NOTES FOR THE COURSE
Introduction to Random Matrix Theory, Riemann–Hilbert problems, and Universality

spring 2022

G. Ruzza — IRMP, UCLouvain, giulio.ruzza@uclouvain.be[1]

WORK IN PROGRESS: last update May 4, 2022

[1] Plaese report typos!

The following texts were consulted in the drafting of these notes.

- *G. W. Anderson, A. Guionnet, & O. Zeitouni. "An introduction to random matrices". Cambridge Studies in Advanced Mathematics, 118.* Cambridge University Press, Cambridge, *2010*
- *P. Deift. "Orthogonal polynomials and random matrices: a Riemann-Hilbert approach". Courant Lecture Notes in Mathematics, 3.* New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, *1999*
- *P. J. Forrester. "Log-gases and random matrices". London Mathematical Society Monographs Series, 34.* Princeton University Press, Princeton, NJ, *2010*
- *A. R. Its. "Large N asymptotics in random matrices: the Riemann–Hilbert approach".* Random matrices, random processes and integrable systems, *351–413, CRM Ser. Math. Phys.,* Springer, New York, *2011.*
- *E. S. Meckes. "The random matrix theory of the classical compact groups". Cambridge Tracts in Mathematics, 218.* Cambridge University Press, Cambridge, *2019*
- *M. L. Mehta. "Random matrices". Pure and Applied Mathematics (Amsterdam) 142.* Elsevier/Academic Press, Amsterdam, *2004*
- *A. Soshnikov. "Determinantal random point fields".* Uspekhi Mat. Nauk **55** *(2000), no. 5(335), 107–160; translation in* Russian Math. Surveys **55** *(2000), no. 5, 923–975 (available at https://arxiv.org/pdf/math/0002099.pdf)*
- *C. Tracy & H. Widom. "Introduction to random matrices". Lecture Notes in Phys., 424,* Springer, Berlin, *1993 (available at https://arxiv.org/pdf/hep-th/9210073.pdf)*

Contents

Chapter 1

Introduction

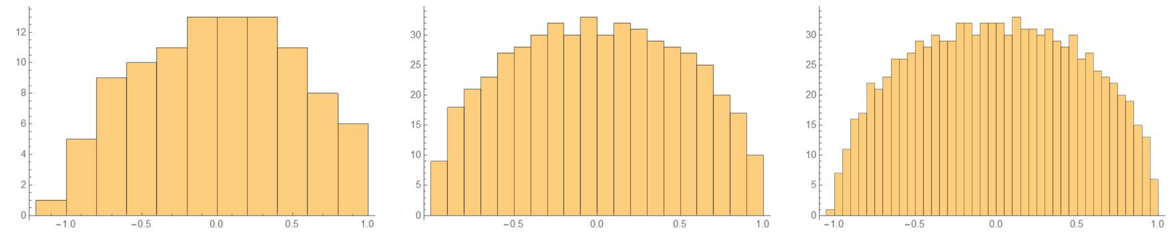
The main question in Random Matrix Theory is: *what can we say about spectral properties (eigenvalues and eigenvectors) of large matrices with random entries?*

Remarkably, we can say a lot, as we will see from a few examples now and later in the course. Moreover, the answer to such question 'extends' far beyond the original context of matrices with random entries.

1.1 A first numerical exploration

1.1.1 Emergence of the semicircle distribution

Consider a random $n \times n$ matrix X with i.i.d.[1] entries $X_{i,j}$, normally distributed[2]. A random matrix distributed as the matrix $(X + X^\top)/2$ is said to be a matrix from the **Gaussian Orthogonal Ensemble** (GOE), or just a GOE matrix. A GOE matrix is surely symmetric, so it has real eigenvalues; let us plot them.

