Lectures on Random Matrix Theory

Govind Menon and Thomas Trogdon

March 12, 2018

Contents

Chapter 1

Overview

1.1 What is a random matrix?

There are two distinct points of view that one may adopt. On one hand, our intuitive ideas of randomness are intimately tied to the notion of sampling a realization of a random variable. Thus, given a random number generator, one may build a random Hermitian matrix, $M \in \mathsf{Her}(n)$, by choosing its real diagonal and complex upper-triangular entries independently at random. It is conventional to assume further that all the diagonal entries have the same law, and that all the upper-triangular entries have the same law. For example, we may assume that the entries on the diagonal are ± 1 with probability $1/2$, and that the upper-triangular entries are $\pm 1 \pm \mathrm{i}$ with probability $1/4$. It is also conventional to have the variance of the diagonal entries to be twice that of the off-diagonal entries. Random matrices of this kind, are said to be drawn from *Wigner ensembles*.

On the other hand, one may adopt a more analytic view. The Hilbert-Schmidt inner product of two Hermitian matrices, $\mathrm{Tr}(MN) = \sum_{j,k=1}^{n} M_{jk}\bar{N}_{jk}$, defines a natural metric $\mathrm{Tr}(\mathrm{d}M^2)$ and volume form $\mathrm{D}M$ on $\mathsf{Her}(n)$ (see Lecture 2). In this text, unless otherwise stated, $\|M\| = \sqrt{\mathrm{Tr}\, M^*M}$). Thus, each positive function $p : \mathsf{Her}(n) \to [0, \infty)$ that decays sufficiently fast as $\|M\| \to \infty$, may be normalized to define a probability measure. A fundamental example is the law of the *Gaussian Unitary Ensemble* (GUE)

$$p_{\mathrm{GUE}}(M)\mathrm{D}M = \frac{1}{Z_n}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(M^2)}\mathrm{D}M. \tag{1.1.1}$$

Here Z_n is a normalization constant that ensures p_{GUE} is a probability density (we use the same notation for different ensembles; thus the numerical value of Z_n must be inferred from the context). The term GUE was introduced by Freeman Dyson [9], and refers to an important invariance property of p_{GUE}. Each $U \in \mathsf{U}(n)$ defines a transformation $\mathsf{Her}(n) \to \mathsf{Her}(n)$, $M \mapsto UMU^*$. It is easily checked that the volume form $\mathrm{D}M$ is invariant under the map $M \mapsto$

UMU^*, as is the measure $p_{\mathrm{GUE}}(M)\mathrm{D}M$. More generally, a probability measure on $\mathsf{Her}(n)$ is said to be *invariant* if $p(M)\,\mathrm{D}M$ remains invariant under the map $M \mapsto UMU^*$, for each $U \in \mathsf{U}(n)$. Important examples of invariant ensembles are defined by polynomials in one-variable of the form

$$g(x) = a_{2N}x^{2N} + a_{2N-1}x^{2N-1} + \ldots + a_0, \quad a_j \in \mathbb{R}, \quad j = 0, 1, \ldots, 2N, \; a_{2N} > 0. \tag{1.1.2}$$

Then the following probability measure is invariant

$$p(M)\mathrm{D}M = \frac{1}{Z_n}\mathrm{e}^{-\operatorname{Tr} g(M)}\mathrm{D}M. \tag{1.1.3}$$

We have assumed that all matrices are Hermitian simply to be concrete. The above notions extend to ensembles of matrices from $\mathsf{Symm}(n)$ and $\mathsf{Quart}(n)$. The notion of invariance in each case is distinct: for $\mathsf{Symm}(n)$, the natural transformation is $M \mapsto QMQ^T$, $Q \in \mathsf{O}(n)$; for $\mathsf{Quart}(n)$ it is $M \mapsto SMS^D$, $S \in \mathsf{USp}(n)$. The standard Gaussian ensembles in these cases are termed GOE (the *Gaussian Orthogonal Ensemble*) and GSE (the *Gaussian Symplectic Ensemble*), and they are normalized as follows:

$$p_{\mathrm{GOE}}(M)\mathrm{d}M = \frac{1}{Z_n}\mathrm{e}^{-\frac{1}{4}\operatorname{Tr}(M^2)}\mathrm{d}M, \quad p_{\mathrm{GSE}}(M)\mathrm{d}M = \frac{1}{Z_n}\mathrm{e}^{-\operatorname{Tr}(M^2)}\mathrm{D}M. \tag{1.1.4}$$

The differing normalizations arise from the different volume forms on $\mathsf{Symm}(n)$, $\mathsf{Her}(n)$ and $\mathsf{Quart}(n)$ as will be explained in Lecture 2. For now, let us note that the densities for all the Gaussian ensembles may be written in the unified form

$$Z_n(\beta)^{-1}\mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr}(M^2)} \tag{1.1.5}$$

where $\beta = 1$,2 and 4 for GOE, GUE and GSE respectively. While it is true that there are no other ensembles that respect fundamental physical invariance (in the sense of Dyson), many fundamental results of random matrix theory can be established for all $\beta > 0$. These results follow from the existence of ensembles of *tridiagonal* matrices, whose eigenvalues have a joint distribution that interpolates those of the $\beta = 1$,2 and 4 ensembles to all $\beta > 0$ [8].

1.2 The main limit theorems

The basic question in random matrix theory is the following: what can one say about the statistics of the eigenvalues of a random matrix? For example, what is the probability that the largest eigenvalue lies below a threshold? Or, what is the probability that there are no eigenvalues in a given interval? The difficulty here is that even if the entries of a random matrix are independent, the eigenvalues are strongly coupled.

Gaussian ensembles play a very special, and important, role in random matrix theory. These are the only ensembles that are both Wigner and invariant (see Theorem 18 below). Pioneering, ingenious calculations by Dyson [9],

Gaudin and Mehta [26, 25], on the Gaussian ensembles served to elucidate the fundamental limit theorems of random matrix theory. In this section we outline these theorems, assuming always that the ensemble is GUE. Our purpose is to explain the form of the main questions (and their answers) in the simplest setting. All the results hold in far greater generality as is briefly outlined at the end of this section.

By the normalization (1.1.1), a GUE matrix has independent standard normal entries on its diagonal (mean zero, variance 1). The off-diagonal entries have mean zero and variance 1/2. We denote the ordered eigenvalues of the GUE matrix by $\lambda_1 \leq \lambda_2 \leq \dots \lambda_n$. A fundamental heuristic for GUE matrices (that will be proven later, and may be easily simulated) is that the largest and smallest eigenvalues have size $O(\sqrt{n})$. In fact, $\lambda_1 \approx -2\sqrt{n}$ and $\lambda_n \approx 2\sqrt{n}$ as $n \to \infty$. Since there are n eigenvalues, the gap between these eigenvalues is typically $O(1/\sqrt{n})$. There are thus two natural scaling limits to consider as $n \to \infty$:

1. Rescale $M \mapsto n^{-1/2}M$ so that the spectral radius is $O(1)$. In this scaling limit, n eigenvalues are contained within a bounded interval, and we obtain a deterministic limit called the *semicircle law*.

2. Rescale $M \mapsto n^{1/2}M$ so that the gaps between eigenvalues are $O(1)$. In this scaling limit, we expect a random limiting point process. The limiting point process is a *determinantal point process* called the Sine_2 process.

In fact, the situation is more subtle. While the expected value of the gap between eigenvalues for a GUE matrix is indeed $O(1/n)$, the gaps are $O(n^{-2/3})$ about the edge of the spectrum. There is an an entirely different scaling limit called the Airy_2 process obtained by rescaling the spectrum of $M \pm 2\sqrt{n}I$.

In all that follows, we consider a sequence of random matrices of size n sampled from GUE(n). To make this explicit, the matrix is denoted M_n, and its ordered eigenvalues are denoted $\lambda_1^{(n)} \leq \lambda_2^{(n)} \leq \dots \leq \lambda_n^{(n)}$.

1.2.1 The semicircle law

Definition 1. The probability density and distribution function

$$p_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{|x|\leq 2}, \quad F_{\mathrm{sc}}(x) = \int_{-\infty}^{x} p_{\mathrm{sc}}(x')\,\mathrm{d}x'. \tag{1.2.1}$$

are called the *semicircle density* and the *semicircle distribution* respectively.

Theorem 2. *Let M_n be a sequence of GUE matrices of size n. The rescaled empirical spectral measures*

$$\mu_n(\mathrm{d}x) = \frac{1}{n}\sum_{j=1}^{n} \delta_{n^{-1/2}\lambda_j^{(n)}}(\mathrm{d}x) \tag{1.2.2}$$

converge weakly to the semicircle density almost surely.

Theorem 2 may also be interpreted as the statement that the empirical spectral distribution of the matrices $M_n/\sqrt{n}$ converges to the semicircle distribution. The shortest proof of Theorem (2) uses the following integral transform.

Definition 3. Assume μ is a measure on $\mathbb{R}$ that satisfies the finiteness condition

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{1+x^2}} \mu(\mathrm{d}x) < \infty. \tag{1.2.3}$$

The Stieltjes transform of μ is the function

$$\hat{\mu}(z) = \int_{-\infty}^{\infty} \frac{1}{x-z} \mu(\mathrm{d}x), \quad z \in \mathbb{C}\backslash\mathbb{R}. \tag{1.2.4}$$

The Stieltjes transform is of fundamental importance in the theory of orthogonal polynomials and spectral theory. This is because there are natural Stieltjes transforms associated to the resolvent $(M-z)^{-1}$, such as

$$\mathrm{Tr}(M-z)^{-1} \quad \text{and} \quad v^*(M-z)^{-1}v, \quad v \in \mathbb{C}^n, \quad |v| = 1. \tag{1.2.5}$$

The general proof of Theorem 2 uses a recursive expression for the law of $\mathrm{Tr}(z - M_n)^{-1}$. As $n \to \infty$, the fixed point of this recursion, R_{sc} solves the quadratic equation

$$R^2 - zR + 1 = 0. \tag{1.2.6}$$

It is then easy to verify that

$$R_{\mathrm{sc}}(z) = \frac{1}{2}\left(-z + \sqrt{z^2-4}\right), \quad z \in \mathbb{C}\backslash[-2,2]. \tag{1.2.7}$$

We recover the semicircle law from $R_{\mathrm{sc}}(z)$ by evaluating the jump in $\mathsf{Im}(R_{\mathrm{sc}}(z))$ across the branch cut $[-2,2]$.

Remark 4. The heuristic to determine the typical spacings is the following. Define $\gamma_j^{(n)} \in [-2,2]$ by the relation

$$\frac{j}{n} = \int_{-\infty}^{\gamma_j^{(n)}} p_{\mathrm{sc}}(x)\mathrm{d}x, \quad j = 1, 2, \ldots, n.$$

Then the approximation $\lambda_j^{(n)} \approx \sqrt{n}\gamma_j^{(n)}$ should hold[1]. We have

$$\frac{1}{n} = \int_{\gamma_j^{(n)}}^{\gamma_{j+1}^{(n)}} p_{\mathrm{sc}}(x)\mathrm{d}x \approx (\gamma_{j+1}^{(n)} - \gamma_j^{(n)}) p_{\mathrm{sc}}(\gamma_j^{(n)}). \tag{1.2.8}$$

If $j = j(n)$ is chosen so that $\gamma_j^{(n)} \to r$, $r \in (-2,2)$ (i.e. in the "bulk") we have

$$\lambda_{j+1}^{(n)} - \lambda_j^{(n)} \approx \frac{1}{\sqrt{n} p_{\mathrm{sc}}(r)}.$$

[1]This is made rigorous and quantitative by Erdős, Yau and Yin [13].

At the edge, consider (noting that $\gamma_1^{(n)} > -2$)

$$\frac{1}{n} = \int_{-2}^{\gamma_1^{(n)}} p_{\mathrm{sc}}(x)\mathrm{d}x \approx \int_{-2}^{\gamma_1^{(n)}} \frac{1}{\pi}\sqrt{2+x}\mathrm{d}x = \frac{2}{3\pi}\left(\gamma_1^{(n)}+2\right)^{3/2},$$
$$\gamma_1^{(n)} + 2 \approx \frac{c}{n^{2/3}},$$
$$2\sqrt{n} + \lambda_1^{(n)} = O(n^{-1/6}), \quad \lambda_n^{(n)} - 2\sqrt{n} = O(n^{-1/6}). \tag{1.2.9}$$

1.2.2 Fluctuations in the bulk: the sine process

We now rescale so that the gaps between eigenvalues is $O(1)$, and the scaling limit is a random process. This random process will always be denoted Sine_2 (and Sine_β for the general β-ensembles). Each realization of the Sine_2 process is a countable set of points $\{x_k\}_{k=-\infty}^{\infty}$. One of the fundamental statistics associated to a point process is the probability of having k points in an interval. In order to state a typical fluctuation theorem that describes these probabilities, we must define the sine-kernel and its Fredholm determinants.

Definition 5. The sine-kernel is the integral kernel on $\mathbb{R}\times\mathbb{R}$ given by

$$K_{\mathsf{sine}}(x,y) = \frac{\sin\pi(x-y)}{\pi(x-y)}, \quad x \neq y, \tag{1.2.10}$$

and $K_{\mathsf{sine}}(x,x,) = 1$.

In the following theorem we will assume that x and y are restricted to a finite interval (a,b). The sine-kernel defines an integral operator on $L^2(a,b)$ that we denote by $K_{\mathsf{sine}}\mathbf{1}_{(a,b)}$. The kernel $K_{\mathsf{sine}}(x,y)$ is clearly continuous, thus bounded, for $x,y\in(a,b)$. Thus, $K_{\mathsf{sine}}\mathbf{1}_{(a,b)}$ defines an integral operator on $L^2(a,b)$ that is trace-class, and it has a well-defined Fredholm determinant

$$\det\left(1 - K_{\mathsf{sine}}\mathbf{1}_{(a,b)}\right) \tag{1.2.11}$$
$$= 1 + \sum_{m=0}^{\infty}\frac{(-1)^m}{m!}\int_{(a,b)^m}\det\left(K_{\mathsf{sine}}(x_j,x_k)_{1\le j,k\le m}\right)\mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_m.$$

Though perhaps mysterious at first sight, the origin of this formula is rather simple. Integral operators with some smoothness and boundedness (in particular, continuous integral operators K whose trace $\int_a^b |K(x,x)|\mathrm{d}x$ is finite) may be approximated on a discrete-grid of size h by a finite-dimensional discretization K_h. The determinant $(I-K_h)$ is then the usual determinant of a matrix and we may use the definition of the determinant to expand $\det(I-K_h)$ in a finite series, which is nothing but the infinite series above in the instance when all terms beyond $m = \mathrm{rank}(K_h)$ vanish. This approach was pioneered by Fredholm in 1900 before the development of functional analysis. From a probabilistic point of view, this formula arises from the Inclusion-Exclusion Principle, taken to the limit. But the operator theory pioneered by Fredholm allows for that limit to be understood.

Theorem 6 (Gaudin-Mehta [26]). *For each finite interval $(a, b) \subset \mathbb{R}$, and $r \in (-2, 2)$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}p_{\mathrm{sc}}(r)\left(\lambda_k^{(n)} - r\sqrt{n}\right) \notin (a,b),\, 1 \le k \le n\right) = \det\left(1 - K_{\mathsf{sine}}\mathbf{1}_{(a,b)}\right). \tag{1.2.12}$$

The probabilities of the Sine_2 process can be expressed without reference to the matrices M_n. For each interval (a, b) let $N_{(a,b)} = \sum_{k=-\infty}^{\infty} \mathbf{1}_{x_k \in (a,b)}$. Then,

$$\mathbb{P}\left(N_{(a,b)} = 0\right) = \det\left(1 - K_{\mathsf{sine}}\mathbf{1}_{(a,b)}\right). \tag{1.2.13}$$

For comparison, if we had a Poisson process $\{\tilde{x}_k\}_{k=-\infty}^{\infty}$ with rate $\lambda(\mathrm{d}x)$, the associated count $\tilde{N}_{(a,b)}$ would satisfy

$$\mathbb{P}\left(\tilde{N}_{(a,b)} = 0\right) = \exp\left(-\int_a^b \lambda(\mathrm{d}x)\right).$$

1.2.3 Fluctuations at the edge: the Airy point process

Remark 4 and Theorem 6 reveal that the gaps between consecutive eigenvalues $\lambda_j^{(n)}$ and $\lambda_{j+1}^{(n)}$ is of $O(n^{-1/2})$. However, the fluctuations at the edge are much larger $O(n^{-1/6})$. The point process of shifted and scaled eigenvalues converges in distribution to a limiting point process, $\{y_k\}_{k=1}^{\infty}$ called the Airy_2 process. In order to describe the law of this process, we must define the Airy function and the Airy kernel.

Definition 7. The Airy function is defined by the oscillatory integral

$$\mathrm{Ai}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}kx}\mathrm{e}^{\mathrm{i}k^3/3}\,\mathrm{d}k. \tag{1.2.14}$$

The Airy function is one of the classical special functions [1]. It admits several alternative definitions. For instance, the oscillatory integral in (1.2.14) may be deformed into an absolutely convergent integral in the complex plane. This argument allows us to establish that the Airy function is entire and to determine its asymptotic expansions as $x \to \pm\infty$.

These properties may also be established using the theory of ordinary differential equations in the complex plane [17]. It is easy to verify from (1.2.14), after deformation, that $\mathrm{Ai}(x)$ satisfies the differential equation

$$\varphi''(x) = x\varphi, \quad -\infty < x < \infty. \tag{1.2.15}$$

Equation (1.2.15) admits two linearly independent solutions, only one of which decays as $x \to \infty$. Up to a (fixed by convention, but otherwise arbitrary) normalization constant, the decaying solution is $\mathrm{Ai}(x)$.

Definition 8. The Airy kernel is the continous integral kernel on $\mathbb{R} \times \mathbb{R}$ given by

$$K_{\mathsf{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}, \quad x \neq y,$$

and by continuity at $x = y$.

Observe that both the sine and Airy kernel have the form

$$K(x,y) = \frac{f(x)f'(y) - f'(x)f(y)}{x-y}, \quad x \neq y \tag{1.2.16}$$

where f solves a second-order linear differential equation. Similar kernels arise in various limiting models in random matrix theory. For instance, the Bessel kernel – corresponding to $f(x) = J_\alpha(x)$, the Bessel function with parameter α – describes fluctuations about the *singular values* of random positive definite Hermitian matrices.

Theorem 9. *For each interval $(a,b) \subset \mathbb{R}$, $-\infty < a < b \leq \infty$,*

$$\lim_{n\to\infty} \mathbb{P}\left(n^{1/6}\left(\lambda_k^{(n)} - 2\sqrt{n}\right) \notin (a,b),\, 1 \leq k \leq n\right) = \det\left(1 - K_{\mathsf{Airy}}\mathbf{1}_{(a,b)}\right). \tag{1.2.17}$$

As in the remarks following Theorem 6, the expression $\det\left(1 - K_{\mathsf{Airy}}\mathbf{1}_{(a,b)}\right)$ gives the probability that no points of a realization of the Airy_2 point process lie in (a,b).

1.2.4 Fredholm determinants, Painlevé equations, and integrable systems

It is immediate from Theorem 6 and Theorem 9 that the Fredholm determinants $\det\left(1 - K_{\mathsf{sine}}\mathbf{1}_{(a,b)}\right)$ and $\det\left(1 - K_{\mathsf{Airy}}\mathbf{1}_{(a,b)}\right)$ are positive for all (a,b). This is astonishing, if one treats (1.2.11) as a starting point, since it is by no means clear that the signed infinite series sums to a positive number! It is in fact, rather challenging to extract meaningful information, such as the asymptotics of tails, from the expression of probabilities as Fredholm determinants. A crucial piece of the puzzle lies in the connection between Fredholm determinants and the theory of integrable systems. More precisely, the Fredholm determinants satisfy differential equations in a and b (or more generally in endpoints of intervals, when one considers the obvious extensions of Theorem 6 and Theorem 9 to a collection of intervals $\prod_{j=1}^m (a_m, b_m)$). These ordinary differential equations have a special, integrable structure, that allows their analysis. The following theorems illustrate this aspect of random matrix theory.

Theorem 10 (Jimbo-Miwa-Mori-Sato [18])**.** *For all $t > 0$,*

$$\det\left(1 - K_{\mathsf{sine}}\mathbf{1}_{(-\frac{t}{2},\frac{t}{2})}\right) = \exp\left(\int_0^t \frac{\sigma(s)}{s}\,\mathrm{d}s\right), \tag{1.2.18}$$

where $\sigma(t)$ is the solution to the Painlevé-5 equation

$$\left(t\sigma''\right)^2 + 4\left(t\sigma' - \sigma\right)\left(t\sigma' - \sigma + \sigma^2\right) = 0, \tag{1.2.19}$$

which satisfies the asymptotic condition

$$\sigma(t) = -\frac{t}{\pi} - \frac{t^2}{\pi^2} - \frac{t^3}{\pi^3}, \quad t \downarrow 0. \tag{1.2.20}$$

Theorem 11 (Tracy-Widom distribution [36])**.** *For all real t,*

$$F_2(t) := \det\left(1 - K_{\mathsf{Airy}}\mathbf{1}_{(t,\infty)}\right) = \exp\left(-\int_t^\infty (s-t)q^2(s)\,\mathrm{d}s\right), \tag{1.2.21}$$

where q is the solution to the Painlevé-2 equation

$$q'' = tq + 2q^3, -\infty < t < \infty \tag{1.2.22}$$

which satisfies the asymptotic condition

$$q(t) \sim \mathrm{Ai}(t), \quad t \to \infty. \tag{1.2.23}$$

The Painlevé differential equations are a special family of *nonlinear* ordinary differential equations that generalize the classical theory of *linear* differential equations in the complex plane and the associated theory of special functions [17]. For example, the Painlevé-2 equation (1.2.22) may be viewed as a nonlinear analogue of the Airy differential equation (1.2.15). Broadly, the Painlevé differential equations represent a complete classification of second-order differential equations with the Painlevé property — their only movable singularities (movable by changing initial conditions) are poles — that are not solvable with elementary functions. The theory of Painlevé equations was developed in the early years 1900's, by Boutroux and Painlevé, but fell into obscurity[2]. It was reborn in the 1970s with the discovery of their importance in integrable systems and exactly solvable models in statistical mechanics, such as the Ising model in 2D [24]. We illustrate these links with a fundamental integrable system: the Korteweg-de Vries (KdV) equation

$$u_t + 6uu_x + u_{xxx} = 0, \quad -\infty < x < \infty,\ t \geq 0. \tag{1.2.24}$$

Despite the fact that KdV is nonlinear, it may be solved explicitly through the inverse scattering transform. We will not discuss this method in detail. But in order to make the connection with random matrix theory, let us note that if one seeks self-similar solutions to KdV of the form

$$u(x,t) = \frac{1}{(3t)^{2/3}} q\left(\frac{x}{(3t)^{2/3}}\right) \tag{1.2.25}$$

[2]Paul Painlevé was rather restless: he began in mathematics, became an early aviation enthusiast, and then turned to politics. He rose to become the Prime Minister of France for part of World War I, and was later the designer of the disastrous Maginot line.

then $q = v^2 + v'$ and v satisfies the Painlevé-2 equation (1.2.22). In fact, it is in this context that Hastings and McLeod established the existence of a solution to (1.2.22) that satisfies the asymptotic condition (1.2.23) [15]. It is remarkable that it is exactly this solution that describes the Tracy-Widom distribution $F_2(t)$!

1.2.5 Universality

We have restricted attention to matrices from GUE to present some of the central theorems in the subject in an efficient manner. One of the main achievements of the past decade has been the establishment of *universality* – informally, this is the notion that the limiting fluctuations in the bulk and edge described by the Sine_2 and Airy_2 processes, hold for both Wigner and invariant ensembles which satisfy natural moment assumptions. The idea of universality is of clear practical importance (we need understand only a few universal limits). It also appears to hold the key to some of the connections between random matrix theory and other areas of mathematics. The explanation of these connections may lie in the fact that determinantal point processes, such as the Sine_2 and Airy_2 process, have the simplest structure of *strongly interacting* point processes. By contrast, Poisson processes, while universal, describe *non-interacting* points.

1.3 Connections to other areas of mathematics

Random matrix theory has deep connections with many areas of mathematics, many of which are still poorly understood. A brief overview of some of these connections is presented below. While some of these notions, such as the connections with stochastic PDE require more background than we assume, some other connections (e.g. with quantum gravity) are in fact more elementary (and fundamental) than one may naively expect. Our purpose here is to present a small sample of the rich set of ideas that make the subject so attractive.

1.3.1 Number theory

The Riemann zeta function is defined by the infinite sum

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \mathsf{Re}(s) > 1. \tag{1.3.1}$$

The function $\zeta(s)$ is central to number theory, since it provides a generating function for the distribution of the prime numbers via Euler's product formula

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}, \quad \mathsf{Re}(s) > 1. \tag{1.3.2}$$

For instance, the divergence of the harmonic series at $s = 1$ provides a proof that there are infinitely many primes. The study of $\zeta(s)$ by complex analysis is

the cornerstone of analytic number theory. The basic facts are as follows. The function $\zeta(z)$ extends to a meromorphic function on $\mathbb{C}$ by analytic continuation, which has a simple pole at $s = 1$ where the residue is 1. A closely related function is the Riemann ξ-function

$$\xi(s) = \frac{1}{2\pi^{s/2}} s(s-1)\Gamma\left(\frac{s}{2}\right)\zeta(s). \tag{1.3.3}$$

Recall that the Γ function is a 'continuous interpolation' of the factorial, defined by the integral

$$\Gamma(s) = \int_0^\infty \mathrm{e}^{-x} x^{s-1}\,\mathrm{d}x, \quad \mathsf{Re}(s) > 0. \tag{1.3.4}$$

The Γ-function extends to a meromorphic function $\mathbb{C}$, which has simple poles at $\ldots, -2, -1, 0$ where the residue is 1. These poles cancel the 'trivial' zeros of the ζ function, and the essential difficulties related to the study of the ζ function are more transparent for the ξ function. It satisfies the functional equation

$$\xi(s) = \xi(1-s), \quad s \in \mathbb{C}. \tag{1.3.5}$$

The celebrated *Riemann Hypothesis* is the conjecture that all zeros of the ξ function lie on the critical line $\mathsf{Re}(s) = 1/2$ (this line is the symmetry axis for the functional equation above). In his fundamental paper on the distribution of prime numbers (translated in [12] and [30]) Riemann presented a series of asymptotic expansions that would imply rigorous bounds on the distribution of primes if the Riemann Hypothesis is true.

The connection between random matrix theory and the Riemann Hypothesis is two-fold. First, if one could construct a Hermitian operator with point spectrum whose eigenvalues coincide with the zeros of $\xi(i(s-1/2)$ then the Riemann Hypothesis would follow immediately (since all eigenvalues of a Hermitian operator are real). The catch, of course, is to determine such an operator. Nevertheless, as we discuss below, random matrix theory has shed new lie on the spectral theory of several operators, deterministic and random. Thus, the theory provides a catalog of 'guesses'. Second, if one assumes the Riemann hypothesis, the *fluctuations* in the zeros of $\zeta(s)$ are described by the sine-kernel! Under the Riemann hypothesis, the non-trivial zeros of $\zeta(s)$ may be written $\gamma_n = \frac{1}{2} \pm it_n$ with $0 < t_1 < t_2 < \ldots$. Let

$$w_n = \frac{t_n}{2\pi}\log\left(\frac{t_n}{2\pi}\right), \quad \text{and} \quad N(x) = \sum_{k=1}^{\infty} \mathbf{1}_{w_n \le x}. \tag{1.3.6}$$

This rescaling is chosen so that $\lim_{x\to\infty} N(x)/x = 1$ in accordance with the Prime Number Theorem.

Despite the fact that the zeros w_n are deterministic, we may introduce probabilistic notions by counting the (rescaled) zeros upto a level $x > 0$. For example, we may define the empirical probability measure

$$\mu_1(\mathrm{d}w; x) = \frac{1}{N(x)} \sum_{k=1}^{N(x)} \delta_{w_k}(\mathrm{d}w). \tag{1.3.7}$$

In order to study the gaps between eigenvalues, we must consider instead the empirical measures

$$\mu_2(\mathrm{d}l; x) = \frac{1}{x} \sum_{1 \le j,k \le N(x); j \neq k} \delta_{w_j - w_k}(\mathrm{d}l). \tag{1.3.8}$$

The expectation of a continuous function with respect to $\mu_2(\mathrm{d}l; x)$ is denoted

$$R_2(f; x) = \int_{-\infty}^{\infty} f(l)\mu_2(\mathrm{d}l; x) = \frac{1}{x} \sum_{1 \le j < k \le N(x)} f\left(w_j - w_k\right). \tag{1.3.9}$$

Under the assumption that f is band-limited, i.e. that its Fourier transform has compact support, Montgomery established the following

Theorem 12 (Montgomery)**.** *Assume the Riemann Hypothesis. Assume f is a Schwartz function whose Fourier transform $\hat{f}$ is supported in $[-1, 1]$. Then*

$$\lim_{x \to \infty} R_2(f; x) = \int_{-\infty}^{\infty} f(l)\mu_2(\mathrm{d}l), \quad \mu_2(\mathrm{d}l) = \left(1 - \left(\frac{\sin \pi l}{\pi l}\right)^2\right) \mathrm{d}l. \tag{1.3.10}$$

The point here is that the right hand side of (1.3.10) is precisely the 2-point function for the sine process. More generally, Montgomery's theorem is now known to hold for the distribution of n-consecutive gaps. That is, the rescaled fluctuations converge to the Sine_2 process in distribution. Bourgade's thesis is an excellent review of the state of the art [4].

1.3.2 Combinatorics and enumerative geometry

We will present two problems of enumerative combinatorics that connect with random matrix theory. As a first example, we note that the $2m$-th moment of the semicircle law

$$\int_{-2}^{2} x^{2m} p_{\mathrm{sc}}(x)\, \mathrm{d}x = \frac{1}{m+1}\binom{2m}{m} = C_m, \tag{1.3.11}$$

the m-the Catalan number. An analytic proof of this identity follows from a comparison between the Stieltjes transform $R_{\mathrm{sc}}(z)$, and the generating function

$$\hat{C}(x) = \sum_{m \ge 0} C_m x^m = \frac{1 - \sqrt{1 - 4x}}{x}. \tag{1.3.12}$$

The Catalan numbers describe the solution to many combinatorial problems [3]. For example, C_m enumerates the number of *Bernoulli excursions* or *Dyck paths* of length $2m$: these are walks S_k, $1 \le k \le 2m$ such that $S_0 = S_{2m} = 0$, $S_k \ge 0$, $0 \le k \le 2m$, and $|S_{k+1} - S_k| = 1$.

[3] Stanley lists 66 examples in [32, Exercise 6.19].

A deeper set of connections between integrals on $\mathsf{Her}(n)$ and geometry was first noticed by the physicist 't Hooft [34]. Ignoring (for now), physicists' motivation, let us illustrate a particular computational technique that underlies their work. Consider a matrix integral of the form

$$Z_n(z) = \int_{\mathsf{Her}(n)} \mathrm{e}^{\mathrm{Tr}(-zM^4)} p_{\mathrm{GUE}}(M)\,\mathrm{D}M, \quad \mathsf{Re}(z) > 0. \tag{1.3.13}$$

The quartic nonlinearity prevents us from expressing this integral in closed form. Nevertheless, this integral may be expanded in a Taylor series

$$Z_n(z) = \sum_{k=0}^{\infty} \frac{(-z)^k}{k!} \int \left(\mathrm{Tr}(M^4)\right)^k p_{\mathrm{GUE}}(M)\,\mathrm{D}M, \quad \mathsf{Re}(z) > 0. \tag{1.3.14}$$

A fundamental lemma on Gaussian integrals (on $\mathbb{R}^N$) (Wick's lemma) allows us to reduce each integral above to a sum over pairings of indices. It is convenient to keep track of these pairings with a graphical description, called a *Feynman diagram*. 't Hooft observed that when $\mathbb{R}^N \equiv \mathsf{Her}(n)$ the Feynman diagram associated to each term in (1.3.14) enumerates embedded graphs on a Riemann surface. This characterization was independently discovered by mathematicians.

Lemma 1 (Harer-Zagier [14])**.** *Let $\varepsilon_g(m)$ denote the number of ways to pair the edges of a symmetric $2m$-gon to form an orientable surface with genus g. Then*

$$f(m,n) = \sum_{g=0}^{\infty} \varepsilon_g(m) n^{m+1-2g} = \int_{\mathsf{Her}(n)} \mathrm{Tr}(M^{2m}) p_{\mathrm{GUE}}(M)\,\mathrm{D}M. \tag{1.3.15}$$

Note that only finitely many terms in the sum are non-zero. The series above is an instance of a *genus*-expansion. It illustrates the beautiful fact that matrix integrals serve as the generating functions for Riemann surfaces with a given combinatorial decomposition!

1.3.3 Random permutations

Consider the symmetric group $\mathsf{S}(n)$ of permutations of size n. We have that $|\mathsf{S}(n)| = n!$ and any element of $\mathsf{S}(n)$ can be represented as a ordering of the integers $1, 2, \ldots, n$. For example, three elements of $\mathsf{S}(5)$ are

$$\pi_1 = 54312, \quad \pi_2 = 12435, \quad \pi_3 = 45123.$$

We define a function $\ell : \mathsf{S}(5) \to \mathbb{N}$ by $\ell(\pi) =$ length of the longest increasing subsequence of π. For example,

$$\ell(\pi_1) = 2, \quad \ell(\pi_2) = 3, \quad \ell(\pi_3) = 4.$$

There is a natural probability distribution $\mathrm{Uni}(n)$ on $\mathsf{S}(n)$, the uniform distribution, or Haar measure. If $\Pi_n \sim \mathrm{Uni}(n)$ then $\mathbb{P}(\Pi_n = \pi) = \frac{1}{n!}$ for any $\pi \in S_n$.

This problem was one of the first, if not the first, problem to be investigated by Monte Carlo simulation on a computer — Ulam performed simulations in the early 60's [37] and conjectured that

$$\frac{1}{\sqrt{n}}\mathbb{E}\left[\ell(\Pi_n)\right] \to c.$$

It was later established by Vershik and Kerov that $c = 2$ [38]. The detailed numerical computations of Odlyzko and Rains [28] indicated

$$\mathbb{E}\left[\ell(\Pi_n)\right] - 2\sqrt{n} = O(n^{-1/6}). \tag{1.3.16}$$

The comparison between (1.2.9) and (1.3.16) should be striking. Indeed, the following is often called the Baik–Deift–Johansson Theorem and it makes this scaling rigorous.

Theorem 13 ([2])**.** *Let* $\mathsf{S}(n)$*,* ℓ *and* Π_n *be as above. Then for all* $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\ell(\Pi_n) - 2\sqrt{n}}{n^{1/6}} \leq t\right) = \det(1 - K_{\mathsf{Airy}}\mathbf{1}_{(t,\infty)}).$$

That is, the limit is the same as the largest eigenvalue of a random Hermitian matrix.

This theorem is discussed in great detail in [3]. This surprising connection was explored further by Johansson [19] leading to many connections to random growth processes and the KPZ equation.

1.3.4 Spectral and inverse spectral theory of operators

While Theorem 2–Theorem 9 associate limits to the spectrum of the operators M_n, it is natural to ask if there are limiting operators that may be naturally associated to the limiting spectra. Thus, for Theorem 2 we ask for a 'natural' operator that has spectral density given by the semicircle law, p_{sc}, and for Theorem 6 and Theorem 9 we seek 'natural' random operators that have pure point spectra with the law of the Sine_2 and Airy_2 point processes. What is a 'natural' operator is, of course, a subjective idea, but convincing candidates operators are suggested by inverse spectral theory.

We say that a matrix $T \in \mathsf{Symm}(n)$ is a Jacobi matrix if all its off-diagonal entries are strictly positive. The spectral measure of a Jacobi matrix is the measure whose Stieltjes transform is $e_1^T(T-z)^{-1}e_1$. There is a $1-1$ correspondence between the space of $n \times n$ Jacobi matrices and probability measures on the line with n atoms. This correspondence extends naturally to semi-infinite Jacobi matrices. The essence of this theory (due to Stieltjes) is that the entries of T may be determined from the *continued fraction expansion* of $e_1^T(T-z)^{-1}e_1$. This correspondence will be considered in detail in Lecture 3, but here is a concrete example. By applying Stieltjes' procedure to the semicircle law, we

discover that $p_{\mathrm{sc}}(x)$ is the spectral density for the seminfinite tridiagonal matrix that is 1 on the off-diagonal, and 0 in all other entries. This follows from the continued fraction expansion

$$R_{\mathrm{sc}}(-z) = \cfrac{1}{z - \cfrac{1}{z - \cfrac{1}{z - \dots}}} \tag{1.3.17}$$

Ensembles of tridiagonal matrices are of practical important in numerical linear algebra. For instance, a key pre-processing step while solving symmetric linear systems is to transform the matrix to tridiagonal form by Householder's procedure (Lecture 3). Dumitriu and Edelman pushed forward the Gaussian measures under this procedure to obtain a family of tridiagonal ensembles, known as the general-β ensembles [8]. Further, Edelman and Sutton made a formal expansion of these operators, and observed that as $n \to \infty$, the tridiagonal operators appeared to converge to the *stochastic Airy operator* [11]:

$$H_\beta = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x + \frac{2}{\sqrt{\beta}}\dot{b}, \quad 0 < x < \infty \tag{1.3.18}$$

with Dirichlet boundary conditions at $x = 0$. Here $\dot{b}$ denotes (formally) white noise (it is not hard to define H_β rigorously).

Theorem 14 (Ramirez-Rider-Virag [29])**.** *The spectrum $\sigma(H_\beta)$ of the operator H_β is almost surely a countably infinite number of eigenvalues $\mu_1 < \mu_2 < \mu_3 < \dots$. Moreover, $\sigma(H_\beta)$ has the same law as the Airy_β point process.*

In particular, for $\beta = 2$, the spectrum of the stochastic Airy operator describes the limiting fluctuations at the edge of the spectrum of GUE matrices. Despite the simplicity of this characterization, it is not know how to recover the explicit determinantal formulas of Tracy and Widom from this formulation.

Chapter 2

Integration on spaces of matrices

In this section, we review the geometry of the classical Lie groups, as well as the spaces $\mathsf{Symm}(n)$, $\mathsf{Her}(n)$ and $\mathsf{Quart}(n)$ and explain how to integrate over these groups and spaces. Given an point on a manifold $M \in \mathcal{M}$, we use $\mathrm{d}M$ to denote the differential of M, i.e., an infintesimal element on the tangent space $T_M\mathcal{M}$ at M. We reserve $\mathrm{D}M$ to refer to a (naturally induced) volume form defined using an inner-product on the tangent space. Note that for $x \in \mathbb{R}$, $\mathrm{d}x = \mathrm{D}x$. Our main goal is the following

Theorem 15 (Weyl's formula)**.**

$$\mathrm{D}M = |\triangle(\Lambda)|^{\beta}\ \mathrm{D}\Lambda\,\mathrm{D}U \tag{2.0.1}$$

where $\triangle(\Lambda)$ *is the Vandermonde determinant*

$$\triangle(\Lambda) = (-1)^{\frac{n(n-1)}{2}} \prod_{1\le j<k\le n} (\lambda_j - \lambda_k), \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \tag{2.0.2}$$

$\mathrm{D}\Lambda$ *is Lebesgue measure on* $\mathbb{R}^n$*, and* $\mathrm{D}U$ *denotes (unnormalized) Haar measure on* $\mathsf{O}(n)$*, and an appropriately defined measure on* $\mathsf{U}(n)/\mathbb{T}^n$*, and* $\mathsf{USp}(n)/\mathbb{T}^n$ *in the cases* $\beta = 1, 2$ *and* 4 *respectively.*

The main strategy to prove Theorem 15 is to treat the mapping from matrices with distinct eigenvalues to their eigenvalues and eigenvectors. Then we identify the tangent spaces, and give a formula that relates the tangent space for the the eigenvalues and the tangent space for the eigenvectors to the tangent space for the matrix. This formula allows one to change variables in the metric tensor and therefore in the volume form.

Remark 16. It is common to normalize the Haar measure such that it is a probability measure. We have ignored this constant here, though is is explored

in the exercises. The essential aspect of (2.0.1) is that the Jacobian for diagonalization is given by $|\triangle(\Lambda)|^\beta$. This has far-reaching consequences for random matrix theory and has the interesting physical interpretation of *eigenvalue repulsion*.

In what follows, we first present a detailed description of integration on $\mathsf{O}(n)$ and $\mathsf{Symm}(n)$. The ideas are then extended to $\mathsf{Her}(n)$ and $\mathsf{Quart}(n)$.

2.1 Integration on $\mathsf{O}(n)$ and $\mathsf{Symm}(n)$

All matrices in this section lie in $\mathbb{R}^{n\times n}$.

There is a natural volume form on each finite-dimensional inner-product space of dimension p. For example, on $\mathbb{R}^p$, the standard inner product defines the metric with infinitesimal length element $\mathrm{d}s^2 = \sum_{j=1}^p \mathrm{d}x_j^2$ and the volume form $\mathrm{D}x = \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_p$ (we follow the notation of [39] for volume forms). More generally, each $g \in \mathsf{Symm}_+(p)$ defines an inner-product and metric on $\mathbb{R}^p$.

$$\langle x, y\rangle_g = \sum_{j,k=1}^p g_{jk}x_j y_k, \quad \mathrm{d}s^2 = \sum_{j,k=1}^p g_{jk}\mathrm{d}x_j\mathrm{d}x_k. \tag{2.1.1}$$

The associated p-dimensional volume form is

$$\mathrm{D}x = \sqrt{\det(g)}\,\mathrm{d}x_1\ldots\mathrm{d}x_p. \tag{2.1.2}$$

The following calculation demonstrates why the choice for the volume form, stemming from an inner-product, is important. Let X and Y be two finite-dimensional inner-product spaces (over $\mathbb{R}$) with inner-products $\langle\cdot,\cdot\rangle_X$, $\langle\cdot,\cdot\rangle_Y$ and let F be a (linear)[1] isometry $F : X \to Y$, i.e. $\langle Fx, Fy\rangle_Y = \langle x, y\rangle_X$ for all $x, y \in X$. Assume the dimension of these spaces is p. Then select bases $(u_i)_{1\le i\le p}$ of X and $(v_i)_{1\le i\le p}$ of Y. We then get the isomorphism

$$T_X : X \to \mathbb{R}^p, \quad T_X x = [\langle x, u_1\rangle_X, \langle x, u_2\rangle_X, \ldots, \langle x, u_p\rangle_X]^T, \tag{2.1.3}$$

and a similar isomorphism for Y. T_X becomes an isometry if we equip $\mathbb{R}^p$ with the inner-product

$$\langle a, b\rangle_{T_X} = \sum_{j,k=1}^p g^X_{ij}a_j b_k, \quad g^X_{jk} = \langle u_j, u_k\rangle_X. \tag{2.1.4}$$

Following the previous discussion, we arrive at two naturally induced volume forms on $\mathbb{R}^p$

$$\mathrm{D}x = \sqrt{\det(g^X)}\,\mathrm{d}x_1\ldots\mathrm{d}x_p, \tag{2.1.5}$$

$$\mathrm{D}y = \sqrt{\det(g^Y)}\,\mathrm{d}y_1\ldots\mathrm{d}y_p. \tag{2.1.6}$$

[1] It is not necessary to assume that the transformation is linear, but it takes more work to prove that an isometry must be affine.

