, these interactions were not well known; but even if they had been, the task of diagonalizing the Hamiltonian to find the energy levels of the nucleus was so formidable that Wigner looked for an alternative. He suggested that we should abandon the idea of finding precisely all energy levels, but rather rephrase the question as a *statistical question*: what is the probability to find an energy level within a certain interval, what is the probability that the distance between two successive levels is equal to a certain value, etc.? The idea of Wigner was that the answer to these questions could be, to some degree, universal, i.e. independent of the specific Hermitian matrix describing the system, provided it was *complex enough*. If this is the case, why not replace the Hamiltonian of the system by a purely random matrix with the correct symmetry properties? In the case of time-reversal invariant quantum systems, the Hamiltonian is a real symmetric matrix (of infinite size). In the presence of a magnetic field, the Hamiltonian is a complex, Hermitian matrix (see Section 3.1.1). In the presence of "spin–orbit coupling", the Hamiltonian is symplectic (see Section 3.1.2).

This idea has been incredibly fruitful and has led to the development of a subfield of mathematical physics called "random matrix theory". In this book we will study the properties of some ensembles of random matrices. We will mostly focus on symmetric matrices with real entries as those are the most commonly encountered in data analysis and statistical physics. For example, Wigner's idea has been transposed to glasses and spin-glasses, where the interaction between pairs of atoms or pairs of spins is often replaced by a real symmetric, random matrix (see Section 13.4). In other cases, the randomness stems from noisy observations. For example, when one wants to measure the covariance matrix of the returns of a large number of assets using a sample of finite length (for example the 500 stocks of the S&P500 using 4 years of daily data, i.e. $4 \times 250 = 1000$ data points per stock), there is inevitably some measurement noise that pollutes the

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

determination of said covariance matrix. We will be confronted with this precise problem in Chapters 4 and 17.

In the present chapter and the following one, we will investigate the simplest of all ensembles of random matrices, which was proposed by Wigner himself in the context recalled above. These are matrices where all elements are Gaussian random variables, with the only constraint that the matrix is real symmetric (the Gaussian orthogonal ensemble, GOE), complex Hermitian (the Gaussian unitary ensemble, GUE) or symplectic (the Gaussian symplectic ensemble, GSE).

2.1 Normalized Trace and Sample Averages

We first generalize the notion of expectation value and moments from classical probabilities to large random matrices. We could simply consider the moments $\mathbb{E}[\mathbf{A}^k]$ but that object is very large ($N \times N$ dimensional). It is not clear how to interpret it as $N \to \infty$. It turns out that the correct analog of the expectation value is the normalized trace operator $\tau(.)$, defined as

$$\tau(\mathbf{A}) := \frac{1}{N}\mathbb{E}[\mathrm{Tr}\,\mathbf{A}]. \tag{2.1}$$

The normalization by $1/N$ is there to make the normalized trace operator finite as $N \to \infty$. For example for the identity matrix, $\tau(\mathbf{1}) = 1$ independently of the dimension and our definition therefore makes sense as $N \to \infty$. When using the notation $\tau(\mathbf{A})$ we will only consider the dominant term as $N \to \infty$, implicitly taking the large N limit.

For a polynomial function of a matrix $F(\mathbf{A})$ or by extension for a function that can be written as a power series, the trace of the function can be computed on the eigenvalues:

$$\frac{1}{N}\,\mathrm{Tr}\,F(\mathbf{A}) = \frac{1}{N}\sum_{k=1}^{N} F(\lambda_k). \tag{2.2}$$

In the following, we will denote as $\langle . \rangle$ the average over the eigenvalues of a single matrix $\mathbf{A}$ (sample), i.e.

$$\langle F(\lambda)\rangle := \frac{1}{N}\sum_{k=1}^{N} F(\lambda_k). \tag{2.3}$$

For large random matrices, many scalar quantities such as $\tau(F(\mathbf{A}))$ do not fluctuate from sample to sample, or more precisely such fluctuations go to zero in the large N limit. Physicists speak of this phenomenon as *self-averaging* and mathematicians speak of *concentration of measure*.

$$\tau(F(\mathbf{A})) = \frac{1}{N}\mathbb{E}[\mathrm{Tr}\,F(\mathbf{A})] \approx \langle F(\lambda)\rangle \text{ for a single } \mathbf{A}. \tag{2.4}$$

When the eigenvalues of a random matrix $\mathbf{A}$ converge to a well-defined density $\rho(\lambda)$, we can write

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

$$\tau(F(\mathbf{A})) = \int \rho(\lambda) F(\lambda) \mathrm{d}\lambda. \tag{2.5}$$

Using $F(\mathbf{A}) = \mathbf{A}^k$, we can define the kth moment of a random matrix by $m_k := \tau(\mathbf{A}^k)$. The first moment m_1 is simply the normalized trace of $\mathbf{A}$, while $m_2 = 1/N \sum_{ij} \mathbf{A}_{ij}^2$ is the normalized sum of the squares of all the elements. The square-root of m_2 satisfies the axioms of a norm and is called the *Frobenius norm* of $\mathbf{A}$:

$$||A||_{\mathrm{F}} := \sqrt{m_2}. \tag{2.6}$$

2.2 The Wigner Ensemble

2.2.1 Moments of Wigner Matrices

We will define a Wigner matrix $\mathbf{X}$ as a symmetric matrix ($\mathbf{X} = \mathbf{X}^T$) with Gaussian entries with zero mean. In a symmetric matrix there are really two types of elements: diagonal and off-diagonal, which can have different variances. Diagonal elements have variance σ_{d}^2 and off-diagonal elements have variance σ_{od}^2. Note that $\mathbf{X}_{ij} = \mathbf{X}_{ji}$ so they are not independent variables.

In fact, the elements in a Wigner matrix do not need to be Gaussian or even to be IID, as there are many weaker (more general) definitions of the Wigner matrix that yield the same final statistical results in the limit of large matrices $N \to \infty$. For the purpose of this introductory book we will stick to the strong Gaussian hypothesis.