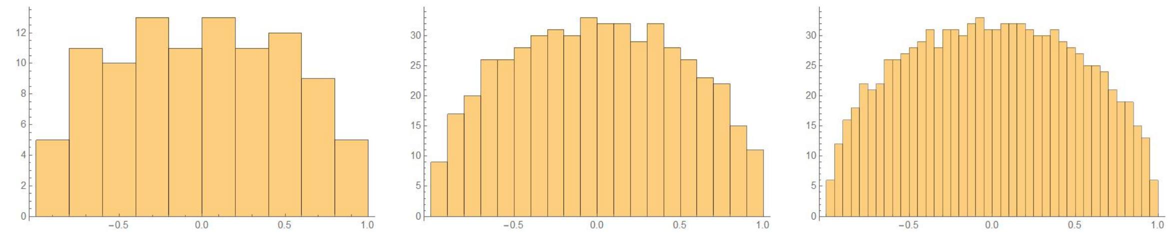


$n = 100$ $\qquad$ $n = 500$ $\qquad$ $n = 1000$

Figure 1.1: Eigenvalues (scaled by $1/\sqrt{2n}$) for GOE random matrices of size $n = 100, 500, 1000$.

A semicircle-shaped distribution emerges. This is called **Wigner semicircle distribution**.

Let us repeat the same construction, this time with $X_{i,j}$ being i.i.d. Bernoulli random variables taking the values ± 1, each with probability $1/2$. The same semicircle law appears!



$n = 100$ $\qquad$ $n = 500$ $\qquad$ $n = 1000$

Figure 1.2: Eigenvalues (scaled by $1/\sqrt{2n}$) for Bernoulli random matrices of size $n = 100, 500, 1000$.

[1]Independent and identically distributed

[2]i.e. their distribution is the standard Gaussian with zero mean and unit variance, with density $(2\pi)^{-1/2} \exp(-x^2/2)$.

Two things should be remarked.

- *Universality*: the same semicircle distribution appears for different specific entry distributions.
- *Concentration of measure*: for *any* realization of a random matrix, the eigenvalue histogram looks like a semicircle distribution (more and more accurately as the size increases). In other words, the average behavior is also the typical behavior; see Figure 1.3 for some different realizations of the GOE of the same size. Thus a random matrix is actually more predictable than a fixed matrix!

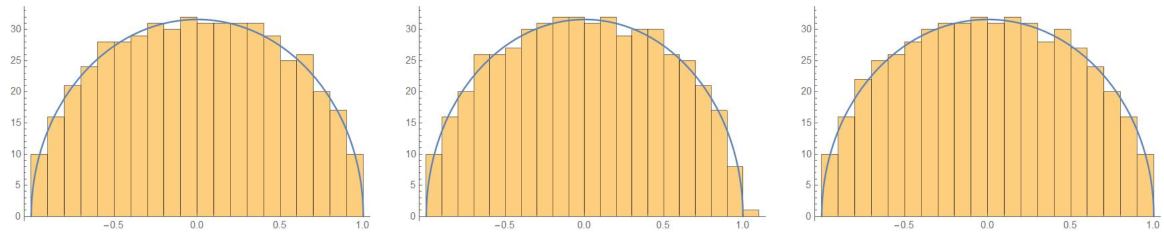


Figure 1.3: Eigenvalues (scaled by $1/\sqrt{2n}$) for three different realization of a GOE matrix size $n = 500$. The semicircle distribution is also shown.

The universality of Wigner semicircle law holds under mild analytic assumptions on the specific entry distribution; in particular these are not satisfied for instance by the Cauchy distribution (the measure $\frac{1}{\pi}\frac{\mathrm{d}x}{1+x^2}$ on $\mathbb{R}$), see Figure 1.4.

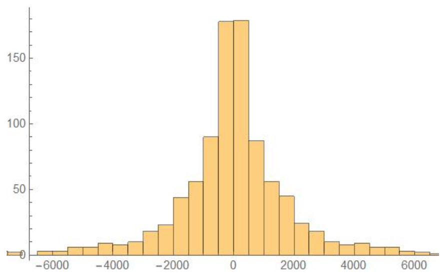


Figure 1.4: Eigenvalues for the symmetric matrix $(X + X^{\top})/2$ obtained from a realization X of a 1000×1000 random matrix with i.i.d. Cauchy entries $X_{i,j}$.

1.1.2 Eigenvalue repulsion

Another class of examples consists of orthogonal matrices; a natural way to generate them is to consider a matrix X with i.i.d. Gaussian entries and then construct a random orthogonal matrix by applying the Gram–Schmidt procedure to X (say, column-wise). The ensemble of such random matrices is called **Circular Orthogonal Ensemble** (COE), and a random matrix from this ensemble is said to be a COE matrix.