Now, let $f : X \to \mathbb{R}$. We define (f is measurable if $f \circ T_X^{-1}$ is Lebesgue measurable)

$$\int_X f\mathrm{D}X = \int_{\mathbb{R}^p} f(T_X^{-1}x)\sqrt{\det(g^X)}\,\mathrm{d}x_1 \ldots \mathrm{d}x_p, \tag{2.1.7}$$

with a similar expression for $g : Y \to \mathbb{R}$. One should check that this definition is independent of the choice of basis. Now choose $f(x) = g(Fx)$. We find

$$\int_X f\mathrm{D}X = \int_{\mathbb{R}^p} g(FT_X^{-1}x)\sqrt{\det(g^X)}\,\mathrm{d}x_1 \ldots \mathrm{d}x_p. \tag{2.1.8}$$

If $(u_i)_{1\le i\le p}$ is a basis for X then $(Fu_i)_{1\le i\le p}$ is a basis for Y and we perform all these calculations for this basis. Consider the transformation

$$S : \mathbb{R}^p \to \mathbb{R}^p, \quad Sx = T_Y F T_X^{-1}. \tag{2.1.9}$$

But, after careful consideration, we see that $S = I$ and $\det(g^Y) = \det(g^X)$. Then, using this change of variables $y = Sx$ we find

$$\int_{\mathbb{R}^p} g(FT_X^{-1}x)\sqrt{\det(g^X)}\,\mathrm{d}x_1 \ldots \mathrm{d}x_p = \int_{\mathbb{R}^p} g(T_Y^{-1}y)\sqrt{\det(g^Y)}\,\mathrm{d}y_1 \ldots \mathrm{d}y_p. \tag{2.1.10}$$

This establishes the change of variables formula

$$\int_X (g \circ F)\mathrm{D}X = \int_Y g\mathrm{D}Y, \tag{2.1.11}$$

whenever F is an isometry from X to Y. In particular, this shows how inner-product structure on X and Y is converted to properties of volume.

The Lie group $\mathsf{O}(n)$ is the group, under composition, of linear transformations of $\mathbb{R}^n$ that preserve the standard metric $g = I$. For each $O \in \mathsf{O}(n)$ and each $x \in \mathbb{R}^n$ we must have $(Ox)^T(Ox) = x^T x$. Thus, $\mathsf{O}(n)$ is equivalent to the group of matrices O such that $O^T O = I$. The group operation is matrix multiplication. It is easy to check that the group axioms are satisfied, but a little more work is required to check that $\mathsf{O}(n)$ is a differentiable manifold, and that the group operation is smooth.

We now introduce the natural volume forms on $\mathsf{Symm}(n)$ and $\mathsf{O}(n)$. We first note that the space $\mathsf{Symm}(n)$ is isomorphic to $\mathbb{R}^p$, $p = n(n+1)/2$ via the map

$$M \mapsto \xi = (M_{11}, \ldots, M_{nn}, M_{12}, \ldots, M_{n-1,n}). \tag{2.1.12}$$

Thus, all that is needed to define integrals over $\mathsf{Symm}(n)$ is a choice of inner-product. We will always use the Hilbert–Schmidt inner product

$$\mathsf{Symm}(n) \times \mathsf{Symm}(n) \to \mathbb{R}, \quad (M, N) \mapsto \mathrm{Tr}(M^T N) = \mathrm{Tr}(MN). \tag{2.1.13}$$

The associated infinitesimal length element is

$$\mathrm{d}s^2 = \mathrm{Tr}(\mathrm{d}M^T \mathrm{d}M) = \sum_{j=1}^{n} \mathrm{d}M_{jj}^2 + 2\sum_{j<k} \mathrm{d}M_{jk}^2. \tag{2.1.14}$$

In ξ coordinates on $\mathbb{R}^p$, the associated metric tensor g is diagonal and takes the value 1 for the first n coordinates (diagonal terms), and the value 2 for all the other coordinates (off-diagonal terms). Thus, the metric tensor $g \in \mathsf{Symm}_+(p)$ has determinant $2^{n(n-1)/2}$. We apply formula (2.1.2) to find the following volume form on $\mathsf{Symm}(n)$,

$$\mathrm{D}M = 2^{n(n-1)/4} \prod_{j=1}^{n} \mathrm{d}M_{jj} \prod_{1 \le j < k \le n} \mathrm{d}M_{jk}. \tag{2.1.15}$$

Each $O \in \mathsf{O}(n)$ defines a map $\mathsf{Symm}(n) \to \mathsf{Symm}(n)$, $M \mapsto OMO^T$. This map is an isometry of $\mathsf{Symm}(n)$ with the metric above. It is in this sense that (2.1.17) is the natural inner-product. Since this map is an isometry, the volume element $\mathrm{D}M$ is also invariant.

$\mathsf{O}(n)$ is a differentiable manifold. Thus, in order to define a volume form on $\mathsf{O}(n)$, we must identify its tangent space $T\mathsf{O}(n)$, and then introduce an inner-product on $T\mathsf{O}(n)$. Further, the 'natural' inner-product must be invariant under the group operations. The tangent space at the identity to $\mathsf{O}(n)$, $T_I\mathsf{O}(n)$, is isomorphic to the Lie algebra, $\mathfrak{o}(n)$, of $\mathsf{O}(n)$. In order to compute $\mathfrak{o}(n)$ we consider smooth curves $(-a, a) \to \mathsf{O}(n)$, $a > 0$, $t \mapsto Q(t)$ with $Q(0) = I$, differentiate the equation $Q(t)^T Q(t) = I$ with respect to t, and evaluate at $t = 0$ to find

$$\dot{Q}(0)^T = -\dot{Q}(0). \tag{2.1.16}$$

Thus, each matrix in $\mathfrak{o}(n)$ is antisymmetric. Conversely, given an antisymmetric matrix A, the curve $t \mapsto e^{tA}$ gives a smooth curve in $\mathsf{O}(n)$ that is tangent to I at $t = 0$. Thus,

$$T_I\mathsf{O}(n) = \mathfrak{o}(n) = \{A \,|\, A = -A^T\}. \tag{2.1.17}$$

The tangent space at arbitrary $O \in \mathsf{O}(n)$ is obtained by replacing (2.2.2) with the condition that $O^T\dot{O}$ is antisymmetric. Thus,

$$T_O\mathsf{O}(n) = \{OA \,|\, A \in \mathfrak{o}(n)\}. \tag{2.1.18}$$

Finally, given $A, \tilde{A} \in \mathfrak{o}(n)$, we define their inner product $\langle A, \tilde{A} \rangle = \mathrm{Tr}(A^T\tilde{A}) = -\mathrm{Tr}(A\tilde{A})$. This inner-product is natural, because it is invariant under left-translation. That is, for two vector $OA, O\tilde{A} \in T_O\mathsf{O}(n)$ we find $\mathrm{Tr}\left(OA\right)^T(O\tilde{A}\Big) = \mathrm{Tr}(A^T\tilde{A})$. The associated volume form on $\mathsf{O}(n)$ is called *Haar measure*. It is unique, up to a normalizing factor, and we write

$$\mathrm{D}O = 2^{n(n-1)/4} \prod_{1 \le j < k \le n} \mathrm{d}A_{jk}. \tag{2.1.19}$$

Now let $f : \mathsf{O}(n) \to \mathbb{R}$ be a bounded, measurable function. Define a neighborhood of $O \in \mathsf{O}(n)$ by $B_\varepsilon(O) = \{\tilde{O} \in \mathsf{O}(n) : \|O - \tilde{O}\| < \varepsilon\}$. Then for $\varepsilon > 0$, sufficiently small, we can find a diffeomorphism (i.e., a chart) $\varphi_O : U_O \to B_\varepsilon(O) \subset \mathsf{O}(n)$, U_O open satisfying

$$0 \in U_O \subset T_O\mathsf{O}(n), \quad \varphi_O(0) = O \tag{2.1.20}$$

Then for such $\varepsilon > 0$ define

$$\int_{B_\varepsilon(O)} f \mathrm{D}O = 2^{n(n-1)/4} \int_{\varphi_O^{-1}(B_\varepsilon(O))} f(\varphi_O(A)) \prod_{1 \le j < k \le n} \mathrm{d}A_{jk}. \tag{2.1.21}$$

It can be verified that this is independent of the choice of φ_O. So, now consider such mapping at the identity, φ_I. And choose

$$\varphi_O(A) = O\varphi_I(O^T A). \tag{2.1.22}$$

We find

$$\int_{B_\varepsilon(O)} f \mathrm{D}O = 2^{n(n-1)/4} \int_{O\varphi_I^{-1}(B_\varepsilon(I))} f(O\varphi_I(O^T A)) \prod_{1 \le j < k \le n} \mathrm{d}A_{jk}. \tag{2.1.23}$$

In comparing with, (2.1.11), we use the fact that O furnishes an isometry from $T_I\mathsf{O}(n)$ to $T_O\mathsf{O}(n)$ so that

$$\int_{B_\varepsilon(O)} f \mathrm{D}O = 2^{n(n-1)/4} \int_{\varphi_I^{-1}(B_\varepsilon(I))} f(O\varphi_I(A)) \prod_{1 \le j < k \le n} \mathrm{d}A_{jk}. \tag{2.1.24}$$

In particular, if we choose $f \equiv 1$, then $\int_{B_\varepsilon(O)} \mathrm{D}O$ does not depend on $O \in \mathsf{O}(n)$, showing that this is indeed uniform measure on $\mathsf{O}(n)$.

2.2 Weyl's formula on $\mathsf{Symm}(n)$

Let us now recall some basic facts about $\mathsf{Symm}(n)$. Each matrix $M \in \mathsf{Symm}(n)$ has n real eigenvalues and an orthonormal basis of real eigenvectors. We write Λ for the matrix $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ of eigenvalues, and Q for a matrix whose k-th column is a normalized eigenvector of M associated to the eigenvalue λ_k, $1 \le k \le n$. Since the columns of Q are orthogonal and normalized to length 1, it is immediate that $Q \in \mathsf{O}(n)$. Thus,

$$MQ = Q\Lambda \quad \text{and} \quad M = Q\Lambda Q^T. \tag{2.2.1}$$

In what follows, we will view the transformation $M \mapsto (\Lambda, Q)$ as a change of variables, from $\mathsf{Symm}(n) \to \mathbb{R}^n \times \mathsf{O}(n)$. Strictly speaking, this change of variables is not well-defined since (2.2.1) is unaffected if we replace the k-th column Q_k of Q by $-Q_k$. This issue is considered more carefully in Lemma 3 and Lemma 5 below. In a loose sense, diagonalization is analogous to polar coordinates in $\mathbb{R}^n$,

$$\mathbb{R}^n \to [0, \infty) \times S^{n-1}, \quad x \mapsto (r, u), \quad r = |x|, u = \frac{x}{r}. \tag{2.2.2}$$

Polar coordinates are natural for rotation invariant probability density on $\mathbb{R}^n$. For example, the standard Gaussian measure on $\mathbb{R}^n$ may be written

$$e^{-\frac{|x|^2}{2}} \mathrm{D}x = C_n e^{-\frac{r^2}{2}} r^{n-1} \, \mathrm{d}r \, \mathrm{D}u, \tag{2.2.3}$$

where Du denotes the normalized $n-1$-dimensional measure on S^{n-1} and C_n is a universal constant. The factor r^{n-1} is the Jacobian of this transformation. Weyl's formula shows that the Jacobian for (2.2.1) is $|\triangle(\Lambda)|$. The proof of Weyl's formula relies on an orthogonal decomposition of $T_M\mathsf{Symm}(n)$.

Lemma 2. *Let M have distinct eigenvalues. Then*

$$T_M\mathsf{Symm}(n) \cong \mathbb{R}^n \oplus \mathfrak{o}(n), \tag{2.2.4}$$

and these subspaces are orthogonal.

Proof. We first assume that $M = \Lambda$ is diagonal. Consider a smooth curve $(-a, a) \to \mathsf{Symm}(n)$, $a > 0$, $t \mapsto M(t) = Q(t)\Lambda(t)Q(t)^T$ such that $M(0) = \Lambda(0) = \Lambda$, and $Q(0) = I$. We differentiate[2] this expression with respect to t and evaluate it at $t = 0$ to find the following expression for a tangent vector in $T_\Lambda\mathsf{Symm}(n)$:

$$\dot{M} = \dot{\Lambda} + [\dot{Q}, \Lambda]. \tag{2.2.5}$$

Here $\dot{\Lambda}$ can be an arbitrary diagonal matrix, and $\dot{Q}$ an arbitrary antisymmetric matrix. By the assumption of distinct eigenvalues, given $\dot{M}$, $\dot{\Lambda} = \text{diagonal}(\dot{M})$ and Λ, $\dot{Q}$ is uniquely determined. Since the diagonal terms of $\dot{Q}$ vanish these two matrices are orthogonal . Thus,

$$T_\Lambda\mathsf{Symm}(n) \cong \mathbb{R}^n \oplus \mathfrak{o}(n). \tag{2.2.6}$$

When $M = QMQ^T$ is not diagonal, we consider a curve $M(t)$ as above, with $M(0) = M$, $\Lambda(0) = \Lambda$ and $Q(0) = Q$. Now equation (2.2.5) is replaced by

$$\dot{M} = Q\left(\dot{\Lambda} + [Q^T\dot{Q}, \Lambda]\right)Q^T. \tag{2.2.7}$$

The matrices $Q^T\dot{Q}$ are antisymmetric and span $\mathfrak{o}(n)$. Again, $\dot{Q}$ is uniquely determined by $\dot{M}, \dot{\Lambda}$ and Λ. For arbitrary $\dot{\Lambda}$ and A we find $M(t) := Q\mathrm{e}^{tA}(\Lambda + t\dot{\Lambda})\mathrm{e}^{-tA}Q^T$ is a smooth curve in $\mathsf{Symm}(n)$, satisfying $M(0) = M$.

□

Remark 17. The proof of Lemma 2 reveals that all matrices of the form

$$M + \int_0^t Q(s)\left([Q(s)^T\dot{Q}(s), \Lambda]\right)Q(s)^T\mathrm{d}s \tag{2.2.8}$$

lie on an *isospectral manifold* – i.e. a manifold of matrices in $\mathsf{Symm}(n)$ with the same spectrum as Λ. And if one makes the ansatz $Q(t) = e^{tA}$ for an antisymmetric matrix A, one has

$$\dot{M} = [A, M]. \tag{2.2.9}$$

[2]Differentiability is guaranteed by classical perturbation theory [21, Theorem 5.4].

Proof of Weyl's formula for $\beta = 1$. We now have two coordinate systems on $T_M\mathsf{Symm}(n)$ provided that the eigenvalues of M are distinct. We take up this issue below and show that the set of all symmetric matrices with distinct eigenvalues is open. The coordinates ξ_α, $1 \le \alpha \le p$ give the metric (2.1.14). The second coordinate system, which is always locally defined, is $\left(\dot{\Lambda}, \dot{A}\right)$, where $\dot{\Lambda}$ is a diagonal matrix and $\dot{A}$ is an antisymmetric matrix. We use (2.2.7) to find the infinitesimal length element in this coordinate system. On the subset of $\mathsf{Symm}(n)$ consisting of matrices with distinct eigenvalues, using that M is symmetric, and $Q^T \mathrm{d}Q = \mathrm{d}A, \quad A \in \mathfrak{o}(n)$,

$$\begin{aligned} \operatorname{Tr} \mathrm{d}M^2 &= \operatorname{Tr}(\mathrm{d}M)^T \mathrm{d}M = \operatorname{Tr} Q(\mathrm{d}\Lambda + [\mathrm{d}A, \Lambda])^T (\mathrm{d}\Lambda + [\mathrm{d}A, \Lambda]) Q^T \\ &= \operatorname{Tr} \mathrm{d}\Lambda^2 + 2 \operatorname{Tr} \mathrm{d}\Lambda [\mathrm{d}A, \Lambda] + \operatorname{Tr}[\mathrm{d}A, \Lambda]^2 \\ &= \operatorname{Tr} \mathrm{d}\Lambda^2 + \operatorname{Tr}[\mathrm{d}A, \Lambda]^2. \end{aligned} \tag{2.2.10}$$

Expanding out this last trace, we find

$$\begin{aligned} \operatorname{Tr}[\mathrm{d}A, \Lambda]^2 &= \operatorname{Tr}(\mathrm{d}A\Lambda - \Lambda \mathrm{d}A)^2 \\ &= \operatorname{Tr} \mathrm{d}A\Lambda \mathrm{d}A\Lambda + \operatorname{Tr} \Lambda \mathrm{d}A\Lambda \mathrm{d}A - \operatorname{Tr} \Lambda \mathrm{d}A^2 \Lambda - \operatorname{Tr} \mathrm{d}A\Lambda^2 \mathrm{d}A \\ &= 2 \sum_{j=1}^n \sum_{k=1}^n \mathrm{d}A_{jk} \mathrm{d}A_{kj} \lambda_k \lambda_j - \sum_{j=1}^n \sum_{k=1}^n \mathrm{d}A_{jk} \mathrm{d}A_{kj} \lambda_j^2 - \sum_{j=1}^n \sum_{k=1}^n \mathrm{d}A_{jk} \mathrm{d}A_{kj} \lambda_k^2 \\ &= 2 \sum_{j<k} (\lambda_j - \lambda_k)^2 \mathrm{d}A_{jk}^2. \end{aligned} \tag{2.2.11}$$

Therefore

$$\mathrm{d}s^2 = \operatorname{Tr}(\mathrm{d}M^2) = \sum_{j=1}^n \mathrm{d}\lambda_j^2 + 2 \sum_{1 \le j < k \le n} (\lambda_j - \lambda_k)^2 \mathrm{d}A_{jk}^2. \tag{2.2.12}$$

Thus, the metric tensor in these coordinates is a diagonal matrix in $\mathsf{Symm}_+(p)$ that takes the value 1 on the first n coordinates, and the value $2(\lambda_j - \lambda_k)^2$ for each term A_{jk}. By (2.1.2), the volume form is

$$\mathrm{D}M = 2^{n(n-1)/4} \prod_{j=1}^n \mathrm{d}\lambda_j \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k| \, \mathrm{d}A_{jk} = |\triangle(\Lambda)| \, \mathrm{D}\Lambda \, \mathrm{D}O. \tag{2.2.13}$$

□

To interpret Weyl's formula, in a neighborhood U_M of a matrix with distinct eigenvalues, one needs to construct a well-defined map $\phi(M) = (\Lambda, Q)$ from symmetric matrices in this neighborhood to these "spectral" variables. Then for f with compact support in U_M

$$\int f(M) \mathrm{D}M = \int_{\phi(U_M)} f(Q\Lambda Q^T) |\triangle(\Lambda)| \mathrm{D}\Lambda \mathrm{D}O. \tag{2.2.14}$$

We now work to understand how to define such a map, and why matrices with repeated eigenvalues do not cause further issues.

2.3 Diagonalization as a change of coordinates

Some care is needed when treating the map $M \to (\Lambda, Q)$ as a change of variables. First, the map is not even well-defined in general, since the sign of each normalized eigenvector is arbitrary. Second, even if we fix the signs, the choice of eigenvectors is degenerate when M has repeated eigenvalues. Third, Λ is not uniquely defined if we do not specify an ordering of the eigenvalues. The following lemmas address this issue. Define the *Weyl chamber*

$$\mathcal{W}^n = \{\Lambda \in \mathbb{R}^n \,|\lambda_1 < \lambda_2 < \ldots < \lambda_n\}. \tag{2.3.1}$$

Lemma 3. *Assume $M_0 \in \mathsf{Symm}(n)$ has distinct eigenvalues. Then there exists $\varepsilon > 0$ such that for each $s \in \{\pm 1\}^n$, there is a C^∞ map*

$$h^{(s)} : B_\varepsilon(M_0) \to \mathcal{W}^n \times \mathsf{O}(n), \quad M \mapsto \left(\Lambda,\, Q^{(s)}\right)$$

that is a C^∞ diffeomorphism onto its image.

Proof of Lemma 3. An outline of the proof is presented. The remaining details are left to the exercises.

Standard perturbation theory (see [21], for example) demonstrates that the map is C^∞. The choice of s corresponds to fixing the signs of the eigenvectors as follows. Let a basis of normalized eigenvectors of M_0 be fixed. Call the associated matrix of eigenvectors Q_0. For each s, let $Q_0^{(s)} = \operatorname{diag}(s_1, \ldots, s_n)Q_0$. Each $Q_0^{(s)}$ is also an eigenvector matrix for M_0. Since the eigenvalues of M are distinct, we may use the implicit function theorem to solve the algebraic equations that determine the eigenvalues and eigenvectors, in a way that is consistent with the choice of s. □

Lemma 4 (Weilandt–Hoffman inequality)**.** *Let $M_1, M_2 \in \mathsf{Symm}(n)$ and use $\lambda_j(M_i)$ to denote the jth eigenvalue (in increasing order) of M_i. Then*

$$\sum_{j=1}^{n} |\lambda_j(M_1) - \lambda_j(M_2)|^2 \leq \|M_1 - M_2\|^2.$$

Proof. See [35, Section 1.3] for a particularly nice proof. □

Lemma 5. *Assume that $M \in \mathsf{Symm}(n)$ has a repeated eigenvalue. Then for every $\varepsilon > 0$ there exists $M_\varepsilon \in \mathsf{Symm}(n)$, such that $\|M - M_\varepsilon\| < \varepsilon$ and M_ε has distinct eigenvalues. Furthermore, the set of all matrices in $\mathsf{Symm}(n)$ with distinct eigenvalues is open.*

Proof. Exercise. □

Lemma 3 shows that the map $M \mapsto (\Lambda, Q)$ provides a local coordinate system near each matrix with distinct eigenvalues. Lemma 5 shows that set of such matrices is dense. More is true. The set of all matrices with both distinct eigenvalues and non-vanishing first entries in its eigenvectors is of full measure.

This follows from (3.3.3) and Lemma 8 below. Truly, one has to note that the procedure of reducing a full matrix to a tridiagonal matrix that is used to establish (3.3.3) does not affect the first row of the eigenvector matrix.

2.4 Independence and Invariance implies Gaussian

Fix $M \in \mathsf{Symm}(n)$ with spectrum $\sigma(M)$. Fix an interval $(a,b) \subset \mathbb{R}$ and let $\mathsf{Symm}(n)_{(a,b)}$ denote the set of $M \in \mathsf{Symm}(n)$ with spectrum $\sigma(M) \subset (a,b)$. Each function $f : (a,b) \to \mathbb{R}$ extends naturally to a map $\mathsf{Symm}(n)_{(a,b)} \to \mathsf{Symm}(n)$ as follows:

$$f(M) = Qf(\Lambda)Q^T, \quad M = Q\Lambda Q^T, \quad f(\Lambda) = \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n)). \tag{2.4.1}$$

Clearly, $\mathrm{Tr}(f(M)) = \mathrm{Tr}(f(\Lambda)) = \sum_{j=1}^n f(\lambda_j)$. Each $f : \mathbb{R} \to \mathbb{R}$ that grows sufficiently fast as $x \to \pm\infty$ defines an invariant distribution on $\mathsf{Symm}(n)$

$$\mu(\mathrm{D}M) = \frac{1}{Z} \exp\left(-\mathrm{Tr}(f(M))\right) \mathrm{D}M. \tag{2.4.2}$$

This is the most general form of an invariant probability distribution.

By contrast, a Wigner distribution relies on independence of the entries of M. This means that if a Wigner distribution has a density, then it must be of the form

$$\mu(\mathrm{D}M) = \frac{1}{Z} \left(\prod_{j=1}^n f_j\left(M_{jj}\right) \prod_{1 \le j < k \le n} f_{jk}\left(M_{jk}\right) \right) \mathrm{D}M. \tag{2.4.3}$$

Theorem 18. *Assume a probability measure μ on $\mathsf{Symm}(n)$ is both a Wigner distribution and an invariant distribution. Assume further that $\mu(\mathrm{D}M)$ has a strictly positive, smooth density of the form (2.4.2) and (2.4.3). Then $\mu(\mathrm{D}M)$ is a Gaussian ensemble,*

$$\mu(\mathrm{D}M) = \frac{1}{Z} \mathrm{e}^{-\frac{1}{2\sigma^2}\mathrm{Tr}(M-\gamma I)^2} \mathrm{D}M, \tag{2.4.4}$$

with variance σ^2 and mean $\gamma\, I$, for some $\gamma \in \mathbb{R}$.

Proof. We first illustrate the essential calculation for 2×2 matrices. Suppose

$$\mu(\mathrm{D}M) = p(M)\,\mathrm{D}M = \frac{1}{Z} f(M_{11}) g(M_{22}) h(M_{12}) \mathrm{d}M_{11} \mathrm{d}M_{12} \mathrm{d}M_{22}. \tag{2.4.5}$$

We compute the variation in μ along an isospectral curve (see Remark 17). Consider the curve $M(t) = Q(t) M Q(t)^T$ with

$$Q(t) = \mathrm{e}^{tR}, \quad R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{2.4.6}$$

The matrix R spans $\mathfrak{so}(2)$. We differentiate $M(t)$ with respect to t to obtain

$$\dot{M}(0) = [R, M] = \begin{pmatrix} -2M_{12} & M_{11} - M_{22} \\ M_{11} - M_{22} & 2M_{12} \end{pmatrix} \tag{2.4.7}$$

Thus, the infinitesimal change in the density $p(M(t))$ is

$$\begin{aligned} \frac{1}{p}\frac{\mathrm{d}p}{\mathrm{d}t}\bigg|_{t=0} &= \frac{f'(M_{11})}{f(M_{11})}\dot{M}_{11} + \frac{g'(M_{22})}{g(M_{22})}\dot{M}_{22} + \frac{h'(M_{12})}{h(M_{12})}\dot{M}_{12} \qquad (2.4.8) \\ &= -2M_{12}\left(\frac{f'(M_{11})}{f(M_{11})} - \frac{g'(M_{22})}{g(M_{22})}\right) + (M_{11} - M_{22})\frac{h'(M_{12})}{h(M_{12})}. \end{aligned}$$

On the other hand, since $\mu(\mathrm{D}M)$ is invariant, $p(M(t)) = p(M)$ and

$$\frac{\mathrm{d}p}{\mathrm{d}t}\bigg|_{t=0} = 0. \tag{2.4.9}$$

We equate (2.4.8) and (2.4.9), and separate variables to obtain

$$\frac{1}{M_{11} - M_{22}}\left(\frac{f'(M_{11})}{f(M_{11})} - \frac{g'(M_{22})}{g(M_{22})}\right) = c = \frac{1}{2M_{12}}\frac{h'(M_{12})}{h(M_{12})}, \tag{2.4.10}$$

for some constant $c \in \mathbb{R}$. Equation (2.4.10) immediately implies that

$$h(M_{12}) = h(0)\mathrm{e}^{cM_{12}^2}. \tag{2.4.11}$$

Separating variables again in (2.4.10), we find with a second constant $b \in \mathbb{R}$,

$$\frac{f'}{f} = cM_{11} + b, \quad \frac{g'}{g} = cM_{22} + b, \tag{2.4.12}$$

which integrates to

$$f(M_{11}) = f(0)\mathrm{e}^{\frac{cM_{11}^2}{2}}\mathrm{e}^{bM_{11}}, \quad g(M_{22}) = g(0)e^{\frac{cM_{22}^2}{2}}\mathrm{e}^{bM_{22}}. \tag{2.4.13}$$

We combine all the terms to obtain

$$p(M) = f(0)g(0)h(0)\mathrm{e}^{c\frac{\mathrm{Tr}(M^2)}{2}}\mathrm{e}^{b\,\mathrm{Tr}(M)}. \tag{2.4.14}$$

Since $p(M)$ integrates to 1, we must have $c < 0$, say $c = -1/\sigma^2$. The scalar b is arbitrary and contributes a shift in the mean that is a scalar multiple of I. The combination of constants $f(0)g(0)h(0)$ may be absorbed into the normalization constant Z^{-1}. We have thus proved Theorem 18 for $n = 2$.

In order to prove Theorem 18 for arbitrary n we generalize the above argument as follows. Fix a pair of off-diagonal indices $1 \leq l < m \leq n$. We consider a rotation in $\mathbb{R}^n$ that rotates the $x_l x_m$ plane as above, and leaves the other coordinates invariant. This entails replacing the matrix R in the argument above with the matrix $R^{lm} \in \mathfrak{so}(n)$ with coordinates $R^{lm}_{jk} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}$. The

argument above now shows that the density of p in the M_{ll}, M_{lm} and M_{mm} coordinates is a Gaussian distribution of the form (2.4.14):

$$p(M^{lm}) = \mathrm{e}^{c\frac{\mathrm{Tr}((M^{lm})^2)}{2}} \mathrm{e}^{b\,\mathrm{Tr}(M^{lm})}, \tag{2.4.15}$$

where M^{lm} denotes the 2×2 matrix

$$M^{lm} = \begin{pmatrix} M_{ll} & M_{lm} \\ M_{lm} & M_{mm} \end{pmatrix}.$$

At this stage, the constants c and b depend on l and m. But now note that since the same argument applies to *every* pair of indices $1 \leq l < m \leq n$, the constants c and b must be independent of l and m. □

2.5 Integration on $\mathsf{Her}(n)$ and $\mathsf{U}(n)$

The space of Hermitian matrices $\mathsf{Her}(n)$ is a vector-space of real dimension n^2, as may be seen by the isomorphism $\mathsf{Her}(n) \to \mathbb{R}^{n^2}$,

$$M \mapsto \xi = (M_{11}, \ldots, M_{nn}, \mathsf{Re}M_{12}, \ldots, \mathsf{Re}M_{n-1,n}, \mathsf{Im}M_{12}, \ldots, \mathsf{Im}M_{n-1,n}). \tag{2.5.1}$$

The Hilbert-Schmidt inner product on $\mathsf{Her}(n)$ is

$$\mathsf{Her}(n) \times \mathsf{Her}(n) \to \mathbb{C}, \quad (M, N) \mapsto \mathrm{Tr}(M^* N). \tag{2.5.2}$$

The associated infinitesimal length element is

$$\mathrm{d}s^2 = \mathrm{Tr}(\mathrm{d}M^2) = \sum_{j=1}^{n} \mathrm{d}M_{jj}^2 + 2 \sum_{1 \leq j < k \leq n} \left(\mathrm{d}\,\mathsf{Re}M_{jk}^2 + \mathrm{d}\,\mathsf{Im}M_{jk}^2 \right). \tag{2.5.3}$$

Thus, in the coordinates ξ, the metric is an $n^2 \times n^2$ diagonal matrix whose first n entries are 1 and all other entries are 2. We apply (2.2.1) to obtain the volume form on $\mathsf{Her}(n)$

$$\mathrm{D}M = 2^{n(n-1)/2} \prod_{j=1}^{n} \mathrm{d}M_{jj} \prod_{1 \leq j < k \leq n} \mathrm{d}\,\mathsf{Re}M_{jk}\,\mathrm{d}\,\mathsf{Im}M_{jk}. \tag{2.5.4}$$

The unitary group, $\mathsf{U}(n)$ is the group of linear isometries of $\mathbb{C}^n$ equipped with the standard inner-product $\langle x, y \rangle = x^* y$. Thus, $\mathsf{U}(n)$ is equivalent to the group of matrices $U \in \mathbb{C}^{n \times n}$ such that $U^* U = I$. The inner-product (2.5.3) and volume form (2.5.4) are invariant under the transformation $M \mapsto UMU^*$.

The Lie algebra $\mathfrak{u}(n)$ is computed as in Section 2.1. We find

$$\mathfrak{u}(n) = T_I \mathsf{U}(n) = \left\{ A \in \mathbb{C}^{n \times n} \,|\, A = -A^* \right\}, \quad T_U \mathsf{U}(n) = \{ UA \,|\, A \in \mathfrak{u}(n) \}. \tag{2.5.5}$$

The transformation $M \mapsto \mathrm{i}M$ is an isomorphism between Hermitian and anti-Hermitian matrices. In fact, the map $\mathsf{Her}(n) \to \mathsf{U}(n)$, $M \mapsto \mathrm{e}^{\mathrm{i}M}$ is onto and

locally one-to-one. The inner-product $A, \tilde{A} \mapsto \mathrm{Tr}(A^*\tilde{A})$ is invariant under left application of $\mathsf{U}(n)$. Thus, we obtain the volume form for Haar measure on $\mathsf{U}(n)$

$$\mathrm{D}\tilde{U} = 2^{n(n-1)/2} \prod_{j=1}^{n} \mathrm{d}A_{jj} \prod_{1\le j<k\le n} \mathrm{d}\,\mathsf{Re}A_{jk}\,\mathrm{d}\,\mathsf{Im}A_{jk}. \tag{2.5.6}$$

However, when viewing diagonalization $M \mapsto U\Lambda U^*$ as a change of variables on $\mathsf{Her}(n)$, it is necessary to quotient out the following degeneracy: For each $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{T}^n$, the diagonal matrix $D = \mathrm{diag}\left(e^{i\theta_1}, \ldots, e^{i\theta_n}\right)$ is unitary and $M = U\Lambda U^*$ if and only if $M = UD\Lambda D^*U^*$. Thus, for $\mathsf{Her}(n)$, the measure $D\tilde{U}$ must be replaced a measure on $\mathsf{U}(n)/\mathbb{T}^n$. The form of this measure on $\mathsf{U}(n)/\mathbb{T}^n$ follows from the following assertion, which is proved as in Section 2.1.

Lemma 6. *Each matrix $\dot{M} \in T_M\mathsf{Her}(n)$ is of the form*

$$\dot{M} = U\left(\dot{\Lambda} + [U^*\dot{U}, \Lambda]\right)U^*, \quad \dot{\Lambda} \in T_\Lambda\mathbb{R}^n, \quad \dot{U} \in T_U\mathsf{U}(n), \quad \mathrm{diagonal}(U^*\dot{U}) = 0. \tag{2.5.7}$$

The matrices $\dot{\Lambda}$ and $U^*\dot{U}$ are orthogonal under the inner-product (2.5.2).

Thus, the volume form on the quotient $\mathsf{U}(n)/\mathbb{T}^n$ is locally equivalent to a volume form on the subspace of $\mathsf{Her}(n)$ consisting of matrices with zero diagonal:

$$\mathrm{D}U = 2^{n(n-1)/2} \prod_{1\le j<k\le n} \mathrm{d}\,\mathsf{Re}A_{jk}\,\mathrm{d}\,\mathsf{Im}A_{jk}. \tag{2.5.8}$$

Furthermore, $B \mapsto \varphi(B) = U\mathrm{e}^{U^*B}$ provides a locally one-to-one mapping from $PT_U\mathsf{U}(n) = UT_I\mathsf{U}(n)$ to $\mathsf{U}(n)/\mathbb{T}^n$.

Lemma 6 shows that the mapping $\mathbb{R}^n \oplus PT_I\mathsf{U}(n) \to T_M\mathsf{Her}(n)$, $PT_I\mathsf{U}(n) = \{A \in T_I\mathsf{U}(n) \mid \mathrm{diag}(A) = 0\}$, defined by $(\dot{\Lambda}, \dot{A}) \mapsto U(\dot{\Lambda} + [\dot{A}, \Lambda])U^*$ maps onto $T_M\mathsf{Her}(n)$. Again, the two spaces are isomorphic if M has distinct eigenvalues.

Proof of Weyl's formula for $\beta = 2$. We write, on the subset of $\mathsf{Symm}(n)$ consisting of matrices with distinct eigenvalues, using that M is Hermitian, and $U^*\mathrm{d}U = \mathrm{d}A, \quad A \in T_I\mathsf{U}(n), \mathrm{diag}(A) = 0$,

$$\begin{aligned} \mathrm{Tr}\,\mathrm{d}M^2 &= \mathrm{Tr}\,\mathrm{d}\Lambda^2 + 2\,\mathrm{Tr}\,\mathrm{d}\Lambda[\mathrm{d}A, \Lambda] + \mathrm{Tr}[\mathrm{d}A, \Lambda]^*[\mathrm{d}A, \Lambda] \\ &= \mathrm{Tr}\,\mathrm{d}\Lambda^2 + \mathrm{Tr}[\mathrm{d}A, \Lambda]^*[\mathrm{d}A, \Lambda]. \end{aligned} \tag{2.5.9}$$

Expanding out this last trace, using that $\mathrm{d}A = \mathrm{d}\mathsf{Re}A + \mathrm{i}\,\mathrm{d}\mathsf{Im}A$, we need only collect the real part

$$\begin{aligned} \mathrm{Tr}[\mathrm{d}A, \Lambda]^*[\mathrm{d}A, \Lambda] &= \mathrm{Tr}(\mathrm{d}\,\mathsf{Re}A)\Lambda(\mathrm{d}\,\mathsf{Re}A)\Lambda + \mathrm{Tr}\,\Lambda(\mathrm{d}\,\mathsf{Re}A)\Lambda(\mathrm{d}\,\mathsf{Re}A) \\ &\quad - \mathrm{Tr}\,\Lambda(\mathrm{d}\,\mathsf{Re}A)^2\Lambda - \mathrm{Tr}(\mathrm{d}\,\mathsf{Re}A)\Lambda^2(\mathrm{d}\,\mathsf{Re}A) \\ &\quad + \mathrm{Tr}(\mathrm{d}\,\mathsf{Im}A)\Lambda(\mathrm{d}\,\mathsf{Im}A)\Lambda + \mathrm{Tr}\,\Lambda(\mathrm{d}\,\mathsf{Im}A)\Lambda(\mathrm{d}\,\mathsf{Im}A) \\ &\quad - \mathrm{Tr}\,\Lambda(\mathrm{d}\,\mathsf{Im}A)^2\Lambda - \mathrm{Tr}(\mathrm{d}\,\mathsf{Im}A)\Lambda^2(\mathrm{d}\,\mathsf{Im}A) \\ &= 2\sum_{j<k}(\lambda_j - \lambda_k)^2\mathrm{d}\mathsf{Re}A_{jk}^2 + 2\sum_{j<k}(\lambda_j - \lambda_k)^2\mathrm{d}\mathsf{Im}A_{jk}^2. \end{aligned} \tag{2.5.10}$$

Then it follows that the associated volume form satisfies

$$\mathrm{D}M = |\triangle(\Lambda)|^2 \mathrm{D}\Lambda \mathrm{D}U. \tag{2.5.11}$$

□

2.6 Integration on Quart(n) and USp(n)

The field of quaternions, $\mathbb{H}$, is the linear space

$$x = c_0 + c_1\mathfrak{e}_1 + c_2\mathfrak{e}_2 + c_3\mathfrak{e}_3, \quad c_i \in \mathbb{R},\ i = 0, 1, 2, 3, \tag{2.6.1}$$

equipped with the non-commutative rules of multiplication

$$\mathfrak{e}_1^2 = \mathfrak{e}_2^2 = \mathfrak{e}_3^2 = \mathfrak{e}_1\mathfrak{e}_2\mathfrak{e}_3 = -1. \tag{2.6.2}$$

These rules ensure that the product of any two quaternions is again a quaternion. Each $x \in \mathbb{H}$ has a complex conjugate $\bar{x} = c_0 - c_1\mathfrak{e}_1 - c_2\mathfrak{e}_2 - c_3\mathfrak{e}_3$, and its absolute value $|x|$ is determined by

$$|x|^2 = \bar{x}x = c_0^2 + c_1^2 + c_2^2 + c_3^2. \tag{2.6.3}$$

Each non-zero $x \in \mathbb{H}$ has a multiplicative inverse $1/x = \bar{x}/|x|^2$. Thus, $\mathbb{H}$ is indeed a field.

The normed linear vector space $\mathbb{H}^n$ consists of vectors $x = (x_1, \ldots, x_n)^T$ with inner product $\langle x, y\rangle = \sum_{j=1}^n \bar{x}_j y_j$. The adjoint, $M^\dagger$ of a linear transformation $M : \mathbb{H}^n \to \mathbb{H}^n$ is defined by the inner-product

$$\langle M^\dagger x, y\rangle := \langle x, My\rangle. \tag{2.6.4}$$

It follows that the entries of $M^\dagger$ are $M_{jk} = \bar{M}_{kj}$. We say that an operator is self-adjoint if $M = M^\dagger$. It is anti self-adjoint if $M = -M^\dagger$. The space of self-adjoint operators is denoted Quart(n). We equip this space with the Hilbert-Schmidt norm as before.

The group USp(n) is the set of linear transformations of $\mathbb{H}^n$ that preserve this inner product. We thus require that for each $x, y \in \mathbb{H}^n$

$$\langle x, y\rangle = \langle Ux, Uy\rangle = \langle U^\dagger Ux, y\rangle. \tag{2.6.5}$$

Thus, USp(n) is equivalent to $U \in \mathbb{H}^{n\times n}$ such that $U^\dagger U = I$. As for U(n) we find that its Lie algebra $\mathfrak{usp}(n)$ is the space of anti self-adjoint matrices. The inner-product on $\mathfrak{usp}(n)$ and Haar measure are defined exactly as in Section 2.5, as is the analogue of Lemma 6 and the Weyl formula. It is also clear from how the proof of Weyl's formula extends to $\beta = 2$, that because the field of quarternions is a four-dimensional space, $|\triangle(\Lambda)|^4$ will arise, see (2.5.10).

Exercises

2.1. Show that

$$\triangle(\Lambda) = \det \begin{pmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_n \\ \vdots & & \vdots \\ \lambda_1^{n-1} & \dots & \lambda_n^{n-1}. \end{pmatrix}. \tag{2.6.6}$$

2.2. The *Pauli matrices*,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.6.7}$$

allow a representation of the quarternions in terms of Hermitian matrices.

(a) Show that the Pauli matrices together with the identity matrix span $\mathsf{Her}(2)$.

(b) Show that the matrices $\{\mathrm{i}\sigma_1, \mathrm{i}\sigma_2, \mathrm{i}\sigma_3\}$ form a basis of $\mathfrak{su}(2)$. (This is the subalgebra of $\mathfrak{u}(2)$ consisting of trace-free matrices).

(c) Verify that if $\mathfrak{e}_j = i\sigma_j$, the rules (2.6.2) hold (replace 1 by I_2).

2.3. The *canonical symplectic matrix* of size $2n \times 2n$ denoted J_n, or simply J, is the matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \tag{2.6.8}$$

where 0 and I denote the $n \times n$ zero and identity matrices. The *symplectic group* $\mathsf{Sp}(2n, \mathbb{R})$ (not to be confused with the *unitary* symplectic group $\mathsf{USp}(n)$!) is

$$\mathsf{Sp}(2n, \mathbb{R}) = \left\{ S \in \mathbb{R}^{n \times n} \,\middle|\, S^T J S = J \right\}. \tag{2.6.9}$$

Verify that $\mathsf{Sp}(2n, \mathbb{R})$ is a group and compute its Lie algebra $\mathfrak{sp}(2n, \mathbb{R})$.

2.4. Use the Gaussian integral

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\frac{|x|^2}{2}} \, \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

to compute the $n-1$-dimensional volume ω_{n-1} of the unit sphere S^{n-1}. Determine the asymptotic behavior of ω_{n-1} as $n \to \infty$.
Hint: Do the integral two ways– once in Cartesian coordinates, and once in polar coordinates.

2.5. Assume given a C^1 function $f : (a, b) \to \mathbb{R}$, and extend it to a function $f : \mathsf{Symm}(n) \to \mathsf{Symm}(n)$ as in (2.4.1). Compute the Jacobian of this transformation. Apply this formula to the function $f(x) = e^{ix}$ to compute the analogue of Weyl's formula on $\mathsf{U}(n)$ (note that each $U \in \mathsf{U}(n)$ is of the form e^{iM} for some $M \in \mathsf{Her}(n)$).

2.6. Prove Lemma 4.

2.7. Let $A \in \mathbb{R}^{m\times n}$ for $m < n$. Show that $\{x \mid Ax = 0\} \subset \mathbb{R}^n$ has zero Lebesgue measure.