The first few moments of our Wigner matrix $\mathbf{X}$ are given by

$$\tau(\mathbf{X}) = \frac{1}{N}\mathbb{E}[\mathrm{Tr}\,\mathbf{X}] = \frac{1}{N}\,\mathrm{Tr}\,\mathbb{E}[\mathbf{X}] = 0, \tag{2.7}$$

$$\tau(\mathbf{X}^2) = \frac{1}{N}\mathbb{E}[\mathrm{Tr}\,\mathbf{X}\mathbf{X}^T] = \frac{1}{N}\mathbb{E}\left[\sum_{ij=1}^{N} \mathbf{X}_{ij}^2\right] = \frac{1}{N}[N(N-1)\sigma_{\mathrm{od}}^2 + N\sigma_{\mathrm{d}}^2]. \tag{2.8}$$

The term containing σ_{od}^2 dominates when the two variances are of the same order of magnitude. So for a Wigner matrix we can pick any variance we want on the diagonal (as long as it is small with respect to $N\sigma_{\mathrm{od}}^2$). We want to normalize our Wigner matrix so that its second moment is independent of the size of the matrix (N). Let us pick

$$\sigma_{\mathrm{od}}^2 = \sigma^2/N. \tag{2.9}$$

For σ_{d}^2 the natural choice seems to be $\sigma_{\mathrm{d}}^2 = \sigma^2/N$. However, we will rather choose $\sigma_{\mathrm{d}}^2 = 2\sigma^2/N$, which is easy to generate numerically and more importantly respects rotational invariance for finite N, as we show in the next subsection. The ensemble described here (with the choice $\sigma_{\mathrm{d}}^2 = 2\sigma_{\mathrm{od}}^2$) is called the *Gaussian orthogonal ensemble* or GOE.[1]

[1] Some authors define a GOE matrix to have $\sigma^2 = 1$ others as $\sigma^2 = N$. For us a GOE matrix can have any variance and is thus synonymous with the Gaussian rotationally invariant Wigner matrix.

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

To generate a GOE matrix numerically, first generate a non-symmetric random square matrix $\mathbf{H}$ of size N with IID $\mathcal{N}(0, \sigma^2/(2N))$ coefficients. Then let the Wigner matrix $\mathbf{X}$ be $\mathbf{X} = \mathbf{H} + \mathbf{H}^T$. The matrix $\mathbf{X}$ will then be symmetric with diagonal variance twice the off-diagonal variance. The reason is that off-diagonal terms are sums of two independent Gaussian variables, so the variance is doubled. Diagonal elements, on the other hand, are equal to twice the original variables H_{ii} and so their variance is multiplied by 4.

With any choice of σ_{d}^2 we have

$$\tau(\mathbf{X}^2) = \sigma^2 + O(1/N), \tag{2.10}$$

and hence we will call the parameter σ^2 the variance of the Wigner matrix.

The third moment $\tau(\mathbf{X}^3) = 0$ from the fact that the Gaussian distribution is even. Later we will show that

$$\tau(\mathbf{X}^4) = 2\sigma^4. \tag{2.11}$$

For standard Gaussian variables $\mathbb{E}[x^4] = 3\sigma^4$, this implies that the eigenvalue density of a Wigner is **not** Gaussian. What is this eigenvalue distribution? As we will show many times over in this book, it is given by the semi-circle law, originally derived by Wigner himself:

$$\rho(\lambda) = \frac{\sqrt{4\sigma^2 - \lambda^2}}{2\pi\sigma^2} \text{ for } -2\sigma < \lambda < 2\sigma. \tag{2.12}$$

2.2.2 Rotational Invariance

We remind the reader that to rotate a vector $\mathbf{v}$, one applies a rotation matrix $\mathbf{O}$: $\mathbf{w} = \mathbf{O}\mathbf{v}$ where $\mathbf{O}$ is an orthogonal matrix $\mathbf{O}^T = \mathbf{O}^{-1}$ (i.e. $\mathbf{O}\mathbf{O}^T = \mathbf{1}$). Note that in general $\mathbf{O}$ is not symmetric. To rotate the basis in which a matrix is written, one writes $\widetilde{\mathbf{X}} = \mathbf{O}\mathbf{X}\mathbf{O}^T$. The eigenvalues of $\widetilde{\mathbf{X}}$ are the same as those of $\mathbf{X}$. The eigenvectors are $\{\mathbf{O}\mathbf{v}\}$ where $\{\mathbf{v}\}$ are the eigenvectors of $\mathbf{X}$.

A rotationally invariant random matrix ensemble is such that the matrix $\mathbf{O}\mathbf{X}\mathbf{O}^T$ is as probable as the matrix $\mathbf{X}$ itself, i.e. $\mathbf{O}\mathbf{X}\mathbf{O}^T \overset{\text{in law}}{=} \mathbf{X}$.

Let us show that the construction $\mathbf{X} = \mathbf{H} + \mathbf{H}^T$ with a Gaussian IID matrix $\mathbf{H}$ leads to a rotationally invariant ensemble. First, note an important property of Gaussian variables, namely that a Gaussian IID vector $\mathbf{v}$ (a white multivariate Gaussian vector) is rotationally invariant. The reason is that $\mathbf{w} = \mathbf{O}\mathbf{v}$ is again a Gaussian vector (since sums of Gaussians are still Gaussian), with covariance given by

$$\mathbb{E}[w_i w_j] = \sum_{k\ell} O_{ik} O_{j\ell} \mathbb{E}[v_k v_\ell] = \sum_{k\ell} O_{ik} O_{j\ell} \delta_{k\ell} = [\mathbf{O}\mathbf{O}^T]_{ij} = \delta_{ij}. \tag{2.13}$$