The eigenvalues of a COE matrix are random points on the circle; in Figure 1.5 you can see the difference between a realization of the eigenvalues of a COE matrix and i.i.d. realizations of the random uniform distribution on the circle.

In this example we can clearly observe **repulsion** for the eigenvalues of a random matrix.

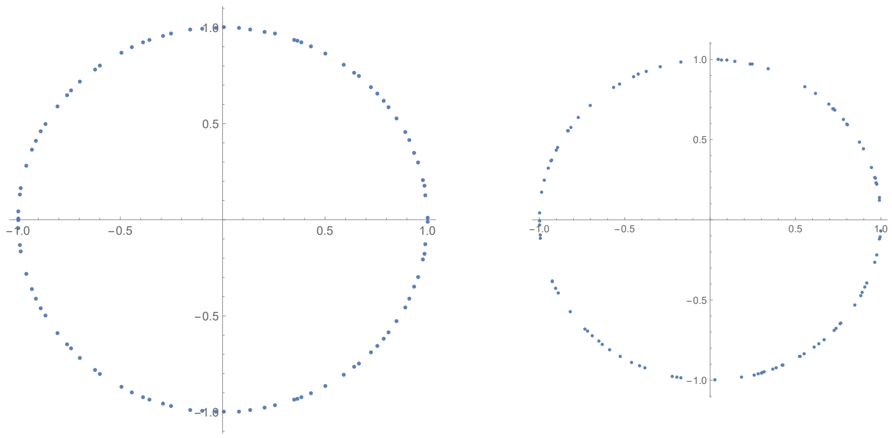


Figure 1.5: On the left: eigenvalues of a 100×100 COE matrix. On the right: 100 random independent points on the circle (sampled according to the uniform probability distribution on the circle).

Chapter 2

Invariant ensembles

2.1 Background material on matrix groups

2.1.1 Notations

Definition 2.1.1. We denote $\mathbb{C}^{m\times n}$ (resp. $\mathbb{R}^{m\times n}$) the set of $m\times n$ matrices with complex (resp. real) entries. Given $M\in\mathbb{C}^{m\times n}$ the **transposed matrix** is the matrix $M^\top\in\mathbb{C}^{n\times n}$ with entries

$$(M^\top)_{ab}=M_{ba},\qquad 1\le a\le m,\ 1\le b\le n. \tag{2.1.1}$$

Given $M\in\mathbb{C}^{m\times n}$, the **adjoint matrix** $M^\dagger\in\mathbb{C}^{n\times m}$ is the conjugate transpose matrix, namely its entries are related to those of M by

$$(M^\dagger)_{ab}=\overline{M_{ba}},\qquad 1\le a\le m,\ 1\le b\le n, \tag{2.1.2}$$

where $\overline{z}=x-\mathrm{i}y$ is the complex conjugate of $z=x+\mathrm{i}y\in\mathbb{C}$. //

Definition 2.1.2. A matrix $M\in\mathbb{C}^{n\times n}$ is **Hermitian** if and only if $M^\dagger=M$. A matrix $U\in\mathbb{C}^{n\times n}$ is **unitary** if and only if $UU^\dagger=\mathbf{1}$. We denote:

$$\mathbf{H}_n:=\{M\in\mathbb{C}^{n\times n}:\ M^\dagger=M\}, \tag{2.1.3}$$

$$\mathbf{U}_n:=\{U\in\mathbb{C}^{n\times n}:\ UU^\dagger=\mathbf{1}\}. \tag{2.1.4}$$

//

If $(,)$ is the standard complex inner product, $(v,w)=\sum_{i=1}^n v_i\overline{w_i}$, then for all $M\in\mathbb{C}^{n\times n}$

$$(v,Mw)=(M^\dagger v,w),\qquad \forall v,w\in\mathbb{C}^n. \tag{2.1.5}$$

Hence

$$M\in\mathbf{H}_n \iff (v,Mw)=(Mv,w)\text{ for all }v,w\in\mathbb{C}^n, \tag{2.1.6}$$

$$U\in\mathbf{U}_n \iff (Uv,Uw)=(v,w)\text{ for all }v,w\in\mathbb{C}^n. \tag{2.1.7}$$

In particular $U\in\mathbf{U}_n$ if and only if its rows (or, equivalently, its columns) are orthonormal.