2.8. Assume $f : \mathbb{R} \to (0, \infty)$ satisfies the *functional equation*

$$f(x + y) = f(x)f(y), \quad x, y \in \mathbb{R}. \tag{2.6.10}$$

It is easy to check that for each $a \in \mathbb{R}$ functions of the form $f(x) = e^{ax}$ solve (2.6.10). Show that these are the only solutions to (2.6.10) assuming only that f is continuous. (Do *not* assume that f is differentiable).

Remark 19. The use of row operations in Problem (1) underlies the introduction of orthogonal polynomials. Problems (2) and (3) may be combined to show that $\mathsf{Sp}(2n, \mathbb{C}) \cap \mathsf{U}(n) \cong \mathsf{USp}(n)$. The approach in Problem (4) yields the volume of $\mathsf{O}(n)$, $\mathsf{U}(n)$ and $\mathsf{USp}(n)$ when applied to GOE, GUE and GSE. The assumptions of Problem (7) may be weakened further – measurability is enough! You could try to develop a similar approach for the functional equation implicit in the proof of Theorem 18. That is, can you establish a stronger form of Theorem 18 that does not assume differentiability ?

2.9. Show that the mapping $A \mapsto (I-A)(A+I)^{-1}$ from $\mathfrak{o}(n)$ to $\mathsf{O}(n)$ is bijective in a neighborhood of 0 to a neighborhood of the identity. Construct an atlas of $\mathsf{O}(n)$ using this mapping.

2.10. Using the *Submersion Theorem* [5, Proposition 3.42] (also called the Regular Value theorem) show that $\mathsf{O}(n)$ is a smooth manifold.
Hint: Consider $\phi : \mathbb{R}^{n\times n} \to \mathsf{Symm}(n)$ defined by $\phi(X) = X^T X$. Then show that I is a regular value and therefore $\phi^{-1}(I) = \mathsf{O}(n)$ is a smooth manifold.

Chapter 3

Jacobi matrices and tridiagonal ensembles

3.1 Jacobi ensembles

The space of real $n \times n$ tridiagonal matrices is denoted $\mathsf{Tridiag}(n)$. A typical matrix in $\mathsf{Tridiag}(n)$ is written

$$T = \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & b_2 & & 0 \\ 0 & b_2 & a_3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & b_{n-1} \\ 0 & 0 & \dots & b_{n-1} & a_n \end{pmatrix}. \tag{3.1.1}$$

Jacobi matrices, and their closure within the space $\mathsf{Tridiag}(n)$ are the manifolds

$$\begin{aligned} \mathsf{Jac}(n) &= \{T \in \mathsf{Tridiag}(n) \mid b_j > 0,\, 1 \le j \le n\}, \\ \overline{\mathsf{Jac}(n)} &= \{T \in \mathsf{Tridiag}(n) \mid b_j \ge 0,\, 1 \le j \le n\}. \end{aligned} \tag{3.1.2}$$

Jacobi matrices, or more generally Jacobi operators, are of fundamental importance in spectral theory. A self-adjoint operator K on a Hilbert space can be decomposed using its cyclic subspaces. On each of these cyclic subspaces an orthonormal basis for $\mathrm{span}\{K^j x \mid j = 0, 1, 2, \dots\}$ can be found and the operator K becomes tridiagonal in this basis. This is an idea used by conjugate gradient algorithm [16]. They also play an important role in approximation theory, the theory of orthogonal polynomials, and more widely in numerical linear algebra. An essential step in the symmetric eigenvalue problem is the reduction of a full symmetric matrix to an isospectral tridiagonal matrix (*tridiagonalization*) by a sequence of orthogonal reflections. Under this procedure, the Gaussian ensembles push forward to ensembles of tridiagonal matrices whose laws have the following simple description.

Definition 20 (Dumitriu–Edelman [8])**.** For each $\beta > 0$, the Hermite(β) ensemble consists of $T \in \mathsf{Tridiag}(n)$ such that a_k, $1 \leq k \leq n$, are iid normal random variables with mean zero and variance $2/\beta$, and b_k, $1 \leq k \leq n-1$ where are independent $\chi_{(n-k)\beta}(1/\beta)$ random variables.

Then density for $\chi_k(\sigma^2)$ is supported on $[0, \infty)$ and is proportional to

$$t^{k-1} \mathrm{e}^{-\frac{t^2}{2\sigma^2}}.$$

The point here is that the Hermite(β) ensembles are the push-forwards of the Gaussian ensembles when $\beta = 1$, 2 or 4. Then they interpolate Dyson's classification of ensembles to every $\beta > 0$. When combined with classical spectral theory, they provide a distinct, and important, perspective on the limit theorems of random matrix theory. Our immediate goal in this chapter is the following

Theorem 21. *Fix $\beta > 0$ and assume $T \sim$ Hermite(β). Then the marginal distribution of its eigenvalues is*

$$p_{\mathrm{Hermite}(\beta)}(\Lambda) \mathrm{D}\Lambda = \frac{1}{Z_{n,\beta}} \mathrm{e}^{-\frac{\beta}{4} \mathrm{Tr}(\Lambda)^2} |\triangle(\Lambda)|^\beta \, \mathrm{D}\Lambda. \tag{3.1.3}$$

The chapter concludes with a more refined version of Theorem 21 that includes the distribution of the *spectral measure* of matrices $T \sim$ Hermite(β).

3.2 Householder tridiagonalization on $\mathsf{Symm}(n)$

Each $M \in \mathsf{Symm}(n)$ may be diagonalized $M = Q \Lambda Q^T$. However, the computation of Λ depends on the solvability of the characteristic polynomial $\det(zI - M) = 0$. For $n \geq 5$, there is no general closed form solution for the characteristic polynomial[1]. Nevertheless, every matrix always admits the following reduction.

Theorem 22. *For every $M \in \mathsf{Symm}(n)$ there exists a tridiagonal matrix T and $Q \in \mathsf{O}(n)$ such that*

$$M = QTQ^T. \tag{3.2.1}$$

The transformation (3.2.1) is given by a change of variables

$$\mathsf{Symm}(n) \to \overline{\mathsf{Jac}(n)} \times \left(S^{n-2} \times S^{n-3} \times \ldots S^1 \right). \tag{3.2.2}$$

under which the volume form $\mathrm{D}M$ on $\mathsf{Symm}(n)$ transforms as follows:

$$\mathrm{D}M = C_n \prod_{j=1}^{n} \mathrm{d}a_j \prod_{k=1}^{n-1} b_k^{(n-k)-1} \mathrm{d}b_k \prod_{l=1}^{n-2} \mathrm{D}\omega_l \tag{3.2.3}$$

where $\mathrm{D}\omega_l$ denotes uniform measure on the sphere S^l, and C_n is a universal constant.

[1] Practical numerical schemes for eigenvalue decomposition are unaffected by this algebraic obstruction, since they rely on iteration.

The set of attainable matrices Q are given by a mapping

$$h : S^{n-2} \times S^{n-3} \times \cdots \times S^1 \mapsto \mathsf{O}(n). \tag{3.2.4}$$

This mapping is given below, explicitly in terms of Householder reflections. As the dimension of the domain for this mapping is less than $\frac{1}{2}n(n-1)$, the dimension of $\mathsf{O}(n)$, not all matrices in $\mathsf{O}(n)$ are attainable.

Remark 23. The space $\mathsf{Tridiag}(n)$ clearly inherits the inner-product $\mathrm{Tr}(T^2) = \sum_{j=1}^{n} a_j^2 + 2\sum_{j=1}^{n-1} b_j^2$ from $\mathsf{Symm}(n)$. However, the volume form obtained from this metric is *not* the same as the volume form in (3.2.3) above.

Remark 24. (For algebraists!) The proof will also show that T and Q may be computed with a finite number of the following algebraic operations: addition, multiplication and square-roots.

Definition 25. Suppose $v \in \mathbb{R}^n$ is a unit vector. The *Householder reflection* in v is the matrix

$$P_v = I - 2vv^T. \tag{3.2.5}$$

Lemma 7. *The matrix P_v has the following properties:*

(a) $P_v^2 = I$.

(b) $P_v \in \mathsf{O}(n)$.

Proof. Decompose $\mathbb{R}^n$ into the orthogonal subspaces $\mathrm{span}\{v\}$ and $v^\perp$. Then $P_v v = -v$ and $P_v|_{v^\perp} = I$. Thus, $P_v^2 = I$. This proves (a). By construction $P_v^T = P_v$. Thus, by (a), we also have $P_v^T P_v = I$. □

Proof of Theorem 22. 1. The proof relies on a sequence of Householder reflections that progressively introduce zeros in a sequence of matrices similar to M. The first such matrix is the following. Let $w_1 = (M_{21}, \ldots, M_{n1})^T \in \mathbb{R}^{n-1}$ denote the last $n-1$ entries of the first column of M. If the first coordinate of w_1 is non-negative, and all other coordinates vanish there is nothing to do. If not, we may choose a Householder reflection (in $\mathbb{R}^{n-1}$) that maps w_1 to $|w_1|e_1^{(n-1)}$ (here the superscript $n-1$ denotes that we consider the basis vector $e_1 \in \mathbb{R}^{n-1}$). Geometrically, such a reflection is obtained by choosing v_1 to be the unit vector that lies in between w_1 and $|w_1|e_1^{(n-1)}$. Explicitly, we set[2]

$$\tilde{v}_1 = |w_1|e_1^{(n-1)} - w_1, \quad v_1 = \frac{\tilde{v}_1}{|\tilde{v}_1|}, \quad P^{(1)} = P_{v_1}. \tag{3.2.6}$$

By Lemma 7, $P^{(1)} \in \mathsf{O}(n-1)$ is a Householder reflection that maps w_1 to $|w_1|e_1^{(n-1)}$. It may be extended to a Householder reflection in $\mathsf{O}(n)$, by defining

$$Q^{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & P^{(1)} \end{pmatrix}. \tag{3.2.7}$$

[2]If one is using this method numerically and $|\tilde{v}_1|$ is small, instabilities can be introduced. In this case one should use $-|w_1|e_1^{(n-1)} - w_1$.

Then the matrix

$$M^{(1)} := Q^{(1)} M \left(Q^{(1)}\right)^T = Q^{(1)} M Q^{(1)}, \tag{3.2.8}$$

is similar to M. By construction, the first row of $M^{(1)}$ is $(M_{11}, |w_1|, 0, \ldots, 0)$, and the first column is $(M_{11}, |w_1|, 0, \ldots, 0)^T$. Thus, we may write

$$M^{(1)} = \begin{pmatrix} T^{(1)} & |w_1|(e_1^{(n-1)})^T \\ |w_1| e_1^{(n-1)} & N^{(1)} \end{pmatrix}, \tag{3.2.9}$$

where $T^{(1)}$ is a (trivial) 1×1 tridiagonal matrix and $N^{(1)} \in \mathrm{Symm}(n-1)$.

2. The proof is completed by induction. Assume that $M^{(k)} \in \mathsf{Symm}(n)$ has the form

$$M^{(k)} = \begin{pmatrix} T^{(k)} & |w_k|(e_1^{(n-k)})^T \\ |w_k| e_1^{(n-k)} & N^{(k)}) \end{pmatrix}, \tag{3.2.10}$$

where $T^{(k)} \in \mathsf{Tridiag}(k)$ and $N^{(k)} \in \mathrm{Symm}(n-k)$, $1 \le k \le n-1$. We apply the procedure of step 1 to $N^{(k)}$ to obtain a vector v_k, a Householder reflection $P^{(k)} = P_{v_k}$, and an orthogonal transformation of $M^{(k)}$,

$$Q^{(k)} = \begin{pmatrix} I_k & 0 \\ 0 & P^{(k)} \end{pmatrix} \in \mathsf{O}(n), \quad M^{(k+1)} = Q^{(k)} M^{(k)} Q^{(k)}. \tag{3.2.11}$$

Note that $Q^{(k)}$ leaves the first k rows and columns of $M^{(k)}$ unchanged, thus it does not destroy the tridiagonal structure of the first k rows and columns. Thus, $M^{(k+1)}$ has the form (3.2.10) with the index k replaced by $k+1$.

The procedure terminates when $k = n-2$, and yields

$$M = QTQ^T, \quad Q = Q^{(n-2)} Q^{(n-3)} \ldots Q^{(1)}. \tag{3.2.12}$$

3. It is simplest to prove (3.2.3) probabilistically. Informally, the k-th step of the procedure above is a change to polar coordinates in $\mathbb{R}^{n-k}$, with $b_k \ge 0$ playing the role of the radius, and the factor $b_k^{n-k-1} \mathrm{d}b_k \mathrm{D}\omega_{n-1-k}$ being the pushforward of Lebesgue measure in $\mathbb{R}^{n-k}$ to polar coordinates. More precisely, assume that $M \sim \mathrm{GOE}(n)$. We note that the first step of the above procedure leaves M_{11} alone. Thus, $a_1 = M_{11} \sim \mathcal{N}(0,1)$. Moreover, the term b_1 is the length of the first column of M, not including the diagonal term M_{11}. Since a χ_m^2 random variable has the same law as the length of a vector in $\mathbb{R}^m$ whose entries are iid $\mathcal{N}(0,1)$ random variables, we see that $b_1 \sim \chi_{n-1}$. Further, the vector $\omega_1 = w_1/|w_1|$ is uniformly distributed on S^{n-2} and independent of both a_1 and b_1 (see Exercise 3.1). We next observe that by the independence and invariance of the Gaussian ensembles, the matrix $N^{(1)}$ in (3.2.9) $\sim \mathrm{GOE}(n-1)$. Indeed, $\tilde{M}_1$, the lower-right $(n-1) \times (n-1)$ block of M, is a $\mathrm{GOE}(n-1)$ matrix, and the reflection $P^{(1)}$ is independent of $\tilde{M}_1$. Thus, $N^{(1)} = P^{(1)} \tilde{M}_1 P^{(1)}$ has law $\mathrm{GOE}(n-1)$ and is independent of b_1, a_1 and ω_1 (see Exercise 3.2). Thus, $a_2 \sim \mathcal{N}(0,1)$ and $b_2 \sim \chi_{n-2}$. An obvious induction now shows that if

$M \sim$ GOE then $T \sim$ Hermite(1), and the vectors $\omega_k = w_k/|w_k|$, are uniformly distributed on S^{n-1-k}, $1 \le k \le n-2$. Comparing the two laws, we find (with $\beta = 1$)

$$\mathrm{e}^{-\frac{\beta \operatorname{Tr}(M^2)}{2}} \mathrm{D}M = C_n \mathrm{e}^{-\frac{\beta \operatorname{Tr}(T^2)}{2}} \mathrm{d}a_j \prod_{k=1}^{n-1} b_k^{n-k-1} \mathrm{d}b_k \prod_{l=1}^{n-2} \mathrm{D}\omega_l \tag{3.2.13}$$

The exponential weights cancel, and yield the Jacobian formula (3.2.3). □

3.3 Tridiagonalization on Her(n) and Quart(n)

Theorem 22 admits a natural extension to Her(n) and Quart(n).

Theorem 26. *For every $M \in$ Her(n) (resp. Quart(n)) there exists a tridiagonal matrix $T \in$ Jac(n) and $Q \in$ U(n) (resp. USp(n)) such that*

$$M = QTQ^*. \tag{3.3.1}$$

The transformation (3.3.1) is given by a change of variables

$$\mathsf{Her}(n) \to \overline{\mathsf{Jac}(n)} \times \left(S_{\mathbb{F}}^{n-2} \times S_{\mathbb{F}}^{n-3} \times \dots S_{\mathbb{F}}^{1}\right), \tag{3.3.2}$$

where $S_{\mathbb{F}}^l$ denotes the unit sphere in $\mathbb{F}^l$, with $\mathbb{F} = \mathbb{C}$ (resp. $\mathbb{H}$). The volume form DM on Her(n) *(resp.* Quart(n)*) transforms as follows:*

$$\mathrm{D}M = C_n \prod_{j=1}^{n} \mathrm{d}a_j \prod_{k=1}^{n-1} b_k^{\beta(n-k)-1} \mathrm{d}b_k \prod_{l=1}^{n-2} \mathrm{D}\omega_l \tag{3.3.3}$$

where $\mathrm{D}\omega_l$ denotes uniform measure on the sphere $S_{\mathbb{F}}^l$, and C_n is a universal constant.

For a vector $w \in \mathbb{C}^n$ with independent standard normal complex entries, $w_j \sim \frac{1}{\sqrt{2}}(N_1 + \mathrm{i}N_2)$, where $N_1, N_2 \sim \mathcal{N}(0,1)$ are independent, $|w| \sim \frac{1}{\sqrt{2}}\chi_{2n}$. For a quarternion vector w, one finds $|w| \sim \frac{1}{2}\chi_{4n}$. So, β is introduced in this way.

Remark 27. Note that the matrix T is always real, whereas the entries of M and Q are in $\mathbb{C}$ or $\mathbb{H}$.

The proof of Theorem 26 is in the same vein as that of Theorem 22. It is only necessary to replace the Householder projections in O(n) with projections in U(n) and USp(n). For example, given $v \in \mathbb{C}^n$ with $|v| = 1$, the associated Householder projection in U(n) is $P_v = I - 2vv^*$. Step 3 in the proof of Theorem 26 also explains the role of the parameter β in the definition of the Hermite-β ensembles. The k-th step of the Householder transformation maps a standard Gaussian vector in $\mathbb{C}^{n-k}$ to its magnitude and direction. The law of the magnitude is now $\chi_{2(n-k)}$ (or $\chi_{\beta(n-k)}$ with $\beta = 2$). Similarly, the direction of the Gaussian vector is uniformly distributed on the unit sphere in $\mathbb{C}^{n-k-1}$.

3.4 Inverse spectral theory for Jacobi matrices

Bounded Jacobi operators admit a complete and beautiful spectral theory that is intimately tied to orthogonal polynomials and continued fractions. We first introduce this theory for finite Jacobi matrices, since it underlies Theorem 21. As usual, write

$$T = Q\Lambda Q^T, \quad Q \in \mathsf{O}(n), \tag{3.4.1}$$

for the diagonalization of T. We also recall the *Weyl chamber*

$$\mathcal{W}^n = \{\Lambda \in \mathbb{R}^n \,|\lambda_1 < \lambda_2 < \ldots < \lambda_n\}. \tag{3.4.2}$$

For each $\Lambda \in \mathcal{W}^n$, its *isospectral manifold* is the set

$$\mathcal{M}_\Lambda = \{T \in \overline{\mathsf{Jac}(n)} \,\big|\, T = Q\Lambda Q^T, \text{ for some } Q \in \mathsf{O}(n)\}. \tag{3.4.3}$$

The following theorem shows that the interior of the isospectral manifold is diffeomorphic to the positive orthant $S^{n-1}_+ = \{u \in \mathbb{R}^n \,|\, |u| = 1,\ u_j > 0,\ j = 1, 2, \ldots, n\}$ of the unit sphere. Given T, we uniquely define Q by forcing the first non-zero entry in each column to be positive.

Theorem 28. *The spectral mapping*

$$\mathcal{S} : \mathsf{Jac}(n) \to \mathcal{W}^n \times S^{n-1}_+, \quad T \mapsto (\Lambda, Q^T e_1), \tag{3.4.4}$$

is an analytic diffeomorphism.

We prove this in stages below. See Figure 3.4.1.

The isospectral manifold admits several distinct parametrizations. First, it is clear that we could use the simplex Σ_n instead of the orthant S^{n-1}_+. Indeed, let $u = Q^T e_1$ denote the first row of the matrix of eigenvectors and define $c_j = u_k^2$, $1 \le k \le n$. Since $Q \in \mathsf{O}(n)$, $\sum_{k=1}^n u_k^2 = 1$. Thus, $u \in S^{n-1}$ and $c \in \Sigma_n$. But, we shall use S^{n-1}_+. Lemma 8 below shows that u_k can be chosen to be strictly positive, which allows us to restrict attention to the positive orthant S^{n-1}_+

Theorem 28 may also be viewed as a mapping to the *spectral measure*

$$T \mapsto \mu = \sum_{j=1}^n u_j^2 \delta_{\lambda_j} = \sum_{j=1}^n c_j \delta_{\lambda_j}. \tag{3.4.5}$$

It is often more convenient to work with the Cauchy transform of the spectral measure, μ. Define the *τ-function*,

$$\mu \mapsto \tau(z) = \int_{\mathbb{R}} \frac{1}{x - z} \mu(\mathrm{d}x) = \sum_{j=1}^n \frac{u_j^2}{\lambda_j - z}, \quad z \in \mathbb{C}\backslash\{\lambda_1, \ldots, \lambda_n\}. \tag{3.4.6}$$

The inverse $\tau \mapsto \mu$ is obtained by computing the poles and residues of τ.

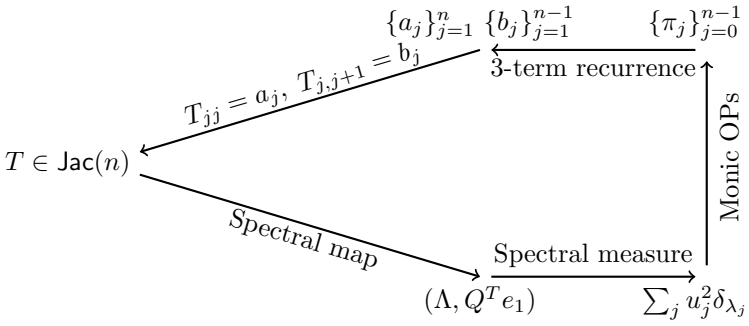

Figure 3.4.1: The construction of the spectral map and its inverse. The transformation to spectral variables is computed by computing eigenvalues and taking the first component of the (normalized) eigenvectors. Then a spectral measure (3.4.5) is created from this data and is used to define monic orthogonal polynomials (3.4.16). These polynomials satisfy a three-term recurrence relation (see Lemma 11) and the coefficients in the relation allow for the (unique) reconstruction of T, see (3.4.21). This shows the spectral map from $\mathsf{Jac}(n)$ to $\mathcal{W}^n \times S_+^{n-1}$ is invertible.

The τ-function may also be written as a ratio of polynomials of degree $n-1$ and n respectively. Let $T_k \in \mathsf{Jac}(k)$ denote the lower-right $k \times k$ submatrix of T, $1 \le k \le n$. It follows from Cramer's rule that

$$\tau(z) = e_1^T (T - z)^{-1} e_1 = \frac{\det(T_{n-1} - zI)}{\det(T - zI)} = \frac{\prod_{j=1}^{n-1}(\lambda_j^{(n-1)} - z)}{\prod_{j=1}^{n}(\lambda_j^{(n)} - z)}, \tag{3.4.7}$$

where $\Lambda^{(k)}$ denotes the diagonal matrix of eigenvalues of T_k and $\Lambda^{(n)} = \Lambda$. We will show that the ordered eigenvalues of T_{n-1} and T_n interlace, i.e.

$$\lambda_1^{(n)} < \lambda_1^{(n-1)} < \lambda_2^{(n)} < \ldots < \lambda_{n-1}^{(n-1)} < \lambda_n^{(n)}. \tag{3.4.8}$$

Thus, *interlacing sequences* provide another parametrization of $\mathsf{Jac}(n)$. A convenient visal description of interlacing sequences, called *diagrams*, was introduced by Kerov and Vershik [22]. The importance of these alternate parametrizations (spectral measures, τ-function, diagrams) is that they provide a transparent framework for the analysis of the limit $n \to \infty$.

The surprising aspect of Theorem 28 is that the spectral data (Λ, u) provides enough information to reconstruct the matrix T. There are three reconstruction procedures. The first involves orthogonal polynomials, the second uses the theory of continued fractions and a third involves the explicit solution of the equation $TQ = \Lambda Q$ for T. We explain the use of orthogonal polynomials below, and outline the theory of continued fractions in the exercises. In order to develop these procedures, it is first necessary to establish basic properties of the eigenvalues of Jacobi matrices.

Lemma 8. *Assume $T \in \mathsf{Jac}(n)$. Then*

1. *The first entry of each non-zero eigenvector is non-zero. In particular, we may normalize the eigenvectors to ensure $u_k > 0$ for $1 \leq k \leq n$.*
2. *The eigenvalues of T are distinct.*

Proof. We write the eigenvalue equation $Tv = zv$ in coordinates.

$$b_{k-1}v_{k-1} + (a_k - z)\, v_k + b_k v_{k+1} = 0, \quad 1 \leq k \leq n, \tag{3.4.9}$$

with the convention $b_0 = b_n = 0$. Since the off-diagonal terms b_k are strictly positive, we may solve this linear system recursively. Given v_1, we find

$$v_2 = \frac{v_1(z - a_1)}{b_1}, \quad v_3 = \frac{v_1}{b_1 b_2}\left((a_2 - z)(a_1 - z) - b_1^2\right), \text{ etc.} \tag{3.4.10}$$

Thus, $v \equiv 0 \in \mathbb{R}^n$ if $v_1 = 0$. Further, the solution space to the eigenvalue equation $Tv = \lambda v$ has dimension at most 1. □

Lemma 9. *The characteristic polynomials $d_k(z) = \det(zI - T_k)$ satisfy the recurrence relations*

$$d_{k+1}(z) = (z - a_{n-k})d_k(z) - b_{n-k}^2\, d_{k-1}(z), \quad 1 \leq k \leq n-1, \tag{3.4.11}$$

with the initial condition $d_0(z) \equiv 1$ and the convention $b_n = 0$.

Proof. Expand the determinant $\det(zI - T_k)$ about the k-th row, and compute the minors associated to $z - a_{n-k}$ and b_{n-k}. □

Lemma 10. *The eigenvalues of T_k and T_{k+1} interlace, $1 \leq k \leq n-1$.*

Proof. We consider the τ-functions for the minors T_k,

$$\tau_k(z) = \frac{\det(T_k - zI)}{\det(T_{k+1} - zI)} = -\frac{d_k(z)}{d_{k+1}(z)}. \tag{3.4.12}$$

By the recurrence relation (3.4.11), we have

$$-\frac{1}{\tau_k(z)} = z - a_{n-k} + b_{n-k}^2 \tau_{k-1}(z). \tag{3.4.13}$$

We claim that on the real line, $\tau_k(x)$ is strictly increasing between the zeros of d_k. Indeed, it is clear that $\tau_1(x) = (a_n - x)^{-1}$ has this property, and upon differentiating (3.4.13) we find that

$$\frac{1}{\tau_k^2}\tau_k' = 1 + b_{n-k}^2 \tau_{k-1}' > 0,$$

except at poles. The claim follows by induction.

Since τ_k is strictly increasing between poles, by the intermediate value theorem, it has exactly one zero between any two poles. By (3.4.12), the zeros of τ_k are the eigenvalues of T_k, and the poles of τ_k are the eigenvalues of T_{k+1}. Thus, they interlace. □

A remarkable feature of the spectral theory of Jacobi matrices is that the orthogonal polynomials associated to the spectral measure $\mu(T)$ may be used to reconstruct T. In order to state this assertion precisely, let us recall some basic facts about orthogonal polynomials. Assume given a probability measure μ on $\mathbb{R}$ that has finite-moments of all orders, i.e.,

$$\int_{\mathbb{R}} |x|^{\alpha}\, \mu(\mathrm{d}x) < \infty, \quad \alpha > 0. \tag{3.4.14}$$

We may apply the Gram-Schmidt procedure to the monomials $\{x^k\}_{k=0}^{\infty}$ to construct a sequence of polynomials that are orthogonal in $L^2(\mathbb{R}, \mu)$. There are two standard normalizations that one may adopt.

Orthonormal polynomials, denoted $\{p_k\}_{k=0}^{\infty}$, have the property that p_k is of degree k, $k = 0, 1, 2, \ldots$, and

$$\int_{\mathbb{R}} p_k(x) p_l(x)\, \mu(\mathrm{d}x) = \delta_{kl}. \tag{3.4.15}$$

Monic polynomials, denoted $\{\pi_k\}_{k=0}^{\infty}$ have the property that $\pi_k(x)$ is of degree k and the coefficient of x^k is 1. Further,

$$\int_{\mathbb{R}} \pi_k(x) \pi_l(x)\, \mu(\mathrm{d}x) = 0, \quad k \neq l. \tag{3.4.16}$$

Lemma 11 (Three-term recurrence for orthogonal polynomials)**.** *Given $(\Lambda, u) \in \mathcal{W}^n \times S_+^{n-1}$, let $\mu(\Lambda, u) = \sum_{k=1}^n u_k^2 \delta_{\Lambda_k}$. Then the associated monic orthogonal polynomials $\{\pi_k\}_{k=0}^n$, satisfy the three-term recurrence (3.4.17)*

$$\pi_k(z) = (z - a_k)\pi_{k-1}(z) - b_{k-1}^2\, \pi_{k-2}(z), \quad 1 \leq k \leq n, \tag{3.4.17}$$

where the coefficients a_k and b_k are given by

$$a_k = \frac{\int_{\mathbb{R}} x \pi_{k-1}^2\, \mu(\mathrm{d}x)}{\int_{\mathbb{R}} \pi_{k-1}^2(x)\, \mu(\mathrm{d}x)}, \quad b_k^2 = \frac{\int_{\mathbb{R}} x \pi_k(x) \pi_{k-1}(x)\, \mu(\mathrm{d}x)}{\int_{\mathbb{R}} \pi_{k-1}^2(x)\, \mu(\mathrm{d}x)}, \quad k = 1, \ldots, n, \tag{3.4.18}$$

with $\pi_{-1} = 0$ and hence $b_0 = 0$. Recall that $\pi_1 = 1$. The recurrence (3.4.18) defines a Jacobi matrix $T(\mu)$.

Remark 29. If μ is not a discrete measure of the form (3.4.5), but has bounded support, the recurrence (3.4.17) defines a bounded Jacobi operator on $l^2(\mathbb{C})$.

Proof. Given any μ as in (3.4.14), we obtain the sequence $\{\pi_k\}$ using the Gram-Schmidt procedure. When μ is of the form (3.4.5) with (3.4.5), the vector space $L^2(\mathbb{R}, \mu)$ has dimension n and the Gram-Schmidt procedure yields an orthogonal basis $\{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$ for $L^2(\mathbb{R}, \mu)$.

The three-term recurrence for the orthogonal polynomials is obtained as follows. Since $x\pi_k(x)$ is a polynomial of degree $k+1$ it can be expressed as a

linear combination $x\pi_k(x) = \sum_{j=0}^{k+1} c_{j,k}\pi_j(x)$. Since the π_j are monic, we must have $c_{k+1,k} = 1$. Moreover, for $j = 0, \dots, k-2$

$$\int_{\mathbb{R}} x\pi_k(x)\pi_j(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} \pi_k(x)x\pi_j(x)\,\mu(\mathrm{d}x) = 0,$$

since $x\pi_j$ lies in the span of $\{\pi_0, \dots, \pi_{k-1}\}$. Thus, $c_{j,k} = 0$ for $j = 0, \dots, k-2$ and we find

$$x\pi_k(x) = \pi_{k+1}(x) + c_{k,k}\pi_k(x) + c_{k-1,k}\pi_{k-1}(x). \tag{3.4.19}$$

It remains to show that $c_{k-1,k} > 0$. By orthogonality, $\int_{\mathbb{R}} x\pi_k(x)\pi_{k+1}(x)\,\mu(\mathrm{d}x) = \int_{\mathbb{R}} \pi_{k+1}^2(x)\,\mu(\mathrm{d}x)$. Thus, $c_{k,k-1} > 0$ for all k such that $\pi_{k-1}(x)$ does not vanish in $L^2(\mathbb{R}, \mu)$: Assume π_l does not vanish in $L^2(\mathbb{R}, \mu)$ for $l = 0, 1, 2, \dots, k-1 < n-1$. Then this recurrence defines π_k which is not the zero polynomial since it is monic. For $\Lambda \in \mathcal{W}^n$, it has distinct diagonal entries, so $p(x) \neq 0$ implies $\int p^2(x)\mu(\mathrm{d}x) > 0$ if p is a polynomial of degree less than n. This is (3.4.17) aside from a change in notation. □

Proof of Theorem 28. We have defined a forward map $T \mapsto \mu(T)$ as follows. The matrix T defines a τ-function $\tau(z) = e_1^T(T - zI)^{-1}e_1$, which is expressed as a ratio of characteristic polynomials in (3.4.7). The poles of $\tau(z)$ are the eigenvalues of T. The norming constrants are the residues at the poles, and are given by

$$u_k^2 = -\frac{d_{n-1}(\lambda_k)}{d_n'(\lambda_k)}, \quad 1 \le k \le n. \tag{3.4.20}$$

The inverse map $\mu \to T(\mu)$ is given by Lemma 11. The orthogonal polynomials defined by μ satisfy a three-term recurrence whose coefficients determine T.

We only need to show that the map $\mu \mapsto T(\mu) \mapsto \mu\left(T(\mu)\right)$ is the identity. Let $\mu \cong (\Lambda, u)$ be given and define $T(\mu)$ by the recurrence relations. We will show that

$$e_1^T(T - zI)^{-1}e_1 = \int_{\mathbb{R}} \frac{1}{x - z}\mu(\mathrm{d}x) = \sum_{k=1}^{n} \frac{u_k^2}{\lambda_k - z}. \tag{3.4.21}$$

We first show that the eigenvalues of T coincide with $\{\lambda_k\}$. Define $p_j(x) = \pi_j(x)\prod_{k=1}^{j} b_k^{-1}$, $\pi_0(x) = p_0(x)$, then

$$\begin{aligned} xp_0(x) &= a_1p_0(x) + b_1p_1(x), \\ xp_k(x) &= b_kp_{k-1}(x) + a_{k+1}p_k(x) + b_{k+2}p_{k+1}(x), \quad k > 0. \end{aligned}$$

Because $p_n(\lambda_j) = 0$ for all j, we conclude that

$$(p_0(\lambda_j), p_2(\lambda_j), \dots, p_{n-1}(\lambda_j))^T$$

is a non-trivial eigenvector for eigenvalue λ_j. We expand both sides of (3.4.21) for large z, to see that all we have to establish is the relation

$$e_1^T T^k e_1 = \int_{\mathbb{R}} x^k \mu(\mathrm{d}x), \quad 0 \le k \le n-1. \tag{3.4.22}$$

To see this, consider

$$\begin{aligned} Te_1 &= a_1 e_1 + b_1 e_2, \\ Te_k &= b_{k-1} e_{k-1} + a_k e_k + b_k e_{k+1}, \quad k > 1. \end{aligned}$$

Define new basis vectors $f_j = e_j \prod_{k=1}^{j-1} b_k$, $f_1 = e_1$ because $b_j > 0$ for all $j = 1, 2, \ldots, n-1$. We then have

$$\begin{aligned} Tf_1 &= a_1 f_1 + f_2, \\ Tf_k &= b_{k-1}^2 f_{k-1} + a_k f_k + f_{k+1}, \quad k > 1. \end{aligned}$$

We then diagonalize this, setting $T = Q\Lambda Q^T$, $\hat{f}_j = Q^T f_j$ to find

$$\begin{aligned} \Lambda \hat{f}_1 &= a_1 \hat{f}_1 + b_1^2 \hat{f}_1, \\ \Lambda \hat{f}_k &= b_{k-1}^2 \hat{f}_{k-1} + a_k \hat{f}_k + \hat{f}_{k+1}, \quad k > 1. \end{aligned}$$

Component-wise, this is the same three-term recurrence as the monic polynomials. So, taking into account $f_1 = e_1$, we find

$$\hat{f}_j = \pi_{j-1}(\Lambda) Q^T e_1, \quad f_j = \pi_{j-1}(T) e_1.$$

Then because $x^k = \sum_{j=0}^k c_{jk} \pi_j(x)$ we have $T^k e_1 = \sum_{j=0}^k c_{jk} \pi_j(T) e_1 = \sum_{j=0}^k c_{jk} e_{j+1}$ and

$$e_1^T T^k e_1 = c_{0k}.$$

Similarly,

$$\int_{\mathbb{R}} x^k \mu(\mathrm{d}x) = \sum_{j=0}^k c_{jk} \int_{\mathbb{R}} \pi_j(x) \mu(\mathrm{d}x) = c_{0k}.$$

This proves the theorem and this approach extends to the bi-infinite Jacobi operators [6]. □

Remark 30. Observe that the recurrence relation (3.4.17) may be rewritten as the matrix equation,

$$\begin{pmatrix} a_1 - z & 1 & 0 & \ldots & 0 \\ b_1^2 & a_2 - z & 1 & \ldots & 0 \\ 0 & b_2^2 & a_3 - z & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \ldots & b_{n-1}^2 & a_n - z \end{pmatrix} \begin{pmatrix} \pi_0(z) \\ \pi_1(z) \\ \vdots \\ \pi_{n-1}(z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -\pi_n(z) \end{pmatrix}. \tag{3.4.23}$$

Since $\pi_0(z) = 1$, each zero of $\pi_k(z)$ is an eigenvalue of the matrix above. Thus, $\pi_k(z) = \det(zI - \tilde{T}_k)$ where $\tilde{T}_k$ denotes the *upper-left* $k \times k$ submatrix of T (compare with T_k and $d_k(z) = \det(zI - T_k)$).

Thus, given μ, the entries of T are obtained from "top to bottom". However, given T, the τ-function is the limit of τ-functions $-d_k(z)/d_{k+1}(z)$ computed 'bottom-to-top'.

Remark 31. Consider the sequence of orthogonal polynomials

$$p_k(x) = \left(\prod_{j=1}^{k} b_k\right)^{-1} \pi_k(x), \quad k = 1, 2, \ldots, n-1. \tag{3.4.24}$$

This is actually an orthonormal sequence which satisfies the three-term recurrence

$$b_k p_k(x) = (z - a_k)p_{k-1}(z) - b_{k-1}p_{k-2}(x). \tag{3.4.25}$$

3.5 Jacobians for tridiagonal ensembles

We can now combine Theorem 28 with the definition of Hermite-β ensembles to state a refined version of Theorem 21.

Theorem 32. *For each $\beta > 0$, the law of the* Hermite(β) *ensembles in spectral variables $(\Lambda, u) \in \mathcal{W}^n \times S_+^{n-1}$ is given by*

$$p_{\text{Hermite}}(\Lambda, u)\mathrm{D}\Lambda\mathrm{D}u = \frac{1}{Z_{n,\beta}} \left(e^{-\frac{\beta}{4}\operatorname{Tr}(\Lambda^2)} |\triangle(\Lambda)|^\beta \mathrm{D}\Lambda\right) \left(\prod_{j=1}^{n} u_j^{\beta-1}\right) \mathrm{D}u, \tag{3.5.1}$$

where $\left(\prod_{j=1}^n u_j^{\beta-1}\right)$ $\mathrm{D}u$ refers to the joint density for n independent χ_β random variables, normalized so that the sum of their squares is one. In particular, Λ and u are independent.

Theorem 32 follows from a computation of the Jacobian of the spectral map $\mathcal{S} : \mathsf{Jac}(n) \to \mathcal{W}^n \times S_+^{n-1}$.

Theorem 33. *The volume forms on $\mathsf{Jac}(n)$ and $\mathcal{W}^n \times S_+^{n-1}$ are related by*

$$\mathrm{D}T = \prod_{j=1}^{n} \mathrm{d}a_j \prod_{k=1}^{n-1} b_k^{n-k-1}\, \mathrm{d}b_k = C_n \triangle(\Lambda)\mathrm{D}\Lambda \left(\prod_{k=1}^{n} u_j\right) \mathrm{D}u. \tag{3.5.2}$$

where C_n is a universal constant.

Remark 34. We have suppressed the explicit form of the universal constants in the statement of the lemma to focus on the marginals on $\mathcal{W}^n$ and S_+^{n-1} respectively. The computation of the constants is an interesting exercise (see [8]).

While Theorem 33 is an analytic/geometric assertion, the simplest proof uses probabilistic reasoning, as in step 3 of the proof of Theorem 22. Since we have computed the Jacobian for the diagonalizing map $\mathsf{Symm}(n) \to \mathbb{R}^n \times \mathsf{O}(n)$ (Weyl's formula) and the tridiagonalizing map $\mathsf{Symm}(n) \to \mathsf{Jac}(n)$ (Theorem 22), the ratio of these Jacobians may be used to compute the Jacobian of the spectral map $\mathsf{Jac}(n) \to \mathcal{W}^n \times S_+^{n-1}$. The main point is that by the $\mathsf{O}(n)$ invariance

of GOE, the top row of the eigenvector matrix must be uniformly distributed on S^{n-1} and is independent of Λ. This gives the term $\prod_{k=1}^{n} u_j \, \mathrm{d}u_j$ in equation (3.5.2). As Dumitriu and Edelman remark, this is a 'true random matrix theory' calculation. Another approach to (3.5.2) uses symplectic geometry.

Lemma 12 (Vandermonde determinant in (a, b) coordinates)**.**

$$\triangle(\Lambda) = \prod_{j<k} (\lambda_j - \lambda_k) = \frac{\prod_{k=1}^{n-1} b_k^{n-k}}{\prod_{j=1}^{n} u_j}. \tag{3.5.3}$$

Proof. 1. Recall that $\Lambda^{(l)}$ denotes the diagonal matrix of eigenvalues of T_l and that $d_l(x) = \prod_{j=1}^{l} (x - \lambda_j^{(l)})$. Therefore, we have the identity

$$\prod_{j=1}^{l} \prod_{k=1}^{l-1} \left| \lambda_j^{(l)} - \lambda_k^{(l-1)} \right| = \left| \prod_{j=1}^{l} d_{l-1} \left(\lambda_j^{(l)} \right) \right| = \left| \prod_{k=1}^{l-1} d_l \left(\lambda_k^{(l-1)} \right) \right|. \tag{3.5.4}$$

Since d_{l-1} and d_l are related through the three-term recurrence

$$d_l(x) = (x - a_l) d_{l-1}(x) - b_{n-l+1}^2 d_{l-2}(x),$$

we have

$$\left| \prod_{k=1}^{l-1} d_l(\lambda_k^{(l-1)}) \right| = b_{n-l+1}^{2(l-1)} \left| \prod_{k=1}^{l-1} d_{l-2}(\lambda_k^{(l-1)}) \right| = b_{n-l+1}^{2(l-1)} \left| \prod_{j=1}^{l-2} d_{l-1}(\lambda_j^{(l-2)}) \right|.$$

We apply this identity repeatedly, starting with $l = n$ to obtain

$$\begin{aligned} \left| \prod_{k=1}^{n-1} d_n(\lambda_k^{(n-1)}) \right| &= b_1^{2(n-1)} \left| \prod_{j=1}^{n-2} d_{n-1}(\lambda_j^{(n-2)}) \right| \\ &= b_1^{2(n-1)} b_2^{2(n-2)} \left| \prod_{k=1}^{n-3} d_{n-2}(\lambda_k^{(n-3)}) \right| = \cdots = \prod_{k=1}^{n-1} b_k^{2(n-k)}. \end{aligned}$$

2. The coefficients u_j^2 are the residue of $\tau_n(z)$ at the poles λ_j, i.e.

$$u_k^2 = \left| \frac{d_{n-1}(\lambda_k)}{d_n'(\lambda_k)} \right|, \quad 1 \le k \le n. \tag{3.5.5}$$

Observe also that

$$d_n'(\lambda_k) = \prod_{j \neq k} (\lambda_j - \lambda_k), \text{ and } \prod_{k=1}^{n} d_n'(\lambda_k) = \triangle(\Lambda)^2. \tag{3.5.6}$$

Therefore,

$$\prod_{j=1}^{n} u_j^2 = \frac{1}{\triangle(\lambda)^2} \prod_{k=1}^{n} |d_{n-1}(\lambda_k)| = \frac{\prod_{k=1}^{n-1} b_k^{2(n-k)}}{\triangle(\lambda)^2}. \tag{3.5.7}$$

□

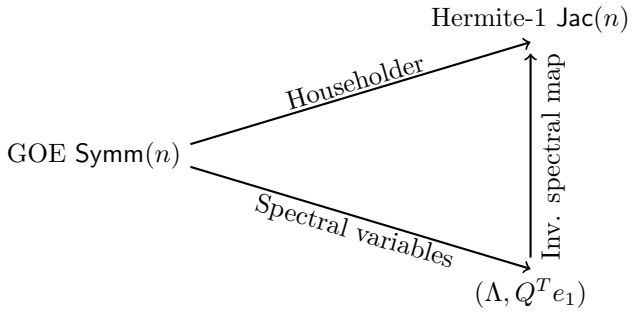

Figure 3.5.1: We have already computed the push-forward of GOE under Householder reflections (3.2.13) and the push-forward of GOE onto spectral variables via Weyl's formula (2.2.13). The composition of the map to spectral variables and the inverse spectral map must give us the reduction to tridiagonal form via Householder reflections. This allows the computation of the Jacobian of the inverse spectral map.