Now, write

$$\mathbf{X} = \mathbf{H} + \mathbf{H}^T, \tag{2.14}$$

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

where $\mathbf{H}$ is a square matrix filled with IID Gaussian random numbers. Each column of $\mathbf{H}$ is rotationally invariant: $\mathbf{OH} \overset{\text{in law}}{=} \mathbf{H}$ and the matrix $\mathbf{OH}$ is row-wise rotationally invariant: $\mathbf{OHO}^T \overset{\text{in law}}{=} \mathbf{OH}$. So $\mathbf{H}$ is rotationally invariant as a matrix. Now

$$\mathbf{OXO}^T = \mathbf{O}(\mathbf{H} + \mathbf{H}^T)\mathbf{O}^T \overset{\text{in law}}{=} \mathbf{H} + \mathbf{H}^T = \mathbf{X}, \tag{2.15}$$

which shows that the Wigner ensemble with $\sigma_{\text{d}}^2 = 2\sigma_{\text{od}}^2$ is rotationally invariant for any matrix size N. More general definitions of the Wigner ensemble (including non-Gaussian ensembles) are only asymptotically rotationally invariant (i.e. when $N \to \infty$).

Another way to see the rotational invariance of the Wigner ensemble is to look at the joint law of matrix elements:

$$P(\{X_{ij}\}) = \left(\frac{1}{2\pi\sigma_{\text{d}}^2}\right)^{N/2} \left(\frac{1}{2\pi\sigma_{\text{od}}^2}\right)^{N(N-1)/4} \exp\left\{-\sum_{i=1}^{N} \frac{X_{ii}^2}{2\sigma_{\text{d}}^2} - \sum_{i<j}^{N} \frac{X_{ij}^2}{2\sigma_{\text{od}}^2}\right\}, \tag{2.16}$$

where only the diagonal and upper triangular elements are independent variables. With the choice $\sigma_{\text{od}}^2 = \sigma^2/N$ and $\sigma_{\text{d}}^2 = 2\sigma^2/N$ this becomes

$$P(\{X_{ij}\}) \propto \exp\left\{-\frac{N}{4\sigma^2} \operatorname{Tr} \mathbf{X}^2\right\}. \tag{2.17}$$

Under the change of variable $\mathbf{X} \to \widetilde{\mathbf{X}} = \mathbf{OXO}^T$ the argument of the exponential is invariant, because the trace of a matrix is independent of the basis, and because the Jacobian of the transformation is equal to 1 (see Section 1.2.7), therefore $\mathbf{X} \overset{\text{in law}}{=} \mathbf{OXO}^T$.

By the same argument any matrix whose joint probability density of its elements can be written as $P(\{M_{ij}\}) \propto \exp\{-N \operatorname{Tr} V(\mathbf{M})\}$, where $V(.)$ is an arbitrary function, will be rotationally invariant. We will study such matrix ensembles in Chapter 5.

2.3 Resolvent and Stieltjes Transform

2.3.1 Definition and Basic Properties

In this section we introduce the Stieltjes transform of a matrix. It will give us information about all the moments of the random matrix and also about the density of its eigenvalues in the large N limit. First we need to define the matrix resolvent.

Given an $N \times N$ real symmetric matrix $\mathbf{A}$, its *resolvent* is given by

$$\mathbf{G}_{\mathbf{A}}(z) = (z\mathbf{1} - \mathbf{A})^{-1}, \tag{2.18}$$

where z is a complex variable defined away from all the (real) eigenvalues of $\mathbf{A}$ and $\mathbf{1}$ denotes the identity matrix. Then the Stieltjes transform of $\mathbf{A}$ is given by[2]

$$g_N^{\mathbf{A}}(z) = \frac{1}{N} \operatorname{Tr}\left(\mathbf{G}_{\mathbf{A}}(z)\right) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{z - \lambda_k}, \tag{2.19}$$

[2] In mathematical literature, the Stieltjes transform is more commonly defined as $s_{\mathbf{A}}(z) = -(1/N) \operatorname{Tr} \mathbf{G}_{\mathbf{A}}(z)$, i.e. with an extra minus sign. Some authors prefer the name Cauchy transform.

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

where λ_k are the eigenvalues of $\mathbf{A}$. The subscript N indicates that this is the finite N Stieltjes transform of a single realization of $\mathbf{A}$. When it is clear from context which matrix we consider we will drop the superscript $\mathbf{A}$ and write $g_N(z)$.

Let us see why the Stieltjes transform gives useful information about the density of eigenvalues of $\mathbf{A}$. For a given random matrix $\mathbf{A}$, we can define the empirical spectral distribution (ESD) also called the sample eigenvalue density:

$$\rho_N(\lambda) = \frac{1}{N}\sum_{k=1}^{N}\delta(\lambda - \lambda_k), \tag{2.20}$$

where $\delta(x)$ is the Dirac delta function. Then the Stieltjes transform can be written as

$$g_N(z) = \int_{-\infty}^{+\infty}\frac{\rho_N(\lambda)}{z-\lambda}\mathrm{d}\lambda. \tag{2.21}$$

Note that $g_N(z)$ is well defined for any $z \notin \{\lambda_k : 1 \le k \le N\}$. In particular, it is well behaved at ∞:

$$g_N(z) = \sum_{k=0}^{\infty}\frac{1}{z^{k+1}}\frac{1}{N}\operatorname{Tr}(\mathbf{A}^k), \qquad \frac{1}{N}\operatorname{Tr}(\mathbf{A}^0) = 1. \tag{2.22}$$

We will consider random matrices $\mathbf{A}$ such that, for large N, the normalized traces of powers of $\mathbf{A}$ converge to their expectation values, which are deterministic numbers:

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{Tr}(\mathbf{A}^k) = \tau(\mathbf{A}^k). \tag{2.23}$$

We then expect that, for large enough z, the function $g_{\mathbf{A}}(z)$ converges to a deterministic limit $\mathfrak{g}(z)$ defined as $\mathfrak{g}(z) = \lim_{N\to\infty}\mathbb{E}[g_N(z)]$, whose Taylor series is

$$\mathfrak{g}(z) = \sum_{k=0}^{\infty}\frac{1}{z^{k+1}}\tau(\mathbf{A}^k), \tag{2.24}$$

for z away from the real axis.