We shall assume that vectors in $\mathbb{C}^n$ are column vectors, namely $\mathbb{C}^n=\mathbb{C}^{n\times 1}$, so that if $v\in\mathbb{C}^n$, $v^\dagger$ is a row vector; in particular $(v,w)=w^\dagger v$. We also denote $\|v\|:=(v,v)^{1/2}$ the norm of a vector $v\in\mathbb{C}^n$.

We shall denote $e_1,\dots,e_n$ the standard basis of $\mathbb{C}^n$, and $E_{ij}=e_ie_j^\top\in\mathbb{C}^{n\times n}$ the elementary matrices with 1 at the entry (i,j) and 0 elsewhere, leaving the size n implicit as it will always be clear from the context.

2.1.2 Spectral theorem for normal matrices

Definition 2.1.3. $M \in \mathbb{C}^{n\times n}$ is a **normal matrix** if and only if $MM^\dagger = M^\dagger M$. //

Unitary and Hermitian matrices are examples of normal matrices.

Theorem 2.1.4 (Spectral Theorem for normal matrices)**.** *$M \in \mathbb{C}^{n\times n}$ is normal if and only if $M = UDU^\dagger$ for $U \in \mathbf{U}_n$ and D diagonal.*

To prove this we first prove

Lemma 2.1.5 (Schur decomposition)**.** *For any $M \in \mathbb{C}^{n\times n}$ there exist $U \in \mathbf{U}_n$ and an upper triangular $T \in \mathbb{C}^{n\times n}$ such that $M = UTU^\dagger$.*

Proof of lemma. We use induction on $n \geq 1$. The case $n = 1$ is obvious so we assume $n \geq 2$. $M \in \mathbb{C}^{n\times n}$ has (at least) one eigenvalue $\lambda \in \mathbb{C}$, so let $v \in \mathbb{C}^n$ be such that $(v, v) = 1$ and $Mv = \lambda v$. By completing v to a basis of $\mathbb{C}^n$ and using Gram–Schmidt we can construct $V \in \mathbf{U}_n$ whose first column is v, namely $Ve_1 = v$; therefore

$$MVe_1 = \lambda Ve_1, \tag{2.1.8}$$

hence

$$V^\dagger MV = \left(\begin{array}{c|c} \lambda & \cdots \\ \hline \mathbf{0} & \widetilde{M} \end{array}\right), \tag{2.1.9}$$

for some $\widetilde{M} \in \mathbb{C}^{(n-1)\times(n-1)}$ and the dots stand for entries whose value is inconsequential to our argument. By induction, there exists $\widetilde{U} \in \mathbf{U}_{n-1}$ and an upper triangular $\widetilde{T} \in \mathbb{C}^{(n-1)\times(n-1)}$ such that $\widetilde{M} = \widetilde{U}\widetilde{T}\widetilde{U}^\dagger$, and so

$$\left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{U}^\dagger \end{array}\right) V^\dagger MV \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{U} \end{array}\right) = \left(\begin{array}{c|c} \lambda & \cdots \\ \hline \mathbf{0} & \widetilde{T} \end{array}\right), \tag{2.1.10}$$

and so $U := V \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & \widetilde{U} \end{array}\right)$ is unitary and satisfies the conditions of the theorem. □

Proof of theorem. Let $M \in \mathbb{C}^{n\times n}$ be normal, and write $M = UTU^\dagger$ with $U \in \mathbf{U}_n$ and T upper triangular. T has to be normal as well; indeed

$$\mathbf{0} = MM^\dagger - M^\dagger M = U(TT^\dagger - T^\dagger T)U^\dagger. \tag{2.1.11}$$