Proof of Theorem 33. 1. Our goal is to compute the Jacobian of the spectral mapping $\mathcal{S}$,

$$\mathrm{D}T = \frac{\partial(T(a,b))}{\partial(\Lambda,u)}\mathrm{D}\Lambda\mathrm{D}u, \tag{3.5.8}$$

where $\mathrm{D}u$ is uniform measure on $\{u \in \mathbb{R}^n \mid |u| = 1, u_j > 0 \text{ for all } j\}$. Rather than compute the change of variables directly, we will compute the push-forward of GOE onto $\mathsf{Jac}(n)$ and $\mathcal{W}^n \times S_+^{n-1}$ separately, and obtain the Jacobian above, see Figure 3.5.1.

2. Consider the push-forward of GOE under the map $M \mapsto (\Lambda, u)$, where $M = Q\Lambda Q^T$ is the diagonalization of M, with the normalization that the first non-zero entry in each column is positive. Since Λ and the matrix of eigenvalues Q are independent, Λ and $u = Q^T e_1$ are independent. Since Q is distributed according to Haar measure on $\mathsf{O}(n)$, the vector u is uniformly distributed on S_+^{n-1} and the push-forward of GOE is the measure

$$p(\Lambda,u)\mathrm{D}\Lambda\mathrm{D}u = c_n \mathrm{e}^{-\frac{1}{4}\operatorname{Tr}(\Lambda)^2} \triangle(\Lambda)\mathrm{D}\Lambda\mathrm{D}u. \tag{3.5.9}$$

3. Next consider the push-forward of GOE under the map $M \mapsto T$, where $M = QTQ^T$ denotes the tridiagonalization of M. As we have seen in the proof of Theorem 20, T and U are independent, and the marginal distribution of T is given by

$$\tilde{p}(T)\,\mathrm{D}T = C_n \mathrm{e}^{-\frac{1}{4}\operatorname{Tr}(T^2)} \prod_{j=1}^{n} \mathrm{d}a_j \prod_{k=1}^{n-1} b_k^{n-k-1}\,\mathrm{d}b_k. \tag{3.5.10}$$

4. Since $T \in \mathsf{Jac}(n)$ and $(\Lambda, u) \in \mathcal{W}^n \times S_+^{n-1}$ are in bijection, we have

$$p(\Lambda,u) = \tilde{p}(T(\Lambda,u))\frac{\partial(T(a,b))}{\partial(\Lambda,u)}. \tag{3.5.11}$$

We compare the expressions in (3.5.9) and (3.5.10) and use Lemma 12 to obtain

$$\frac{\partial(T(a,b))}{\partial(\Lambda,u)} = \frac{C_n}{c_n}\frac{\prod_{k=1}^{n-1} b_k}{\prod_{j=1}^{n} u_j}. \tag{3.5.12}$$

The constants are computed in [8]. □

Proof of Theorem 32. The law of We change variables using the spectral mapping and Theorem 33 to obtain the following identity for the law of the Hermite$-\beta$ ensembles

$$C_{n,\beta}\mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr}(T^2)}\prod_{k=1}^{n-1} b_k^{(\beta-1)(n-k)}\,\mathrm{D}T \tag{3.5.13}$$

$$= C_{n,\beta}\left(\mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr}(\Lambda^2)}\triangle(\Lambda)^\beta\,\mathrm{D}\Lambda\right)\left(\prod_{j=1}^{n} u_j^{\beta-1}\right)\mathrm{D}u. \tag{3.5.14}$$

Since the distribution factors, Λ and u are independent with the laws stated in Theorem 32. □

Exercises

3.1. Let $w \in \mathbb{R}^n$ have iid $\mathcal{N}(0,1)$ components. Show that $|w|$ and $w/|w|$ are independent.

3.2. Let $U \in \mathsf{O}(n)$ be a random orthogonal matrix. For example U could be a Householder reflection associated to a random vector w. Then assume $A \sim$ GOE. Show that $B := UAU^T \sim$ GOE and B is independent of U.

3.3. Write a numerical code to sample matrices from both GOE and the Hermite$-$1 ensemble. Verify numerically that a suitably normalized density of eigenvalues for the GOE matrix approaches the semicircle law as n increases ($n = 100$ should be ample). Is the same true for the Hermite $-$ 1 ensemble? Why or why not?

3.4. Consider the tridiagonal matrix $T \in \mathsf{Jac}(n)$ that has entries $a_j = 0$, $1 \le j \le n$, $b_k = 1$, $1 \le k \le n-1$.

(a) Compute explicitly the spectral measure using Chebyshev polynomials (compare T with the recurrence relations for the Chebyshev polynomials).

(b) Plot histograms of two distributions related to T for $n = 100$: (i) the empirical distribution of eigenvalues ($\frac{1}{n}\sum_{k=1}^{n}\delta_{\lambda_k}$); (ii) the spectral density $\sum_{k=1}^{n} u_k^2\delta_{\lambda_k}$. Can you identify the limit in (i)?

(This exercise will be relevant for an enumeration problem relating Brownian excursion to the Riemann-ζ function).

3.5. Establish uniqueness and smoothness in the proof of Theorem 28.

3.6. Use equation (3.4.12) to recursively expand τ_n as a continued fraction. Combine this with the uniqueness step in Q.2 to deduce an alternative approach to Theorem 28 that avoids the theory of orthogonal polynomials.

3.7. The following property of the function $-z^{-1}$ is relevant in the continued fraction scheme. Symmetric matrices have a partial order: Given A, $B \in \mathsf{Symm}(n)$ we say that $A \geq B$ if $u^T A u \geq u^T B u$ for every $u \in \mathbb{R}^n$. Suppose $A \geq B \geq 0$. Show that $-A^{-1} \geq -B^{-1}$.

3.8. This problem is a follow-up to exercise 5 in HW 1. Given a map f as in that exercise, compute an (explicit) expression for its derivative Df.

3.9. Compute the following normalization constants:

(a) The normalization constants $Z_{n,\beta}$ in the standard definitions of GOE, GUE and GSE with exponential weight $e^{-\frac{\beta}{4}\operatorname{Tr}(M^2)}$.

(b) The constant $C_{n,\beta}$ in (3.5.13).

(c) The constant C_n in the Jacobian for ensembles (3.2.3) (compare with your calculation of the volume of the unit sphere in HW1).

3.10. The proofs of Dumitriu and Edelman finesse the following issue: given $T \in \mathsf{Jac}(n)$ it requires some care to find a decomposition for the tangent space $T_T\mathsf{Jac}(n)$, especially the isospectral manifold, $\mathcal{M}_T$, that is analogous to Lemma 2. As in that lemma, we may split $T_T\mathsf{Jac}(n)$ into orthogonal subspaces that correspond to diagonal matrices $\dot{\Lambda}$ and $Q^T\dot{Q} \in \mathfrak{o}(n)$. However, while each $Q^T\dot{Q} \in \mathfrak{o}(n)$ generates a curve in $T_T\mathsf{Symm}(n)$, not all $Q^T\dot{Q}$ give rise to curves in $T_T\mathsf{Jac}(n)$. Verify this. Explore this issue further by trying to find a basis for the isospectral manifold $\mathcal{M}_T$ (see equation (3.4.3)).

3.6 Notes

To include in improved version.

1. Tridiagonal matrices as weights in enumerations problems.

2. Example: Chebyshev polynomials, Brownian excursion as a scaling limit of Dyck paths and relation with ζ-function.

Chapter 4

Determinantal formulas: From Vandermonde to Fredholm

Our purpose in this section is to present the elegant determinantal formulas of Dyson, Gaudin and Mehta for invariant matrix ensembles on $\mathsf{Her}(n)$. These formulas combine three distinct elements: (i) the Weyl formula on $\mathsf{Her}(n)$; (ii) the theory of orthogonal polynomials; (iii) Fredholm determinants. We first introduce these formulas for GUE. We then use the asymptotic properties of Hermite polynomials to establish their scaling limits (Theorem 2, Theorem 6 and Theorem 9). While the eigenvalues of GOE and GSE do not have a determinantal structure, they have a related Pfaffian structure, which is described in a later chapter.

4.1 Probabilities as determinants

In what follows we will adopt the following notation. In order to avoid confusion, we let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ denote the *unordered* eigenvalues of M, and $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathcal{W}^n$ denote the *ordered* eigenvalues of M. The probability density of x, denoted $P^{(n)}(x_1, \ldots, x_n)$, is obtained from the Weyl's formula

$$P^{(n)}(x_1, \ldots, x_n) = \frac{1}{Z_n} \triangle(x)^2 \mathrm{e}^{-\frac{1}{2}\sum_{k=1}^n x_k^2}. \tag{4.1.1}$$

Observe that $P^{(n)}$ is invariant under permutations $(x_1, \ldots, x_n) \mapsto (x_{\sigma_1}, \ldots, x_{\sigma_n})$, $\sigma \in \mathsf{S}(n)$. In practice, our interest lies not in the joint density of all n eigenvalues, but statistics such as the law of the largest eigenvalue. Thus, what is required is an analytical technique to extract such information from (4.1.1) by integrating out degrees of freedom to obtain information on the joint distribution of m-eigenvalues, $1 \leq m \leq n$. More precisely, given m and a Borel function

$f : \mathbb{R}^m \to \mathbb{R}$ we consider random variables of the type

$$N_f = \sum_{(j_1,\ldots,j_m)\in[\![1,n]\!]^m,\ j_k \text{distinct}} f(x_{j_1},\ldots,x_{j_m}). \tag{4.1.2}$$

Expectations of random variables of the form (4.1.2) are given by

$$\mathbb{E}(N_f) = \int_{\mathbb{R}^m} f(x_1,\ldots,x_m)\, R_m^{(n)}(x_1,\ldots,x_m)\,\mathrm{d}x_1\ldots\mathrm{d}x_m, \tag{4.1.3}$$

where R_m is the *m-point correlation function*

$$\begin{aligned} &R_m^{(n)}(x_1,\ldots,x_m) \\ &\quad = \frac{n!}{(n-m)!}\int_{\mathbb{R}^{n-m}} P^{(n)}(x_1,\ldots,x_m,x_{m+1},\ldots,x_n)\,\mathrm{d}x_{m+1}\ldots\mathrm{d}x_n. \end{aligned} \tag{4.1.4}$$

The combinatorial factor in (4.1.2) arises as follows. There are $\binom{n}{m}$ ways of picking subsets of m distinct indices from $[\![1,n]\!]$. On the other hand,

$$R_m^{(n)}(x_1,\ldots,x_m) = R_m^{(n)}(x_{\sigma_1},x_{\sigma_2},\ldots,x_{\sigma_m}), \quad \sigma\in\mathsf{S}(m). \tag{4.1.5}$$

and the integral on the right hand side of (4.1.5) appears $m!$ times when integrating over the complementary $n-m$ variables for each choice of indices $\{j_1,\ldots,j_m\}\in[\![1,n]\!]^m$. We state the following theorem which is proved in the following sections.

Theorem 35. *The joint density and m-point functions for* GUE(n) *are*

$$P^{(n)}(x_1,\ldots,x_n) = \frac{1}{n!}\det\left(K_n(x_j,x_k)_{1\le j,k\le n}\right), \tag{4.1.6}$$

$$R_m^{(n)}(x_1,\ldots,x_m) = \det\left(K_n(x_j,x_k)_{1\le j,k\le m}\right), \tag{4.1.7}$$

where the integral kernel K_n is defined by the Hermite wave functions

$$K_n(x,y) = \sum_{k=0}^{n-1}\psi_k(x)\psi_k(y). \tag{4.1.8}$$

Remark 36. The kernel K_n may be simplified using identities for the Hermite polynomials. The Christoffel-Darboux formula (B.2.6) allows us to write

$$K_n(x,y) = \sqrt{n}\frac{\psi_n(x)\psi_{n-1}(y)-\psi_n(y)\psi_{n-1}(x)}{x-y}. \tag{4.1.9}$$

Further, eliminating ψ_{n-1} with the identity (B.2.4) yields

$$K_n(x,y) = \frac{\psi_n(x)\psi_n'(y)-\psi_n(x)\psi_{n'}(y)}{x-y} - \frac{1}{2}\psi_n(x)\psi_n(y). \tag{4.1.10}$$

A particular consequence of Theorem 35 is the following fundamental formula. Assume S is a bounded Borel set, let $\mathbf{1}_S$ denote its indicator function, and let $A_m(S)$ denote the probability that the set S contains precisely m eigenvalues for $M \in \mathrm{GUE}(n)$.

Theorem 37. *The generating function of* $\{A_m(S)\}_{m=0}^{\infty}$ *is given by the formula*

$$\det\left(I - zK_n\mathbf{1}_S\right) = \sum_{m=0}^{\infty} A_m(S)(1-z)^m, \quad z \in \mathbb{C}, \tag{4.1.11}$$

where $\det\left(I - zK_n\mathbf{1}_S\right)$ *denotes the Fredholm determinant of the kernel*

$$K_n\mathbf{1}_S(x,y) = \sum_{k=0}^{n-1} \mathbf{1}_S(x)\psi_k(x)\psi_k(y)\mathbf{1}_S(y). \tag{4.1.12}$$

Theorem 35 and Theorem 37 illustrate the general spirit of determinantal formulas in random matrix theory. The density of a joint distribution is expressed as a determinant of an integral operator with finite rank. One may then use the theory of orthogonal polynomials, in particular, results on the asymptotics of orthogonal polynomials, to establish the basic limit theorems outlined in Chapter 1 (see Theorems 38 and Theorem 39 below).

Appendices B and C provide brief introductions to Hermite polynomials and Fredholm determinants respectively.

4.2 The m-point correlation function

Proof of Theorem 35. We form linear combinations of the rows of the Vandermonde matrix to obtain

$$\triangle(x) = \det\begin{pmatrix} \mathfrak{h}_0(x_1) & \mathfrak{h}_0(x_2) & \dots & \mathfrak{h}_0(x_n) \\ \mathfrak{h}_1(x_1) & \mathfrak{h}_1(x_2) & \dots & \mathfrak{h}_1(x_n) \\ \vdots & \vdots & & \vdots \\ \mathfrak{h}_{n-1}(x_1) & \mathfrak{h}_{n-1}(x_2) & \dots & \mathfrak{h}_{n-1}(x_n) \end{pmatrix}. \tag{4.2.1}$$

The calculations above would apply to any set of monic polynomials of degree $0, 1, 2, \dots, n-1$. The Hermite polynomials and wave functions are relevant because they satisfy the orthogonality relations

$$\int_{\mathbb{R}} \mathfrak{h}_j(x)\mathfrak{h}_k(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x = \delta_{jk}k!, \tag{4.2.2}$$

and allow the inclusion of an exponential weight. Precisely, the Hermite wavefunctions

$$\psi_k(x) = \frac{1}{\sqrt{k!}}\mathfrak{h}_k(x)\frac{\mathrm{e}^{-x^2/4}}{(2\pi)^{1/4}}, \tag{4.2.3}$$

satisfy the orthogonality relation

$$\int_{\mathbb{R}} \psi_j(x)\psi_k(x)\,\mathrm{d}x = \delta_{jk}, \tag{4.2.4}$$

and form a basis for $L^2(\mathbb{R})$. Let H denote the matrix with entries $H_{jk} = \psi_{j-1}(x_k)$. It follows from (4.2.1) and (4.2.3) that

$$\mathrm{e}^{-\frac{x^2}{2}} \triangle(x)^2 \propto \det H^2 = \det H^T H = \det\left[K_n(x_j, x_k)\right]_{1\le j,k\le n}, \tag{4.2.5}$$

using the identity

$$\left(H^T H\right)_{jk} = \sum_{l=1}^{n} H_{lj}H_{lk} = \sum_{l=0}^{n-1} \psi_l(x_j)\psi_l(x_k) = K_n(x_j, x_k). \tag{4.2.6}$$

Therefore, the joint density $P^{(n)}(x)$ is proportional to $\det K_n$. To determine the constant of proportionality we recall that the determinant of a matrix $A = [a_{jk}]_{1\le j,k\le n}$ satisfies

$$\det A = \sum_{\sigma\in\mathsf{S}(n)} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma_j j} \tag{4.2.7}$$

where $\mathrm{sgn}(\sigma)$ denotes the sign of the permutation σ. We then evaluate the integral

$$\begin{aligned}
\int_{\mathbb{R}^n} \det(H)^2 \mathrm{d}x_1 \ldots \mathrm{d}x_n &= \int_{\mathbb{R}^n} \left(\det\left[\psi_{j-1}(x_k)\right]_{1\le j,k\le n}\right)^2 \mathrm{d}x_1 \ldots \mathrm{d}x_n \\
&= \sum_{\sigma,\tau\in\mathsf{S}(n)} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \int_{\mathbb{R}^n} \prod_{j=1}^{n} \psi_{\sigma_j-1}(x_j)\psi_{\tau_j-1}(x_j)\,\mathrm{d}x_1 \ldots \mathrm{d}x_n \\
&= \sum_{\sigma,\tau\in\mathsf{S}(n)} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \prod_{j=1}^{n} \delta_{\sigma_j,\tau_j} = \sum_{\sigma,\tau\in\mathsf{S}(n)} \mathbf{1}_{\{\sigma=\tau\}} = n!\,.
\end{aligned} \tag{4.2.8}$$

We combine (4.2.8) and (4.2.6) to obtain the first assertion in Theorem 35:

$$P^{(n)}(x_1, \ldots, x_n) = \frac{1}{n!} \det\left[K_n(x_j, x_k)\right]_{1\le j,k\le n}.$$

The formulas for the correlation functions may be obtained by induction, beginning with

$$R^{(n)}(x_1, \ldots, x_n) = \det\left[K_n(x_j, x_k)\right]_{1\le j,k\le n}. \tag{4.2.9}$$

First, the orthonormality relations (4.2.4) imply

$$\int_{\mathbb{R}} K_n(x,x)\mathrm{d}x = n, \quad \int_{\mathbb{R}} K_n(x,z)K_n(z,y)\,\mathrm{d}z = K_n(x,y). \tag{4.2.10}$$

Assume (4.1.7) holds for an index $m+1 \leq n$. We then have

$$\begin{aligned} R_m^{(n)}(x_1,\dots,x_m) &= \frac{1}{n-m}\int_{\mathbb{R}} R_{m+1}^{(n)}(x_1,\dots,x_m,x_{m+1})\,\mathrm{d}x_{m+1} \\ &= \frac{1}{n-m}\int_{\mathbb{R}} \det\left[K_n(x_j,x_k)\right]_{1\leq j,k\leq m+1}\,\mathrm{d}x_{m+1} \\ = \frac{1}{n-m}\sum_{\sigma\in\mathsf{S}(m+1)} &\mathrm{sgn}(\sigma)\int_{\mathbb{R}} K_n(x_1,x_{\sigma_1})\dots K_n(x_{m+1},x_{\sigma_{m+1}})\,\mathrm{d}x_{m+1}. \end{aligned} \tag{4.2.11}$$

If $\sigma_{m+1} = m+1$ in this sum, then the first equality in (4.2.10) implies

$$\begin{aligned} &\int_{\mathbb{R}} K_n(x_1,x_{\sigma_1})\cdots K_n(x_{m+1},x_{\sigma_{m+1}})\,\mathrm{d}x_{m+1} \\ &\qquad = n\,K_n(x_1,x_{\sigma_1})\cdots K_n(x_m,x_{\sigma_m}). \end{aligned} \tag{4.2.12}$$

If $\sigma_{m+1} \neq m+1$, there exists $j \leq m$ and $k \leq m$ such that $\sigma_j = m+1$ and $\sigma_{m+1} = k$. We then use the second equality in (4.2.10) to find

$$\begin{aligned} &\int_{\mathbb{R}} K_n(x_1,x_{\sigma_1})\cdots K_n(x_{m+1},x_{\sigma_{m+1}})\,\mathrm{d}x_{m+1} \\ &\quad = \int_{\mathbb{R}} K_n(x_1,x_{\sigma_1})\cdots K_n(x_j,x_{m+1})\cdots K_n(x_{m+1},x_k)\,\mathrm{d}x_{m+1} \\ &\quad = K_n(x_1,x_{\sigma'_1})\cdots K_n(x_m,x_{\sigma'_m}). \end{aligned} \tag{4.2.13}$$

where σ' is a permutation of $\{1,\dots,m\}$ such that $\sigma'_j = k$ and $\sigma'_l = \sigma_l$ if $l \neq j$. Each permutation $\sigma' \in S_m$ may come from m permutations $\sigma \in S_{m+1}$. Further, $\mathrm{sgn}(\sigma') = -\mathrm{sgn}(\sigma)$ since these permutations differ by a single swap. Therefore, using equations (4.2.12) and (4.2.13) we have

$$\int_{\mathbb{R}} \det\left[K_n(x_j,x_k)\right]_{1\leq j,k\leq m+1}\,\mathrm{d}x_{m+1} = (n-m)\det\left[K_n(x_j,x_k)\right]_{1\leq j,k\leq m}.$$

□

4.3 Determinants as generating functions

Proof of Theorem 37. The Fredholm determinant $\det(I - zK_n\mathbf{1}_S)$ is an entire function of z. Thus, equation (4.1.11) is equivalent to the statement

$$A_m(S) = \frac{1}{m!}\left(-\frac{\mathrm{d}}{\mathrm{d}z}\right)^m \det(I - zK_n\mathbf{1}_S)\big|_{z=1}. \tag{4.3.1}$$

We first prove formula (4.3.1) in the case $m = 0$. The probability that all

eigenvalues lie outside S is given by

$$\int_{\mathbb{R}^n} \left(\prod_{j=1}^{n} (1 - \mathbf{1}_S(x_j)) \right) P^{(n)}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{4.3.2}$$
$$= \sum_{j=0}^{n} (-1)^j \int_{\mathbb{R}^n} \rho_j^n(\mathbf{1}_S(x_1), \ldots, \mathbf{1}_S(x_n)) P^{(n)}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n,$$

where $\rho_j^n(x_1, \ldots, x_n)$ is the j-th symmetric function in n variables. For example,

$$\rho_0^n(x) = 1, \quad \rho_1^n(x) = \sum_{j=1}^{n} x_j, \quad \rho_2^n(x) = \sum_{j<k}^{n} x_j x_k, \quad \rho_n^n(x) = \prod_{j=1}^{n} x_j.$$

Then, we can express

$$\rho_j^n(x) = \frac{1}{j!} \sum_{(j_1,\ldots,j_m) \in [\![1,n]\!]^j, \, j_k \text{distinct}} \prod_k x_{j_k}.$$

Using the m-point correlation function, we obtain using (4.1.3) with $f(x_1, \ldots, x_m) = \prod_{j=1}^{m} \mathbf{1}_S(x_j)$,

$$\mathbb{E}(N_f) = \int_{\mathbb{R}^n} \rho_j^n(\mathbf{1}_S(x_1), \ldots, \mathbf{1}_S(x_n)) P^{(n)}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \tag{4.3.3}$$
$$= \frac{1}{j!} \int_{\mathbb{R}^j} \det \left[K_n \mathbf{1}_S(x_k, x_l) \right]_{1 \le k,l \le j} \, \mathrm{d}x_1 \ldots \mathrm{d}x_j.$$

In the last equality, we have used (4.1.7) and multiplied the kernel on the left and right by the diagonal matrix $d_S = \mathrm{diag}(\mathbf{1}_S(x_1), \ldots, \mathbf{1}_S(x_j))$, so that

$$\mathbf{1}_S(x_1) \ldots \mathbf{1}_S(x_j) R_j^{(n)}(x_1, \ldots, x_j) = \mathbf{1}_S^2(x_1) \ldots \mathbf{1}_S^2(x_j) R_j^{(n)}(x_1, \ldots, x_j)$$
$$= \det \left(d_S \left[K_n(x_k, x_l) \right]_{1 \le k,l \le j} d_S \right) = \det \left[K_n \mathbf{1}_S(x_k, x_l) \right]_{1 \le k,l \le j},$$

where $K^{(n)} \mathbf{1}_S$ is defined in (4.1.12). We now combine (4.3.2) and (4.3.3) to obtain

$$\sum_{j=0}^{n} (-1)^j \int_{\mathbb{R}^n} \rho_j(\mathbf{1}_S(x_1), \ldots, \mathbf{1}_S(x_n)) P^{(n)}(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$
$$= \det(I - K_n \mathbf{1}_S), \tag{4.3.4}$$

using the infinite series (C.1.8) for the Fredholm determinant (only n terms are non-zero, since $K^{(n)}$ has rank n, see Exercise 4.2).

We now turn to the case $m \ge 1$. Equation (4.3.2) must now be modified to allow exactly m eigenvalues within S and $n - m$ eigenvalues outside S. Define

$$f(x_1, \ldots, x_n) = \prod_{j=1}^{m} \mathbf{1}_S(x_j) \prod_{j=m+1}^{n} (1 - \mathbf{1}_S(x_j)).$$

Then from (4.1.3), when we take into account the $m!$ permutations of the first m elements, and the $(n-m)!$ permutations of the last $n-m$ elements

$$\begin{aligned} A_m(S) &= \frac{1}{m!(n-m)!}\mathbb{E}\left(N_f\right) \\ &= \frac{1}{m!(n-m)!}\int_{\mathbb{R}^n} f(x_1,\ldots,x_n)R_n^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n. \end{aligned}$$

We then write

$$f(x_1,\ldots,x_n) = \prod_{j=1}^{m}\mathbf{1}_S(x_j)\sum_{k=0}^{m-n}(-1)^k\rho_k^{n-m}(\mathbf{1}_S(x_{m+1}),\ldots,\mathbf{1}_S(x_{m+k}))$$

We use the fact that $\rho_k^{n-m}(\mathbf{1}_S(x_{m+1}),\ldots,\mathbf{1}_S(x_n))$ is given by a sum of $\binom{n-m}{k}$ terms, each of which is product of k terms, and all terms integrate to the same value. So,

$$\begin{aligned} &\int_{\mathbb{R}^n}\prod_{j=1}^{m}\mathbf{1}_S(x_j)\rho_k^{n-m}(\mathbf{1}_S(x_{m+1}),\ldots,\mathbf{1}_S(x_{m+k}))R_n^{(n)}(x_1,\ldots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &\quad= \binom{n-m}{k}\int_{\mathbb{R}^{m+k}}\prod_{j=1}^{m+k}\mathbf{1}_S(x_j)\left(\int_{\mathbb{R}^{n-m-k}}R_n^{(n)}(x_1,\ldots,x_{m+k})\mathrm{d}x_{m+k+1}\cdots\mathrm{d}x_n\right) \\ &\quad\times\mathrm{d}x_1\cdots\mathrm{d}x_{m+k} \\ &\quad= \frac{(n-m)!}{k!}\int_{\mathbb{R}^{m+k}}\prod_{j=1}^{m+k}\mathbf{1}_S(x_j)R_{m+k}^{(k)}(x_1,\ldots,x_{m+k})\mathrm{d}x_1\cdots\mathrm{d}x_{m+k} \\ &\quad= \frac{(n-m)!}{k!}\int_{\mathbb{R}^{m+k}}\det\left(K_n\mathbf{1}_S(x_j,x_l)_{1\le j,l\le m+k}\right)\mathrm{d}x_1\cdots\mathrm{d}x_{m+k}. \end{aligned}$$

Then, it follows that

$$\begin{aligned} A_m(S) &= \frac{1}{m!}\sum_{k=0}^{n-m}\frac{(-1)^k}{k!}\int_{\mathbb{R}^{m+k}}\det\left(K_n\mathbf{1}_S(x_j,x_l)_{1\le j,l\le m+k}\right)\mathrm{d}x_1\cdots\mathrm{d}x_{m+k} \\ &= \frac{1}{m!}\left(-\frac{\mathrm{d}}{\mathrm{d}z}\right)^m\det(I-zK_n\mathbf{1}_S)|_{z=1}\,. \end{aligned}$$

□

4.4 Scaling limits of independent points

Recall the semicircle density p_{sc} from (1.2.1). We show in the next section that the global eigenvalue density, or density of states, for GUE(n) is given by p_{sc} as $n\to\infty$. Before we describe this more precisely, we consider a situation of iid points to contrast with the distributions that arise in GUE(n).

Consider an iid vector $\Lambda = \sqrt{n}(\lambda_1, \lambda_2, \ldots, \lambda_n)^T \in \mathbb{R}^n$ where $\mathbb{P}(\lambda_j \in S) = \int_S p_{\mathrm{sc}}(x')\mathrm{d}x'$. We form the empirical measure

$$L_n(\mathrm{d}x) = \frac{1}{n}\sum_{k=1}^{n} \delta_{\Lambda_k}(\mathrm{d}x), \tag{4.4.1}$$

and consider the deterministic measure $\mathbb{E}L_n$ defined by

$$\int f(x)\mathbb{E}L_n(\mathrm{d}x) := \mathbb{E}\langle L_n, f\rangle = \mathbb{E}\frac{1}{n}\sum_{k=1}^{n} f(\lambda_k), \quad f \in C_0(\mathbb{R}). \tag{4.4.2}$$

But, it is clear, and effectively by definition, that $\mathbb{E}L_n(\mathrm{d}x') = p(x')\mathrm{d}x' = \frac{1}{\sqrt{n}} p_{\mathrm{sc}}\left(\frac{x'}{\sqrt{n}}\right)\mathrm{d}x'$ and hence $\sqrt{n}p(\sqrt{n}x')\mathrm{d}x' = p_{\mathrm{sc}}(x')\mathrm{d}x'$.

Next, we consider a gap probability in the "bulk". Let $s \in (-2, 2)$, $I \subset \mathbb{R}$ be an interval and consider the rescaled interval $I_n = \sqrt{n}\left(s + \frac{I}{np_{\mathrm{sc}}(s)}\right)$. Then by independence

$$\mathbb{P}(\text{ no } \lambda_j \in I_n) = \left(1 - \frac{1}{\sqrt{n}}\int_{I_n} p_{\mathrm{sc}}\left(\frac{x'}{\sqrt{n}}\right)\mathrm{d}x'\right)^n. \tag{4.4.3}$$

We directly find that

$$\frac{1}{\sqrt{n}}\int_{I_n} p_{\mathrm{sc}}\left(\frac{x'}{\sqrt{n}}\right)\mathrm{d}x' = \frac{|I|}{n}(1 + o(1)) \text{ as } n \to \infty. \tag{4.4.4}$$

From this it follows that

$$\lim_{n\to\infty} \mathbb{P}(\text{ no } \lambda_j \in I_n) = \exp\left(-\int_I \mathrm{d}x'\right). \tag{4.4.5}$$

This is, of course, the gap probability for a Poisson process.

We now consider the distribution of the maximum, i.e. at the "edge". Let $\hat{\lambda} = \max_j \Lambda_j$. Then, by independence,

$$\mathbb{P}(n^{1/6}(2\sqrt{n} - \hat{\lambda}) > t) = \left(1 - \int_{2-n^{-2/3}t}^{2} p_{\mathrm{sc}}(x')\,\mathrm{d}x'\right)^n.$$

By direct calculation, replacing t with $\pi^{2/3}t^{2/3}(3/2)^{2/3}$ we find, for $t \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}(n^{1/6}(2\sqrt{n} - \hat{\lambda}) > t) = \mathrm{e}^{-\frac{2}{3\pi}t^{3/2}}. \tag{4.4.6}$$

From this we see a (trivial) scaling limit of the density of states after rescaling by $1/\sqrt{n}$, gaps on the order of $1/n$ after this rescaling and a largest "eigenvalue" that satisfies $\hat{\lambda} \sim 2\sqrt{n} + \xi n^{1/6}$ for an appropriate random variable ξ. All of these statements carry over to the random matrix setting, but the actual limits are very different for local statistics.

4.5 Scaling limits I: the semicircle law

The empirical measure of the eigenvalues of GUE(n) is

$$L_n\left(\mathrm{d}x\right) = \frac{1}{n}\sum_{k=1}^{n}\delta_{\lambda_k}(\mathrm{d}x) \tag{4.5.1}$$

has the expected density

$$\mathbb{E}L_n\left(\mathrm{d}x\right) = \frac{1}{n}K_n(x,x)\,\mathrm{d}x. \tag{4.5.2}$$

This density is also referred to as the global eigenvalue density or the density of states. The above expression is somewhat more transparent in its weak form, using unordered $x_1,\ldots,x_n$. For every $f \in C_0(\mathbb{R})$, we have

$$\mathbb{E}\langle L_n, f\rangle = \frac{1}{n}\int_{\mathbb{R}} f(x)R_1^{(n)}(x)\,\mathrm{d}x = \frac{1}{n}\int_{\mathbb{R}} f(x)K_n(x,x)\,\mathrm{d}x, \tag{4.5.3}$$

by Theorem 35 and equation (4.1.3). The value of the kernel K_n on the diagonal is determined by the Christoffel-Darboux relation (4.1.9) and L'Hospital's lemma:

$$K_n(x,x) = \sqrt{n}\left(\psi_n'(x)\psi_{n-1}(x) - \psi_n(x)\psi_{n-1}'(x)\right). \tag{4.5.4}$$

The scaling limit of $\mathbb{E}L_n$ is the semicircle law defined in (1.2.1).

Lemma 13.

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}K_n\left(x\sqrt{n}, x\sqrt{n}\right) = p_{\mathrm{sc}}(x), \quad x\in\mathbb{R}. \tag{4.5.5}$$

Further, for any $\varepsilon > 0$, the convergence is uniform on the set $\{x\,||x-2|\geq\varepsilon\}$.

Proof. The lemma follows from the Plancherel-Rotach asymptotics for the Hermite wave functions (see Cases 1 and 2 and equations (B.5.1)–(B.5.4)) in Appendix B). Define the rescaled wave functions

$$\Psi_{n+p}(x) = n^{\frac{1}{4}}\psi_{n+p}(x\sqrt{n}), \quad p=-2,-1,0. \tag{4.5.6}$$

We use the identity (B.2.4) to eliminate ψ_n' and ψ_{n-1}' from (4.5.4) and find after a few computations that

$$\frac{1}{\sqrt{n}}K_n\left(x\sqrt{n}, x\sqrt{n}\right) = \Psi_{n-1}^2(x) - \sqrt{\frac{n-1}{n}}\Psi_{n-2}(x)\Psi_n(x). \tag{4.5.7}$$

We now use the asymptotic relations (B.5.2) and (B.5.4) depending on whether $|x|<2$ or $|x|>2$. Since the region $|x|>2$ corresponds to exponential decay with a rate proportional to n, we focus on the region $|x|<2$. In order to simplify notation, let

$$\theta = n\left(\varphi - \frac{1}{2}\sin 2\varphi\right) - \frac{1}{2}\varphi - \frac{\pi}{4}. \tag{4.5.8}$$

(This is the argument of the cosine in (B.5.17) when $p = -1$.) Then (4.5.7) and (B.5.2) yield

$$\frac{1}{\sqrt{n}} K_n\left(x\sqrt{n}, x\sqrt{n}\right) \sim \frac{1}{\pi \sin\varphi}\left(\cos^2\theta - \cos(\theta+\varphi)\cos(\theta-\varphi)\right) = \frac{1}{2\pi}\sqrt{4-x^2},$$

using $x = 2\cos\varphi$ and the trigonometric formulae $\cos 2\alpha = 2\cos^2\alpha - 1$ and $2\cos(\theta+\varphi)\cos(\theta-\varphi) = \cos 2\varphi + \cos 2\theta$. A similar calculation with (B.5.4) shows that the limit vanishes outside the set $|x| > 2$. The assertion of uniformity in the convergence follows from the assertion of uniform convergence in the Plancherel-Rotach asymptotics. □

Using Exercise 4.9, Lemma 13 implies that $\mathbb{E}L_n(\mathrm{d}x)$, after rescaling, converges weakly

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k/\sqrt{n}}(\mathrm{d}x) \to p_{\mathrm{sc}}(x)\mathrm{d}x, \quad \text{weakly.} \tag{4.5.9}$$

It is also worth noting that if $f(x) = \mathbf{1}_S$ then

$$\mathbb{E}\,(\text{ fraction of eigenvalues that lie in } S) = \int f(x)\mathbb{E}L_n(\mathrm{d}x) = \frac{1}{n}\int_S K_n(x,x)\mathrm{d}x.$$

4.6 Scaling limits II: the sine kernel

Recall from Definition 5 that K_{sine} is the integral kernel on $\mathbb{R}\times\mathbb{R}$ given by

$$K_{\mathsf{sine}}(x,y) = \frac{\sin\pi(x-y)}{\pi(x-y)}, \quad x \neq y, \tag{4.6.1}$$

and $K_{\mathsf{sine}}(x,x,) = 1$. It defines an integral operator on $L^2(S)$ for every bounded, measurable set S. We can now prove a stronger version of Theorem 6.

Theorem 38. *For each integer $m = 0, 1, 2, \ldots$ and bounded, Borel set S and $r \in (-2, 2)$*

$$\lim_{n\to\infty} \mathbb{P}\left(M \in \mathrm{GUE}(n) \text{ has } m \text{ eigenvalues in } \sqrt{n}\left(r + \frac{S}{np_{\mathrm{sc}}(r)}\right)\right) = \frac{1}{m!}\left(-\frac{d}{dz}\right)^m \det\left(I - zK_{\mathsf{sine}}\mathbf{1}_S\right)\big|_{z=1}. \tag{4.6.2}$$

The proof of Theorem 38 is a consequence of the following

Lemma 14. *Let S be a bounded measurable set. Then for $r \in (-2, 2)$*

$$\lim_{n\to\infty}\sup_{x,y\in S}\left|\frac{1}{p_{\mathrm{sc}}(r)\sqrt{n}} K_n\left(\sqrt{n}r + \frac{x}{p_{\mathrm{sc}}(r)\sqrt{n}}, \sqrt{n}r + \frac{y}{p_{\mathrm{sc}}(r)\sqrt{n}}\right) - K_{\mathsf{sine}}(x,y)\right| = 0. \tag{4.6.3}$$

Proof. For $r \in (-2,2)$ define $\varphi(s)$ by $x = r + \frac{\pi s}{n \sin\varphi(0)} = 2\cos\varphi(s)$. We then note that $\sin\varphi(0)/\pi = p_{\rm sc}(r)$. We expand, for sufficiently large n,

$$\varphi(s) - \frac{1}{2}\sin 2\varphi(s) = \varphi(0) - \frac{1}{2}\sin 2\varphi(0) - \frac{\pi s}{n} + O(n^{-2}). \tag{4.6.4}$$

Define the new functions

$$\Psi_{n,p}(s) = n^{\frac{1}{4}}\psi_{n+p}\left(x\sqrt{n}\right), \tag{4.6.5}$$

From (B.5.2)

$$\Psi_{n,p}(s) \sim \frac{1}{\sqrt{\pi\sin\varphi(0)}}\cos\left[n\left(\varphi(0) - \frac{1}{2}\sin 2\varphi(0)\right) - \pi s + \left(p+\frac{1}{2}\right)\varphi(0) - \frac{\pi}{4}\right] \tag{4.6.6}$$

For fixed r, this is uniform for s in a compact set. We then use (4.1.9) and $y = r - \frac{\pi t}{n\sin\varphi(0)}$ to find, for $s \neq t$,

$$\begin{aligned}
&\frac{\pi}{\sin\varphi(0)\sqrt{n}} K_n(x\sqrt{n}, y\sqrt{n}) \qquad (4.6.7)\\
&= \frac{\pi}{\sin\varphi(0)\sqrt{n}} \frac{\psi_n(x\sqrt{n})\psi_{n-1}(y\sqrt{n}) - \psi_n(y\sqrt{n})\psi_{n-1}(x\sqrt{n})}{x-y}\\
&= \frac{\Psi_{n,0}(s)\Psi_{n,-1}(t) - \Psi_{n,0}(t)\Psi_{n,-1}(s)}{s-t}\\
&\sim \frac{1}{\pi\sin\varphi(0)} \frac{\cos(\theta_n+s)\cos(\theta_n+t-\varphi(0)) - \cos(\theta_n+t)\cos(\theta_n+s-\varphi(0))}{t-s}\\
&= \frac{\sin\pi(s-t)}{\pi(s-t)}. \qquad (4.6.8)
\end{aligned}$$

Here we set $\theta_n = n\left(\varphi(0) - \frac{1}{2}\sin 2\varphi(0)\right) + \frac{1}{2}\varphi(0) - \frac{\pi}{4}$ and used the identity

$$\cos\alpha\cos(\beta+\gamma) - \cos(\alpha+\gamma)\cos\beta = \sin\gamma\sin(\alpha-\beta). \tag{4.6.9}$$

This is uniform for $|t-s| \geq \delta$. For $|t-s| < \delta$, it is convenient to write

$$\frac{\psi_n(x)\psi_{n-1}(y) - \psi_n(y)\psi_{n-1}(x)}{x-y} = \begin{pmatrix}\psi_n(x) & \psi_{n-1}(x)\end{pmatrix}\int_0^1 \begin{pmatrix}-\psi_n'(\ell x + (1-\ell)y)\\ \psi_{n-1}'(\ell x + (1-\ell)y)\end{pmatrix} \mathrm{d}\ell,$$

and establish uniform convergence of this, after rescaling as above, to

$$\frac{\sin\pi(s-t)}{\pi(s-t)} = \begin{pmatrix}\sin\pi s & \cos\pi s\end{pmatrix}\int_0^1 \begin{pmatrix}\sin(\pi\ell s + \pi(1-\ell)t)\\ \cos(\pi\ell s + \pi(1-\ell)t)\end{pmatrix}\mathrm{d}\ell. \tag{4.6.10}$$

□

Proof of Theorem 38. Let $\tilde{K}_n(x,y)$ denote the rescaled kernel $\frac{1}{p_{\rm sc}(r)\sqrt{n}}K_n(x\sqrt{n}, y\sqrt{n})$, $x = r - \frac{s}{np_{\rm sc}(r)}$, $y = r - \frac{t}{np_{\rm sc}(r)}$. It follows from Lemma 14, using Sections C.2.1 and C.2 that

$$\lim_{n\to\infty} \det\left(I - z\tilde{K}_n\mathbf{1}_S\right) = \det\left(I - zK_{\mathsf{sine}}\mathbf{1}_S\right), \quad z \in \mathbb{C}, \tag{4.6.11}$$

and that the convergence is uniform in z for z in a bounded set. In particular, the derivatives at $z = 1$ converge for all m, that is

$$\lim_{n\to\infty} \left(-\frac{d}{dz}\right)^m \det\left(I - z\tilde{K}_n \mathbf{1}_S\right)\Big|_{z=1} = \left(-\frac{d}{dz}\right)^m \det\left(I - zK_{\mathsf{sine}}\mathbf{1}_S\right)\Big|_{z=1}. \tag{4.6.12}$$

By Theorem 37, this is equivalent to (4.6.2). □

4.7 Scaling limits III: the Airy kernel

Recall from Definition 8 that K_{Airy} is the continuous integral kernel on $\mathbb{R} \times \mathbb{R}$ given by

$$K_{\mathsf{Airy}}(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x - y}, \quad x \neq y. \tag{4.7.1}$$

The fluctuations at the edge of the spectrum are described as follows. Let $(x_1, \ldots, x_n)$ denote the unordered eigenvalues of a matrix $M \in \mathrm{GUE}(n)$ and let us consider the shifted and rescaled points

$$s_k = n^{\frac{1}{6}}\left(x - 2\sqrt{n}\right), \quad k = 1, \ldots, n. \tag{4.7.2}$$

For each nonnegative integer m and bounded, measurable set S, let $B_m^{(n)}(S)$ denote the probability that exactly m of the points $s_1, \ldots, s_n$ lie in S when $M \in \mathrm{GUE}(n)$. The following theorem is a consequence of Lemma 15 and the discussion in Section C.2.