Thus $\mathfrak{g}(z)$ is a moment generating function of $\mathbf{A}$. In other words, the knowledge of $\mathfrak{g}(z)$ near infinity is equivalent to the knowledge of all the moments of $\mathbf{A}$. To the level of rigor of this book, the knowledge of all the moments of $\mathbf{A}$ is equivalent to the knowledge of the density of its eigenvalues. For any function $F(x)$ defined over the support of the eigenvalues $[\lambda_-, \lambda_+]$ of $\mathbf{A}$ we can compute its expectation:

$$\tau(F(\mathbf{A})) = \int_{\lambda_-}^{\lambda_+}\rho(\lambda)F(\lambda)\mathrm{d}\lambda; \qquad \rho(\lambda) := \mathbb{E}[\rho_{\mathbf{A}}(\lambda)]. \tag{2.25}$$

Alternatively we can approximate the function $F(x)$ arbitrarily well by a polynomial $Q(x) = a_0 + a_1 x + \cdots + a_K x^K$ and find

$$\tau(F(\mathbf{A})) \approx \tau(Q(\mathbf{A})) = \sum_{k=0}^{K} a_k \tau(\mathbf{A}^k). \tag{2.26}$$

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

To recap, we only need to know $\mathfrak{g}(z)$ in the neighborhood of $|z| \to \infty$ to know all the moments of $\mathbf{A}$ and these moments tell us everything about $\rho(\lambda)$. In computing the Stieltjes transform in concrete cases, we will often make use of that fact and only estimate it for very large values of z.

The Stieltjes transform also gives the negative moments when they exist. If the eigenvalues of $\mathbf{A}$ satisfy $\min_\ell \lambda_\ell > c$ for some constant $c > 0$, then the inverse moments of $\mathbf{A}$ exist and are given by the expansion of $\mathfrak{g}(z)$ around $z = 0$:

$$\mathfrak{g}(z) = -\sum_{k=0}^{\infty} z^k \tau(\mathbf{A}^{-k-1}). \tag{2.27}$$

In particular, we have

$$\mathfrak{g}(0) = -\tau(\mathbf{A}^{-1}). \tag{2.28}$$

Exercise 2.3.1 Stieltjes transform for shifted and scaled matrices

Let $\mathbf{A}$ be a random matrix drawn from a well-behaved ensemble with Stieltjes transform $\mathfrak{g}(z)$. What are the Stieltjes transforms of the random matrices $\alpha\mathbf{A}$ and $\mathbf{A} + \beta\mathbf{1}$ where α and β are non-zero real numbers and $\mathbf{1}$ the identity matrix?

2.3.2 Stieltjes Transform of the Wigner Ensemble

We are now ready to compute the Stieltjes transform of the Wigner ensemble. The first technique we will use is sometimes called the cavity method or the self-consistent equation. We will find a relation between the Stieltjes transform of a Wigner matrix of size N and one of size $N - 1$. In the large N limit, the two converge to the same limiting Stieltjes transform and give us a self-consistent equation that can be solved easily.

We would like to calculate $\mathfrak{g}_N^{\mathbf{X}}(z)$ when $\mathbf{X}$ is a Wigner matrix, with $\mathbf{X}_{ij} \sim \mathcal{N}(0, \sigma^2/N)$ and $\mathbf{X}_{ii} \sim \mathcal{N}(0, 2\sigma^2/N)$. In the large N limit, we expect that $\mathfrak{g}_N^{\mathbf{X}}(z)$ converges towards a well-defined limit $\mathfrak{g}(z)$.

We can use the Schur complement formula (1.32) to compute the $(1,1)$ element of the inverse of $\mathbf{M} = z\mathbf{1} - \mathbf{X}$. Then we have

$$\frac{1}{(\mathbf{G}_{\mathbf{X}})_{11}} = \mathbf{M}_{11} - \sum_{k,l=2}^{N} \mathbf{M}_{1k} (\mathbf{M}_{22})^{-1}_{kl} \mathbf{M}_{l1}, \tag{2.29}$$

where the matrix $\mathbf{M}_{22}$ is the $(N-1)\times(N-1)$ submatrix of $\mathbf{M}$ with the first row and column removed. For large N, we argue that the right hand side is dominated by its expectation value with small ($O(1/\sqrt{N})$) fluctuations. We will only compute its expectation value, but getting a more precise handle on its fluctuations would not be difficult. First, we note

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

that $\mathbb{E}[\mathbf{M}_{11}] = z$. We then note that the entries of $\mathbf{M}_{22}$ are independent of the ones of $\mathbf{M}_{1i} = -\mathbf{X}_{1i}$. Thus we can first take the partial expectation over the $\{\mathbf{X}_{1i}\}$, and get

$$\mathbb{E}_{\{\mathbf{X}_{1i}\}}\left[\mathbf{M}_{1i}(\mathbf{M}_{22})^{-1}_{ij}\mathbf{M}_{1j}\right] = \frac{\sigma^2}{N}(\mathbf{M}_{22})^{-1}_{ii}\delta_{ij} \tag{2.30}$$

so we have

$$\mathbb{E}_{\{\mathbf{X}_{1i}\}}\left[\sum_{k,l=2}^{N}\mathbf{M}_{1k}(\mathbf{M}_{22})^{-1}_{kl}\mathbf{M}_{l1}\right] = \frac{\sigma^2}{N}\,\mathrm{Tr}\left((\mathbf{M}_{22})^{-1}\right). \tag{2.31}$$