Therefore it is sufficient to prove that an upper triangular matrix is normal if and only if it is diagonal; to see it we write

$$T = \left(\begin{array}{c|c} a & b^\dagger \\ \hline 0 & \widetilde{T} \end{array}\right) \tag{2.1.12}$$

for some $a \in \mathbb{C}$, $b \in \mathbb{C}^{(n-1)\times 1}$, and $\widetilde{T} \in \mathbb{C}^{(n-1)\times(n-1)}$ upper triangular. Then

$$TT^\dagger - T^\dagger T = \left(\begin{array}{c|c} a & b^\dagger \\ \hline 0 & \widetilde{T} \end{array}\right)\left(\begin{array}{c|c} \bar{a} & 0 \\ \hline b & \widetilde{T}^\dagger \end{array}\right) - \left(\begin{array}{c|c} \bar{a} & 0 \\ \hline b & \widetilde{T}^\dagger \end{array}\right)\left(\begin{array}{c|c} a & b^\dagger \\ \hline 0 & \widetilde{T} \end{array}\right) = \left(\begin{array}{c|c} (b,b) & \cdots \\ \hline \vdots & \widetilde{T}\widetilde{T}^\dagger - \widetilde{T}^\dagger\widetilde{T} - bb^\dagger \end{array}\right) \tag{2.1.13}$$

so that $b = 0$ and $\widetilde{T}$ is an upper triangular matrix of size $n - 1$ which is normal. Induction on n completes the proof of the theorem. □

Corollary 2.1.6 (Spectral Theorem for Hermitian matrices)**.** *For every $M \in \mathbf{H}_n$ there exist $U \in \mathbf{U}_n$ and $x_1, ..., x_n \in \mathbb{R}$ such that*

$$H = U \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix} U^\dagger. \tag{2.1.14}$$

Corollary 2.1.7 (Spectral Theorem for unitary matrices)**.** *For every $V \in \mathbf{U}_n$ there exist $U \in \mathbf{U}_n$ and $z_1, ..., z_n \in S^1$ such that*

$$V = U \begin{pmatrix} z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & z_n \end{pmatrix} U^\dagger. \tag{2.1.15}$$

Hereafter, the unit circle S^1 is identified with $\{z \in \mathbb{C} : \ |z| = 1\}$.

2.2 Lebesgue measure and unitary-invariant ensembles of Hermitian matrices

2.2.1 The metric

Let us introduce a notation for the standard Hermitian inner product on $\mathbb{C}^{n\times n} \simeq \mathbb{C}^{n^2}$;

$$\langle M, N\rangle := \mathrm{tr}\,(MN^\dagger) = \sum_{1\le i,j\le n} M_{ij}\overline{N}_{ij} \qquad (M, N \in \mathbb{C}^{n\times n}). \tag{2.2.1}$$

Lemma 2.2.1. *For any $U \in \mathbf{U}_n$ the following maps $\mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ are isometries of* (2.2.1)*;*

1. $\mathcal{L}_U : M \mapsto UM$
2. $\mathcal{R}_U : M \mapsto MU$
3. $\mathrm{Ad}_U : M \mapsto UMU^\dagger$ *("adjoint map")*

Proof. **Exercise** (use the cyclic property of the trace). □

2.2.2 Hermitian matrices

Recall that $M \in \mathbf{H}_n$ if and only if $M = M^\dagger$, i.e., if and only if $M = X + \mathrm{i}Y$ for $X, Y \in \mathbb{R}^{n\times n}$ satisfying $X^\top = X$ and $Y = -Y^\top$. Thus, $\mathbf{H}_n$ is a *real* vector space of dimension n^2.

The metric (2.2.1) restricts to an euclidean metric on $\mathbf{H}_n$;

$$\langle M, M'\rangle := \mathrm{tr}\,(MM') = \sum_{1\le i\le n}^{n} X_{ii}X'_{ii} + 2\sum_{1\le i<j\le n}(X_{ij}X'_{ij} + Y_{ij}Y'_{ij}), \tag{2.2.2}$$

for all $M = X + \mathrm{i}Y, M' = X' + \mathrm{i}Y' \in \mathbf{H}_n$. Thus $\mathbf{H}_n$ is a real euclidean vector space. In particular, the adjoint action Ad_U ($U \in \mathbf{U}_n$) is an isometry of $\mathbf{H}_n$. (The other two maps in Lemma 2.2.1 do not map $\mathbf{H}_n$ into itself).