Theorem 39.

$$\lim_{n\to\infty} B_m^{(n)}(S) = \frac{1}{m!}\left(-\frac{\mathrm{d}}{\mathrm{d}z}\right)^m \det\left(I - zK_{\mathsf{Airy}}\mathbf{1}_S\right)\Big|_{z=1}. \tag{4.7.3}$$

Remark 40. The assumption that S is bounded is necessary for K_{sine}. The sine-kernel has a (weak) rate of decay $|x|^{-1}$ as $|x| \to \infty$ and the Fredholm determinant $\det(I - zK_{\mathsf{sine}}\mathbf{1}_S)$ is not finite unless S is bounded. However, the Airy function, and the thus the Airy kernel, has strong decay as x and $y \to \infty$. The Fredholm determinant $\det(I - zK_{\mathsf{Airy}}\mathbf{1}_S)$ is well-defined in $L^2(S)$ for sets S that are bounded below, but not above, such as $S = (a, \infty)$ for any $a \in \mathbb{R}$. Such sets will be considered when we compute the Tracy-Widom distribution. See Exercise 5.

The proof of Theorem 39 follows from the Plancherel-Rotach asymptotics for the Hermite polynomials, in particular the Airy asymptotics in the transition zone (see Case 3 and (B.5.5)–(B.5.6) in Appendix B). The following lemma plays a role analogous to that of Lemma 14 in the proof of Theorem 38.

Lemma 15. *For $x \neq y$*

$$\lim_{n\to\infty} \left| \frac{1}{n^{\frac{1}{6}}} K_n\left(2\sqrt{n} + \frac{x}{n^{\frac{1}{6}}}, 2\sqrt{n} + \frac{y}{n^{\frac{1}{6}}}\right) - K_{\mathsf{Airy}}(x, y)\right| = 0 \tag{4.7.4}$$

and there exists a function $G(x,y) \in L^2([C,\infty)^2)$ *for all* $C \in \mathbb{R}$ *such that*

$$\left| \frac{1}{n^{\frac{1}{6}}} K_n \left(2\sqrt{n} + \frac{x}{n^{\frac{1}{6}}}, 2\sqrt{n} + \frac{y}{n^{\frac{1}{6}}} \right) \right| \leq G(x,y). \tag{4.7.5}$$

Proof. Convergence follows from (B.5.6). The function $G(x,y)$ can be constructed using (B.5.45) and (B.5.46), see Exercise 4.3. □

4.8 The eigenvalues and condition number of GUE

Let $M \sim \mathrm{GUE}(n)$. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of M. A consequence of Theorem 39 is the following, for all $t \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(n^{2/3} \left(\frac{\lambda_n}{\sqrt{n}} - 2 \right) < t \right) = \det(1 - K_{\mathsf{Airy}} \mathbf{1}_{(t,\infty)}) =: F_2(t),$$
$$\lim_{n\to\infty} \mathbb{P}\left(-n^{2/3} \left(2 + \frac{\lambda_1}{\sqrt{n}} \right) < t \right) = F_2(t).$$

Then, Theorem 38 gives for $t \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}|\lambda_j|}{\pi} > t \text{ for all } j \right) = \det(1 - K_{\mathsf{sine}} \mathbf{1}_{(-t,t)}) := S(t). \tag{4.8.1}$$

The singular values $\sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_n$ of a matrix M are the square roots of the non-zero eigenvalues of M^*M. One can rewrite (4.8.1) as

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}\sigma_1}{\pi} > t \right) = S(t). \tag{4.8.2}$$

The condition number is defined as $\kappa(M) := \sigma_n/\sigma_1$.

Lemma 16. *If* $M \sim \mathrm{GUE}(n)$*, then for all* $t > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\pi}{2n} \kappa(M) < t \right) = S(t^{-1}). \tag{4.8.3}$$

Proof. We first show that $\lambda_n/\sqrt{n} \to 2$, $\lambda_1/\sqrt{n} \to -2$ in probability. Fix $\epsilon > 0$, and let $L > 0$. Then

$$1 \leq \mathbb{P}\left(\left| \frac{\lambda_n}{\sqrt{n}} - 2 \right| \leq \epsilon \right) = \mathbb{P}\left(n^{2/3} \left| \frac{\lambda_n}{\sqrt{n}} - 2 \right| \leq n^{2/3}\epsilon \right) \geq \mathbb{P}\left(n^{2/3} \left| \frac{\lambda_n}{\sqrt{n}} - 2 \right| \leq L \right),$$

provided $n^{2/3}\epsilon \geq L$. So we, find

$$1 \leq \liminf_{n\to\infty} \mathbb{P}\left(\left| \frac{\lambda_n}{\sqrt{n}} - 2 \right| \leq \epsilon \right) \geq F_2(L) - F_2(-L).$$

Letting $L \to \infty$ gives convergence in probability for $\lambda_n/\sqrt{n}$. Similar arguments follow for $\lambda_1/\sqrt{n}$. Next, define

$$E_{\epsilon,n} = \left\{ \left| \frac{\lambda_n}{\sqrt{n}} - 2 \right| \le \epsilon, \left| \frac{\lambda_1}{\sqrt{n}} + 2 \right| \le \epsilon \right\}.$$

We know that $\mathbb{P}(E_{\epsilon,n}) \to 1$ as $n \to \infty$. Then

$$\mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t \right) = \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t, E_{\epsilon,n} \right) + \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t, E_{\epsilon,n}^c \right).$$

Because the second term must vanish as $n \to \infty$, we focus on the first term. On $E_{\epsilon,n}$ it follows that $(2-\epsilon)\sqrt{n} \le \sigma_n \le (2+\epsilon)\sqrt{n}$ and

$$\mathbb{P}\left(\frac{\pi(2+\epsilon)}{2n\sigma_1} < t, E_{\epsilon,n} \right) \le \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t, E_{\epsilon,n} \right) \le \mathbb{P}\left(\frac{\pi(2-\epsilon)}{2n\sigma_1} < t, E_{\epsilon,n} \right).$$

We find that for $\epsilon > 0$

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t \right) = \limsup_{n\to\infty} \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t, E_{\epsilon,n} \right) \le S\left(\frac{2-\epsilon}{2} t^{-1} \right),$$
$$\liminf_{n\to\infty} \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t \right) = \limsup_{n\to\infty} \mathbb{P}\left(\frac{\pi}{2n}\kappa(M) < t, E_{\epsilon,n} \right) \ge S\left(\frac{2+\epsilon}{2} t^{-1} \right).$$

If S is continuous at t, send $\epsilon \downarrow 0$ to obtain convergence in distribution. Since $S(t)$ is continuous, the result follows. □

4.9 Notes on universality and generalizations

4.9.1 Limit theorems for $\beta = 1, 4$

4.9.2 Universality theorems

Exercises

4.1. Prove the Christoffel-Darboux identity (B.1.7) for Hermite polynomials. (This is a standard relation and it is easy to find a proof in many texts, but try to do it on your own.)

4.2. Show that

$$\int_{\mathbb{R}^k} \det[K(x_p, x_q)]_{1\le p,q\le k} \, \mathrm{d}x_1 \cdots \mathrm{d}x_k = 0, \tag{4.9.1}$$

for $k > n$, if K is of the form

$$K(x,y) = \sum_{j=0}^{n-1} g_j(y) f_j(x), \quad f_j, g_j \in L^2(\mathbb{R}), \quad j = 0, 2, \ldots, n-1. \tag{4.9.2}$$

4.3. Finish the proof of Lemma 15 by constructing a function $G(x, y)$.

4.4. Establish (4.8.1).

4.5. Use the method of steepest descent to establish the asymptotic formula (A.3.1) for the Airy function. This is an easy application of the method of steepest descent.

4.6. In order to appreciate the power of the Plancherel-Rotach asymptotics, some numerical calculations will help.

(a) Develop a numerical scheme to compute all the roots of the n-th Hermite polynomial $\mathfrak{h}_n$. Plot the empirical distribution of roots for $n = 100$. Can you determine the limiting density of suitably rescaled roots?

(b) Numerically compute the Hermite wave functions for large n, say $n = 100$, and compare the rescaled wave function with the Plancherel-Rotach asymptotic formulas in all three regions (oscillatory, decaying and transition).

4.7. Use the method of steepest descent to establish the Plancherel-Rotach asymptotics in the region of exponential decay (equation (B.5.4)). This requires more care than Q.2.

4.8. Establish the following a priori bound on the Airy kernel. For any $a \in \mathbb{R}$,

$$\sup_{x,y} e^{x+y} |K_{\mathsf{Airy}}(x, y)| < \infty. \tag{4.9.3}$$

Let S be the semi-infinite interval (a, ∞). Use the above estimate to establish that the Fredholm determinant $\det(I - zK_{\mathsf{Airy}}\mathbf{1}_S)$ is an entire function.

4.9. Let $\rho_n(x)$, $n = 1, 2, \ldots$ be probability densities on $\mathbb{R}$ that converge almost uniformly to $\rho(x)$ with respect to Lebesgue measure on $\mathbb{R}$. Assume ρ has compact support. Show that

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\rho_n(x)\mathrm{d}x = \int_{\mathbb{R}} f(x)\rho(x)\mathrm{d}x$$

for every continuous function f with compact support.

4.10 Notes

To include in improved version.

1. Moment estimates to strengthen convergence to semicircle law.
2. Definition of determinantal processes.
3. Pair correlation function for the sine kernel.

Chapter 5

The equilibrium measure

In this section we establish properties of the equilibrium measure for general invariant ensembles. We also relate the equilibrium measure to the classical theory of orthogonal polynomials and Fekete points.

5.1 The log-gas

Let $V : \mathbb{R} \to \mathbb{R}$ denote a potential such that $V(x) \to \infty$ sufficiently rapidly as $|x| \to \infty$. The log-gas with size n and potential nV is a system of n identical charged particles constrained to the line interacting via pairwise Coulomb repulsion and the potential nV (we have scaled the potential V by n in order to ensure a scaling limit). The total energy of the system in any configuration $x \in \mathbb{R}^n$ is given by

$$E(x) = n \sum_{j=1}^{n} V(x_j) + \frac{1}{2} \sum_{j \neq k} \log \frac{1}{|x_j - x_k|}. \tag{5.1.1}$$

A fundamental postulate of equilibrium statistical mechanics is that the probability density of finding the system in a state x at inverse temperature $\beta > 0$ is

$$\frac{1}{Z_{n,V}(\beta)} e^{-\beta E(x)}, \tag{5.1.2}$$

where $Z_{n,V}$ is the partition function

$$Z_{n,V}(\beta) = \int_{\mathbb{R}^n} e^{-\beta E(x)} \, Dx. \tag{5.1.3}$$

The log-gas provides us with a physical caricature of eigenvalue repulsion. On one hand, we see that the energy $E(x)$ has two complementary terms: the logarithmic potential drives charges apart, but the potential V confines them in space. On the other hand, let V define an invariant probability measure of the form (1.1.3) on $\mathsf{Symm}(n)$, $\mathsf{Her}(n)$ or $\mathsf{Quart}(n)$. As a consequence of Weyl's

formula (Theorem 15), the equilibrium density (5.1.2) is precisely the joint law of the eigenvalues for these ensembles at $\beta = 1$, 2 and 4 respectively. It is in this sense that the 'eigenvalues repel'.

We have scaled the energy V with n in (5.1.1) in order to obtain a simple description of the scaling limit when $n \to \infty$. In order to study this limit, we view the energy function as a functional of the empirical measure, L_n, rather than a configuration $x \in \mathbb{R}^n$. For $(r, s) \in \mathbb{R}^2$ let

$$e(r,s) = \frac{1}{2}V(r) + \frac{1}{2}V(s) + \log\frac{1}{|r-s|}, \tag{5.1.4}$$

and given a probability measure μ on the line, define the functional

$$I[\mu] = \int_{\mathbb{R}}\int_{\mathbb{R}} e(r,s)\,\mu(dr)\,\mu(ds). \tag{5.1.5}$$

Observe that if L_n is the empirical measure associated to $x \in \mathbb{R}^n$, then

$$E(x) = n^2\left(\frac{1}{n}\sum_{j=1}^{n} V(x_j) + \frac{1}{n^2}\sum_{j\neq k}\log\frac{1}{|x_j - x_k|}\right) = n^2\tilde{I}[L_n], \tag{5.1.6}$$

and we may rewrite the partition function in the form

$$Z_{n,V}(\beta) = \int_{\mathbb{R}^n} e^{-n^2\beta\tilde{I}[L_n]}\,Dx. \tag{5.1.7}$$

Here $\tilde{I}[L_n]$ denotes the renormalized functional

$$\tilde{I}[\mu] = \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_{r\neq s} e(r,s)\,\mu(dr)\,\mu(ds), \tag{5.1.8}$$

that takes into account all interaction terms in $I[\mu]$, except the singular self-interaction term from $I[\mu]$. The logarithmic singularity in $e(r, s)$ is integrable if $\mu(ds)$ has an absolutely continuous density. Thus, if the particles in the log-gas spread out sufficiently as $n \to \infty$, we expect that μ has a smooth density, and

$$\lim_{n\to\infty}\frac{1}{n^2}\log Z_{n,V}(\beta) = \min_{\mu} I[\mu]. \tag{5.1.9}$$

In order to establish this relation, it is first necessary to obtain a precise analytical understanding of this minimization problem. We first prove such results under the formal assumption that there exists an $R > 0$ such that $V(x) = +\infty$ for $|x| > R$. This simply means that we first restrict attention to measures with support within the interval $[-R, R]$. Once the ideas are clear in this setting, we turn to measures with support on the line.

5.2 Energy minimization for the log-gas

5.2.1 Case 1: bounded support

Let $\mathcal{P}_R$ denote the set of probability measures on the interval $[-R, R]$. Recall that the natural topology on $\mathcal{P}_R$ is the weak topology (we adopt the probabilists convention for what is conventionally termed the weak-$*$ topology). A sequence of measures $\{\mu_k\}_{k=1}^{\infty} \in \mathcal{P}_R$ converges weakly to $\mu \in \mathcal{P}_R$ if

$$\lim_{n\to\infty} \langle \mu_n, f\rangle = \langle \mu, f\rangle, \tag{5.2.1}$$

for every function $f \in C(\mathbb{R})$. This topology is natural, because it yields compactness by Helly's theorem: Each sequence $\{\mu_k\}_{k=1}^{\infty} \in \mathcal{P}_R$ has a subsequence that converges weakly to a measure in $\mathcal{P}_R$.

Theorem 41. *Assume V is a continuous function on $[-R, R]$. There exists a unique probability measure $\mu_* \in \mathcal{P}_R$ such that*

$$I[\mu_*] = \min_{\mu\in\mathcal{P}_R} I[\mu]. \tag{5.2.2}$$

The proof of Theorem 41 is a demonstration of the classical method of the calculus of variations. There are two distinct ideas at work: existence follows from the fact that the functional $I[\mu]$ is weakly lower semicontinuous; uniqueness follows from the fact that $I[\mu]$ is a strictly convex function on $\mathcal{P}_R$.

Lemma 17. *Suppose the sequence $\{\mu_n\}_{n=1}^{\infty} \in \mathcal{P}_R$ converges weakly to $\mu \in \mathcal{P}_R$. Then*

$$I[\mu] \le \liminf_{n\to\infty} I[\mu_n]. \tag{5.2.3}$$

Lemma 18. *Let $\mu_0 \neq \mu_1$ be two measures in $\mathcal{P}_R$ and let $\mu_\theta = (1-\theta)\mu_0 + \theta\mu_1$ denote their convex combination for each $\theta \in (0, 1)$. Then*

$$I[\mu_\theta] < (1-\theta)I[\mu_0] + \theta I[\mu_1]. \tag{5.2.4}$$

Proof of Theorem 41. Existence. Since V is bounded, the function $e(x, y)$ is bounded below on $[-R, R]$. Therefore, $\inf_{\mu\in\mathcal{P}_R} I[\mu] > -\infty$. Further, since the logarithmic singularity is integrable, $I[\mu] < \infty$ for any measure that is absolutely continuous. Thus, we may assume that there is a sequence of measures $\{\mu_k\}_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} I[\mu_k] = \inf_{\mu\in\mathcal{P}_R} I[\mu] < infty. \tag{5.2.5}$$

Since $\mathcal{P}_R$ is compact in the weak topology, we may extract a convergent subsequence, also labeled $\{\mu_k\}_{k=1}^{\infty}$ for simplicity. Let μ_* denote the weak limit of this subsequence. We then use Lemma 17 to obtain the chain of inequalities

$$\inf_{\mu\in\mathcal{P}_R} I[\mu] \le I[\mu_*] \le \liminf_{k\to\infty} I[\mu_k] = \inf_{\mu\in\mathcal{P}_R} I[\mu]. \tag{5.2.6}$$

Thus, μ_* is a minimizer.

Uniqueness. Assume μ_* and ν_* are two distinct minimizers. We apply Lemma 18 to their convex combination with $\theta = 1/2$ to obtain the contradiction

$$\inf_{\mu\in\mathcal{P}_R} I[\mu] \leq I[\frac{1}{2}\mu_* + \frac{1}{2}\nu_*] < \frac{1}{2}\left(I[\mu_*] + I[\nu_*]\right) = \inf_{\mu\in\mathcal{P}_R} I[\mu]. \tag{5.2.7}$$

□

5.2.2 Weak lower semicontinuity

We now turn to the proof of Lemma 17. We first observe that for each monomial $r^j s^k$ in the variables r and s, the quadratic functional

$$\mu \mapsto \int_{-R}^{R}\int_{-R}^{R} r^j s^k \,\mu(dr)\,\mu(ds) = \left(\int_{-R}^{R} r^j \mu(dr)\right)\left(\int_{-R}^{R} s^k \mu(ds)\right)$$

is weakly continuous since it is the product of two bounded linear functionals on $\mathcal{P}_R$. Since each polynomial $p(r,s)$ in the variables (r,s) is a finite sum of monomials, the functional

$$\mu \mapsto \int_{-R}^{R}\int_{-R}^{R} p(r,s)\mu(dr)\,\mu(ds)$$

is also weakly continuous. Finally, since each continuous function $f \in C([-R,R]^2)$ may be uniformly approximated by polynomials, the quadratic functional

$$\mu \mapsto \int_{-R}^{R}\int_{-R}^{R} f(r,s)\mu(dr)\,\mu(ds)$$

is weakly continuous.

The function $e(s,t)$ defined in (5.1.4) is *not* continuous on $[-R,R]^2$ since the logarithmic term is unbounded on the diagonal $s=t$. However, for any $M>0$, the truncated function $e_M(r,s) = \min(e(r,s),M)$ is continuous. Thus, given a weakly convergent sequence of measures $\{\mu_k\}_{k=1}^{\infty}$ with limit $\mu \in \mathcal{P}_R$ we find

$$\begin{aligned}\int_{-R}^{R}\int_{-R}^{R} e_M(r,s)\mu(dr)\,\mu(ds) &= \lim_{k\to\infty}\int_{-R}^{R}\int_{-R}^{R} e_M(r,s)\mu_k(dr)\,\mu_k(ds)\\ &\leq \liminf_{k\to\infty}\int_{-R}^{R}\int_{-R}^{R} e(r,s)\mu_k(ds)\mu_k(ds) = \liminf_{k\to\infty} I[\mu_k].\end{aligned}$$

We let $M\to\infty$ on the left hand side and use the monotone convergence theorem to obtain (5.2.3).

5.2.3 Strict convexity

Lemma 18 is a particular consequence of a general fact in potential theory. The essential idea is to recognize that the function $z \mapsto -\log|z|$ is the fundamental

solution to Laplace's equation in $\mathbb{C} \cong \mathbb{R}^2$. More precisely, given a signed measure μ with a smooth density $\rho(z)$, supported in the ball $B_R \subset \mathbb{C}$ the unique solution to Poisson's equation with Dirichlet boundary condition

$$-\triangle\psi = \mu, \quad z \in \mathbb{C}\backslash\Omega, \quad \psi(z) = 0, \quad |z| = R, \tag{5.2.8}$$

is given by the integral formula

$$\psi(z) = \int_{B_R} G(z,w)\rho(w)\, Dw, \quad z \in B_R, \tag{5.2.9}$$

where Dw denotes the two-dimensional area element in $\mathbb{C}$ and $G(z,w)$ is the Green's function for Poisson's equation in the ball B_R with Dirichlet boundary conditions,

$$G(z,w) = \frac{1}{2\pi}\log\left(\frac{|w|}{R}\frac{|z-w^R|}{|z-w|}\right), \quad w^R = \frac{R^2 w}{|w|^2}, \quad z,w \in B_R. \tag{5.2.10}$$

The function $G(z,w)$ is obtained by the method of images: the image point w^R is the reflection of the point $w \in B_R$ in the circle ∂B_R [20, §4.1]. What matters here is that the dominant term in the Green's function is the logarithmic term $-\log|z-w|$, just as in equation (5.1.5), and the positivity of

$$\int_{B_R}\int_{B_R} G(z,w)\,\mu(dz)\,\mu(dw) = -\int_{B_R}\psi(w)\triangle\psi(w)\,ds = \int_{B_R}|\nabla\psi(w)|^2\,Dw > 0. \tag{5.2.11}$$

However, in contrast with (5.1.5) here we have assumed that $\mu(dw)$ has a smooth density $\rho(w)$, whereas the measures of interest in (5.1.5) are concentrated on an interval, and may have no regularity. Thus, some care is needed in formulating and proving a theorem on positivity analogous to (5.2.11).

Recall that a signed Borel measure μ on the line may be uniquely decomposed into two positive measures $\mu_\pm$ respectively such that $\mu = \mu_+ - \mu_-$. The Fourier transform of a measure is defined by

$$\hat{\mu}(u) = \int_{\mathbb{R}} e^{-ius}\,\mu(ds), \quad u \in \mathbb{R}. \tag{5.2.12}$$

The Fourier transform is a well-defined distribution. If $\mu_\pm$ are finite measures on $[-R,R]$, the Fourier transform is a continuous function of u that decays to zero as $|u| \to \infty$ by the Riemann-Lebesgue lemma.

Lemma 19. *Assume $\mu = \mu_+ - \mu_-$ is a signed measure on $[-R,R]$ such that*

$$\int_{-R}^{R}\mu_+(dr) = \int_{-R}^{R}\mu_-(dr) < \infty. \tag{5.2.13}$$

Then we have the identity

$$\begin{aligned}\int_{-R}^{R}\int_{-R}^{R}\log\frac{1}{|r-s|}\,(\mu_+(dr)\mu_+(ds) + \mu_-(dr)\mu_-(ds)) \qquad (5.2.14)\\ = \int_{-R}^{R}\int_{-R}^{R}\log\frac{1}{|r-s|}\,(\mu_+(dr)\mu_-(ds) + \mu_-(dr)\mu_+(ds)) + \int_0^\infty \frac{|\hat{\mu}(u)|^2}{u}\,du.\end{aligned}$$

In particular, $I[\mu] > 0$ if μ is non-zero and satisfies (5.2.13).

Remark 42. Equation (5.2.14) simply says that

$$\int_{-R}^{R}\int_{-R}^{R} \log\frac{1}{|r-s|}\,\mu(dr)\,\mu(ds) = \int_0^\infty \frac{|\hat{\mu}(u)|^2}{u}\,du. \tag{5.2.15}$$

for a signed measure μ with $\int_{-R}^{R}\mu(ds) = 0$. This identity has been written in the form (5.2.14) in order to ensure that there are no ill-defined terms of the form $\infty - \infty$. It is now clear from (5.1.4) and (5.1.5) that $I[\mu] > 0$ for such measures.

Proof. This proof is from [7, p.142]. We first regularize the logarithm at 0 and use the following integral representation. For any real s and $\varepsilon > 0$

$$\log(s^2+\varepsilon^2) = \log\varepsilon^2 + 2\,\mathsf{Im}\int_0^\infty e^{-\varepsilon u}\frac{e^{isu}-1}{iu}\,du. \tag{5.2.16}$$

We apply this integral representation to the following regularization of $I[\mu]$, and use the fact that $\int_{-R}^{R}\mu(dr) = 0$, to obtain

$$\begin{aligned}
&\int_{-R}^{R}\int_{-R}^{R} \log\left((r-s)^2+\varepsilon^2\right)\mu(dr)\mu(ds)\\
&\quad= 2\,\mathsf{Im}\int_0^\infty e^{-\varepsilon u}\int_{-R}^{R}\int_{-R}^{R}\frac{e^{i(r-s)u}-1}{iu}\,\mu(dr)\mu(ds)\,du\\
&\quad= 2\,\mathsf{Im}\int_0^\infty e^{-\varepsilon u}\frac{|\hat{\mu}(u)|^2}{iu}\,du = -2\int_0^\infty e^{-\varepsilon u}\frac{|\hat{\mu}(u)|^2}{u}\,du.
\end{aligned}$$

We may rewrite this identity in terms of $\mu_\pm$ as follows:

$$\begin{aligned}
&\int_{-R}^{R}\int_{-R}^{R} \log\frac{1}{\sqrt{(r-s)^2+\varepsilon^2}}\left(\mu_+(dr)\mu_+(ds)+\mu_-(dr)\mu_-(ds)\right) \qquad (5.2.17)\\
&\quad= \int_{-R}^{R}\int_{-R}^{R} \log\frac{1}{\sqrt{(r-s)^2+\varepsilon^2}}\left(\mu_+(dr)\mu_-(ds)+\mu_-(dr)\mu_+(ds)\right) + \int_0^\infty e^{-\varepsilon u}\frac{|\hat{\mu}(u)|^2}{u}\,du.
\end{aligned}$$

We now let $\varepsilon \downarrow 0$ and use the monotone convergence theorem to obtain (5.2.14) □

Finally, let us prove Lemma 18. Suppose μ_0 and μ_1 be two measures in $\mathcal{P}_R$ as in (5.2.4). The difference

$$(1-\theta)I[\mu_0]+\theta I[\mu_1]-I[\mu_\theta] = \theta(1-\theta)\int\int \log\frac{1}{|r-s|}\,(\mu_0-\mu_1)\,(dx)\;(\mu_0-\mu_1)\,(dx)$$

in the sense of signed measures. Thus, it is strictly positive when $\mu_0 \neq \mu_1$ by Lemma 19.

5.2.4 Case 2: Measures on the line

Having explained the main ideas behind Theorem 41 for finite measures, let us turn to the measures on the line. The proof of uniqueness requires no change, since it is easily verified that Lemma 19 holds for measures in $\mathcal{P}_{\mathbb{R}}$. However, it is necessary to modify the proof of existence to account for a possible loss of compactness: a sequence of measures in $\mathcal{P}_{\mathbb{R}}$ may drift off to infinity (e.g. $\mu_k = \delta_k$, $k \in \mathbb{Z}$). The appropriate condition required for compactness here is the following.

Definition 43. A sequence of measures $\{\mu_k\}_{k=1}^{\infty} \in \mathcal{P}_{\mathbb{R}}$ is *tight* if for every $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that

$$\sup_{k \geq 1} \mu_k \left(\mathbb{R} \backslash [-M_\varepsilon, M_\varepsilon] \right) < \varepsilon. \tag{5.2.18}$$

Compactness of measures in $\mathcal{P}_{\mathbb{R}}$ is provided by the Prokhorov-Varadarajan criterion: the sequence $\{\mu_k\}_{k=1}^{\infty} \in \mathcal{P}_{\mathbb{R}}$ has a subsequence that converges to a measure $\mu \in \mathcal{P}_{\mathbb{R}}$ if and only if the sequence $\{\mu_k\}_{k=1}^{\infty}$ is tight [33]. In practice, application of this criterion requires a uniform estimate on the tails of the measures $\{\mu_k\}_{k=1}^{\infty}$. Such a bound is possible only if the growth of the confining potential $V(x)$ as $|x| \to \infty$ is faster than the divergence of $\log |x|$ as $|x| \to \infty$. We formalize this requirement as follows. For any $\varepsilon > 0$, observe that

$$|r - s| = |r - 1 - (s - 1)| \leq \sqrt{r^2 + 1}\sqrt{s^2 + 1}. \tag{5.2.19}$$

Therefore, we have the lower bound

$$\log \frac{1}{r - s} \geq \frac{1}{2} \left(\log \frac{1}{r^2 + 1} + \log \frac{1}{s^2 + 1} \right). \tag{5.2.20}$$

Let us define the function

$$l(s) = \frac{1}{2} \log \frac{1}{s^2 + 1} + \frac{1}{2} V(s). \tag{5.2.21}$$

If $l(s)$ is bounded below, then by adding a constant to V if necessary, we can ensure that $l(s) \geq 0$ for all s. Clearly, this does not change the nature of the minimization problem.

Theorem 44. *Assume $V(s)$ is a continuous function such that $l(s)$ is bounded below and $l(s) \to \infty$ as $|s| \to \infty$.*

(a) There exists a unique probability measure $\mu_* \in \mathcal{P}_{\mathbb{R}}$ such that

$$I[\mu_*] \leq \min_{\mu \in \mathcal{P}_{\mathbb{R}}} I[\mu]. \tag{5.2.22}$$

(b) The support of the measure μ_* is contained within a finite interval.

Proof. (a) Since V is bounded below and the addition of a constant to V does not change the minimization problem, we may assume that $l(s) \geq 0$. Then

$$e(r,s) = \log \frac{1}{|r-s|} + \frac{1}{2}V(r) + \frac{1}{2}V(s) \geq l(r) + l(s) \geq 0, \tag{5.2.23}$$

and $c := \inf_{\mu \in \mathcal{P}_{\mathbb{R}}} I[\mu] \geq 0$. Suppose $\mu_k{}_{k=1}^{\infty}$ is an infimizing sequence: i.e. $\lim_{k\to\infty} I[\mu_k] = c$. Without loss of generality, we may assume that $I[\mu_k] \leq c+1$ for all k. Tightness of the sequence $\{\mu_k\}_{k=1}^{\infty}$ follows from the following (Chebyshev) inequality. For any $M > 0$,

$$\begin{aligned} c+1 \geq I[\mu_k] &= \int_{\mathbb{R}} \int_{\mathbb{R}} e(r,s)\mu_k(dr)\mu_k(ds) \\ &\geq 2\int_{\mathbb{R}} l(s)\mu_k(ds) \geq 2l_M \int_{|s|>M} \mu_k(ds) = 2l_M \mu_k(\mathbb{R}\backslash[-M,M]), \end{aligned} \tag{5.2.24}$$

where $l_M = \inf_{|s|\geq M} l(s)$. Since $\lim_{|s|\to\infty} l(s) = \infty$, $l_M \to \infty$ as $M \to \infty$. Thus, for any $\varepsilon > 0$, we may choose $M = M_\varepsilon$ large enough so that (5.2.18) holds. The rest of the proof of part (a) follows that of Theorem 41.

(b) For any $M > 0$, let S_M denote the set $(-\infty, M) \cup (M, \infty)$. We will show that $\mu_*(S_M) = 0$ if M is large enough. The proof relies on varying the measure μ_* by adding more mass proportional to μ_* in the set S_M. More precisely, let ν denote the restriction of μ_* to the set S_M, and for any $t \in (-1,1)$, define the measures

$$\mu_t = \frac{\mu_* + t\nu}{1 + t\nu(S_M)}. \tag{5.2.25}$$

We then find that $I[\mu_t]$ is a differentiable function of t, with

$$0 = \left.\frac{dI[\mu_t]}{dt}\right|_{t=0} = 2\int_{S_M} \nu(ds) \int_{\mathbb{R}} \mu_*(dr)e(r,s) - 2\nu(S_M)I[\mu_*]. \tag{5.2.26}$$

The estimate (5.2.23) and positivity of l yields the lower bound

$$\begin{aligned} &2\int_{S_M} \nu(ds) \int_{\mathbb{R}} \mu_*(dr)e(r,s) \\ &\quad \geq \int_{S_M} l(s)\nu(ds) + \int_{\mathbb{R}} l(r)\mu_*(dr) \geq \int_{S_M} l(s)\nu(ds) \geq l_M \nu(S_M). \end{aligned} \tag{5.2.27}$$

As in part (a), $l_M \to \infty$ as $M \to \infty$. Thus, for M sufficiently large, we have $l_M - I[\mu_*]) > 0$ and since ν is a positive measure, we have the (trivial) estimate

$$2(l_M - I[\mu_*])\nu(S_M) \geq 0. \tag{5.2.28}$$

On the other hand, the inequalities (5.2.26) and (5.2.27) yield the opposite inequality

$$2(l_M - I[\mu_*])\nu(S_M) \leq 0. \tag{5.2.29}$$

Thus, $\nu(S_M) = 0$ for all M such that $l_M > I[\mu_*]$. □

5.3 Fekete points

A second approach to the energy minimization problem relies on a study of the minimizers of the function $E(x)$ defined in (5.1.1) for $x \in \mathbb{R}^n$, and a potential V that satisfies the assumptions of Theorem 44. For any such potential, $0 \leq E(x) < \infty$ for any $x \in \mathbb{R}^n$ such that $x_j \neq x_k$, $j \neq k$. Thus, for each n, there exists a set of points $F_n \subset \mathbb{R}^n$, such that

$$E(x_*) = \min_{x \in \mathbb{R}^n} E(x), \quad x_* \in F_n. \tag{5.3.1}$$

The set F_n is called the set of n-Fekete points. The Fekete points are naturally connected to the minimization problem for the functional $I[\mu]$ through the modified functional $H[L_n]$, where $L_n(x)$ is the empirical measure associated to a point $x \in \mathbb{R}^n$. Let δ_n denote the rescaled energy of Fekete points

$$\delta_n = \frac{1}{n(n-1)} E(x^{(n)}). \tag{5.3.2}$$

The main result is then the following

Theorem 45. *Assume V satisfies the assumptions of Theorem 44. Let $\{x^{(n)}\}_{n=1}^{\infty}$ be a sequence of points $x^{(n)} \in F_n$ and Then*

(a) The rescaled energy of Fekete points increases monotonically to $I[\mu_*]$.

$$0 \leq \delta_n \leq \delta_{n+1} \leq I[\mu_*]. \tag{5.3.3}$$

(b) The empirical measures $L(x^{(n)})$ converge weakly to μ_*.

Proof of (a). We first prove the estimates (5.3.3). The uniform upper bound on $E(x^{(n)})$ is obtained as follows. Fix a positive integer n and a point $x^{(n)} \in F_n$. By definition, for any $s = (s_1, \ldots, s_n) \in \mathbb{R}^n$,

$$E(x^{(n)}) \leq E(s) = \frac{1}{2} \sum_{j,k=1}^{n} (V(s_j) + V(s_k)) + \sum_{j \neq k=1}^{n} \log \frac{1}{|s_j - s_k|}. \tag{5.3.4}$$

Let $\mu(ds)$ be any probability measure on the line. We integrate (5.3.4) with respect to the n-fold tensorized probability measure $\mu \otimes \mu \cdots \otimes \mu$ on $\mathbb{R}^n$ to obtain

$$\begin{aligned} &E(x^{(n)}) \qquad\qquad (5.3.5)\\ &\leq \int_{\mathbb{R}^n} \left[\frac{1}{2} \sum_{j,k=1}^{n} (V(s_j) + V(s_k)) + \sum_{j \neq k=1}^{n} \log \frac{1}{|s_j - s_k|} \right] \mu(ds_1)\mu(ds_2) \cdots \mu(ds_n) \\ &= n(n-1) \int_{\mathbb{R}} \int_{R} e(r,s)\mu(ds)\mu(dr) = I[\mu], \end{aligned}$$

since for each value of the indices j and k only the integrals over $\mu(ds_j)$ and $\mu(ds_k)$ give contributions that are not unity and there are $n(n-1)$ possible unordered pairings of j and k. In particular, $E(x^{(n)}) \leq n(n-1)I[\mu_*]$.

The monotonicity of δ_n follows from the following argument. Suppose $x^{(n+1)} = (x_1, \ldots, x_{n+1})$ is point in the Fekete set F_{n+1}. We fix an index m, $1 \le m \le n+1$ and use the definition of E in (5.1.1) to obtain

$$e^{-\frac{1}{n(n+1)}E(x^{(n+1)})} = \left(\prod_{1\le j\ne k\le n+1} |x_j - x_k|\, e^{-\frac{V(x_j)}{2}} e^{-\frac{V(x_k)}{2}} \right)^{\frac{1}{n(n+1)}} \tag{5.3.6}$$

$$= \left(\prod_{j\ne m} |x_j - x_m|\, e^{-\frac{V(x_j)}{2} - \frac{V(x_m)}{2}} \right)^{\frac{2}{n(n+1)}} \left(\prod_{j,k\ne m} |x_j - x_k|\, e^{-\frac{V(x_j)}{2}} e^{-\frac{V(x_k)}{2}} \right)^{\frac{1}{n(n+1)}}$$

$$\le \left(\prod_{j\ne m} |x_j - x_m|\, e^{-\frac{V(x_j)}{2}} e^{-\frac{V(x_m)}{2}} \right)^{\frac{2}{n(n+1)}} e^{-\delta_n \frac{n-1}{n+1}}$$

since the second term is the energy $E(\hat{x})$ of the point $\hat{x} \in \mathbb{R}^n$ obtained from $x^{(n)}$ by projecting out the coordinate x_m.

Since m is arbitrary, we take the product over $1 \le m \le n+1$ to obtain

$$e^{-\frac{1}{n}E(x^{(n+1)})} \le e^{-(n-1)\delta_n} \left(\prod_{1\le m\le n+1} \prod_{1\le j\le n+1, j\ne m} |x_j - x_m|\, e^{-\frac{V(x_j)}{2} - \frac{V(x_m)}{2}} \right)^{\frac{2}{n(n+1)}}$$

$$= e^{-(n-1)\delta_n} e^{-\frac{2}{n(n+1)}E(x^{(n+1)})}. \tag{5.3.7}$$

This inequality simplifies to $\delta_n \le \delta_{n+1}$. □

Proof of (b). While the self-energy of all the Fekete points is infinite, inequality (5.3.3) shows that a suitably renormalized energy is finite, and bounded above by $I[\mu_*]$. This inequality, in combination with an easy modification of the Chebyshev inequality (5.2.24) also shows that the empirical measures $L(x^{(n)})$ are tight. Thus, there exists a convergent subsequence and a limiting probability measure $\nu \in \mathcal{P}_{\mathbb{R}}$ such that the empirical measures $L^{(n)}$ defined by the Fekete points $x^{(n)}$ converge weakly to ν as $n \to \infty$.

For any $M > 0$, we introduce the cut-off energy $e_M(r,s) = \min(M, e(r,s))$ and observe that

$$\delta_n = \frac{1}{n(n-1)}E(x^{(n)}) = \frac{n^2}{n(n-1)} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{r\ne s} e(r,s) L^{(n)}(dr) L^{(n)}(ds)$$

$$\ge \frac{n^2}{n(n-1)} \int_{\mathbb{R}} \int_{\mathbb{R}} e_M(r,s) L^{(n)}(dr) L^{(n)}(ds) - \frac{M}{n-1}.$$

Since the function $e_M(r,s)$ is continuous and $0 \le e_M(r,s) \le M$, we may interchange limits as $n \to \infty$, and use Theorem 45(a) to obtain

$$I[\mu_*] \ge \liminf_{n\to\infty} \delta_n \ge \int_{\mathbb{R}} \int_{\mathbb{R}} e_M(r,s) \nu(dr) \nu(ds). \tag{5.3.8}$$

We now let $M \to \infty$ and use the monotone convergence theorem and the fact that μ_* is a minimizer to obtain

$$I[\mu_*] \geq I[\mu] \geq I[\mu_*]. \tag{5.3.9}$$

Since μ_* is unique, it follows that $\mu_* = \nu$.

This argument proves that every subsequential limit of $L^{(n)}$ is μ_*. Thus, the entire sequence converges to μ_*. □

5.4 Exercises

The first three questions are related. The goal is to formulate and analyze the equation for the equilibrium measure μ_* associated to the potential $V(x)$. In order to simplify your calculations, assume that μ_* has a continuous density ψ, in all the problems below. The last two questions discuss enumeration problems related to the Catalan numbers.

1. Basics of the Hilbert transform. Let $G(z)$ denote the Stieltjes transform

$$G(z) = \int_{-\infty}^{\infty} \frac{1}{s-z} \mu_*(ds) = \int_{-\infty}^{\infty} \frac{1}{s-z} \psi(s)(ds), \quad z \in \mathbb{C} \backslash \mathrm{supp}(\mu_*). \tag{5.4.1}$$

The Hilbert transform of ψ is the limit of the Stieltjes transform as $z \to x \in \mathbb{R}$. The Hilbert transform also differs from the Stieltjes transform by the inclusion of a factor of π (since this makes the Fourier transform of the operator $\mathcal{H}$ particularly simple). That is, given μ_* as above, we set

$$\mathcal{H}\psi(x) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{\psi(s)}{x-s}\, ds := \lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{x-s}{(x-s)^2+\varepsilon^2}\, \psi(s)\, ds. \tag{5.4.2}$$

(a) Show that $\mathcal{H}\psi$ is a bounded function when $\psi(x)$ is continuous.

(b) Show that μ_* may be recovered from G by evaluating the jump in the imaginary part of G across the support of μ_*:

$$\lim_{\varepsilon \to 0} \frac{1}{2\pi i} \left(G(x+i\varepsilon) - G(x-i\varepsilon) \right) = \psi(x). \tag{5.4.3}$$

(c) Compute the Hilbert transform of the following functions to obtain a feel for it (answers are on wikipedia):

$$e^{ix}, \quad \delta_0(x), \quad \mathbf{1}_{[a,b]}(x).$$

2. Integral equation for ψ. Assume V is differentiable and satisfies the assumptions of Theorem 44 so that μ_* has compact support. Show that if μ_* has a density ψ as above, then it satisfies the integral equation

$$\mathcal{H}\psi(x) = \frac{1}{2\pi} V'(x) \quad \text{on} \quad \mathrm{supp}(\mu_*). \tag{5.4.4}$$

3. Fixed point equation for the resolvent. One solution to (5.4.4) uses the Stieltjes transform $G(z)$. Assume that $V(x)$ is a polynomial of degree $d \geq 2$.

(a) Show that G satisfies the quadratic equation

$$G^2(z) + V'(z)G(z) + P(z) = 0, \tag{5.4.5}$$

where $P(z)$ is a polynomial of degree $d-2$ whose coefficients are determined by the moments of μ_* of degree lower than d. The solution branch is determined by the requirement that $G(z) \sim -1/z$ as $z \to \infty$ which is immediate from (5.4.1).

(b) Equation (5.4.5) may be solved by making further assumptions on the form of μ_*. In particular, assume that $V(z)$ is even, that the support of μ_* is a single interval $[-2a, 2a]$, and show that (5.4.5) simplifies to

$$G(z) = Q(z)\sqrt{z^2 - 4a^2} - \frac{1}{2}V'(z) \tag{5.4.6}$$

where $Q(z)$ is a polynomial of degree $d-2$ whose coefficients are determined by the condition that $G(z) \sim -1/z$ as $z \to \infty$.