Another observation is that $1/(N-1)\,\mathrm{Tr}\left((\mathbf{M}_{22})^{-1}\right)$ is the Stieltjes transform of a Wigner matrix of size $N-1$ and variance $\sigma^2(N-1)/N$. In the large N limit, the Stieltjes transform should be independent of the matrix size and the difference between N and $(N-1)$ is negligible. So we have

$$\mathbb{E}\left[\frac{1}{N}\,\mathrm{Tr}\left((\mathbf{M}_{22})^{-1}\right)\right] \to \mathfrak{g}(z). \tag{2.32}$$

We therefore have that $1/(\mathbf{G}_{\mathbf{X}})_{11}$ equals a deterministic number with negligible fluctuations; hence in the large N limit we have

$$\mathbb{E}\left[\frac{1}{(\mathbf{G}_{\mathbf{X}})_{11}}\right] = \frac{1}{\mathbb{E}[(\mathbf{G}_{\mathbf{X}})_{11}]}. \tag{2.33}$$

From the rotational invariance of $\mathbf{X}$ and therefore of $\mathbf{G}_{\mathbf{X}}$, all diagonal entries of $\mathbf{G}_{\mathbf{X}}$ must have the same expectation value:

$$\mathbb{E}\left[(\mathbf{G}_{\mathbf{X}})_{11}\right] = \frac{1}{N}\mathbb{E}[\mathrm{Tr}(\mathbf{G}_{\mathbf{X}})] = \mathbb{E}[g_N] \to \mathfrak{g}. \tag{2.34}$$

Putting all the pieces together, we find that in the large N limit Eq. (2.29) becomes

$$\frac{1}{\mathfrak{g}(z)} = z - \sigma^2\mathfrak{g}(z). \tag{2.35}$$

Solving (2.35) we obtain that

$$\sigma^2\mathfrak{g}^2 - z\mathfrak{g} + 1 = 0 \Rightarrow \mathfrak{g} = \frac{z \pm \sqrt{z^2 - 4\sigma^2}}{2\sigma^2}. \tag{2.36}$$

We know that $\mathfrak{g}(z)$ should be analytic for large complex z but the square-root above can run into branch cuts. It is convenient to pull out a factor of z and express the square-root as a function of $1/z$ which becomes small for large z:

$$\mathfrak{g}(z) = \frac{z \pm z\sqrt{1 - 4\sigma^2/z^2}}{2\sigma^2}. \tag{2.37}$$

We can now choose the correct root: the $+$ sign gives an incorrect $\mathfrak{g}(z) \sim z/\sigma^2$ for large z while the $-$ sign gives $\mathfrak{g}(z) \sim 1/z$ for any large complex z as expected, so we have:

$$\mathfrak{g}(z) = \frac{z - z\sqrt{1 - 4\sigma^2/z^2}}{2\sigma^2}. \tag{2.38}$$

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

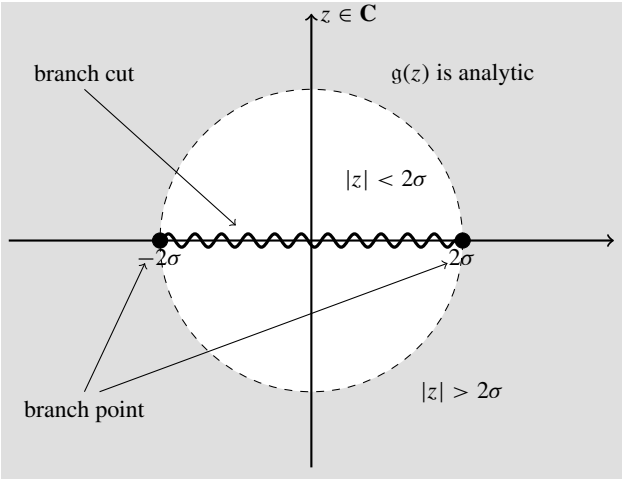

Figure 2.1 The branch cuts of the Wigner Stieltjes transform.

Note, for numerical applications, it is very important to pick the correct branch of the square-root. The function $\mathfrak{g}(z)$ is analytic for $|z| > 2\sigma$, the branch cuts of the square-root must therefore be confined to the interval $[-2\sigma, 2\sigma]$ (Fig. 2.1). We will come back to this problem of determining the correct branch of Stieltjes transform in Section 4.2.3.

It might seem strange that $\mathfrak{g}(z)$ given by Eq. (2.38) has no poles but only branch cuts. For finite N, the sample Stieltjes transform

$$g_N(z) := \frac{1}{N} \sum_{k=1}^{N} \frac{1}{z - \lambda_k} \tag{2.39}$$

has poles at the eigenvalues of $\mathbf{X}$. As $N \to \infty$, the poles fuse together and

$$\frac{1}{N} \sum_{k=1}^{N} \delta(x - \lambda_k) \sim \rho(x). \tag{2.40}$$

The density $\rho(x)$ can have extended support and/or isolated Dirac masses. Then as $N \to \infty$, we have

$$\mathfrak{g}(z) = \int_{\text{supp}\{\rho\}} \frac{\rho(x)\mathrm{d}x}{z - x}, \tag{2.41}$$

which is the Stieltjes transform of the limiting measure $\rho(x)$.

2.3.3 Convergence of Stieltjes near the Real Axis

It is natural to ask the following questions: how does $g_N(z)$ given by Eq. (2.39) converge to $\mathfrak{g}(z) = \int \frac{\rho(x)\mathrm{d}x}{z-x}$, and how do we recover $\rho(x)$ from $\mathfrak{g}(z)$?

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

We have argued before that $g_N(z)$ converges to $\mathfrak{g}(z)$ for very large complex z such that the Taylor series around infinity is convergent. The function $\mathfrak{g}(z)$ is not defined on the real axis for $z = x$ on the support of $\rho(x)$, nevertheless, immediately below (and above) the real axis the random function $g_N(z)$ converges to $\mathfrak{g}(z)$. (The case where z is right on the real axis is discussed in Section 2.3.6.) Let us study the random function $g_N(z)$ just below the support of $\rho(x)$.