The volume form associated with the euclidean metric (2.2.2) takes the form

$$\Omega_{\mathbf{H}_n}(\mathrm{d}M) = 2^{n(n-1)/2} \prod_{1\le i\le j\le n} \mathrm{d}X_{ij} \prod_{1\le i<j\le n} \mathrm{d}Y_{ij}, \tag{2.2.3}$$

(again, writing $M = X + \mathrm{i}Y$) and will be referred to as **Lebesgue measure** on $\mathbf{H}_n$.

Remark 2.2.2. In general, given a finite dimensional vector space V with a basis $e_1, \ldots, e_n$ and a scalar product $(,)$, the associated volume form is given by $\Omega := \sqrt{\det_{1\le i,j\le n}(e_i, e_j)}\,\epsilon^1 \wedge \cdots \wedge \epsilon^n$, where $\epsilon^1, \ldots, \epsilon^n$ is the dual basis of V^*, $\epsilon^i(e_j) = \delta^i_j$. This is independent of the choice of basis. This explains the normalization $2^{n(n-1)/2}$ in (2.2.3). //

Lemma 2.2.3. *For any $U \in \mathbf{U}_n$ the adjoint map $\mathrm{Ad}_U : M \mapsto UMU^\dagger$ from $\mathbf{H}_n$ into itself preserves the volume form $\Omega_{\mathbf{H}_n}$.*

Proof. It follows from Lemma 2.2.1. □

Example 2.2.4. When $n = 2$ a general element of $\mathbf{H}_2$ is $M = \begin{pmatrix} p & q+\mathrm{i}r \\ q-\mathrm{i}r & s \end{pmatrix}$, for $(p,q,r,s) \in \mathbb{R}^4$. The metric is

$$\mathrm{tr}\,(\mathrm{d}M^2) = \mathrm{tr}\left[\begin{pmatrix} \mathrm{d}p & \mathrm{d}q+\mathrm{id}r \\ \mathrm{d}q-\mathrm{id}r & \mathrm{d}s \end{pmatrix}^2\right] = \mathrm{d}p^2 + \mathrm{d}s^2 + 2\mathrm{d}q^2 + 2\mathrm{d}r^2. \tag{2.2.4}$$

The volume form is

$$\Omega_{\mathbf{H}_2}(\mathrm{d}M) = 2\mathrm{d}p\mathrm{d}s\mathrm{d}q\mathrm{d}r. \tag{2.2.5}$$

//

2.2.3 Unitary-invariant ensembles of Hermitian matrices

Definition 2.2.5. A **unitary-invariant ensemble** of Hermitian matrices is given by the probability distribution

$$\frac{1}{Z_n}\exp(-\mathrm{tr}\,V(M))\,\Omega_{\mathbf{H}_n}(\mathrm{d}M) \tag{2.2.6}$$

where $V : \mathbb{R} \mapsto \mathbb{R} \cup \{+\infty\}$ (called "potential") is a sufficiently regular and growing sufficiently fast at $\pm\infty$. (We set $\mathrm{e}^{-\infty} := 0$). More precisely we always assume that V is piecewise smooth and such that

$$\lim_{x\to\pm\infty}\frac{V(x)}{\log|x|} = +\infty. \tag{2.2.7}$$

The expression $V(M)$ is defined via the spectral theorem, namely $\mathrm{tr}\,V(M) = V(x_1) + \cdots + V(x_n)$ where x_i are the eigenvalues of M. The normalization constant Z_n is

$$Z_n = \int_{\mathbf{H}_n}\exp(-\mathrm{tr}\,V(M))\,\Omega_{\mathbf{H}_n}(\mathrm{d}M) \tag{2.2.8}$$

and we will see that (at least when V has no singularities) $Z_n < +\infty$ thanks to (2.2.7). //

Example 2.2.6. **1.** When $V(x) = x^2/2$ the unitary-invariant ensemble is called **Gaussian Unitary Ensemble** (GUE).