(c) Apply these ideas to compute the equilibrium measure for the quartic potential

$$V(x) = \frac{1}{2}x^2 + \frac{g}{4}x^4. \tag{5.4.7}$$

Show that

$$G(z) = \left(\frac{1}{2} + \frac{g}{2}x^2 + ga^2\right)\sqrt{x^2 - 4a^2} - \frac{1}{2}\left(x + gx^3\right), \tag{5.4.8}$$

where a^2 solves the quadratic equation

$$3ga^4 + a^2 - 1 = 0. \tag{5.4.9}$$

(d) Compute the associated density $\psi(x)$ and plot it as g varies.

4. Establish the identity (1.3.11).

5. Show that the Catalan numbers enumerate the number of Dyck paths as discussed below equation (1.3.12).

5.5 Notes

To include in improved version.

1. Fixed point equation for equilibrium measure.
2. Properties of Hilbert transform.
3. Convergence of k-point distribution to tensor product of equilibrium measure.

Chapter 6

Other random matrix ensembles

In this chapter we discuss other random matrix ensembles that differ fundamentally from GUE, GOE and GSE. For this discussion we concentrate on real and complex matrices. The first ensembles we consider are the real and complex *Ginibre ensembles*[1], $\mathrm{Gin}_{\mathbb{R}}(m,n)$ on $\mathbb{R}^{m\times n}$ and $\mathrm{Gin}_{\mathbb{C}}(m,n)$ on $\mathbb{C}^{m\times n}$. These are ensembles of real and complex matrices of size $m\times n$. without symmetry conditions. Their densities are given by

$$p_{\mathrm{Gin},\mathbb{R}}(Y)\mathrm{D}Y = \frac{1}{Z_{\mathbb{R},m,n}}\mathrm{e}^{-\frac{1}{4}\operatorname{Tr}Y^TY}\mathrm{D}Y, \quad p_{\mathrm{Gin},\mathbb{C}}(X)\mathrm{D}X = \frac{1}{Z_{\mathbb{C},m,n}}\mathrm{e}^{-\frac{1}{2}\operatorname{Tr}X^*X}\mathrm{D}X.$$

Thus, the entries are distributed as independent (real or complex) normal random variables. The definition $\mathrm{D}Y$ and $\mathrm{D}X$ in each case follows directly from the volume forms associated to the length elements $\mathrm{Tr}(\mathrm{d}Y^T\mathrm{d}Y)$ and $\mathrm{Tr}(\mathrm{d}X^*\mathrm{d}X)$. When $m=n$ we use the notation $\mathrm{Gin}_{\mathbb{C}}(n)$ and $\mathrm{Gin}_{\mathbb{R}}(n)$ and $Z_{\mathbb{R},n}$ and $Z_{\mathbb{C},m}$.

The Ginibre ensembles allow us to define the Laguerre ensembles as transformations of $\mathrm{Gin}_{\mathbb{C}}(m,n)$ and $\mathrm{Gin}_{\mathbb{R}}(m,n)$. These are ensembles of positive (semi-) definite matrices defined by X^*X where $X\sim\mathrm{Gin}_{\mathbb{C}}(m,n),\mathrm{Gin}_{\mathbb{R}}(m,n)$. The Laguerre ensembles are often referred to as Wishart matrices and they get their name from the close connection to Laguerre polynomials.

We end this chapter with a discussion of the so-called Jacobi ensembles. It is important to note that these ensembles are *not* ensembles of Jacobi matrices, rather, they get their name from their close connection to Jacobi polynomials. Jacobi polynomials are polynomials orthogonal on the interval $[-1,1]$, and so Jacobi matrices have eigenvalues that all lie in the same interval $[-1,1]$.

[1] Often, the term Ginibre ensemble is reserved for square matrices, but we find it convenient to keep it for all rectangular matrices.

6.1 The Ginibre ensembles

Our first task is to generalize Weyl's formula to the Ginibre ensembles $\mathrm{Gin}_{\mathbb{R}}(n)$ and $\mathrm{Gin}_{\mathbb{C}}(n)$. To compute this, we use the Schur decomposition. The Schur decomposition is often seen as a numerical tool to perform a spectral decomposition of non-normal matrices. The eigenvalue decomposition is unstable to compute: matrices with distinct eigenvalues are dense and so, computing a Jordan block of a non-normal matrix is a precarious task when round-off errors are present. An arbitrarily small perturbation will lead to an $O(1)$ change in the eigenvalue matrix.

Theorem 46. *All matrices $Y \in \mathbb{R}^{n\times n}$ and $X \in \mathbb{C}^{n\times n}$ have decompositions*

$$Y = OSO^T, \quad X = UTU^*,$$

where $O \in \mathsf{O}(n)$, $U \in \mathsf{U}(n)$. Here $T \in \mathbb{C}^{n\times n}$ is upper-triangular and $S \in \mathbb{R}^{n\times n}$ is block-upper triangular with blocks of size 1 or 2. These 2×2 blocks have the form

$$\begin{pmatrix} \alpha & -\gamma \\ \delta & \alpha \end{pmatrix}, \ \alpha \in \mathbb{R}, \ \delta, \gamma > 0. \tag{6.1.1}$$

Furthermore, if the eigenvalues are distinct with a given ordering, and the eigenvectors are normalized (say, first non-zero component is positive), the decomposition is unique.

This can be proved by first performing an eigenvalue decomposition and second, performing a QR factorization of the eigenvector matrix. We now describe the QR decomposition algorithm, using Householder reflections, for real matrices. Another numerically viable algorithm is the modified Gram–Schmidt procedure. Both algorithms extend to complex matrices in a straightforward way. Given a matrix $Y \in \mathbb{R}^{m\times n}$, $Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$, define a transformation $Y \mapsto P(Y)Y$ by

$$\begin{aligned} P(Y)Y &= \begin{pmatrix} |y_1| & P_v y_2 & \cdots & P_v y_n \end{pmatrix}, \quad P_v = I - 2vv^T, \\ v &= \tilde{v}/|\tilde{v}|, \quad \tilde{v} = |y_1|e_1 - y_1. \end{aligned} \tag{6.1.2}$$

If $y_1 = 0$, we use $P = I$. Let I_j be the $j \times j$ identity matrix, and let $[Y]_{j,k}$ be the lower-right $j \times k$ sub-block of Y. The QR factorization of a matrix Y is then given via

$$\begin{aligned} Y_0 &= Y, \\ Y_1 &= Q_1 Y_0 := P(Y_0)Y_0, \\ Y_2 &= Q_2 Y_2 := \begin{pmatrix} I_1 & 0 \\ 0 & P\left([Y_1]_{m-1,n-1}\right) \end{pmatrix} Y_1, \\ &\vdots \\ Y_j &= Q_j Y_{j-1} := \begin{pmatrix} I_{j-1} & 0 \\ 0 & P\left([Y_{j-1}]_{m-j+1,n-j+1}\right) \end{pmatrix} Y_{j-1}. \end{aligned} \tag{6.1.3}$$

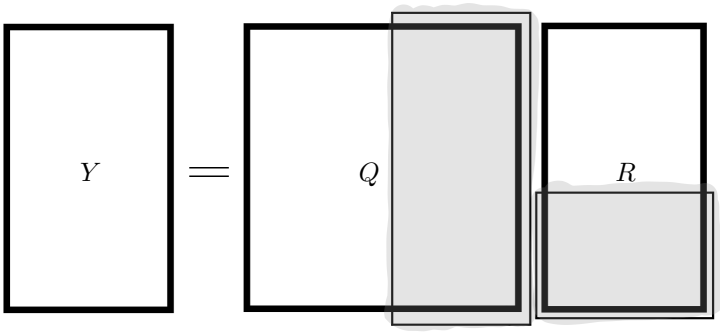

Figure 6.1.1: The full QR decomposition in the case $m > n$. The shaded area columns and rows are removed to create the reduced QR decomposition.

It follows that $R = Y_{\min\{m,n\}}$ is upper-triangular and $Y = QR$ where $Q = (Q_{\min\{m,n\}} \cdots Q_2 Q_1)^T$. We arrive at the following.

Theorem 47. *Every matrix $Y \in \mathbb{R}^{m\times n}$, $X \in \mathbb{C}^{m\times n}$ has a factorization $Y = QR$, $X = UT$ such that $Q \in \mathsf{O}(m)$, $U \in \mathsf{U}(m)$ where R, T are upper-triangular with non-negative diagonal entries. The factorization is unique if X (resp. Y) is invertible. This is called the QR factorization, or decomposition, of the matrix.*

This theorem gives the *full* QR decomposition. If $m > n$, then a $m - n$ columns of Q, U are redundant, and $m - n$ rows of R, T are as well, see Figure 6.1.1. After dropping these columns and rows, one obtains the *reduced* QR decomposition.

If $m > n$, one can count the number of degrees of freedom to see that neither Q nor U could ever be distributed according to Haar measure on $\mathsf{U}(m)$ or $\mathsf{O}(n)$ for $X \sim \mathrm{Gin}_{\mathbb{C}}(m, n)$ or $Y \sim \mathrm{Gin}_{\mathbb{R}}(m, n)$, respectively. So, we instead consider the QR factorization of the augmented matrices

$$\begin{pmatrix} X & X' \end{pmatrix} \text{ and } \begin{pmatrix} Y & Y' \end{pmatrix}, \quad X' \sim \mathrm{Gin}_{\mathbb{C}}(m, m-n), \quad Y' \sim \mathrm{Gin}_{\mathbb{R}}(m, m-n), \tag{6.1.4}$$

for X' and Y' independent of X and Y, respectively. This can be performed even if X and Y are deterministic matrices. So, in the real case, and similarly in the complex case,

$$Y \mapsto \begin{pmatrix} Y & Y' \end{pmatrix} = QR' \mapsto QR := QR' \begin{pmatrix} I_n \\ 0 \end{pmatrix} = Y.$$

Since it is a non-classical theorem for the Schur decomposition, we state the following.

Theorem 48. *Let $X(t)$, $X : (-a, a) \to \mathbb{F}^{n\times n}$, $a > 0$, be a C^k matrix function. Assume $X(0)$ has distinct eigenvalues. Then the induced factors $X(t) \mapsto (T(t), U(t))$ or $X(t) \mapsto (S(t), O(t))$ obtained by the Schur decomposition for $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$ are also C^k in a neighborhood of $t = 0$.*

Finally, before we proceed to pushing forward measure via these decompositions, we prove an elementary result for Ginibre ensembles using the QR factorization.

Theorem 49. *If $X \sim \mathrm{Gin}_{\mathbb{C}}(m,n)$, $Y \sim \mathrm{Gin}_{\mathbb{R}}(m,n)$, $m \geq n$ then*

$$\mathbb{P}\left(\operatorname{rank} X < n\right) = 0 \quad \textit{and} \quad \mathbb{P}\left(\operatorname{rank} Y < n\right) = 0.$$

Proof. We use induction on n for the real case. The complex case is similar. If $n = 1$, then a Gaussian vector in $\mathbb{R}^n$ is non-zero with probability one. If $n > 1$, $n \leq m-1$, assume

$$\mathbb{P}\left(\operatorname{rank} Y < n\right) = 0, \quad Y \sim \mathrm{Gin}_{\mathbb{R}}(m,n).$$

Let $b \in \mathbb{R}^m$ be an independent Gaussian vector ($b \sim \mathrm{Gin}_{\mathbb{R}}(m,1)$). Then

$$\mathbb{P}\left(\operatorname{rank}\begin{pmatrix} Y & b \end{pmatrix} < n+1\right) = \mathbb{E}\left[\mathbb{P}\left(\operatorname{rank}\begin{pmatrix} Y & b \end{pmatrix} < n+1 \,|\, Y\right)\right].$$

On a set of full probability $\operatorname{rank} Y = n$. For such a matrix consider

$$\mathbb{P}\left(\operatorname{rank}\begin{pmatrix} Y & b \end{pmatrix} < n+1 \,|\, Y\right).$$

Solve

$$Yx = b = QRx = b, \quad Rx = Q^T b =: \tilde{b},$$

and therefore $\tilde{b} \sim \mathrm{Gin}_{\mathbb{R}}(m,1)$. For this equation to have a solution x, $Rx = \tilde{b}$, since $R \in \mathbb{R}^{m\times n}$, triangular, and $n < m$, the last entry of $\tilde{b}$ must vanish. Thus

$$\mathbb{P}\left(\operatorname{rank}\begin{pmatrix} Y & b \end{pmatrix} < m+1 \,|\, Y\right) = 0$$

almost surely. This proves the claim. □

Finally, we want to know that the probability of finding a Ginibre matrix with an eigenvector that has a zero first component is zero.

Theorem 50. *Assume $X \sim \mathrm{Gin}_{\mathbb{C}}(n)$, $Y \sim \mathrm{Gin}_{\mathbb{R}}(n)$. Then*

$$\mathbb{P}\left(\exists \lambda \in \mathbb{C}, v \in \mathbb{C}^n, Xv = \lambda v \text{ and } v_1 = 0\right) = 0,$$
$$\mathbb{P}\left(\exists \lambda \in \mathbb{C}, v \in \mathbb{R}^n, Yv = \lambda v \text{ and } v_1 = 0\right) = 0.$$

Proof. We prove this for Y. The proof for X is similar. First, we write

$$Y = \begin{pmatrix} y_0 & y_1^T \\ y_2 & Y' \end{pmatrix},$$
$$y_0 \sim \mathrm{Gin}_{\mathbb{R}}(1),\ y_1, y_2 \sim \mathrm{Gin}_{\mathbb{R}}(n-1,1),\ Y' \sim \mathrm{Gin}_{\mathbb{R}}(n-1,n-1),$$

all independent. Let

$$E = \left\{\exists \lambda \in \mathbb{C}, v \in \mathbb{R}^{n-1}, Y'v = \lambda v \text{ and } Y\begin{pmatrix} 0 \\ v \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ v \end{pmatrix}\right\}.$$

It then follows that

$$\mathbb{P}\left(\exists \lambda \in \mathbb{C}, v \in \mathbb{R}^n, Yv = \lambda v \text{ and } v_1 = 0\right) = \mathbb{P}(E) = \mathbb{E}\left[\mathbb{P}(E|Y')\right].$$

Then

$$\mathbb{P}(E|Y') = \mathbb{P}\left(\exists v \in \mathbb{R}^n,\ y_1^T v = 0,\ v \text{ is an eigenvector of } Y'|Y'\right).$$

For the eigenvalue λ_j of Y', let $V_j = \left(v^{(1)}, \ldots, v^{(\ell)}\right)$, $\ell \leq n-1$ be a basis of eigenvectors for this eigenvalue. Then

$$\mathbb{P}\left(\exists \{c_j\} \text{ so that } \ y_1^T \left(\sum_{j=1}^{\ell} c_j v^{(j)}\right) = 0 \middle| X'\right) = 0,\ a.s.$$

Because, given X', perform a QR factorization of $V_j = QR$, and consider $y_1^T QRc = 0$, $c = (c_1, \ldots, c_j)^T$. But as R has rank ℓ, this amounts to the condition that (at least) one component of the Gaussian vector $x^T = y_1^T Q$ has to vanish, a probability zero event. A union bound over all the distinct eigenvalues proves the result. □

This theorem has an interesting implication. If a matrix Y has a repeated eigenvalue and two linearly independent eigenvectors, then an eigenvector can be constructed that has a zero first component. By the theorem, this event occurs with probability zero for $\mathrm{Gin}_{\mathbb{R}}(n)$, $\mathrm{Gin}_{\mathbb{C}}(n)$. And so, if one shows that Y is diagonalizable with probability one, then Y has distinct eigenvalues with probability one. Nevertheless, it is actually easier to directly show this.

Theorem 51. *Assume $X \sim \mathrm{Gin}_{\mathbb{C}}(n)$, $Y \sim \mathrm{Gin}_{\mathbb{R}}(n)$. Then*

$$\mathbb{P}\left(X \textit{ has distinct eigenvalues }\right) = 1,$$
$$\mathbb{P}\left(Y \textit{ has distinct eigenvalues }\right) = 1.$$

Proof. We show that the Vandermonde squared $\triangle(\Lambda)^2$ is a polynomial in the entries of the matrix. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of Y and consider

$$V = (V_{ij}), \quad V_{jk} = \lambda_j^{k-1}.$$

Then

$$\triangle(\Lambda)^2 = \det(V)^2 = \det(V^T V), \quad (V^T V)_{jk} = \sum_{\ell=1}^{n} \lambda_\ell^{j+k-2} = \mathrm{Tr}\, Y^{j+k-2}.$$

Now consider a rectangle $R = [a,b]^{n^2} \subset \mathbb{R}^{n^2}$, and assume that

$$\int_R \mathbf{1}_{\{Y \in \mathbb{R}^{n\times n} \ |\ |\triangle(\Lambda)|=0\}} \mathrm{D}Y > 0.$$

Since the set of matrices with distinct eigenvalues is dense, $\triangle(\Lambda) \neq 0$ for some Y. But the only way for the zero locus of a polynomial in n variables to have positive n-dimensional Lebesgue measure is for the polynomial to vanish identically. The theorem follows. □

6.1.1 Schur decomposition of $\mathrm{Gin}_{\mathbb{C}}(n)$

Theorems 46 and 48 allow us to compute the distribution induced on U and T in the Schur decomposition. We first identify the tangent space.

Theorem 52. *Assume $X \in \mathbb{C}^{n\times n}$ has distinct eigenvalues. Then*

$$T_X\mathbb{C}^{n\times n} \cong \mathbb{R}^{n(n-1)} \oplus PT_I\mathsf{U}(n).$$

Proof. A straightforward computation, using the differentiability of the Schur decomposition gives

$$\dot{X} = U(\dot{T} + [U^*\dot{U}, T])U^*, \tag{6.1.5}$$

after using $X(t)$, $t \in (-a, a)$, $a > 0$, differentiating and evaluating at $t = 0$. It follows that $S := U^*\dot{U}$ is skew-symmetric. We then decompose $T = \Lambda + T_+$ and $S = S_0 + S_- + S_+$, where the $\pm$ refers to strict upper- and lower- triangular parts. We can first solve for S_- of S in the following way. Define $S_- \mapsto \zeta \in \mathbb{C}^{n(n-1)/2}$ by ordering the entries of using the following relations:

$$\begin{aligned} &(i,j) < (i',j') \ \text{ if } \ i - j < i' - j', \\ &(i,j) < (i',j') \ \text{ if } \ i - j = i' - j' \text{ and } i < i'. \end{aligned} \tag{6.1.6}$$

The first inequality orders entries by which diagonal they lie on. The second orders within the diagonal. Then

$$\dot{X}_- = [S_-, \Lambda] + [S_-, T_+].$$

With the chosen ordering

$$\zeta \mapsto [S_-, T_+] =: M_-\zeta \tag{6.1.7}$$

is strictly lower triangular. Thus provided $\lambda_i \neq \lambda_j$ for $i \neq j$, we can solve this for S_-. If we then make the choice that $S_0 = 0$, we can clearly solve for $\dot{T}$ once S is known. Finally, by adjusting $\dot{T}$ accordingly, it is clear that any $\dot{X}$ can be achieved with $S_0 = 0$. □

Now, we give the analogue of Weyl's formula for $\mathbb{C}^{n\times n}$.

Theorem 53. *For $X \in \mathbb{C}^{n\times n}$,*

$$\mathrm{D}X = |\triangle(\Lambda)|^2 \,\mathrm{D}T\,\mathrm{D}U, \tag{6.1.8}$$

where $\mathrm{D}T = \prod_{j=1}^n \mathsf{dRe}\lambda_j \mathsf{dIm}\lambda_j \prod_{j<k} \mathsf{dRe}T_{jk}\mathsf{dIm}T_{jk}$ and $\mathrm{D}U$ refers to the same distribution as that of the eigenvectors of $\mathrm{GUE}(n)$.

Proof. We first map X to $\mathbb{C}^{n^2}$ in a consistent way. We order X_- using (6.2.2) giving ζ^{X_-}. We then order diagonal(X) in the usual way. Then, finally we order X_+ using

$$(i,j) \prec (i',j') \text{ if and only if } (j,i) < (j',i'),$$

giving ζ^{X_+}, and $X \mapsto [\zeta^{X_-}, \eta, \zeta^{X_+}]^T$. We use ζ^{S_-} and ζ^{T_+} in same way for S_- and T_+, respectively. It then follows that, after ordering $U^*\mathrm{d}XU$,

$$U^*\mathrm{d}XU = \begin{pmatrix} \tilde{\Lambda} + M_- & 0 & 0 \\ D & I & 0 \\ M_+ & 0 & I \end{pmatrix} \begin{pmatrix} \mathrm{d}\zeta^{S_-} \\ \mathrm{d}\Lambda \\ \mathrm{d}\zeta^{T_+} \end{pmatrix}.$$

where $\tilde{\Lambda}\zeta^{S_-}$ is defined through $\zeta^{S_-} \mapsto [S_-, \Lambda]$, which is diagonal, $S_{jk} \mapsto (\lambda_k - \lambda_j)S_{jk}$. M_- and D are matrices whose exact form is irrelevant. Decomposing all differentials into real and imaginary parts and computing the metric tensor

$$\mathrm{Tr}\, \mathrm{d}X^* \mathrm{d}X,$$

we find (6.1.8) by using $\det(\tilde{\Lambda} + M_-) = \prod_{j<k}(\lambda_k - \lambda_j)$ and computing the associate volume form. Here one has to use that if $A : \mathbb{C}^n \to \mathbb{C}^n$ induces $B : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ (by separating real and imaginary parts), then $\det B = |\det A|^2$. □

Theorem 54. *The Schur decomposition of* $\mathrm{Gin}_{\mathbb{C}}(n)$ *is given by*

$$p_{\mathrm{Gin},\mathbb{C}}(X)\mathrm{D}X = \frac{1}{Z_{\mathbb{C},n}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}\, T^*T} |\triangle(\Lambda)|^2 \,\mathrm{D}T\,\mathrm{D}U. \tag{6.1.9}$$

Note that this implies that the strict upper-triangular entries of T are all iid complex normal random variables.

6.1.2 QR decomposition of $\mathrm{Gin}_{\mathbb{C}}(m, n)$

We now consider the distribution induced on U and T by $\mathrm{Gin}_{\mathbb{C}}(m, n)$. Following the discussion in (6.1.4), we assume $n \geq m$. We follow the push forward of the distributions under the algorithm in (6.1.3). If $X \sim \mathrm{Gin}_{\mathbb{C}}(m, n)$ then it follows that if we replace Q_j with U_j and Y_j with X_j in (6.1.4) then X_j and U_j are independent for every j using the fact that the length of a Gaussian vector is independent of its angle and UX is independent of $U \in \mathsf{U}(m)$ if U is independent of X. And therefore, for $X = UT$, U is independent of T.

From the discussion in Section 3.2 it follows that the induced volume form on T is

$$\propto \mathrm{e}^{-\frac{\beta}{4}\mathrm{Tr}\, T^*T} \prod_{j=1}^{m} T_{jj}^{2m-2j+1} \mathrm{D}T, \quad \beta = 2,$$

where $\mathrm{D}T$ refers to standard Lebesgue measure on $\mathbb{R}_+^m \times \mathbb{C}^{m(m-1)/2+m(m-n)}$. Note that all the strictly upper-triangular entries are standard complex normal random variables and the entries on the diagonal are all chi-distributed. To understand the distribution on U all we need to do use to use that for $O \in \mathsf{U}(m)$, $OX \sim \mathrm{Gin}_{\mathbb{C}}(m, n)$ if $X \sim \mathrm{Gin}_{\mathbb{C}}(m, n)$. Then factorize

$$X = UT \quad OX = U'T'.$$

From the uniqueness of the QR factorization (on set of full probability where X is invertible), $T = T'$ and $U = O^T U'$. But U and U' have the same distribution and this distribution must therefore be invariant under left multiplication by any element of $\mathsf{U}(m)$. We conclude U is distributed according to Haar measure on $\mathsf{U}(m)$ [27] and to proportionality constants:

$$\mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr} X^* X}\mathrm{D}X \xrightarrow{\mathrm{QR}} \mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr} T^* T}\prod_{j=1}^{\tilde{n}} T_{jj}^{2m-2j+1}\mathrm{D}T\,\mathrm{D}\,\tilde{U}, \quad \tilde{n} = \min\{m,n\},$$

where $\mathrm{D}\tilde{U}$ is defined in (2.5.8). The normalization constant is easily computed in terms of Γ-functions. This can be seen as an equality when $m \leq n$. For $m \geq n$, we add additional degrees of freedom to find $\mathrm{D}\tilde{U}$, and so this is the push-forward under a random transformation.

6.1.3 Eigenvalues and eigenvectors of $\mathrm{Gin}_{\mathbb{R}}(n)$

Computing the analogue of Weyl's formula for $\mathrm{Gin}_{\mathbb{R}}(n)$ is much more complicated. This comes from the fact that complex eigenvalues must arise as complex conjugate pairs. Furthermore, for finite n there is a non-zero probability that the matrix with have k real eigenvalues. Thus the distribution on the eigenvalues is not absolutely continuous with respect to Lebesgue measure on $\mathbb{C}$. We first compute the tangent space, under the assumption of k real eigenvalues.

Theorem 55. *Assume that Y has exactly k real eigenvalues. Assume further that the real part of all the eigenvalues of $Y(0) = Y$ in the closed upper-half plane are distinct. Finally, assume that each 2×2 block in the real Schur factorization has $\gamma \neq \delta$ in* (6.1.1)*. Then*

$$T_Y\mathbb{R}^{n\times n} \cong \mathbb{R}^{n(n-1)/2} \oplus \mathfrak{o}(n).$$

Proof. Assume $Y(t)$ is a smooth curve in $\mathbb{R}^{n\times n}$ such that $Y(t)$ has k real eigenvalues for all t. As before, we have the relation

$$\dot{Y} = O(\dot{S} + [O^T\dot{O}, S])O^T.$$

We need to show that the entries of $\dot{S}$ and $\dot{O}$ are uniquely determined by this relation. We assume

$$S = \left(\begin{array}{cccc|ccc} R_1 & \times & \cdots & & & \cdots & \times \\ 0 & R_2 & \times & \cdots & & \cdots & \times \\ \vdots & \ddots & \ddots & & & & \vdots \\ 0 & \cdots & 0 & R_\ell & \times & \cdots & \times \\ \hline 0 & \cdots & \cdots & 0 & \lambda_1 & \cdots & \times \\ \vdots & & & & \ddots & \ddots & \\ 0 & \cdots & & & \cdots & 0 & \lambda_k \end{array}\right), \quad R_j = \begin{pmatrix} \alpha_j & -\gamma_j \\ \delta_j & \alpha_j \end{pmatrix},$$

where $\ell = (n-k)/2$ and $n-k$ is assumed to be even. The ordering is fixed by $\alpha_j < \alpha_{j+1}$ and $\lambda_j < \lambda_{j+1}$. We also refer to the location of all the imposed zeros in S as the *generalized lower-triangular* part of S, denoted $L_G(S)$. Similarly, $U_G(S) = (L_G(S^T))^T$ and $D_G(S) = S - U_G(S) - L_G(S)$. So, we have

$$L_G(O^T \dot{Y} O) = L_G\left([A,S]\right), \quad A^T = -A.$$

After careful consideration, we find

$$L_G\left([A,S]\right) = L_G\left([L_G(A), U_G(S)] + [L_G(A), D_G(S)]\right)$$

by noting that

$$\begin{aligned}[A,S] = {} & [L_G(A), L_G(S)] + [D_G(A), L_G(S)] + [U_G(A), L_G(S)] \\ & + [L_G(A), D_G(S)] + [D_G(A), D_G(S)] + [U_G(A), D_G(S)] \\ & + [L_G(A), U_G(S)] + [D_G(A), U_G(S)] + [U_G(A), U_G(S)],\end{aligned}$$

$L_G(S) = 0$, and any term involving only D_G and U_G or only U_G does not contribute to $L_G([A,S])$. Then, it is a non-trivial but straightforward calculation to find that $L_G([D_G(A), D_G(S)]) = 0$. This gives a linear system of equations for $L_G(A)$. Since it will be of use in computing the metric tensor below, we compute the determinant of this matrix in the following lemma.

Lemma 20. *There exists a trivial mapping $L_G(A) \to \xi \in R^{n(n-1)/2-\ell}$ defined by ordering the elements of $L_G(A)$ so that when M is the matrix representation for $\xi \mapsto L_G\left([A,S]\right)$ we have*

$$\det M = \triangle_k(\Lambda) := \left(\prod_{1\le i<j\le k} (\lambda_j - \lambda_i)\right)\left(\prod_{1\le j<k\le \ell} \triangle_{ij}^{(1)}\right)\left(\prod_{1\le i\le k, 1\le j\le \ell} \triangle_{ij}^{(2)}\right)$$

where $\lambda_1, \ldots, \lambda_k$ are the real eigenvalues, $\mu_j = \alpha_j + \mathrm{i}\beta_j$, $\beta_j > 0$ are the complex eigenvalues (in the upper half plane) and

$$\begin{aligned}\triangle_{ij}^{(1)} &= |\mu_j - \mu_i|^2 |\mu_j - \bar{\mu}_i|^2 = |\mu_j - \mu_i|^2 |\bar{\mu}_j - \mu_i|^2, \\ \triangle_{ij}^{(2)} &= |\mu_j - \lambda_i|^2.\end{aligned}$$

Proof of Lemma 20. The important aspect of this is to choose the ordering. First split

$$L_G(A) = \left(\begin{array}{c|c} A^{(1,1)} & 0 \\ A^{(2,1)} & A^{(2,2)} \end{array}\right).$$

We order the 2×2 blocks of $A^{(1,1)}$ according to (6.2.2). Within each block we use this same ordering. We then order the entries of $A^{(2,2)}$ according to (6.2.2). Finally, we order the 1×2 blocks of $A^{(2,1)}$ according to (6.2.2) and within each block we use this same ordering. This defines $L_G(A) \mapsto \xi \in \mathbb{R}^{n(n-1)/2-\ell}$. Define $L = L_G(L_G(A), U_G(S))$ and decompose L into $L^{(i,j)}$, $i = 1,2$, $j = 1,2$

in the same was as for $L_G(A)$. From the reasoning[2] that went into (6.1.7), we have that the (i,j) block of $L^{(1,1)}$ depends only on blocks (i',j') of $A^{(1,1)}$ for $(i',j') > (i,j)$ and entries in $A^{(2,1)}$. Similarly, the (i,j) entry of $L^{(2,2)}$ depends only on entries (i',j') of $A^{(2,2)}$ for $(i',j') > (i,j)$ and entries in $A^{(2,1)}$. Lastly, one checks that block (i,j) of $L^{(2,1)}$ depends only on blocks (i',j') of $A^{(2,1)}$ for $(i',j') > (i,j)$. This gives a strong form of strict lower-triangularity for $\xi \mapsto L$.

We now show that $\xi \mapsto K := L_G(L_G(A), D_G(S))$ is block-diagonal in a way that does not overlap with this strict lower-triangularity. First, decompose K into $K^{(i,j)}$, $i = 1, 2,\ j = 1, 2$ in the same was as for $L_G(A)$ and L. We obtain the following relations for blocks of size 2×2, 1×1 and 1×2, respectively:

$$\begin{aligned} K_{ij}^{(1,1)} &= A_{ij}^{(1,1)} R_j - R_i A_{ij}^{(1,1)}, \\ K_{ij}^{(2,2)} &= A_{ij}^{(2,2)} (\lambda_j - \lambda_i), \\ K_{ij}^{(2,1)} &= A_{ij}^{(2,1)} R_j - \lambda_i A_{ij}^{(2,1)}. \end{aligned}$$

The determinants of each of these linear transformations are

$$\begin{gathered} (\alpha_j - \alpha_i)^4 + (\delta_j \gamma_j - \delta_i \gamma_i)^2 + 2(\alpha_j - \alpha_i)^2 (\delta_j \gamma_j + \delta_i \gamma_i), \\ (\lambda_j - \lambda_i), \\ (\alpha_j - \lambda_i)^2 + \delta_j \gamma_j, \end{gathered}$$

respectively. For the non-real eigenvalues in the upper-half plane, we have $\mu_j = \alpha_j + \mathrm{i}\sqrt{\gamma_j \delta_j}$. This proves the lemma. □

From this lemma, with our assumptions, we can uniquely find $L_G(A)$. But as A is skew-symmetric, we have ℓ entries left undetermined. So, we consider

$$\begin{aligned} (O^T \dot{Y})_{2j,2j} &= (\dot{S} + [A,S])_{2j,2j} = (\dot{\alpha}_j + (\gamma_j - \delta_j)\dot{s}_{2j+1,2j}) + f_{2j}(L_G(A)), \\ (O^T \dot{Y})_{2j+1,2j+1} &= (\dot{S} + [A,S])_{2j+1,2j+1} = (\dot{\alpha}_j + (\delta_j - \gamma_j)\dot{s}_{2j+1,2j}) + f_{2j}(L_G(A)). \end{aligned} \tag{6.1.10}$$

for some functions f_j. As $L_G(A)$ is known, this gives a solvable system for $\dot{\alpha}_j$ and $s_{2j+1,2j}$, with determinant $2^\ell \prod_{j=1}^{\ell} (\gamma_j - \delta_j)$. The remaining entries of $\dot{S}$ are given through the relation

$$\dot{S} = O^T \dot{Y} O - [A,S].$$

□

We now can compute the volume form.

[2]The commutator of lower-triangular and upper triangular matrices at entry (i,j) only depends on entries (i',j') of the lower-triangular matrix for $j' \leq j$ with $i = i'$ and $i' \geq i$ with $j = j'$. With strict triangularity, fewer dependencies occur.

Theorem 56. *For $Y \in \mathbb{R}^{n\times n}$ with k real eigenvalues,*

$$\mathrm{D}Y = 2^{\ell}|\triangle_k(\Lambda)| \left(\prod_{j=1}^{\ell} |\gamma_j - \delta_j|\right) \mathrm{D}S\,\mathrm{D}O, \tag{6.1.11}$$

where

$$\mathrm{D}S = \prod_{j=1}^{\ell} \mathrm{d}\alpha_j \mathrm{d}\gamma_j \mathrm{d}\delta_j \prod_{j=1}^{k} \mathrm{d}\lambda_j \prod_{s\in U_G(S)} \mathrm{d}s, \tag{6.1.12}$$

and $\mathrm{D}O$ *refers to the same distribution as that of the eigenvectors of* $\mathrm{GOE}(n)$*, i.e., Haar measure on* $\mathsf{O}(n)$*.*

When we restrict to k real eigenvalues we use the notation

$$p_{\mathrm{Gin},\mathbb{R},k}(Y)\mathrm{D}Y = \frac{1}{Z^{(k)}_{\mathbb{R},n}} \mathrm{e}^{-\frac{1}{4}Y^TY} \mathbf{1}_{\{Y \text{ has } k \text{ real eigenvalues}\}} \mathrm{D}Y. \tag{6.1.13}$$

Theorem 57. *The real Schur decomposition of* $\mathrm{Gin}_{\mathbb{R}}(n)$ *given k real eigenvalues is*

$$p_{\mathrm{Gin},\mathbb{R},k}(Y)\mathrm{D}Y = \frac{2^{\ell}}{Z^{(k)}_{\mathbb{R},n}} \mathrm{e}^{-\frac{1}{4}\operatorname{Tr} S^TS} |\triangle_k(\Lambda)| \left(\prod_{j=1}^{\ell} |\gamma_j - \delta_j|\right) \mathrm{D}S\,\mathrm{D}O. \tag{6.1.14}$$

Note that this implies that the generalized upper-triangular entries of S are all iid normal random variables.

6.1.4 QR decomposition of $\mathrm{Gin}_{\mathbb{R}}(m,n)$

It follows from the discussion in Section 6.1.2 that up to proportionality constants

$$\mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr} Y^TY}\mathrm{D}Y \xrightarrow{\mathrm{QR}} \mathrm{e}^{-\frac{\beta}{4}\operatorname{Tr} R^TR} \prod_{j=1}^{\tilde{n}} R_{jj}^{m-j+1} \mathrm{D}R\,\mathrm{D}Q, \quad \beta = 1, \quad \tilde{n} = \min\{m,n\},$$

where $\mathrm{D}R$ refers to standard Lebesgue measure on $\mathbb{R}^m_+ \times \mathbb{R}^{m(m-1)/2+m(m-n)}$, and $\mathrm{D}Q$ is Haar measure on $\mathsf{O}(n)$.

6.2 Singular value decomposition and the Laguerre (Wishart) ensembles

Next, we turn to understanding the singular value decomposition of $\mathrm{Gin}_{\mathbb{C}}(m,n)$ and $\mathrm{Gin}_{\mathbb{R}}(m,n)$. This is done by means of describing the eigenvalue and eigenvector distributions of the so-called Laguerre ensembles. The following gives the singular value decomposition.

Theorem 58. *Every matrix $Y \in \mathbb{R}^{m\times n}$ and $X \in \mathbb{C}^{m\times n}$ has a decomposition*

$$Y = Q\Sigma O^T, \quad X = U\Sigma V^*,$$

where $Q \in \mathsf{O}(m)$, $O \in \mathsf{O}(n)$, $U \in \mathsf{U}(m)$, $V \in \mathsf{U}(m)$ and $\Sigma \in \mathbb{R}^{m\times n}$ is a diagonal matrix with non-negative diagonal entries.

The entries in Σ_j of Σ are called the *singular values* of the matrix in question.

6.2.1 The Cholesky decomposition

To compute the singular value decomposition of $\mathrm{Gin}_{\mathbb{R}}(m,n)$ and $\mathrm{Gin}_{\mathbb{C}}(m,n)$ we follow the approach of Edelman [10] and first compute the Cholesky decomposition.

Theorem 59. *Every strictly positive definite matrix $A \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) has a unique decomposition*

$$A = LL^T$$

where $L \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) is a lower-triangular matrix with positive diagonal entries.

Proof. We concentrate on the real case and we first show uniqueness. Assume $A = LL^T = L_1L_1^T$ for two different factorizations. Then

$$L_1^{-1}L = L_1^TL^{-T}, \quad \text{where} \quad L^{-T} = (L^{-1})^T.$$

Since the non-singular upper- and lower-triangular matrices for groups, the left-hand (right-hand) side is lower-triangular (upper-triangular). Therefore $L_1^{-1}L$ is a diagonal matrix that is equal to its own transpose-inverse: $e_jL_1^{-1}Le_j = \pm 1$. Positivity of the diagonal entries gives $L_1 = L$. Now, by Gaussian elimination, without pivoting[3] $A = \tilde{L}U$ where $\tilde{L}$ is lower-triangular and U is upper-triangular. Here $\tilde{L}$ has ones on the diagonal. We know that $e_j^TAe_j > 0$ and therefore $e_j^T\tilde{L}Ue_j = U_{jj} > 0$. Then Let $U_d = \mathrm{diagonal}(U)^{1/2}$ and $A = \tilde{L}U_dU_d^{-1}U$. It follows from the symmetry of A that $L = \tilde{L}U_d$ gives the Cholesky factorization. Similar considerations follow for $A \in \mathbb{C}^{m\times n}$. □

Change of variables for $\mathrm{Gin}_{\mathbb{C}}(m,n)$

We now consider the change of variables that closely resembles the singular value decomposition, but differs in a fundamental way. For $X \in \mathbb{C}^{m\times n}$, full rank, define

$$X = UT \overset{QR}{\mapsto} (U,T) \overset{\text{Inv. Cholesky}}{\mapsto} (U, A = T^*T) = (U, V\Lambda V^*) \overset{\text{Spectral map}}{\mapsto} (U,\Lambda,V). \tag{6.2.1}$$

[3] Pivoting is not required for strictly positive definite matrices because the upper left $\ell \times \ell$ blocks are non-singular for every ℓ.

This is a well-defined, invertible mapping, provided that the first row of V contains non-vanishing entries. It will follow from Section 6.2.2 that the probability of this is one. But we emphasize that for this decomposition $X \neq U\Lambda V^*$, generally. We now show that if $X \sim \mathrm{Gin}_{\mathbb{C}}(m,n)$ then U, Λ, V are independent and we then characterize the distribution of Λ and V.

Lemma 21 (Spectral variables for $\mathsf{Her}_+(n)$)**.** *If $A \in \mathsf{Her}_+(n)$ is non-singular with distinct eigenvalues then*

$$T_A\mathsf{Her}_+(n) \cong \mathbb{R}^n \oplus PT_I\mathsf{U}(n).$$

Proof. The proof is essentially the same as Lemma 6, just using that the set of strictly positive definite matrices is open. □

We define $\mathrm{D}A$ in the natural way as the volume form induced by the metric tensor $\mathrm{Tr}\,\mathrm{d}A^2$. We then have the analogous formula to Theorem 15:

$$\mathrm{D}A = |\triangle(\Lambda)|^2\,\mathrm{D}\Lambda\,\mathrm{D}U.$$

Next, we compute the volume form associated with the change Cholesky change of variables.