We let $z = x - \mathrm{i}\eta$, with $x \in \mathrm{supp}\{\rho\}$ and η is a small positive number. Then

$$g_N(x - \mathrm{i}\eta) := \frac{1}{N}\sum_{k=1}^{N} \frac{1}{x - \mathrm{i}\eta - \lambda_k} = \frac{1}{N}\sum_{k=1}^{N} \frac{x - \lambda_k + \mathrm{i}\eta}{(x - \lambda_k)^2 + \eta^2}. \tag{2.42}$$

We focus on the imaginary part of $g_N(x - \mathrm{i}\eta)$ (the real part is discussed in Section 19.5.2). Note that it is a convolution of the empirical spectral density $\rho_N(\lambda)$ and π times the Cauchy kernel:

$$\pi K_\eta(x) = \frac{\eta}{x^2 + \eta^2}. \tag{2.43}$$

The Cauchy kernel $K_\eta(x)$ is strongly peaked around zero with a window width of order η (Fig. 2.2). Since there are N eigenvalues lying inside the interval $[\lambda_-, \lambda_+]$, the typical eigenvalue spacing is of order $(\lambda_+ - \lambda_-)/N = O(N^{-1})$.

(1) Suppose $\eta \ll N^{-1}$. Then there are typically 0 or 1 eigenvalue within a window of size η around x. Then $\mathrm{Im}\, g_N$ will be affected by the fluctuations of single eigenvalues of $\mathbf{X}$, and hence it cannot converge to any deterministic function. (see Fig. 2.3).

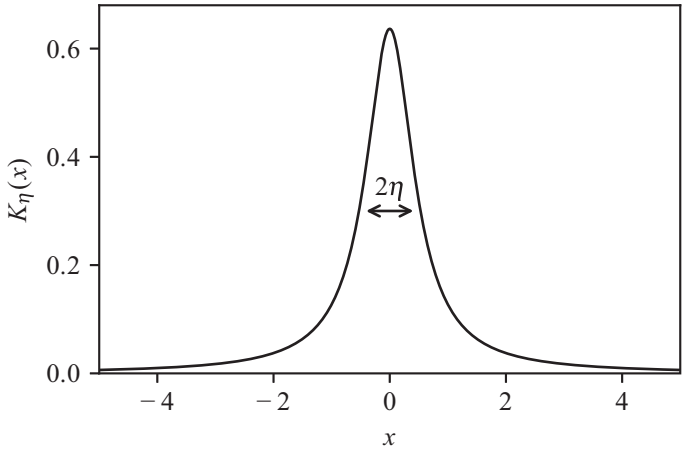

Figure 2.2 The Cauchy kernel for $\eta = 0.5$. It is strongly peaked around zero with a window width of order η. When $\eta \to 0$, the Cauchy kernel is a possible representation of Dirac's δ-function.

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

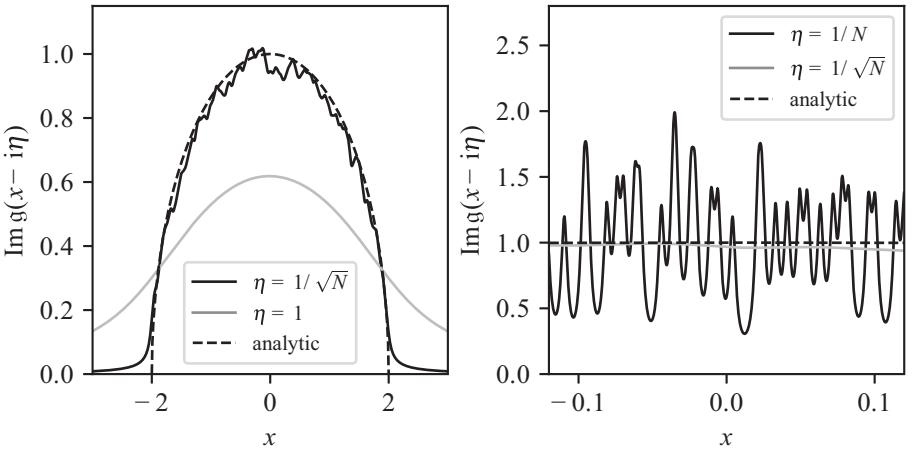

Figure 2.3 Imaginary part of $\mathfrak{g}(x - \mathrm{i}\eta)$ for the Wigner ensemble. The analytic result for $\eta \to 0^+$ is compared with numerical simulations ($N = 400$). On the left for $\eta = 1/\sqrt{N}$ and $\eta = 1$. Note that for $\eta = 1$ the density is quite deformed. On the right (zoom near $x = 0$) for $\eta = 1/\sqrt{N}$ and $\eta = 1/N$. Note that for $\eta = 1/N$, the density fluctuates wildly as only a small number of (random) eigenvalues contribute to the Cauchy kernel.

(2) Suppose $N^{-1} \ll \eta \ll 1$ (e.g. $\eta = N^{-1/2}$). Then on a small scale $\eta \ll \Delta x \ll 1$, the density ρ is locally constant and there are a great number n of eigenvalues inside:

$$n \sim N\rho(x)\Delta x \gg N\eta \gg 1. \tag{2.44}$$

The law of large numbers allows us to replace the sum with an integral; we obtain that

$$\frac{1}{N} \sum_{k:\lambda_k \in [x-\Delta x, x+\Delta x]} \frac{\mathrm{i}\eta}{(x-\lambda_k)^2 + \eta^2} \to \mathrm{i} \int_{x-\Delta x}^{x+\Delta x} \frac{\rho(x)\eta \mathrm{d}y}{(x-y)^2 + \eta^2} \to \mathrm{i}\pi\rho(x), \tag{2.45}$$

where the last limit is obtained by writing $u = (y - x)/\eta$ and noting that as $\eta \to 0$ we have

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}u}{u^2 + 1} = \pi. \tag{2.46}$$

Exercise 2.3.2 Finite N approximation and small imaginary part

Im $g_N(x - \mathrm{i}\eta)/\pi$ is a good approximation to $\rho(x)$ for small positive η, where $g_N(z)$ is the sample Stieltjes transform ($g_N(z) = (1/N)\sum_k 1/(z - \lambda_k)$). Numerically generate a Wigner matrix of size N and $\sigma^2 = 1$.