2. When $V(x) = \begin{cases} x - \alpha\log x, & x > 0, \\ +\infty, & x \le 0, \end{cases}$ the unitary-invariant ensemble is called **Laguerre Unitary Ensemble** (LUE); in this case α is a real number satisfying $\alpha > -1$, in order to have $Z_n < +\infty$. By the definition of V the measure is supported on the cone of positive-definite Hermitian matrices. Note that in this case $\exp(-\mathrm{tr}\,V(M)) = (\det M)^\alpha\exp(-\mathrm{tr}\,M)$ (recall that $\mathrm{e}^{\mathrm{tr}\,L} = \det(\mathrm{e}^L)$ for any matrix L). //

Exercise 2.2.7. Prove that the matrix entries of the GUE are independent (i.e., their joint probability distribution function factorizes). Compute Z_n for the GUE. //

According to the first part of this exercise, the GUE is a Wigner ensemble; we shall see later that it is the unique unitary-invariant ensemble to have this property.

Exercise 2.2.8. Prove that if Z is a square random matrix of size n with entries $Z_{ij} \sim \mathcal{N}(0,1) + \mathrm{i}\,\mathcal{N}(0,1)$ then $\frac{1}{2}(Z + Z^\dagger)$ is distributed as a GUE matrix. //

Here, for $\mu \in \mathbb{R}, \sigma \in \mathbb{R}_{>0}$, $\mathcal{N}(\mu,\sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2, defined by the probability density function

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right). \tag{2.2.9}$$

It is useful for the previous exercise to recall that if $X_1 \sim \mathcal{N}(\mu_1,\sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2,\sigma_2^2)$ are two Gaussian random variables, then their sum is also Gaussian, $X + Y \sim \mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$, and that if $X \sim \mathcal{N}(\mu,\sigma^2)$ then $cX \sim \mathcal{N}(c\mu,(c\sigma)^2)$, for any $c \in \mathbb{R}$.

2.3 Invariant measure on the unitary group and the Circular Unitary Ensemble

2.3.1 Generalities on invariant measures

Recall that a topological group is a group G equipped with a topology such that the map $(g,h) \mapsto gh$ is continuous as a map from $G \times G$ (with the product topology) to G, and the map $g \mapsto g^{-1}$ is continuous as a map from G to G.

Recall also that for a topological space X, a Borel measure is a measure defined on all the open sets of X (hence on all closed sets of X, and on all the Borel sigma-algebra of X).

Definition 2.3.1. Let G be a topological group. A Borel measure μ is said to be **left-invariant** if and only if

$$\mu(gE) = \mu(E), \tag{2.3.1}$$

for all $g \in G$ and all Borel subsets $E \subseteq G$, where we denote $gE := \{h \in G : g^{-1}h \in E\}$. Similarly, it is said to be **right-invariant** if and only if

$$\mu(Eg) = \mu(E), \tag{2.3.2}$$

for all $g \in G$ and all Borel subsets $E \subseteq G$, where we denote $Eg := \{h \in G : hg^{-1} \in E\}$. In case it is both left- and right-invariant it is then said to be **bi-invariant**, or even just **invariant** when no confusion should arise. //

Note that if μ is an invariant measure, so is $c\mu$ for any $c \in \mathbb{R}_{\geq 0}$. We say that two measures μ_1 and μ_2 are proportional if there are nonnegative constants $(c_1, c_2) \neq (0,0)$ such that $c_1\mu_1 = c_2\mu_2$.

Theorem 2.3.2 (Haar, 1933)**.** *Let G be a Hausdorff locally compact topological group. There exists a nonzero left-invariant measure on G; moreover any two left-invariant measures on G are proportional.*

The proof in this generality goes far beyond the scope of the course. Moreover, for Lie groups the situations simplifies a lot (and was know long before Haar's Theorem, at least since Hurwitz, around 1897), see for instance Remark 2.3.10. Since we will mostly work with $\mathbf{U}_n$, we will consider only this case explicitly.