Lemma 22. *Let $A = LL^*$ be the Cholesky decomposition for a non-singular $A \in \mathsf{Her}_+(n)$. Let $\mathrm{D}L$ be the natural volume form induced by $\mathrm{Tr}(\mathrm{d}L^*\mathrm{d}L)$. Then*

$$\mathrm{D}A = 2^n \prod_{j=1}^{n} L_{jj}^{2(n-j)+1}\mathrm{D}L.$$

Proof. We prove this by identifying that the Jacobian of the transformation is triangular, and computing the diagonal entries. We first compute for $j \geq k$

$$\frac{\partial A}{\partial \mathsf{Re}L_{jk}} = e_j e_k^T L^* + L e_k e_j^T, \quad \frac{\partial A}{\partial \mathsf{Im}L_{jk}} = e_j e_k^T L^* - L e_k e_j^T.$$

Examine the structure of these matrices. Since $e_j e_k^T L^*$ is the matrix that contains the kth row of L^* in its jth row, with all other row being zero we find the following picture

$$\frac{\partial A}{\partial \mathsf{Re}L_{jk}} = \left(\begin{array}{ccccccc|c|c} & & & & & & & 0 & \\ & & & & & & & \vdots & \\ & & & & & & & 0 & \\ & & & & & & & L_{kk} & \\ & & & & & & & L_{k+1,k} & \\ & & & & & & & \vdots & \\ & & & & & & & L_{j-1,k} & \\ \hline 0 & \cdots & 0 & \overline{L}_{kk} & \overline{L}_{k+1,k} & \cdots & \overline{L}_{j-1,k} & 2\mathsf{Re}L_{jk} & \cdots \\ \hline & & & & & & & \vdots & \end{array}\right).$$

Here only the jth row and jth column have non-zero entries. Here $2\mathsf{Re}L_{jk}$ is in the (j,j) entry. A similar picture holds for $\frac{\partial A}{\partial \mathsf{Im}L_{jk}}$, with $2\mathsf{Im}L_{jk}$ in the (j,j) entry. We define a mapping $\mathsf{Re}L \mapsto \xi \in \mathbb{R}^{n(n-1)/2}$ and $\mathsf{Im}L \mapsto \eta \in \mathbb{R}^{n(n-3)/2}$ by the ordering of the non-zero elements of L:

$$\begin{aligned} (j,k) &< (j',k') \text{ if } j < j', \\ (j,k) &< (j\,,k') \text{ if } k < k'. \end{aligned} \tag{6.2.2}$$

This orders first by row, and then by columns within each row. Assume $(i,\ell) < (j,k)$, $j \geq k$, $i \geq \ell$. Then

$$\frac{\partial A_{i\ell}}{\partial \mathsf{Re}L_{jk}} = 0, \quad \frac{\partial A_{i\ell}}{\partial \mathsf{Im}L_{jk}} = 0.$$

because either $i < j$ or $\ell < k'$ if $i = j$. And, it is clear that

$$\begin{aligned} \frac{\partial A_{jk}}{\partial \mathsf{Re}L_{jk}} &= L_{kk}, \quad j > k, \quad \frac{\partial A_{jk}}{\partial \mathsf{Re}L_{jk}} = 2L_{kk}, \quad j = k, \\ \frac{\partial A_{jk}}{\partial \mathsf{Im}L_{jk}} &= L_{kk}, \quad j > k. \end{aligned}$$

Then, if we define $L \mapsto \zeta$ where $\zeta = (\xi_1, \eta_1, \xi_2, \eta_2, \ldots)^T$ we find that the Jacobian is triangular and

$$\frac{\partial A}{\partial L} = 2^n \prod_{j=1}^{n} L_{jj}^{2(n-j)+1}.$$

□

This theorem allows one to understand transformations of $\mathrm{Gin}_{\mathbb{C}}(m,n)$. Following the transformation (6.2.1), with $X \in \mathbb{C}^{m\times n}$ with $m \geq n$ using $T = L^*$ noting that

$$T = \begin{pmatrix} \tilde{T} \\ 0 \end{pmatrix}.$$

where $\tilde{T}$ is a upper-triangular matrix with positive diagonal entries. Then

$$\mathrm{D}X \xrightarrow{\mathrm{QR}} \prod_{j=1}^{n} T_{jj}^{2(m-j)+1} \mathrm{D}T\, \mathrm{D}\tilde{U} = 2^{-n} \prod_{j=1}^{n} T_{jj}^{2(m-n)} \mathrm{D}A\, \mathrm{D}\tilde{U} \tag{6.2.3}$$

$$= 2^{-n} \prod_{j=1}^{n} \sigma_j^{2(m-n)} |\triangle(\Sigma^2)|^2 \mathrm{D}\Sigma\, \mathrm{D}\tilde{U}\, \mathrm{D}V. \tag{6.2.4}$$

Here $\mathrm{D}\tilde{U}$ is Haar measure on $\mathsf{U}(n)$ and $\mathrm{D}V$ represents the same distribution as the eigenvectors of $\mathrm{GUE}(n)$. Also, $\mathrm{D}\Sigma$ is Lebesgue measure on $\mathbb{R}^n_+$. As noted

below (6.2.1), this is not the singular value decomposition for X, but we claim, it is in a distributional sense. For $X \sim \mathrm{Gin}_{\mathbb{C}}(m,n)$, $m \geq n$ and consider

$$X = U_1 \Sigma V, \quad \tilde{X} := U \Sigma V$$

where (U, V, Σ) are independent with joint distribution (6.2.4), U_1 is the matrix of left singular vectors for X, and U is independent of U_1. Then $\tilde{X} = U U_1^* X$, but then by the invariance of U, for measureable sets $S_1 \subset \mathsf{U}(m)$, $S_2 \subset \mathbb{C}^{m \times m}$,

$$\begin{aligned}
&\mathbb{P}(UU_1^* \in S_1) = \mathbb{P}(U \in S_1 U_1) = \mathbb{P}(U \in S_1), \\
&\mathbb{P}(UU_1^* \in S_1, X \in S_2) = \mathbb{P}(U \in S_1 U_1, X \in S_2) \\
&\qquad = \int_{S_2} \left(\int_{S_1 U_1} \mathrm{D}U \right) p_{\mathrm{Gin},\mathbb{C}}(X) \mathrm{D}X = \mathbb{P}(U \in S_1)\mathbb{P}(X \in S_2).
\end{aligned}$$

So, UU_1^* is independent of X and therefore $\tilde{X}$ must have the same distribution as X. This implies the singular value decomposition of $\mathrm{Gin}_{\mathbb{C}}(m,n)$ is given by (6.2.4).

Remark 60. If one wants to match of dimensions, then $\mathrm{D}U$ should be replaced by the push-forward of uniform measure on $S_{\mathbb{C}}^{m-1} \times S_{\mathbb{C}}^{m-2} \times \cdots \times S_{\mathbb{C}}^{m-n-1}$ onto $\mathsf{U}(m)$ via Householder reflections.

Change of variables for $\mathrm{Gin}_{\mathbb{R}}(m,n)$

Similar considerations show for $Y = Q \Sigma O^T \sim \mathrm{Gin}_{\mathbb{R}}(m,n)$ the singular value distributions are given by

$$\mathrm{D}Y \xrightarrow{\mathrm{QR}} 2^{-n} \prod_{j=1}^{n} \Sigma_j^{m-n} |\triangle(\Sigma^2)| \mathrm{D}\Sigma \, \mathrm{D}Q \, \mathrm{D}O$$

where $\mathrm{D}O$ is Haar measure on $\mathsf{U}(n)$, $\mathrm{D}Q$ is Haar measure on $\mathsf{O}(m)$ and $\mathrm{D}\Sigma$ is as before.

In both cases, $\mathrm{Gin}_{\mathbb{R}}(m,n)$ or $\mathrm{Gin}_{\mathbb{C}}(m,n)$, if $m < n$, then same distributional description holds with the addition of $n - m$ point masses at zero for $\Sigma_1, \ldots, \Sigma_{n-m}$ (depending one's ordering convention) to indicate the deficiency of the matrix.

6.2.2 Bidiagonalization of Ginibre

Consider $X \sim \mathrm{Gin}_{\mathbb{C}}(m,n)$ or $Y \sim \mathrm{Gin}_{\mathbb{R}}(m,n)$, $m \geq n$, and consider the sample covariance matrices X^*X/m and Y^TY/m.

Theorem 61. *Let $x_1, \ldots, x_n$ be the unordered eigenvalues of X^*X/m ($\beta = 2$) or Y^TY/m ($\beta = 1$). The following gives their joint marginal distribution*

$$\frac{1}{Z_n(\beta)} \prod_{j=1}^{n} x_j^{\frac{\beta}{2}(m-n)-\frac{1}{2}} \prod_{j<k} |x_j - x_k|^{\beta} \mathrm{e}^{-\frac{\beta m}{4} \sum_{j=1}^{n} x_j} \mathbf{1}_{\{x_j \geq 0, \text{ for all } j\}} \mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{6.2.5}$$

We now consider the reduction of $\mathrm{Gin}_{\mathbb{C}}(m,n)$ and $\mathrm{Gin}_{\mathbb{R}}(m,n)$ to bidiagonal matrices and, in the process, find a generalization of (6.2.5) to general β. This is sometimes called Golub–Kahan bidiagonalization. The aim here is not to preserve eigenvalues, but to preserve singular values as transformations are performed. So, we can perform independent Householder reflections from the left and the right. Recall the definition of $P(Y)$ from (6.1.2). Let $Y \sim \mathrm{Gin}_{\mathbb{R}}(m,n)$ for $m \geq n$. Consider the iterative method

$$\begin{aligned}
Y_0 &= Y, \\
\tilde{Y}_1 &= Q_1 Y_0 := P(Y_0) Y_0, \\
Y_1^T = \tilde{Q}_1 \tilde{Y}_1^T &:= \begin{pmatrix} 1 & 0 \\ 0 & P\left([\tilde{Y}_1^T]_{n-1,m-1}\right) \end{pmatrix} \tilde{Y}_1^T, \\
\tilde{Y}_2 = Q_2 Y_1 &:= \begin{pmatrix} 1 & 0 \\ 0 & P\left([Y_1]_{m-1,n-1}\right) \end{pmatrix} Y_1, \\
&\vdots \\
\tilde{Y}_j &= Q_j Y_{j-1} \begin{pmatrix} I_{j-1} & 0 \\ 0 & P\left([Y_{j-1}]_{m-j+1,n-j+1}\right) \end{pmatrix} Y_{j-1}, \\
Y_j^T = \tilde{Q}_j \tilde{Y}_{j-1}^T &:= \begin{pmatrix} I_j & 0 \\ 0 & P\left([\tilde{Y}_j^T]_{n-j,m-j}\right) \end{pmatrix} \tilde{Y}_{j-1}^T.
\end{aligned} \tag{6.2.6}$$

The algorithm terminates when $j = n-1$, returning Y_{n-1} which is a bidiagonal matrix. Let $(Y_n)_{jj} = c_j$ and $(Y_n)_{j,j+1} = d_j$ for $j = 1, 2, \ldots$. We find that $(Q_j, \tilde{Q}_j, c_j, d_j)_{j \geq 1}$ is an independent set of random variables, with Q_j being defined by $v_j \in S_{\mathbb{R}}^{n-j}$ and $\tilde{Q}_j$ being defined by $\tilde{v}_j \in S_{\mathbb{R}}^{n-j-1}$ (Q_{n-1} gives a sign flip of one entry). Under this change of variables, following the arguments for (3.3.3), we have

$$\mathrm{D}Y \propto \prod_{j=1}^{n} c_j^{m-j} \mathrm{d}c_j \prod_{k=1}^{n-1} d_k^{n-k-1} \mathrm{d}d_k \prod_{l=1}^{n-2} \mathrm{D}\tilde{\omega}_k \prod_{p=1}^{n-1} \mathrm{D}\omega_p,$$

where $\mathrm{D}\tilde{\omega}_l$ and $\mathrm{D}\omega_p$ denote uniform measure on $S_{\mathbb{R}}^l$ and $S_{\mathbb{R}}^p$, respectively. Similarly, by applying this algorithm to $X \sim \mathrm{Gin}_{\mathbb{C}}(m,n)$ we find

$$\mathrm{D}X \propto \prod_{j=1}^{n} c_j^{2(m-j)+1} \mathrm{d}c_j \prod_{k=1}^{n-1} d_k^{2(n-k)-1} \mathrm{d}d_k \prod_{l=1}^{n-2} \mathrm{D}\tilde{\omega}_k \prod_{p=1}^{n-1} \mathrm{D}\omega_p,$$

where $\mathrm{D}\tilde{\omega}_l$ and $\mathrm{D}\omega_p$ denote uniform measure on $S_{\mathbb{C}}^l$ and $S_{\mathbb{C}}^p$, respectively.

6.2.3 Limit theorems

Circular law

We now describe the global eigenvalue distribution for $\mathrm{Gin}_{\mathbb{C}}(n)$ as $n \to \infty$. We have the following distribution on the (unordered) eigenvalues $Z = (z_1, z_2, \ldots, z_n)$

from (6.1.8)

$$\hat{P}^{(n)}(z_1,\dots,z_n)\mathrm{D}z = \frac{1}{Z_n}|\triangle(Z)|^2 \mathrm{e}^{-\frac{1}{2}\sum_{j=1}^n |z_j|^2} \prod_{j=1}^n \mathsf{dRe}\, z_j \mathsf{dIm}\, z_j.$$

Owing to the calculations that result in Theorem 35 we have

$$\begin{aligned}
\hat{P}^{(n)}(z_1,\dots,z_n) &= \frac{1}{n!}\det(\hat{K}_n(z_j,z_k)_{1\le j,k\le n}),\\
\hat{R}^{(n)}_m(z_1,\dots,z_m) &= \det(\hat{K}_n(z_j,z_k)_{1\le j,k\le m}), \quad 1\le m\le n,\\
\hat{K}_n(z,w) &= \sum_{j=0}^{n-1} c_j \Phi_j(z)\overline{\Phi_j(w)}, \quad \Phi_j(z) = c_j z^j \mathrm{e}^{-\frac{1}{4}|z|^2}.
\end{aligned}$$

where $\hat{R}^{(n)}_m$ is the m-point correlation function defined by (4.1.4) with $\hat{P}^{(n)}$ instead of $P^{(n)}$ and $\mathsf{dRe}\, z_j \mathsf{dIm}\, z_j$ instead of $\mathrm{d}x_j$. To show that this is the correct choice for $\hat{K}_n$ and to determine c_j we need to show that $\{\Phi_j\}_{j=0}^{n-1}$ are orthogonal and choose $c_j > 0$ to normalize the functions. Consider for $j < k$

$$\begin{aligned}
\int_{\mathbb{C}} \Phi_j(z)\overline{\Phi_k(z)}\,\mathsf{dRe}z\,\mathsf{dIm}z &= c_j\bar{c}_k \int_{\mathbb{C}} \bar{z}^{k-j}|z|^{2j}\mathrm{e}^{-\frac{1}{2}|z|^2}\mathsf{dRe}z\,\mathsf{dIm}z\\
&= c_j\bar{c}_k \int_0^\infty \left(\int_0^{2\pi} (\cos\theta + \mathrm{i}\sin\theta)^{k-j}\mathrm{d}\theta\right) r^{k+j+1}\mathrm{e}^{-\frac{1}{2}|r|^2}\mathsf{dRe}z\,\mathsf{dIm}z = 0.
\end{aligned}$$

If $j = k$ we find

$$\int_{\mathbb{C}} |\Phi_j(z)|^2\,\mathsf{dRe}z\,\mathsf{dIm}z = |c_j|^2 \int_{\mathbb{C}} |z|^{2j}\mathrm{e}^{-\frac{1}{2}|z|^2}\mathsf{dRe}z\,\mathsf{dIm}z,$$

and using $r = \sqrt{2s}$

$$\begin{aligned}
\int_{\mathbb{C}} |z|^{2j}\mathrm{e}^{-\frac{1}{2}|z|^2}\mathsf{dRe}z\,\mathsf{dIm}z &= 2\pi\int_0^\infty r^{2j+1}\mathrm{e}^{-\frac{1}{2}r^2}\mathrm{d}r = 2^{j+1}\pi\int_0^\infty s^j\mathrm{e}^{-s}\mathrm{d}s\\
&= 2^{j+1}\pi\Gamma(j+1) = 2^{j+1}j!\pi
\end{aligned}$$

so

$$c_j = \frac{1}{2^{j/2+1/2}\sqrt{\pi j!}}, \quad c_j\frac{1}{\sqrt{2}\sqrt{j+1}} = c_{j+1}.$$

So, we find a simple two-term recurrence formula

$$\Phi_{j+1}(z) = \frac{z}{\sqrt{2}\sqrt{j+1}}\Phi_j(z), \quad \Phi_0(z) = \frac{1}{\sqrt{2\pi}}.$$

The corresponding Christoffel-Darboux-type formula is

$$\hat{K}_n(z,w) = \frac{\mathrm{e}^{z\bar{w}/2}}{2\pi}\frac{\Gamma(n, z\bar{w}/2)}{(n-1)!}\mathrm{e}^{-\frac{1}{4}(|z|^2+|w|^2)}.$$

where $\Gamma(n,z)=\int_z^\infty t^{n-1}\mathrm{e}^{-t}\mathrm{d}t$ is the incomplete Gamma function. To see this let $f_n(z)=\mathrm{e}^z\Gamma(n,z)$, and we find

$$f_n^{(j)}(0)=(n-1)!,\quad j=0,1,2,\dots,n-1,$$
$$f_n^{(j)}(0)=0,\quad j\geq n,$$

so that

$$f_n(z)=\sum_{j=0}^{n-1}\frac{(n-1)!}{j!}z^j.$$

Define the rescaled empirical spectral measure

$$\hat{L}_n(\mathrm{D}z)=\frac{1}{n}\sum_{k=1}^n\delta_{\lambda_k/\sqrt{2n}}(\mathrm{D}z),\quad \mathrm{D}z=\mathrm{d}\mathsf{Re}z\,\mathrm{d}\mathsf{Im}z.$$

It then follows that for $f\in C_0(\mathbb{C})$ by (4.1.3)

$$\mathbb{E}\left(\int f(z)\hat{L}_n(\mathrm{D}z)\right)=:\int f(z)\mathbb{E}L_n(\mathrm{D}z)=2\int f(z)\hat{K}_n(z\sqrt{2n},z\sqrt{2n})\mathrm{D}z.$$

We then perform the asymptotic analysis of this density. Consider

$$\Gamma(n,z\bar{z}/4)\overset{z\mapsto z\sqrt{2n}}{\longrightarrow}\int_{n|z|^2}^\infty t^{n-1}\mathrm{e}^{-t}\mathrm{d}t.$$

Then

$$\int_{n|z|^2}^\infty t^{n-1}\mathrm{e}^{-t}\mathrm{d}t=n^n\int_{|z|^2}^\infty t^{n-1}\mathrm{e}^{-nt}\mathrm{d}t=n^n\int_{|z|^2}^\infty t^{-1}\mathrm{e}^{-ng(t)}\mathrm{d}t,$$

where $g(t)=t-\log t$. The stationary phase point here is $t=1$, $g'(1)=0$ and $g''(1)=1$. So, if $|z|\leq 1-\epsilon$, the stationary phase point is in the interval of integration and

$$n^n\int_{|z|^2}^\infty t^{-1}\mathrm{e}^{-ng(t)}\mathrm{d}t=\mathrm{e}^{-n}n^{n-1/2}\sqrt{2\pi}(1+O(n^{-1}))=\mathrm{e}^{-n}n^{n-1}\sqrt{2\pi n}(1+O(n^{-1}))$$
$$=(n-1)!(1+O(n^{-1}))$$

uniformly as $n\to\infty$ by Stirling's approximation. Then for $|z|\geq 1+\epsilon$, by integrating by parts

$$I_n(z):=\int_{|z|^2}^\infty t^{n-1}\mathrm{e}^{-nt}\mathrm{d}t=\frac{1}{n}|z|^{2n-2}\mathrm{e}^{-n|z|^2}+\frac{n-1}{n}\int_{|z|^2}^\infty t^{n-2}\mathrm{e}^{-nt}\mathrm{d}t$$
$$\leq\frac{1}{n}|z|^{2n-2}\mathrm{e}^{-n|z|^2}+\frac{n-1}{n}I_n(z).$$

Therefore

$$I_n(z)\leq|z|^{2n-2}\mathrm{e}^{-n|z|^2}.$$

From these estimates, the following lemma follows.

Lemma 23. *Fix $0 < \epsilon < 1$. As $n \to \infty$, for $|z| \le 1 - \epsilon$*

$$2\hat{K}_n(z\sqrt{2n}, z\sqrt{2n}) = \frac{1}{\pi} + O(n^{-1}),$$

uniformly. As $n \to \infty$, for $|z| \ge 1 + \epsilon$

$$2\hat{K}_n(z\sqrt{2n}, z\sqrt{2n}) = O(n^{-1}),$$

uniformly.

This shows that (see Exercise 4.9)

$$\mathbb{E}L_n(\mathrm{D}z) \to \frac{1}{\pi}\mathbf{1}_{\{|z|\le 1\}}\mathrm{D}z$$

weakly. This is the *averaged circular law*.

Marchenko–Pastur law

Again, consider $X \sim \mathrm{Gin}_{\mathbb{C}}(m, n)$ or $Y \sim \mathrm{Gin}_{\mathbb{R}}(m, n)$, $m \ge n$, and consider the sample covariance matrices X^*X/m and Y^TY/m, and let $x_1, \dots, x_n$ be their unordered eigenvalues. Define the empirical spectral measure

$$L_n(\mathrm{d}x) = \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j}(\mathrm{d}x).$$

Assume further that $n/m \to d \in (0, 1]$. The Marchenko–Pastur law states that

$$\mathbb{E}L_n(\mathrm{d}x) \to p_{\mathrm{MP}}(x; d)\mathrm{d}x := \frac{1}{2\pi d}\sqrt{\frac{|(\lambda_+ - x)(x - \lambda_-)|\mathbf{1}_{[\lambda_-,\lambda_+]}}{x^2}}\mathrm{d}x, \quad \lambda_\pm = (1 \pm \sqrt{d})^2,$$

weakly as $n \to \infty$. If $d = 0$, then the limiting law is $\delta_0(\mathrm{d}x)$, by the Law of Large Numbers.

Chapter 7

Sampling random matrices

Chapter 8

Additional topics

Appendix A

The Airy function

A.1 Integral representation

There are several different conventions for the definition of the Airy function. The standardization adopted here follows [1]. The Airy function, $\mathrm{Ai}(x)$ is defined as the oscillatory integral

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos\left(\frac{t^3}{3} + xt\right)\,\mathrm{d}t = \frac{1}{\pi}\lim_{b\to\infty}\int_0^b \cos\left(\frac{t^3}{3} + xt\right)\,\mathrm{d}t. \qquad \text{(A.1.1)}$$

This is an improper integral, that is, the integral converges conditionally, not absolutely. In order to obtain an absolutely convergent integral, it is necessary to work in the complex plane. Let C denote a contour in the complex plane that starts and ends at the point at infinity, and is asymptotically tangent to the rays $e^{-i\pi/3}$ and $e^{+i\pi/3}$ respectively. Then first setting $t = -iz$ and then deforming the contour, we have

$$\mathrm{Ai}(x) = \frac{1}{2\pi\mathrm{i}}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\left(\frac{z^3}{3} - xz\right)}\,\mathrm{d}z = \frac{1}{2\pi\mathrm{i}}\int_C \mathrm{e}^{\mathrm{i}\left(\frac{z^3}{3} - xz\right)}\,\mathrm{d}z. \qquad \text{(A.1.2)}$$

The integral is absolutely convergent for every $x \in \mathbb{C}$ on the contour C. Indeed, with $z = re^{i\theta}$,

$$\left|\mathrm{e}^{\mathrm{i}\left(\frac{z^3}{3} - xz\right)}\right| \le \mathrm{e}^{|x|r}\mathrm{e}^{-r^3\cos(3\theta)/3} \sim \mathrm{e}^{-r^3/3}\mathrm{e}^{r|x|} \qquad \text{(A.1.3)}$$

as $z \to \infty$ along the rays $\theta = \pm\pi/3$. Thus, $\mathrm{Ai}(x)$ is an entire function.

A.2 Differential equation

We differentiate under the integral sign (justified by (A.1.3)) and integrate by parts to obtain

$$\begin{aligned}\mathrm{Ai}''(x) &= \frac{1}{2\pi i}\int_C z^2 e^{\left(\frac{z^3}{3}-xz\right)}\,dz \\ &= \frac{1}{2\pi i}\int_C \frac{d}{dz}e^{\frac{z^3}{3}}\,e^{-xz}\,dz = -\frac{1}{2\pi i}\int_C e^{\frac{z^3}{3}}\,\frac{d}{dz}e^{-xz}\,dz = x\mathrm{Ai}(x).\end{aligned} \tag{A.2.1}$$

Thus, $\mathrm{Ai}(x)$ satisfies the *Airy differential equation*

$$y'' = xy, \quad x \in \mathbb{C}. \tag{A.2.2}$$

This differential equation has a scaling invariance: if $y(x)$ is a solution, so are $y(\omega x)$ and $y(\omega^2 x)$ where $\omega = e^{2\pi i/3}$ is a cube root of unity. Thus, both $\mathrm{Ai}(\omega x)$ and $\mathrm{Ai}(\omega^2 x)$ solve (A.2.2). Each of these solutions is linearly independent of $\mathrm{Ai}(x)$. A solution to (A.2.2) that is real when x is real, and is linearly independent from $\mathrm{Ai}(x)$, is obtained from the linear combination

$$\mathrm{Bi}(x) = e^{\pi i/6}\mathrm{Ai}(\omega x) + e^{\pi - i/6}\mathrm{Ai}(\omega^2 x). \tag{A.2.3}$$

A.3 Asymptotics

The functions $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$ have the following asymptotic properties.

Asymptotics as $x \to \infty$.

$$\zeta = \frac{2}{3}x^{\frac{3}{2}}, \quad \mathrm{Ai}(x) \sim \frac{e^{-\zeta}}{2x^{\frac{1}{4}}\sqrt{\pi}}, \quad \mathrm{Bi}(x) \sim \frac{x^{\frac{1}{4}}}{\sqrt{\pi}}e^{\zeta}. \tag{A.3.1}$$

Asymptotics as $x \to -\infty$.

$$\zeta = \frac{2}{3}(-x)^{\frac{3}{2}}, \quad \mathrm{Ai}(x) \sim \frac{1}{x^{\frac{1}{4}}\sqrt{\pi}}\sin\left(\zeta + \frac{\pi}{4}\right), \quad \mathrm{Bi}(x) \sim \frac{1}{x^{\frac{1}{4}}\sqrt{\pi}}\cos\left(\zeta + \frac{\pi}{4}\right). \tag{A.3.2}$$

Appendix B

Hermite polynomials

In this chapter, μ denotes the weight function

$$\mu(\mathrm{d}x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\, \mathrm{d}x. \tag{B.0.1}$$

The (probablilists') Hermite polynomials $\{\mathfrak{h}_k\}_{k=0}^{\infty}$ are the monic family of polynomials of degree k orthogonal with respect to the weight μ.

B.1 Basic formulas

$$\mathfrak{h}_k(x) = \mathrm{e}^{\frac{x^2}{2}} \left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^k e^{\frac{-x^2}{2}}. \tag{B.1.1}$$

$$\mathfrak{h}_k(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-\mathrm{i}\xi)^k \mathrm{e}^{-\frac{1}{2}(\xi - \mathrm{i}x)^2}\, \mathrm{d}\xi. \tag{B.1.2}$$

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathfrak{h}_k(x)\mathfrak{h}_l(x) e^{-\frac{x^2}{2}}\, dx = \sqrt{2\pi} k! \delta_{kl}. \tag{B.1.3}$$

$$x\mathfrak{h}_k(x) = \mathfrak{h}_{k+1}(x) + k\mathfrak{h}_{k-1}(x), \quad k \geq 1. \tag{B.1.4}$$

$$\mathfrak{h}_k'(x) = k\mathfrak{h}_{k-1}(x). \tag{B.1.5}$$

$$\mathfrak{h}_k''(x) - x\mathfrak{h}_k'(x) + k\mathfrak{h}_k(x) = 0. \tag{B.1.6}$$

$$\sum_{j=0}^{k-1} \frac{1}{j!} \mathfrak{h}_j(x)\mathfrak{h}_j(y) = \frac{(\mathfrak{h}_k(x)\mathfrak{h}_{k-1}(y) - \mathfrak{h}_{k-1}(x)\mathfrak{h}_k(y))}{(k-1)!(x-y)}, \quad x \neq y. \tag{B.1.7}$$

Relation (B.1.1) may be treated as an alternate definition of the Hermite polynomials. On the other hand, since we have defined the Hermite polynomials as the monic orthogonal polynomials obtained by applying the Gram-Schmidt procedure to the set $\{1, x, x^2, \ldots\}$ in $L^2(\mathbb{R}, \mu)$, equation (B.1.1) may be verified as follows. First, it is clear from (B.1.1) that $\mathfrak{h}_k(x)$ is a monic polynomial of degree k and that $\mathfrak{h}_0(x) = 1$, $\mathfrak{h}_1(x) = x$. By induction, if it has been established that property (B.1.1) defines the Hermite polynomials for $j \leq k-1$, then it is only necessary to show that the monic polynomial

$$P_k(x) = \mathrm{e}^{\frac{x^2}{2}} \left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \mathrm{e}^{-\frac{x^2}{2}},$$

is the same as $\mathfrak{h}_k$. The polynomial P_k is orthogonal to $\mathfrak{h}_j$, $0 \leq j \leq k-1$ because, using integration by parts,

$$\int_{\mathbb{R}} P_k(x)\mathfrak{h}_j(x)\mu(\mathrm{d}x) = \int_{\mathbb{R}} \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \mathfrak{h}_j(x)\mu(\mathrm{d}x) = 0,$$

since H_j has degree less than k. Since P_k is monic, it must be $\mathfrak{h}_k$. The same calculation serves to establish (B.1.3).

The integral representation (B.1.2) follows from the formula for the Fourier transform of a Gaussian

$$\mathrm{e}^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi x} \mathrm{e}^{-\frac{\xi^2}{2}}\, \mathrm{d}\xi, \tag{B.1.8}$$

and the identity (B.1.1).

The two-term recurrence relation follows from (3.4.18) and (B.1.3) (see also Remark 29). The coefficient a_k vanishes because equation (B.1.1) shows that $\mathfrak{h}_k^2$ is an even polynomial for all k. The coefficient b_k^2 may be rewritten

$$b_k^2 = \frac{\int x\mathfrak{h}_{k-1}(x)\mathfrak{h}_k(x)\mu(\mathrm{d}x)}{\int \mathfrak{h}_{k-1}^2\mu(\mathrm{d}x)} = \frac{\int x\mathfrak{h}_{k-1}(x)\mathfrak{h}_k(x)\mu(\mathrm{d}x)}{\int \mathfrak{h}_k^2\mu(\mathrm{d}x)} \frac{\int \mathfrak{h}_k^2(x)\mu(\mathrm{d}x)}{\int \mathfrak{h}_{k-1}^2\mu(\mathrm{d}x)} = 1 \cdot k^2, \tag{B.1.9}$$

by (B.1.3).

The differential equation (B.1.5) is obtained by rewriting (B.1.1) in the form

$$\mathrm{e}^{-\frac{x^2}{2}}\mathfrak{h}_k(x) = (-1)^k \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \mathrm{e}^{-\frac{x^2}{2}},$$

differentiating both sides, and then multiplying by $\mathrm{e}^{\frac{x^2}{2}}$. Equation (B.1.6) is obtained by differentiating (B.1.5) and using (B.1.4). The proof of the Christoffel-Darboux identity is left as an exercise to the reader.

B.2 Hermite wave functions

The Hermite wave functions $\{\psi\}_{k=0}^{\infty}$ are defined by

$$\psi_k(x) = \frac{1}{\sqrt{k!}} \frac{\mathrm{e}^{-x^2/4}}{(2\pi)^{1/4}} \mathfrak{h}_k(x), \quad k = 0, 1, 2, \ldots \tag{B.2.1}$$

The following properties of the Hermite wave-functions follow immediately from the corresponding properties of the Hermite polynomials.

$$\int_{\mathbb{R}} \psi_k(x)\psi_l(x)\,\mathrm{d}x = \delta_{kl}. \tag{B.2.2}$$

$$x\psi_k(x) = \sqrt{k+1}\psi_{k+1}(x) + \sqrt{k}\psi_{k-1}(x). \tag{B.2.3}$$

$$\psi_k'(x) = -\frac{x}{2}\psi_k(x) + \sqrt{k}\psi_{k-1}(x). \tag{B.2.4}$$

$$\psi_k''(x) + \left(k + \frac{1}{2} - \frac{x^2}{4}\right)\psi_k(x) = 0. \tag{B.2.5}$$

$$\sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n}\frac{(\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(x))}{x-y}. \tag{B.2.6}$$

B.3 Small x asymptotics

The following classical formulas capture the asymptotics of the Hermite polynomials near the origin [1, §22.15].

$$\lim_{n\to\infty} \frac{(-1)^n}{2^n}\frac{\sqrt{n}}{n!}\,\mathfrak{h}_{2n}\left(\frac{x}{\sqrt{2n}}\right) = \frac{1}{\sqrt{\pi}}\cos x. \tag{B.3.1}$$

$$\lim_{n\to\infty} \frac{(-1)^n}{2^n\,n!}\,\mathfrak{h}_{2n+1}\left(\frac{x}{\sqrt{2n}}\right) = \sqrt{\frac{2}{\pi}}\,\sin x. \tag{B.3.2}$$

Further, the convergence to the limit is uniform over x in a bounded interval.

In comparing equations (B.3.1) and (B.3.2) with a standard reference such as [1], the reader should note that there are two conventions in the definition of Hermite polynomials. The exponential weight in earlier sources was chosen to be e^{-x^2}, which differs from our choice (B.0.1). The relation between the Hermite polynomials, $\{H_n(x)\}$ in [1], and those used here are:

$$H_n(x) = 2^{\frac{n}{2}}\,\mathfrak{h}_n(x\sqrt{2}), \quad \mathfrak{h}_n(x) = 2^{-\frac{n}{2}}\,H_n\left(\frac{x}{\sqrt{2}}\right). \tag{B.3.3}$$

These formulas may be immediately translated into asymptotic formulas for the Hermite wave functions, using Stirling's approximation for the factorial:

$$n! = \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n (1 + O(n^{-1})) \text{ as } n \to \infty. \tag{B.3.4}$$

$$\lim_{n\to\infty} (2n)^{1/4}(-1)^n\,\psi_{2n}\left(\frac{x}{\sqrt{2n}}\right) = \frac{1}{\sqrt{\pi}}\cos x. \tag{B.3.5}$$

$$\lim_{n\to\infty} (2n)^{1/4}(-1)^n\psi_{2n+1}\left(\frac{x}{\sqrt{2n}}\right) = \frac{1}{\sqrt{\pi}}\,\sin x. \tag{B.3.6}$$

The asymptotic formulas (B.3.1) and (B.3.2) are proved by applying Laplace's method to the integral formula (B.1.2). We only explain how to prove (B.3.1) since equation (B.3.2) is similar. Since $(\mathrm{i})^{2n} = (-1)^n$, we take the real part of (B.1.2) to find

$$\begin{aligned}(-1)^{2n}\mathfrak{h}_{2n}\left(\frac{x}{\sqrt{2n}}\right) &= \sqrt{\frac{2}{\pi}}\,\mathrm{e}^{\frac{x^2}{4n}}\int_0^\infty \xi^{2n}\mathrm{e}^{-\frac{\xi^2}{2}}\cos\left(\frac{x\xi}{\sqrt{2n}}\right)\,\mathrm{d}\xi \\ &= \frac{2^{n+1}n^{n+\frac{1}{2}}}{\sqrt{\pi}}\int_0^\infty \mathrm{e}^{-n\left(t^2-2\log t\right)}\,\cos xt\,\mathrm{d}t,\end{aligned} \tag{B.3.7}$$

by rescaling $\xi = \sqrt{n}\,t$. We now apply Laplace's method to the integral above. The function $g(t) = t^2 - 2\log t$ has a single minimum on the interval $(0,\infty)$ at $t = 1$. At this point

$$g(1) = 1, \quad g'(1) = 0, \quad g''(1) = 4. \tag{B.3.8}$$

Laplace's approximation now yields

$$\int_0^\infty \mathrm{e}^{-ng(t)}\cos xt\,\mathrm{d}x \sim \mathrm{e}^{-n}\sqrt{\frac{\pi}{2n}}\cos x, \tag{B.3.9}$$

which when combined with (B.3.7) implies

$$(-1)^{2n}\mathfrak{h}_{2n}\left(\frac{x}{\sqrt{2n}}\right) \sim 2^{n+\frac{1}{2}}n^n\mathrm{e}^{-n}\cos x. \tag{B.3.10}$$

Equation (B.3.10) is equivalent to (B.3.1) by Stirling's approximation (B.3.4). Further, it is easy to check that the error is uniformly small for x in a bounded set.

B.4 Steepest descent for integrals

Consider the integral

$$\int_\Gamma f(t)\mathrm{e}^{-n\Phi(t)}\mathrm{d}t \tag{B.4.1}$$

where f and Φ are entire functions. Assume $\Phi(t^*) = 0$, $\Phi'(t^*) = 0$, $\Phi''(t^*) \neq 0$, $\mathsf{Im}\Phi(t) = 0$ for $t \in \Gamma$. Further assume Γ is the path of steepest ascent for Φ, i.e. the path of steepest descent for $-\Phi(t)$. Having $\mathsf{Im}\Phi(t) = 0$ is enough to

ensure that Γ is either the path of steepest ascent (locally) or steepest descent: Let $t = x(s) + \mathrm{i}y(s)$ be a smooth local parameterization of Γ, then by the Cauchy–Riemann equations

$$0 = \frac{\mathrm{d}}{\mathrm{d}s}\mathsf{Im}\,\Phi(t) = \mathsf{Im}\,\Phi_x(t)\frac{\mathrm{d}x}{\mathrm{d}s} + \mathsf{Im}\,\Phi_y(t)\frac{\mathrm{d}y}{\mathrm{d}s} = -\mathsf{Re}\,\Phi_x(t)\frac{\mathrm{d}x}{\mathrm{d}s} + \mathsf{Re}\,\Phi_x(t)\frac{\mathrm{d}y}{\mathrm{d}s}.$$

This shows that $\nabla\mathsf{Re}\,\Phi$ is orthogonal to the tangent vector $(-y'(s), x'(s))$, implying that Γ is in the direction of greatest increase/decrease for $\mathsf{Re}\,\Phi$.

Performing a Taylor expansion, we have

$$\Phi(t) = \frac{\Phi''(t^*)}{2}(t-t^*)^2(1+O(|t-t^*|). \tag{B.4.2}$$

The point is that Φ is locally quadratic at t^* and we use this to inform the change of variables. But if we naïvely looked to solve

$$\Phi(t^*+v) = s^2,$$

for v as a function of s, $v(0)=0$, we would fail. The implicit function theorem fails because we have two solution branches! Instead we consider

$$\frac{\Phi(t^*+sv)}{s^2} - 1 = 0 = \frac{\Phi''(t^*)}{2}v^2 - 1 + O(|sv^3|). \tag{B.4.3}$$

We can choose $v = \pm R^{-1/2}\mathrm{e}^{-\mathrm{i}\phi/2}$ where $\frac{\Phi''(t^*)}{2} = R\mathrm{e}^{\mathrm{i}\phi}$. For either choice, we can apply the implicit function theorem (the derivative with respect to v, evaluated at $(s,v) = (0, \pm R^{-1/2}\mathrm{e}^{-\mathrm{i}\phi/2})$ does not vanish). We use $v = \pm R^{-1/2}\mathrm{e}^{-\mathrm{i}\phi/2}$ to obtain $v(s)$, and our local parameterization of Γ: $t(s) = t^* + sv(s)$. We use this a change of variables, within a neighborhood $B(t^*,\epsilon)$ on which the implicit function theorem applies (here we assume the orientation of Γ is the same as the induced orientation on $t((-\delta_1,\delta_2))$)

$$\int_{\Gamma\setminus B(t^*,\epsilon)} f(t)\mathrm{e}^{-n\Phi(t)}\mathrm{d}t = \int_{-\delta_1}^{\delta_2} f(t^*+sv(s))\mathrm{e}^{-ns^2}\frac{\mathrm{d}s}{v(s)+sv'(s)}, \quad \delta_1,\delta_2 > 0. \tag{B.4.4}$$

Now let $\delta = \min\{\delta_1,\delta_2\}$. It follows that on $\Gamma_\delta = \Gamma \setminus t(-\delta,\delta)$, $\Phi(t) \geq \delta^2$. Then

$$\left|\int_{\Gamma_\delta} f(t)\mathrm{e}^{-n\Phi(t)}\mathrm{d}t\right| \leq \mathrm{e}^{-n\delta^2}\int_{\Gamma_\delta}|f(t)|\mathrm{e}^{-n(\Phi(t)-\delta^2)}\mathrm{d}t. \tag{B.4.5}$$

For $n \geq 1$, we have

$$\int_{\Gamma_\delta}|f(t)|\mathrm{e}^{-n(\Phi(t)-\delta^2)}\mathrm{d}t \leq \int_{\Gamma_\delta}|f(t)|\mathrm{e}^{-(\Phi(t)-\delta^2)}\mathrm{d}t := M. \tag{B.4.6}$$

And therefore (B.4.5) is exponentially small in n, less than $M\mathrm{e}^{-n\delta^2}$. Now, consider

$$\int_{-\delta}^{\delta} f(t^*+sv(s))\mathrm{e}^{-ns^2}\left(v(s)+sv'(s)\right)\mathrm{d}s \tag{B.4.7}$$

and we can directly apply Laplace's method. Taylor expand the function

$$f(t^* + sv(s))(v(s) + sv'(s))$$

at $s = 0$, and term by term integration gives an expansion in powers of $n^{-1/2}$ with the leading order term being

$$\begin{aligned}\int_\Gamma f(t)e^{-n\Phi(t)}\mathrm{d}t &\\ &= \int_{-\delta}^{\delta} f(t^* + sv(s))\mathrm{e}^{-ns^2}(v(s) + sv'(s))\mathrm{d}s + O(n^{-\alpha}) \\ &= \int_{-\delta}^{\delta} f(t^*)v(0)(1 + O(s))\mathrm{e}^{-ns^2}\mathrm{d}s + O(n^{-\alpha}), \quad \text{for all } \alpha > 0. \quad \text{(B.4.8)}\end{aligned}$$

Performing a change of variables $s = y/\sqrt{2n}$ we have

$$\int_{-\delta}^{\delta} \mathrm{e}^{-ns^2}\mathrm{d}s = \int_{-\sqrt{2n}\delta}^{\sqrt{2n}\delta} \mathrm{e}^{-y^2/2}\mathrm{d}y = \sqrt{\frac{\pi}{n}} + O(n^{-\alpha}), \quad \text{for all } \alpha > 0,$$

$$\int_{-\delta}^{\delta} |s|\mathrm{e}^{-ns^2}\mathrm{d}s = \frac{1}{\sqrt{2n}} \int_{-\sqrt{2n}\delta}^{\sqrt{2n}\delta} |y|\mathrm{e}^{-y^2/2}\mathrm{d}y = \frac{C}{n} + O(n^{-\alpha}), \quad \text{for all } \alpha > 0.$$

So, we have

$$\int_\Gamma f(t)e^{-n\Phi(t)}\mathrm{d}t = \sqrt{\frac{2\pi}{n}} f(t^*)|\Phi''(t^*)|^{-1/2}\mathrm{e}^{-\mathrm{i}\phi/2} + O(n^{-1}) \text{ as } n \to \infty. \quad \text{(B.4.9)}$$

B.5 Plancherel–Rotach asymptotics

Another asymptotic regime is obtained when we consider $x = O(\sqrt{n})$ and let $n \to \infty$. Plancharel–Rotach asymptotics refer to the asymptotics of polynomials scaled by their largest zero. The limit is oscillatory or exponential depending on the range of x. This is to be expected: for each n, the polynomial $\mathfrak{h}_n(x)$, and thus the wave function $\psi_n(x)$, has n zeros. The largest and smallest of the zeros are approximately $\pm\sqrt{(n + 1/2)}$. The oscillatory regime is obtained when $x(n + 1/2)^{-1/2}$ lies well within the interval $(-1, 1)$. Outside this interval, the Hermite wave function decays exponentially fast. A more delicate calculation, using the Airy function, is required to understand the transition from oscillatory to exponential behavior.

We will prove a weaker version of the Plancherel-Rotach formulas, that suffices for our needs. These formula are as follows.

Case 1. Oscillatory behavior.

$$x = 2\cos\varphi, \qquad 0 < \varphi < \pi. \quad \text{(B.5.1)}$$

$$n^{\frac{1}{4}}\psi_{n+p}\left(x\sqrt{n}\right) \sim \frac{1}{\sqrt{\pi\sin\varphi}} \cos\left[n\left(\varphi - \frac{1}{2}\sin 2\varphi\right) + \left(p + \frac{1}{2}\right)\varphi - \frac{\pi}{4}\right]. \quad \text{(B.5.2)}$$

The convergence is uniform for φ in a closed subset of $(0, \pi)$.

Case 2. Exponential decay.

$$|x| = 2\cosh\varphi, \qquad 0 < \varphi. \tag{B.5.3}$$

$$n^{\frac{1}{4}}\psi_{n+p}\left(x\sqrt{n}\right) \sim \frac{\mathrm{e}^{(p+\frac{1}{2})\varphi}}{\sqrt{2\pi\sinh\varphi}}\mathrm{e}^{-\frac{n}{2}\left(\mathrm{e}^{2\varphi}+1-2\varphi\right)}. \tag{B.5.4}$$

The convergence is uniform for φ in a compact subset of $(0,\infty)$. Observe that $\mathrm{e}^{2\varphi} - 2\varphi - 1 > 0$ when $\varphi > 0$, ensuring exponential decay.

Case 3. The transition region.

$$x = 2\sqrt{n} + \frac{s}{n^{\frac{1}{6}}} \quad s \in \mathbb{C}, \tag{B.5.5}$$

$$n^{1/12-p/2}\psi_n(x\sqrt{n}) \tag{B.5.6}$$

$$= \sqrt{\frac{n!}{(n+p)!}}\left(\mathrm{Ai}(s) + n^{-1/3}\left(\frac{1}{2} - p\right)\mathrm{Ai}'(s) + O(n^{-2/3})\right). \tag{B.5.7}$$

The convergence is uniform for s in a compact subset of $\mathbb{C}$.