(a) For three values of η, $\{1/N, 1/\sqrt{N}, 1\}$, plot Im $g_N(x - \mathrm{i}\eta)/\pi$ and the theoretical $\rho(x)$ on the same plot for x between -3 and 3.

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

(b) Compute the error as a function of η where the error is $(\rho(x) - \operatorname{Im} g_N(x - \mathrm{i}\eta)/\pi)^2$ summed for all values of x between -3 and 3 spaced by intervals of 0.01. Plot this error for η between $1/N$ and 1. You should see that $1/\sqrt{N}$ is very close to the minimum of this function.

2.3.4 Stieltjes Inversion Formula

From the above discussions, we extract the following important results:

(1) The Stieltjes inversion formula (also called the Sokhotski–Plemelj formula):

$$\lim_{\eta \to 0+} \operatorname{Im} \mathfrak{g}(x - \mathrm{i}\eta) = \pi \rho(x). \tag{2.47}$$

(2) When applied to finite size Stieltjes transform $g_N(z)$, we should take $N^{-1} \ll \eta \ll 1$ for $g_N(x - \mathrm{i}\eta)$ to converge to $\mathfrak{g}(x - \mathrm{i}\eta)$ and for (2.47) to hold. Numerically, $\eta = N^{-1/2}$ works quite well.

We discuss briefly why $\eta = N^{-1/2}$ works best. First, we want η to be as small as possible such that the local density $\rho(x)$ is not blurred. If η is too large, one introduces a systematic error of order $\rho'(x)\eta$. On the other hand, we want $N\eta$ to be as large as possible such that we include the statistics of a sufficient number of eigenvalues so that we measure $\rho(x)$ accurately. In fact, the error between g_N and g is of order $\frac{1}{N\eta}$. Thus we want to minimize the total error $\mathcal{E}$ given by

$$\mathcal{E} = \rho'(x)\eta + \frac{1}{N\eta}, \qquad \rho'(x)\eta : \text{ systematic error}, \quad \frac{1}{N\eta} : \text{ statistical error}. \tag{2.48}$$

Then it is easy to see that the total error is minimized when η is of order $1/\sqrt{N\rho'(x)}$.

2.3.5 Density of Eigenvalues of a Wigner Matrix

We go back to study the Stieltjes transform (2.38) of the Wigner matrix. Note that for $z = x - \mathrm{i}\eta$ with $\eta \to 0$, $\mathfrak{g}(z)$ can only have an imaginary part if $\sqrt{x^2 - 4\sigma^2}$ is imaginary. Then, using (2.47), we get the Wigner semi-circle law:

$$\rho(x) = \frac{1}{\pi} \lim_{\eta \to 0+} \operatorname{Im} \mathfrak{g}(x - \mathrm{i}\eta) = \frac{\sqrt{4\sigma^2 - x^2}}{2\pi\sigma^2}, \quad -2\sigma \le x \le 2\sigma. \tag{2.49}$$

Note the following features of the semi-circle law (see Fig. 2.4): (1) asymptotically there is no eigenvalue for $x > 2\sigma$ and $x < -2\sigma$; (2) the eigenvalue density has square-root singularities near the edges: $\rho(x) \sim \sqrt{x + 2\sigma}$ near the left edge and $\rho(x) \sim \sqrt{2\sigma - x}$ near the right edge. For finite N, some eigenvalues are present in a small region of width $N^{-2/3}$ around the edges, see Section 14.1.

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

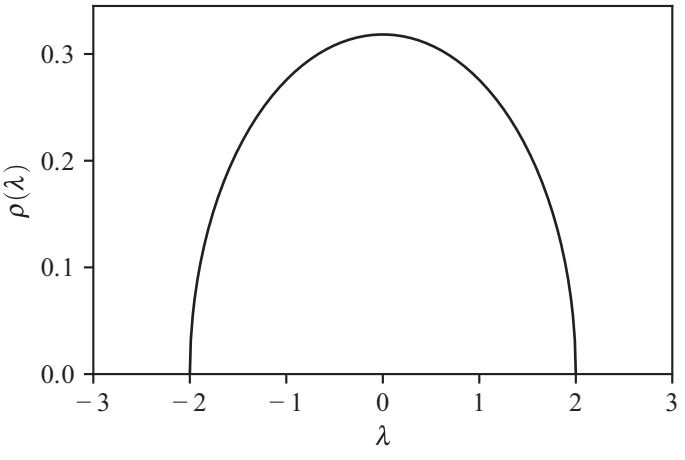

Figure 2.4 Density of eigenvalues of a Wigner matrix with $\sigma = 1$: the *semi-circle law*.

Exercise 2.3.3 From the moments to the density

A large random matrix has moments $\tau(\mathbf{A}^k) = 1/k$.

(a) Using Eq. (2.24) write the Taylor series of $\mathfrak{g}(z)$ around infinity.
(b) Sum the series to get a simple expression for $\mathfrak{g}(z)$. Hint: look up the Taylor series of $\log(1 + x)$.
(c) Where are the singularities of $\mathfrak{g}(z)$ on the real axis?
(d) Use Eq. (2.47) to find the density of eigenvalues $\rho(\lambda)$.
(e) Check your result by recomputing the moments and the Stieltjes transform from $\rho(\lambda)$.
(f) Redo all the above steps for a matrix whose odd moments are zero and even moments are $\tau(\mathbf{A}^{2k}) = 1$. Note that in this case the density $\rho(\lambda)$ has Dirac masses.

2.3.6 Stieltjes Transform on the Real Axis

What about computing the Stieltjes transform when z is real and inside the spectrum? This seems dangerous at first sight, since $g_N(z)$ diverges when z is equal to one of the eigenvalues of $\mathbf{X}$. As these eigenvalues become more and more numerous as N goes to infinity, z will always be very close to a pole of the resolvent $g_N(z)$. Interestingly, one can turn this predicament on its head and actually exploit these divergences. In a hand-waving manner, the probability that the difference $d_i = |z - \lambda_i|$ between z (now real) and a given eigenvalue λ_i is very small, is given by

$$\mathbb{P}[d_i < \epsilon/N] = 2\epsilon\rho(z), \tag{2.50}$$

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

where $\rho(z)$ is the normalized density of eigenvalues around z. But as $\epsilon \to 0$, i.e. when z is extremely close to λ_i, the resolvent becomes dominated by a unique contribution – that of the λ_i term, all other terms $(z - \lambda_j)^{-1}$, $j \neq i$, become negligible. In other words, $g_N(z) \approx \pm(Nd_i)^{-1}$, and therefore

$$\mathbb{P}[|g| > \epsilon^{-1}] = \mathbb{P}[d_i < \epsilon/N] = 2\epsilon\rho(z). \tag{2.51}$$

Hence, $g_N(z)$ does not converge for $N \to \infty$ when z is real, but the tail of its distribution decays precisely as $\rho(z)/g^2$. Studying this tail thus allows one to extract the eigenvalue density $\rho(z)$ while working directly on the real axis. Let us show how this works in the case of Wigner matrices.

The idea is that, for a rotationally invariant problem, the distribution of a randomly chosen diagonal element of the resolvent (say $\mathbf{G}_{11}$) is the same as the distribution $P(g)$ of the normalized trace.[3] With this assumption, Eq. (2.29) can be interpreted as giving the evolution of $P(g)$ itself, i.e.

$$P^{(N)}(g) = \int_{-\infty}^{+\infty} \mathrm{d}g' P^{(N-1)}(g')\delta\left(g - \frac{1}{z - \sigma^2 g'}\right), \tag{2.52}$$

where we have used the fact that, for large N, $\sum_{k,\ell=2}^{N} \mathbf{M}_{1k}(\mathbf{M}_{22})^{-1}_{k\ell}\mathbf{M}_{\ell\ell} \to \sigma^2 g^{(N-1)}$. Now, this functional iteration admits the following Cauchy distribution as a fixed point:

$$P^{\infty}(g) = \frac{\rho(z)}{(g - \frac{z}{2\sigma^2})^2 + \pi^2\rho(z)}. \tag{2.53}$$

This simple result, that the resolvent of a Wigner matrix on the real axis is a Cauchy variable, calls for several comments. First, one finds that $P^{\infty}(g)$ indeed behaves as $\rho(z)/g^2$ for large g, as argued above. Second, it would have been entirely natural to find a Cauchy distribution for g had the eigenvalues been independent. Indeed, since g is then the sum of N random variables (i.e. the $1/d_i$'s) distributed with an inverse square power, the generalized CLT predicts that the resulting sum is Cauchy distributed. In the present case, however, the eigenvalues are strongly correlated – see Section 5.1.4. It was recently proven that the Cauchy distribution is in fact *super-universal* and holds for a wide class of point processes on the real axis, in particular for the eigenvalues of random matrices. It is in fact even true when these eigenvalues are strictly equidistant, with a random global shift.

Bibliographical Notes

- The overall content of this chapter is covered in many books and textbooks, see for example
  - M. L. Mehta. *Random Matrices*. Academic Press, San Diego, 3rd edition, 2004,
  - T. Tao. *Topics in Random Matrix Theory*. American Mathematical Society, Providence, Rhode Island, 2012,
  - G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, Cambridge, 2010,
  - G. Blower. *Random Matrices: High Dimensional Phenomena*. Cambridge University Press, Cambridge, 2009,

[3] This is not a trivial statement but it can be proven along the lines of Aizenman and Warzel [2015].

https://doi.org/10.1017/9781108768900.004 Published online by Cambridge University Press

- B. Eynard, T. Kimura, and S. Ribault. Random matrices. *preprint arXiv:1510.04430*, 2006,
- G. Livan, M. Novaes, and P. Vivo. *Introduction to Random Matrices: Theory and Practice*. Springer, New York, 2018.

- For some historical papers, see
  - E. P. Wigner. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Mathematical Proceedings of the Cambridge Philosophical Society*, 47(4):790–798, 1951,
  - F. J. Dyson. The threefold way: algebraic structure of symmetry groups and ensembles in quantum mechanics. *Journal of Mathematical Physics*, 3(6):1199–1215, 1962.
- About the Stieltjes transform on the real line and the super-universality of the Cauchy distribution, see
  - M. Aizenman and S. Warzel. On the ubiquity of the Cauchy distribution in spectral problems. *Probability Theory and Related Fields*, 163:61–87, 2015,
  - Y. V. Fyodorov and D. V. Savin. Statistics of impedance, local density of states, and reflection in quantum chaotic systems with absorption. *Journal of Experimental and Theoretical Physics Letters*, 80(12):725–729, 2004,
  - Y. V. Fyodorov and I. Williams. Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity. *Journal of Statistical Physics*, 129(5-6):1081–1116, 2007,
  - J.-P. Bouchaud and M. Potters. Two short pieces around the Wigner problem. *Journal of Physics A: Mathematical and Theoretical*, 52(2):024001, 2018,

  and for a related discussion see also
  - M. Griniasty and V. Hakim. Correlations and dynamics in ensembles of maps: Simple models. *Physical Review E*, 49(4):2661, 1994.