All three asymptotic relations are obtained by the method of steepest descent for integrals. Assume $x \in \mathbb{R}$. We fix an integer p, use the integral identity (B.1.2) with $k = n + p$, and rescale $\xi = nt$ to obtain

$$\mathfrak{h}_{n+p}\left(x\sqrt{n}\right) = \left(-\mathrm{i}\sqrt{n}\right)^{n+p}\sqrt{\frac{n}{2\pi}}\int_{-\infty}^{\infty} t^{n+p}\mathrm{e}^{-\frac{n}{2}(t-\mathrm{i}x)^2}\,dt \tag{B.5.8}$$

$$= \left(-\mathrm{i}\sqrt{n}\right)^{n+p}\sqrt{\frac{n}{2\pi}}\left(\int_0^{\infty} t^{n+p}\mathrm{e}^{-\frac{n}{2}(t-\mathrm{i}x)^2}\,dt + (-1)^{n+p}\int_0^{\infty} t^{n+p}\mathrm{e}^{-\frac{n}{2}(t+\mathrm{i}x)^2}\,dt\right)$$

$$:= \left(-\mathrm{i}\sqrt{n}\right)^{n+p}\sqrt{\frac{n}{2\pi}}\left(I_{n,p}(x) + (-1)^{n+p}I_{n,p}(-x)\right). \tag{B.5.9}$$

The integral $I_{n,p}(x)$ may be rewritten in the form

$$I_{n,p}(x) = \int_0^{\infty} t^p\mathrm{e}^{-ng(t)}\,\mathrm{d}t, \quad g(t) = \frac{1}{2}(t - \mathrm{i}x)^2 - \log t. \tag{B.5.10}$$

As is usual, the first step is to determine the critical points where $g'(t) = 0$. This reduces to the quadratic equation $t^2 - \mathrm{i}xt - 1 = 0$. The three distinct asymptotic limits arise from the three distinct possibilities for the roots.

(a) $|x| < 2$. The function g has two critical points on the unit circle, given by

$$t_{\pm} = \frac{\mathrm{i}x \pm \sqrt{4 - x^2}}{2} = \mathrm{i}\mathrm{e}^{\mp\mathrm{i}\varphi}, \tag{B.5.11}$$

where x and φ are related through (B.5.1).

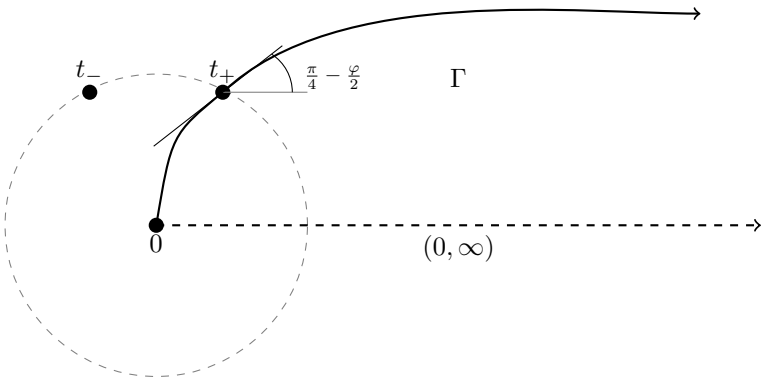

Figure B.5.1:

(b) $|x| > 2$. The two critical points lie on the imaginary axis, and may be written in the form

$$t_\pm = \mathrm{i}\left(\frac{x \pm \sqrt{x^2-4}}{2}\right) = \mathrm{i}\,\mathrm{sgn}(x)\mathrm{e}^{\pm\varphi}, \tag{B.5.12}$$

where each branch of φ is defined through the relation (B.5.3).

(c) $|x| = 2$. The two critical points coalesce into a single value $t = \mathrm{i}$. A further blow-up is necessary to obtain the Airy asymptotics (B.5.6).

Let us first consider the integral $I_{n,p}(x)$ in case (a), and assume that $x > 0$ to be concrete. We deform the integral over $(0, \infty)$ a contour Γ which is the path of steepest descent that passes through the critical point t_+ as shown in Figure B.5.1. The existence of such a contour may be deduced by continuity, beginning with the observation that when $x = 0$, Γ is simply the segment $(0, \infty)$ along the real line. While in general, Γ is given by the equation $\mathsf{Im}(g(t)) = \mathsf{Im}(g(t_+))$. It is not important for us to solve for the contour explicitly: all that is required is to understand the phase of $g''(t_+)$, check that $0 \in \Gamma$ and the integral over $(0, \infty)$ can be deformed to an integral over Γ.

It is easy to check that when $|x| < 2$

$$g''(t_+) = 1 + \frac{1}{t_+^2} = 1 - e^{2\mathrm{i}\varphi} = \left(-\mathrm{i}\mathrm{e}^{\mathrm{i}\varphi}\right)(2\sin\varphi)\,. \tag{B.5.13}$$

Thus, we have

$$\begin{aligned} I_{n,p}(x) &= \int_0^\infty t^p \mathrm{e}^{-ng(t)}\,\mathrm{d}t = \mathrm{e}^{-ng(t_+)}\int_\Gamma t^p \mathrm{e}^{-n(g(t)-g(t_+))}\,\mathrm{d}t \\ &= \mathrm{e}^{-ng(t_+)} t_+^p \left.\frac{\mathrm{d}t}{\mathrm{d}s}\right|_{t_+} \int_{-\infty}^\infty \mathrm{e}^{-\frac{n}{2}|g''(t_+)|s^2}\,\mathrm{d}s + O(n^{-1}). \end{aligned} \tag{B.5.14}$$

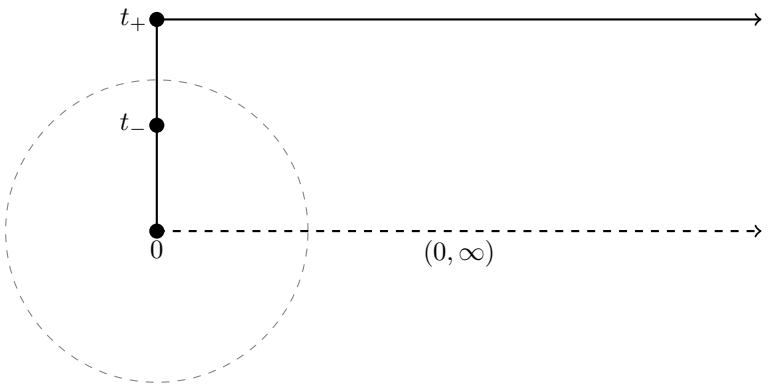

Figure B.5.2:

In the second line, we have used the fact that $\mathsf{Im}(g(t)-g(t_+))=0$ on Γ, and we have further approximated the integral over Γ by an integral over the tangent to Γ at t_+. More precisely, the approximation here is

$$g''(t_+)(t-t_+)^2 = |g''(t_+)|s^2,$$

which implies

$$\left.\frac{\mathrm{d}t}{\mathrm{d}s}\right|_{t_+} = \mathrm{e}^{\mathrm{i}\left(\frac{\pi}{4}-\frac{\varphi}{2}\right)}. \tag{B.5.15}$$

We now combine the values

$$t_+ = \mathrm{i}\mathrm{e}^{-\mathrm{i}\varphi}, \quad g(t_+) = -\frac{\mathrm{e}^{2\mathrm{i}\varphi}}{2} + \mathrm{i}\left(\varphi - \frac{\pi}{2}\right),$$

with (B.5.14) and (B.5.15) to obtain

$$I_{n,p}(x) \sim \mathrm{e}^{\frac{n}{2}\cos 2\varphi}\sqrt{\frac{\pi}{n\sin\varphi}}\mathrm{e}^{\mathrm{i}\left(\frac{n}{2}\sin 2\varphi+(n+p+\frac{1}{2})\left(\frac{\pi}{2}-\varphi\right)\right)}. \tag{B.5.16}$$

Finally, since x is real, we have $\overline{I_{n,p}(x)} = I_{n,p}(-x)$. We combine (B.5.9) with (B.5.16) to obtain

$$\mathfrak{h}_{n+p}(x\sqrt{n}) \sim n^{\frac{n+p}{2}}\sqrt{\frac{2}{\sin\varphi}}\mathrm{e}^{\frac{n}{2}\cos 2\varphi}\cos\left[n\left(\varphi-\frac{1}{2}\sin 2\varphi\right)+\left(p+\frac{1}{2}\right)\varphi-\frac{\pi}{4}\right], \tag{B.5.17}$$

where x and φ are related via (B.5.1). We now use (B.2.1) and Stirling's approximation (B.3.4) to obtain (B.5.2).

The asymptotics in case (b) are obtained as follows. We make the change of variables (B.5.3), and deform the domain of integration for $I_{n,p}$ to the contour consisting of two straight lines shown in Figure B.5.2. So see that this is enough, we have that $g''(t_+) > 0$ while $g''(t_-) < 0$. So the path of steepest ascent

(descent for $-g$) through t_- is the imaginary axis. Then, the steepest ascent path through t_+ makes a right angle with the imaginary axis. And one can check that the real part of $g(t)$ is strictly increasing along the contour $t_+ + \mathrm{i}[0,\infty)$. The dominant contribution comes from t_-. The remaining calculations are left to the reader. The final asymptotic relation is

$$\mathfrak{h}_{n+p}(x\sqrt{n}) = n^{\frac{n+p}{2}} \frac{\mathrm{e}^{-\frac{n}{2}}}{\sqrt{\sinh\varphi}} \mathrm{e}^{\left(p+\frac{1}{2}\right)\varphi - \frac{n}{2}(\sinh(2\varphi)-2\varphi)}(1+o(1)), \tag{B.5.18}$$

which combines with (B.2.1) and Stirling's approximation (B.3.4) to yield (B.5.4).

We now turn to case (c). We begin with the integral representation(B.5.8) and substitute

$$t = \mathrm{i} + \frac{r}{n^{\frac{1}{3}}}, \quad x = 2\sqrt{n} + \frac{s}{n^{\frac{1}{6}}}, \tag{B.5.19}$$

moving the integral over $\mathbb{R}$ to an integral over the line $\mathrm{i}+\mathbb{R}$, to obtain

$$\mathfrak{h}_n(x\sqrt{n}) = (-\mathrm{i}\sqrt{n})^n \frac{n^{\frac{1}{6}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{nh(r)}\,\mathrm{d}r, \tag{B.5.20}$$

where

$$\begin{aligned} h(r) &= \log\left(\mathrm{i} + \frac{r}{n^{\frac{1}{3}}}\right) - \frac{1}{2}\left(\left(\mathrm{i} + \frac{r}{n^{\frac{1}{3}}}\right) - \mathrm{i}\left(2 + \frac{s}{n^{\frac{2}{3}}}\right)\right)^2 \\ &= \frac{1}{2} + \log\mathrm{i} + \frac{s}{n^{\frac{2}{3}}} + \frac{1}{n}\left(\mathrm{i}sr + \frac{\mathrm{i}}{3}r^3\right) + \frac{s^2}{2n^{\frac{4}{3}}} + O(n^{-\frac{4}{3}}r^4), \end{aligned} \tag{B.5.21}$$

using the Taylor expansion of the logarithm. The terms that depend on s may be pulled out of the integral and we are left with

$$\begin{aligned} \mathfrak{h}_n(x\sqrt{n}) &\approx \frac{n^{\frac{n}{2}+\frac{1}{6}}\mathrm{e}^{\frac{n}{2}}\mathrm{e}^{sn^{\frac{1}{3}}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\mathrm{i}sr + \frac{\mathrm{i}}{3}r^3}\mathrm{d}r \\ \mathfrak{h}_n(x) &= \sqrt{2\pi}n^{\frac{n}{2}+\frac{1}{6}}\mathrm{e}^{\frac{n}{2}}\mathrm{e}^{sn^{\frac{1}{3}}}\left(\mathrm{Ai}(s) + O(n^{-1/3})\right) \end{aligned} \tag{B.5.22}$$

To make this rigorous, and to obtain the next term in the expansion, We take the integral

$$\mathfrak{h}_{n+p}(x) = \sqrt{\frac{n}{2\pi}}(-\mathrm{i}\sqrt{n})^{n+p} \int_{\mathbb{R}} t^p \mathrm{e}^{-n(\frac{1}{2}(t-\mathrm{i}x)^2 - \log t)}\mathrm{d}t \tag{B.5.23}$$

and deform to $\mathrm{i}+\mathbb{R}$. Then, let $t = \mathrm{i} + r$ and we arrive at

$$\mathfrak{h}_n(x) = \sqrt{\frac{n}{2\pi}}(-\mathrm{i}\sqrt{n})^{n+p} \int_{\mathbb{R}} (\mathrm{i}+r)^p \mathrm{e}^{-n(\frac{1}{2}(r+\mathrm{i}(1-x))^2 - \log(\mathrm{i}+r))}\mathrm{d}r. \tag{B.5.24}$$

Then this can be deformed to a contour $\Gamma = \mathrm{e}^{-\mathrm{i}\pi/6}(-\infty,0] \cup \mathrm{e}^{\mathrm{i}\pi/6}[0,\infty)$.

Then, we perform a Taylor expansion of the logarithm to find, for $H(r) = \frac{1}{2}(r + \mathrm{i}(1-x))^2 - \log(\mathrm{i}+r)$ and $x > 1$,

$$\mathrm{e}^{-\frac{n}{2}(1-x)^2} \int_\Gamma (\mathrm{i}+r)^p \mathrm{e}^{-nH(r)} \mathrm{d}r \tag{B.5.25}$$

$$+ O(\mathrm{e}^{-n(x-1)\delta}) \tag{B.5.26}$$

$$= \mathrm{e}^{\mathrm{i}n\frac{\pi}{2}} \int_{\Gamma \cap B(0,\delta)} \mathrm{e}^{\mathrm{i}n(x-2)r + \mathrm{i}n\frac{r^3}{3}} \left(\mathrm{i}^p + rp\mathrm{i}^{p-1} + O(r^2) - n\mathrm{i}^p \frac{r^4}{4} + nO(r^5) \right) \mathrm{d}r \tag{B.5.27}$$

We compute

$$\int_0^\delta \mathrm{e}^{-n\frac{y^3}{3}} y^\alpha \mathrm{d}y = O(n^{-(\alpha+1)/3}),$$

so that

$$\mathrm{e}^{-\frac{n}{2}(1-x)^2} \int_\Gamma (\mathrm{i}+r)^p \mathrm{e}^{-nH(r)} \mathrm{d}r = \mathrm{e}^{\mathrm{i}n\frac{\pi}{2}} \int_{\Gamma \cap B(0,\delta)} \left(\mathrm{i}^p + p\mathrm{i}^{p-1} r - n\mathrm{i}^p \frac{r^4}{4} \right) \mathrm{e}^{\mathrm{i}n(x-2)r + \mathrm{i}n\frac{r^3}{3}} \mathrm{d}r + O(n^{-1}).$$

Finally, it follows that if $x = 2 + sn^{-2/3}$ and setting $r = k/n^{-1/3}$

$$\int_{\Gamma \cap B(0,\delta)} \mathrm{e}^{\mathrm{i}n(x-2)r + \mathrm{i}n\frac{r^3}{3}} r^\gamma \mathrm{d}r = 2\pi n^{-(\gamma+1)/3} (-\mathrm{i})^\gamma \mathrm{Ai}^{(\gamma)}(s) + O(n^{-\alpha}) \quad \text{for all} \quad \alpha > 0. \tag{B.5.28}$$

This gives

$$\psi_{n+p}(x\sqrt{n}) = (2\pi)^{1/4} \sqrt{n} \frac{\mathrm{e}^{-n\frac{x^2}{4}} \mathrm{e}^{\frac{n}{2}(1-x)^2}}{\sqrt{(n+p)!}} n^{\frac{n}{2}+\frac{p}{2}} n^{-1/3} \tag{B.5.29}$$

$$\times \left(\mathrm{Ai}(s) + n^{-1/3} \left(-p\mathrm{Ai}'(s) - \frac{1}{4}\mathrm{Ai}^{(4)}(s) \right) + O(n^{-2/3}) \right). \tag{B.5.30}$$

We compute

$$\mathrm{e}^{-n\frac{x^2}{4}} \mathrm{e}^{\frac{n}{2}(1-x)^2} = \mathrm{e}^{-\frac{n}{4}(4+4\frac{s}{n^{2/3}}+\frac{s^2}{n^{4/3}})} \mathrm{e}^{\frac{n}{2}(1+2\frac{s}{n^{2/3}}+\frac{s^2}{n^{4/3}})} = \mathrm{e}^{-\frac{n}{2}} \mathrm{e}^{\frac{s^2 n^{-1/3}}{4}}, \tag{B.5.31}$$

and use Stirling's approximation to write

$$\frac{(2\pi)^{1/4}}{\sqrt{(n+p)!}} \mathrm{e}^{-\frac{n}{2}} n^{\frac{n}{2}} = \sqrt{\frac{n!}{(n+p)!}} \frac{(2\pi n)^{1/4}}{\sqrt{n!}} \mathrm{e}^{-\frac{n}{2}} n^{\frac{n}{2}} n^{-1/4} \tag{B.5.32}$$

$$= n^{-1/4} \sqrt{\frac{n!}{(n+p)!}} (1 + O(n^{-1})). \tag{B.5.33}$$

Continuing, we obtain

$$\psi_{n+p}(x\sqrt{n}) = n^{-1/12+p/2}\sqrt{\frac{n!}{(n+p)!}}\mathrm{e}^{\frac{s^2n^{1/3}}{4}} \tag{B.5.34}$$

$$\times\left(\mathrm{Ai}(s)+n^{-1/3}\left(-p\mathrm{Ai}'(s)-\frac{1}{4}\mathrm{Ai}^{(4)}(s)\right)+O(n^{-2/3})\right) \tag{B.5.35}$$

$$= n^{-1/12+p/2}\sqrt{\frac{n!}{(n+p)!}} \tag{B.5.36}$$

$$\times\left(\mathrm{Ai}(s)+n^{-1/3}\left(-p\mathrm{Ai}'(s)+\frac{1}{4}(s^2\mathrm{Ai}(s)-\mathrm{Ai}^{(4)}(s))\right)+O(n^{-2/3})\right) \tag{B.5.37}$$

$$= n^{-1/12+p/2}\sqrt{\frac{n!}{(n+p)!}} \tag{B.5.38}$$

$$\times\left(\mathrm{Ai}(s)+n^{-1/3}\left(\frac{1}{2}-p\right)\mathrm{Ai}'(s)+O(n^{-2/3})\right), \tag{B.5.39}$$

where we used $\mathrm{Ai}^{(4)}(s) = s^2\mathrm{Ai}(s) + 2\mathrm{Ai}'(s)$ in the last line.

B.5.1 Uniform bounds

We need uniform estimates when $x = 2 + sn^{-2/3}$ and $0 \le s \le n^{2/3}$ to allow us to transition into case (b), (B.5.12). We use $\Gamma = \mathrm{e}^{-\mathrm{i}\pi/6}(-\infty, 0] \cup \mathrm{e}^{\mathrm{i}\pi/6}[0, \infty)$.

$$\mathrm{e}^{-\frac{n}{2}(1-x)^2}\int_\Gamma \mathrm{e}^{-nH(r)}\mathrm{d}r = \mathrm{e}^{\mathrm{i}n\frac{\pi}{2}}\int_{\Gamma\cap B(0,\delta)}\mathrm{e}^{-nH(r)}\mathrm{d}r + O(\mathrm{e}^{-n(x-1)\delta}).$$

Then, we deform

$$\mathrm{e}^{-\frac{n}{2}(1-x)^2}\int_{\Gamma\cap B(0,\delta)}\mathrm{e}^{-nH(r)}\mathrm{d}r = \mathrm{e}^{-\frac{n}{2}(1-x)^2}\int_C \mathrm{e}^{-nH(r)}\mathrm{d}r \tag{B.5.40}$$

to a horizontal contour connecting its endpoints. Then on this contour,

$$\mathrm{e}^{-\frac{n}{2}(1-x)^2}\int_C(\mathrm{e}^{-nH(r)} - \mathrm{e}^{\frac{n}{2}(1-x)^2})\mathrm{d}r \tag{B.5.41}$$

$$= \int_C \mathrm{e}^{\mathrm{i}nr(x-2)}\left[\mathrm{e}^{-n\frac{r^2}{2}+\mathrm{i}nr+n\log(\mathrm{i}+r)} - \mathrm{e}^{\mathrm{i}n\frac{r^3}{3}}\right]\mathrm{d}r. \tag{B.5.42}$$

For $r \in C$, $|\mathrm{e}^{\mathrm{i}nr(x-2)}| \le \mathrm{e}^{n\delta/\sqrt{2}(x-2)}$, and we find

$$\left|\mathrm{e}^{-\frac{n}{2}(1-x)^2}\int_C(\mathrm{e}^{-nH(r)} - \mathrm{e}^{\frac{n}{2}(1-x)^2})\mathrm{d}r\right| \le Mn^{-2/3}\mathrm{e}^{n\delta/\sqrt{2}(x-2)}. \tag{B.5.43}$$

Then, define $f_n(s) = \frac{n^{1/3}}{2\pi} \int_C \mathrm{e}^{\mathrm{i}n(x-2)r+\mathrm{i}n\frac{r^3}{3}} \mathrm{d}r$ and we have

$$\left| n^{1/12-p/2} \psi_{n+p}(x\sqrt{n}) - \sqrt{\frac{n!}{(n+p)!}} \mathrm{e}^{\frac{s^2 n^{-1/3}}{4}} f_n(s) \right| \leq M n^{-2/3} \mathrm{e}^{\frac{s^2 n^{1/3}}{4} - n\delta/\sqrt{2}(x-2)}. \tag{B.5.44}$$

Choosing $\delta = \sqrt{2}$, we find $\mathrm{e}^{\frac{s^2 n^{1/3}}{4} - n\delta/\sqrt{2}(x-2)} \leq \mathrm{e}^{-s\frac{3}{4}n^{1/3}}$. A similar estimate follows for ψ_{n+p} and we obtain that there exist a constant $M > 0$ such that for $0 \leq s \leq n^{2/3}$

$$\begin{aligned} \left| n^{1/12} \psi_n(x\sqrt{n}) - \mathrm{e}^{\frac{s^2 n^{-1/3}}{4}} f_n(s) \right| &\leq M n^{-2/3} \mathrm{e}^{-s\frac{3}{4}n^{1/3}}, \\ \left| n^{1/12} \psi_{n-1}(x\sqrt{n}) - \mathrm{e}^{\frac{s^2 n^{-1/3}}{4}} f_n(s) \right| &\leq M n^{-2/3} \mathrm{e}^{-s\frac{3}{4}n^{1/3}}. \end{aligned} \tag{B.5.45}$$

For $s \geq n^{2/3}$, we can use (B.5.4) to find

$$n^{1/4}(|\psi_n(x\sqrt{n})| + |\psi_{n-1}(x\sqrt{n})|) \leq M \mathrm{e}^{-\frac{3}{4}n^{1/3}}. \tag{B.5.46}$$

Appendix C

Fredholm determinants

C.1 Definitions

Our purpose in this section is to explain the notion of a Fredholm determinant and resolvent in a simple and concrete setting. The ideas presented here originated in Fredholm's attempt to find a solution formula akin to Cramer's rule for linear integral equations. The notion of a determinant for an infinite-dimensional linear operator is, of course, of independent interest and has attracted the interest of many mathematicians. Simon's book provides an excellent overview of current knowledge [31].

Assume a given continuous kernel $K : [0,1] \times [0,1] \to \mathbb{R}$ and a continuous function $h : [0,1] \to \mathbb{R}$. Fix a spectral parameter $z \in \mathbb{C}$ and consider the linear integral equation

$$\varphi(x) - z \int_0^1 K(x,y)\varphi(y)\,\mathrm{d}y = h(x), \quad x \in [0,1]. \tag{C.1.1}$$

The integral equation (C.1.1) may be written in the more compact form

$$(I - zK)\varphi = h, \tag{C.1.2}$$

where $I - zK$ denotes the bounded linear operator on $L^2([a,b])$ defined by

$$\varphi \mapsto (I - zK)\varphi, \quad (I - zK)\varphi(x) = \varphi(x) - z \int_a^b K(x,y)\varphi(y)\,\mathrm{d}y \quad x \in [a,b]. \tag{C.1.3}$$

Integral equations such as (C.1.1) may naturally be viewed as continuum limits of linear equations. More precisely, we fix a positive integer n, consider a uniform grid $x_j = j/n$, with uniform weights $w_j = 1/n$, define the vector $h_j^{(n)} = h(x_j)$, matrix $K_{j,k}^{(n)} = w_j K(x_j, x_k)$, $1 \le j,k \le n$ and discretize (C.1.1) by the linear equation

$$\varphi_j^{(n)} - z \sum_{k=1}^{n} K_{j,k}^{(n)} \varphi_k^{(n)} = h_j^{(n)}, \quad 1 \le j \le n. \tag{C.1.4}$$

Equation (C.1.4) has a unique solution if and only if $\det(I_n - zK^{(n)}) \neq 0$. By linearity, the solution for arbitrary $h^{(n)}$ is determined by the resolvent $R^{(n)} = (I_n - zK^{(n)})^{-1}$, which is given by Cramer's rule.

Remark 62. If one wants to compute a Fredholm determinant numerically and K is a smooth function, quadrature rules (such as Gaussian quadrature or Clenshaw–Curtis quadrature) can be used to choose x_j and w_j. See, for example, [**?**].

$$R^{(n)}_{j,k} = (-1)^{j+k} \frac{\det(M_{jk})}{\det(I_n - zK^{(n)})}, \tag{C.1.5}$$

where M_{jk} denotes the matrix obtained from $I_n - zK^{(n)}$ by removing the j-th row and k-th column. Further, if z_j, $j = 1, \dots n$, denote the zeros of the polynomial $\det(I_n - zK^{(n)})$, the eigenvalues of $K^{(n)}$ are given by $1/z_j$. Both these notions may be extended to (C.1.1) via the Fredholm determinant. The basic observation that allows passage to the limit is the identity

$$\begin{aligned} &\det(I_n - zK^{(n)}) = \\ &\quad 1 - \frac{z}{n} \sum_{j_1=1}^{n} K(x_{j_1}, x_{j_1}) + \frac{z^2}{2!} \frac{1}{n^2} \sum_{j_1, j_2 = 1}^{n} \begin{vmatrix} K(x_{j_1}, x_{j_1}) & K(x_{j_1}, x_{j_2}) \\ K(x_{j_2}, x_{j_1}) & K(x_{j_2}, x_{j_2}) \end{vmatrix} + \dots \end{aligned} \tag{C.1.6}$$

The coefficient of z^k in the expansion above may be computed by differentiating the left hand side k times with respect to z, and setting $z = 0$. Since K is continuous, as $n \to \infty$, the k-th term in the sum above converges to the integral

$$\frac{(-z)^k}{k!} \int_{[0,1]^k} \det\left(K(x_p, x_q)_{1 \leq p,q \leq k}\right) \, \mathrm{d}x_1 \dots \mathrm{d}x_k. \tag{C.1.7}$$

Definition-Theorem 63. The Fredholm determinant of the operator $I - zK$ is the entire function of z defined by the convergent series

$$D(z) := \det(I - zK) = 1 + \sum_{k=1}^{\infty} \frac{(-z)^k}{k!} \int_{[0,1]^k} \left(\det(K(x_p, x_q)_{1 \leq p,q \leq k})\right) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k. \tag{C.1.8}$$

Proof. It is only necessary to show that the series(C.1.7) is convergent for all $z \in \mathbb{C}$. The determinant of a $k \times k$ matrix A with columns $a_1, \dots, a_k$ is the (signed) volume of the parallelopiped spanned by the vectors $a_1, \dots, a_k$. Therefore,

$$|\det(A)| \leq |a_1||a_2| \cdots |a_k| \leq \left(\max_{1 \leq j \leq k} |a_j| \right)^k. \tag{C.1.9}$$

We have assumed that K is bounded on $[0,1] \times [0,1]$, say $\max |K| \leq M < \infty$. By the inequality above,

$$|(\det(K(x_p, x_q)_{1 \leq p,q \leq k}))| \leq k^{k/2} M^k. \tag{C.1.10}$$

Thus, the k-term in the series (C.1.8) is dominated by

$$\left|\frac{(-z)^k}{k!}\int_{[0,1]^k}\det\left[K(x_p,x_q)\right]_{1\le p,q\le k}\,\mathrm{d}x_1\cdots\mathrm{d}x_k\right|$$
$$\le(|z|M)^k\frac{k^{k/2}}{k!}\sim\frac{1}{\sqrt{2\pi}}(|z|Me)^k\frac{1}{k^{\frac{k+1}{2}}},$$

where we have used Stirling's approximation in the last step. □

Remark 64. If $[0,1]$ is replaced by a general Borel set S, we assume

$$|K(x,y)|\le M(x),$$

where $M\in L^1(S)$. The same statements about the determinant follow.

Since $D(z)$ is entire, we may differentiate term-by-term to obtain

$$\left(-\frac{\mathrm{d}}{\mathrm{d}z}\right)^m\det(I-zK) \tag{C.1.11}$$
$$=\sum_{k=0}^{\infty}\frac{(-z)^k}{k!}\int_{[0,1]^{m+k}}\det\left[K(x_p,x_q)\right]_{1\le p,q\le m+k}\,\mathrm{d}x_1\cdots\mathrm{d}x_{m+k}$$

for $m\ge 1$. Recall that the zeros of a non-zero entire function form a discrete, countable set. The entire function $\det(I-\lambda^{-1}K)$ is an infinite-dimensional generalization of the characteristic polynomial of the matrix $K^{(n)}$ in the following sense:

Theorem 65 (Eigenvalues of K)**.** *Assume that K is a continuous kernel. The complex number λ is an eigenvalue of K if and only if $D(\lambda^{-1})=0$.*

For more on Fredholm determinants, see [23, Ch.24].

C.2 Convergence

Suppose a kernel $K_n(x,y)\to K_\infty(x,y)$, $(x,y)\in S^2$, pointwise. One needs the additional convergence criteria to conclude

$$\det(1-K_n)\to\det(1-K_\infty). \tag{C.2.1}$$

The following are from [31]. Let $\mathcal{K}_n$ and $\mathcal{K}_\infty$ be the operators on $L^2(S)$ with kernels K_n and K_∞, respectively. Then the trace norm of an operator $\mathcal{K}$ is given by

$$\|\mathcal{K}\|_{\mathrm{Tr}}=\mathrm{Tr}\,\sqrt{\mathcal{K}^*\mathcal{K}}, \tag{C.2.2}$$

where $\mathcal{K}^*$ is the adjoint of $\mathcal{K}$. The general definition of $\sqrt{\mathcal{K}^*\mathcal{K}}$ for general operators is unimportant for us and an operator with finite trace norm is said

to be trace class. But, for example, if $\mathcal{K}$ is a self-adjoint operator with continuous kernel K then

$$\|\mathcal{K}\|_{\mathrm{Tr}} = \int_S K(x,x)\mathrm{d}x. \tag{C.2.3}$$

Theorem 66. *The map $\mathcal{K} \mapsto \det(I+\mathcal{K})$ is a continuous function on the space of trace-class operators (i.e. operators with $\|\mathcal{K}\|_{\mathrm{Tr}} < \infty$) and*

$$|\det(I+\mathcal{K}) - \det(I+\mathcal{L})| \leq \|\mathcal{K}-\mathcal{L}\|_{\mathrm{Tr}} \exp(\|\mathcal{K}\|_{\mathrm{Tr}} + \|\mathcal{L}\|_{\mathrm{Tr}} + 1). \tag{C.2.4}$$

Theorem 67. *Suppose $\mathcal{K}_n, \mathcal{K}$ are trace class. If $\mathcal{K}_n \to \mathcal{K}, |\mathcal{K}_n| \to |\mathcal{K}|$ and $|\mathcal{K}_n^*| \to |\mathcal{K}^*|$ all weakly, then $\|\mathcal{K}_n - \mathcal{K}\|_{\mathrm{Tr}} \to 0$.*

In our cases, $|\mathcal{K}_n| = \mathcal{K}_n = |\mathcal{K}_n^*|$, so to show that $\det(I-K_n) \to \det(I-K)$ it suffices to show for each $f, g \in L^2(S)$ that

$$\int_S \int_S K_n(x,y)f(x)g(y)\mathrm{d}x\mathrm{d}y \to \int_S \int_S K(x,y)f(x)g(y)\mathrm{d}x\mathrm{d}y. \tag{C.2.5}$$

Two such conditions for this to occur are

1. If S is bounded then

$$\sup_{x,y\in S} |K_n(x,y) - K(x,y)| \to 0. \tag{C.2.6}$$

2. If S is unbounded then we require

$$K_n(x,y) \to K(x,y), \tag{C.2.7}$$

for each $x, y \in S$ and there exists $G(x,y) \in L^2(S^2)$ such that $|K_n(x,y)| \leq G(x,y)$. This allows one to use the dominated convergence theorem.

C.2.1 Change of variables and kernel extension

Let $K : S^2 \to \mathbb{R}$ be a kernel. Let $x = r(t)$ and $y = r(s)$ for $s, t \in T$ where r' exists, is continuous and does not vanish. Define

$$\hat{K}(s,t) = \frac{1}{r'(s)} K\left(r(s), r(t)\right), \quad s,t \in T^2. \tag{C.2.8}$$

Then

$$\det(I-K) = \det(I-\hat{K}). \tag{C.2.9}$$

C.3 Computing Fredholm determinants

C.4 Separable kernels

Appendix D

Notation

D.1 Indices

The integers m and n are reserved for the number of rows and columns of a matrix. For square matrices, we use n. The letters j and k are used to denote indices. The letter i is reserved for $\sqrt{-1}$.

D.2 Fields

$\mathbb{R}$ real numbers.

$\mathbb{C}$ complex numbers with imaginary unit $i = \sqrt{-1}$.

$\mathbb{H}$ quaternions with imaginary units $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$.

$\mathbb{F}$ general notation for one of the above fields.

$\mathbb{T}^n$ the n-dimensional real torus.

Σ_n the n-dimensional simplex, $x \in \mathbb{R}^n$, $\sum_{k=1}^n x_k = 1$, $x_k \geq 0$, $1 \leq k \leq n$.

For $x \in \mathbb{C}$ and $x \in \mathbb{H}$ we use $\bar{x}$ to denote the complex and quaternion conjugate respectively. The absolute value of a number $x \in \mathbb{F}$ is always denoted $|x|$. The same notation is used for the Euclidean length of a vector in $\mathbb{F}^n$, but the distinction between the two uses of $|\cdot|$ will be clear from the context. For example, for $x \in \mathbb{F}^n$, $x = (x_1, \ldots, x_n)$, $x_j \in \mathbb{F}$, $1 \leq j \leq n$, we write

$$|x|^2 = \sum_{j=1}^{n} |x_j|^2. \tag{D.2.1}$$

D.3 Matrices

The fundamental spaces of matrices are denoted as follows:

$\mathbb{F}^{m\times n}$ $m \times n$ matrix with entries from a field $\mathbb{F}$.

$\mathbb{F}^{n\times n}$ $n \times n$ matrix with entries from a field $\mathbb{F}$.

$\mathsf{Symm}(n)$ real, symmetric $n \times n$ matrices.

$\mathsf{Her}(n)$ complex, Hermitian $n \times n$ matrices.

$\mathsf{Quart}(n)$ real, self-dual quaternion $n \times n$ matrices.

$\mathsf{Jac}(n)$

We write $M^\dagger$ for the adjoint of a matrix $M \in \mathbb{M}^{\mathbb{F}}_{m\times n}$. Matrices in $\mathsf{Symm}(n)$, $\mathsf{Her}(n)$ and $\mathsf{Quart}(n)$ are self-adjoint: $M = M^\dagger$, but the notion of duality is distint in each setting, since the underlying field is different. For $M \in \mathsf{Symm}(n)$, $M^\dagger = M^T$; if $M \in \mathsf{Her}(n)$, then $M^\dagger = \bar{M}^T = M^*$; and if $M \in \mathsf{Quart}(n)$, then $M^\dagger = \bar{M}^T$. All these matrices have real eigenvalues, and the matrices are said to be positive definite if all eigenvalues are *strictly positive*. We denote the subset of positive definite matrices by $\mathsf{Symm}_+(n)$, $\mathsf{Her}_+(n)$ and $\mathsf{Quart}_+(n)$ respectively. The Hilbert-Schmidt norm of a matrix $M \in \mathbb{F}^{n\times m}$ is denoted

$$\|M\|^2 = \mathrm{Tr}\left(M^\dagger M\right) = \sum_{j,k=1}^{n} |M_{j,k}|^2. \tag{D.3.1}$$

$B_r(M)$ denotes the ball of radius r centered at M in the norm $\|\cdot\|$.

D.4 Lie groups

The classical groups we consider are also Lie groups. For a Lie group, the associated Lie algebra is the tangent space at the identity. We use the following notation for the classical (Lie) groups and their associated Lie algebras, respectively.

$\mathsf{O}(n)$, $\mathfrak{o}(n)$ the real, orthogonal group.

$\mathsf{SO}(n)$, $\mathfrak{so}(n)$ the special orthogonal group.

$\mathsf{U}(n)$, $\mathfrak{u}(n)$ the unitary group.

$\mathsf{USp}(n)$, $\mathfrak{usp}(n)$ the group of unitary symplectic matrices, or the compact symplectic group.

D.5 Banach spaces

The following notation is used for standard Banach spaces.

$C(J)$ The space of continuous functions on an interval J equipped with the supremum norm.

$\mathcal{P}_J$ The space of probability measures on an interval J equipped with the weak topology.

$\langle \mu, f \rangle$ The duality pairing between measures and continuous functions on the interval J given by

$$\langle \mu, f \rangle = \int_J f(x)\, \mu(dx). \tag{D.5.1}$$

$C_0(\mathbb{R})$ The space of continuous function on $\mathbb{R}$ that vanish at infinity, equipped with the supremum norm.

$C_b(\mathbb{R})$ The space of bounded continuous function on $\mathbb{R}$ that vanish at infinity, equipped with the supremum norm.

Bibliography

[1] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Courier Dover Publications, 1972.

[2] J. Baik, P. Deift, and K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations.*, J. Am. Math. Soc., 12 (1999), pp. 1119–1178.

[3] J. Baik, P. Deift, and Suidan, *Combinatorics and Random Matrix Theory*, Amer. Math. Soc., Providence, RI, 2017.

[4] P. Bourgade, *On random matrices and L-functions*, PhD thesis, Ph. D. thesis, 2009.

[5] F. Bullo and A. D. Lewis, *Geometric Control of Mechanical Systems*, Springer, New York, NY, 2005.

[6] P. Deift, *Orthogonal Polynomials and Random Matrices: A Riemann Hilbert Approach*, Courant Lecture Notes in Mathematics, Courant Institute of Mathematical Sciences, New York, First ed., 1999.

[7] P. Deift, C. D. Levermore, and C. E. Wayne, eds., *Dynamical Systems and Probabilistic Methods in Partial Differential Equations*, vol. 31 of Lecture Notes in Applied Mathematics, American Mathematical Society, Providence, First ed., 1996, ch. Integrable Hamiltonian Systems, pp. 103–138.

[8] I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, Journal of Mathematical Physics, 43 (2002), pp. 5830–5847.

[9] F. J. Dyson, *Statistical theory of the energy levels of complex systems. I*, Journal of Mathematical Physics, 3 (1962), pp. 140–156.

[10] A. Edelman, *Eigenvalues and Condition Numbers of Random Matrices*, PhD thesis, Massachusetts Institute of Technology, 1989.

[11] A. Edelman and B. Sutton, *From random matrices to stochastic operators*, Journal of Statistical Physics, 127 (2007), pp. 1121–1165.

[12] H. M. Edwards, *Riemann's zeta function*, vol. 58, Academic Press, 1974.

[13] L. Erds, H.-T. Yau, and J. Yin, *Rigidity of eigenvalues of generalized Wigner matrices*, Adv. Math. (N. Y)., 229 (2012), pp. 1435–1515.

[14] J. Harer and D. Zagier, *The euler characteristic of the moduli space of curves*, Inventiones Mathematicae, 85 (1986), pp. 457–485.

[15] S. Hastings and J. McLeod, *A boundary value problem associated with the second Painlevé transcendent and the Korteweg-de Vries equation*, Archive for Rational Mechanics and Analysis, 73 (1980), pp. 31–51.

[16] M. Hestenes and E. Steifel, *Method of Conjugate Gradients for solving Linear Systems*, J. Res. Nat. Bur. Stand., 20 (1952), pp. 409–436.

[17] E. Hille, *Ordinary differential equations in the complex domain*, Courier Dover Publications, 1997.

[18] M. Jimbo, T. Miwa, Y. Môri, and M. Sato, *Density Matrix of an impenetrable Bose Gas and the fifth Painleve Transcendent*, Physica 1D, 1 (1980), pp. 80–158.

[19] K. Johansson, *Shape Fluctuations and Random Matrices*, Commun. Math. Phys., 209 (2000), pp. 437–476.

[20] F. John, *Partial differential equations*, vol. 1 of Applied Mathematical Sciences, Springer-Verlag, New York, fourth ed., 1991.

[21] T. Kato, *Perturbation theory for linear operators*, Springer, New York, NY, 1995.

[22] S. V. Kerov, *Asymptotic representation theory of the symmetric group and its applications in analysis*, American Mathematical Society, Providence, RI, 2003.

[23] P. Lax, *Functional Analysis*, John Wiley and Sons, 2002.

[24] B. M. McCoy and T. T. Wu, *The two-dimensional Ising model*, vol. 22, Harvard University Press, 1973.

[25] M. L. Mehta, *Random Matrices*, Elsevier, San Diego and London, third ed., 2004.

[26] M. L. Mehta and M. Gaudin, *On the density of eigenvalues of a random matrix*, Nuclear Physics, 18 (1960), pp. 420–427.

[27] L. Nachbin, *The Haar integral*, R.E. Krieger Pub. Co, 1976.

[28] A. M. Odlyzko and E. M. Rains, *On Longest Increasing Subsequences in Random Permutations*, Amer. Math. Soc., Contemp. Math., 251 (2000), pp. 439–451.

[29] J. Ramírez, B. Rider, and B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, Journal of the American Mathematical Society, 24 (2011), pp. 919–944.

[30] B. Riemann, *Collected Works of Bernhard Riemann*, Dover Publications, 1953.

[31] B. Simon, *Trace ideals and their applications*, vol. 120 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, second ed., 2005.

[32] R. P. Stanley, *Enumerative combinatorics*, vol. 2, Cambridge University Press, 2011.

[33] D. W. Stroock, *Probability theory: an analytic view*, Cambridge university press, 2010.

[34] G. 't Hooft, *A planar diagram theory for strong interactions*, Nuclear Physics B, 72 (1974), pp. 461–473.

[35] Tao, T, *Topics in random matrix theory*, AMS, Providence, RI, 2011.

[36] C. A. Tracy and H. Widom, *Level–Spacing Distributions and the Airy Kernel*, Communications in Mathematical Physics, 159 (1994), pp. 151–174.

[37] S. M. Ulam, *Monte Carlo Calculations in Problems of Mathematical Physics*, in Mod. Math. Eng., E. F. Bechenbach, ed., McGraw-Hill, Inc., 1961, pp. 261–281.

[38] A. M. Vershik and S. V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux*, Dokl. AN SSSR, 233 (1977), pp. 1024–1027.

[39] J.-B. Zuber, *Introduction to random matrices*. Lectures at ICTS, Bangalore, 2012.