Example 2.3.3. **1.** Any finite group is a topological group with the discrete topology. Any bi-invariant measure is a multiple of the counting measure.
2. For $(\mathbb{R}^d, +)$ any invariant measure is a scalar multiple of the Lebesgue measure.
3. In general, left- and right- invariance are distinct, and so there might not be a bi-invariant measure. As an example, consider the multiplicative group of 2×2 matrices of the form $\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ $(a \in \mathbb{R}_{>0}, b \in \mathbb{R})$; a left-invariant measure is $a^{-2}\mathrm{d}a\mathrm{d}b$ (and it is not right-invariant) and a right-invariant measure is $a^{-1}\mathrm{d}a\mathrm{d}b$ (but it is not left-invariant). For *compact* Lie groups however, left- and right-invariant measures are the same. //

2.3.2 Invariant measure on $\mathbf{U}_n$

$\mathbf{U}_n$ as a Lie group. $\mathbf{U}_n$ is the level set of $\{\mathbf{1}_n\}$[1] under the map

$$f : \mathbb{C}^{n\times n} \to \mathbf{H}_n \tag{2.3.3}$$

defined by

$$f(M) = MM^\dagger. \tag{2.3.4}$$

(Note that $MM^\dagger$ is always Hermitian.)

$\mathbf{U}_n$ is compact: the diagonal entries of $UU^\dagger = \mathbf{1}_n$ imply $\sum_{j=1}^n |U_{ij}|^2 = 1$ for all $i = 1, \ldots, n$, so that $\mathbf{U}_n$ is closed (level set of a continuous map f) and bounded ($|U_{ij}| \leq 1$) and so $\mathbf{U}_n$ is compact by the Heine–Borel theorem.

[1] We shall denote $\mathbf{1}_n$ the identity matrix of size n.

$\mathbf{U}_n$ is (path-)connected: by the spectral theorem (see Corollary 2.1.7) for any $U \in \mathbf{U}_n$ we have $U = V(\exp(\mathrm{i}\varphi_1), \ldots, \exp(\mathrm{i}\varphi_n))V^\dagger$ for some $V \in \mathbf{U}_n$ and some $\varphi_1, \ldots, \varphi_n \in \mathbb{R}$, and so the map $[0,1] \to \mathbf{U}_n$ given by $t \mapsto U(t)$ with $U(t) := V(\exp(\mathrm{i}t\varphi_1), \ldots, \exp(\mathrm{i}t\varphi_n))V^\dagger$ is a continuous path in $\mathbf{U}_n$ such that $U(0) = \mathbf{1}_n$ and $U(1) = U$.

Lemma 2.3.4. *For all $M \in \mathbb{C}^{n\times n}$ the differential $\mathrm{d}f_M : \mathbb{C}^{n\times n} \to \mathbf{H}_n$ is given by the formula*

$$\mathrm{d}f_M(X) = XM^\dagger + MX^\dagger. \tag{2.3.5}$$

For any M with $\det M \neq 0$, $\mathrm{d}f_M$ has full-rank (i.e., it is surjective).

Proof. The differential is given by the formula

$$\mathrm{d}f_M(X) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}(f(M+\epsilon X) - f(M)) = \lim_{\epsilon\to 0} \frac{1}{\epsilon}((M+\epsilon X)(M+\epsilon X)^\dagger - MM^\dagger) = XM^\dagger + MX^\dagger. \tag{2.3.6}$$

Equivalently $\mathrm{d}f_M(X) = XM^\dagger + (XM^\dagger)^\dagger$. Next, suppose Y is an Hermitian matrix and $M \in \mathbb{C}^{n\times n}$ satisfies $\det M \neq 0$; then

$$Y = XM^\dagger + (XM^\dagger)^\dagger \tag{2.3.7}$$

always admits the solution $X = \frac{1}{2}YM^{-\dagger}$. Therefore, $\mathrm{d}f_M$ is surjective whenever $\det M \neq 0$. □

It follows by the constant rank theorem that $\mathbf{U}_n = f^{-1}(\{\mathbf{1}_n\})$ is locally diffeomorphic to a domain in $\mathbb{R}^{n^2}$ (note that $\mathbb{C}^{n\times n}$ has real dimension $2n^2$ and $\mathbf{H}_n$ has real dimension n^2). In other words, one can locally parametrize $\mathbf{U}_n$ in terms of n^2 real coordinates, which we shall in general denote as $p_1, \ldots, p_{n^2}$.

Remark 2.3.5. In differential-geometric terms, we have just proved that $\mathbf{U}_n$ is a *real smooth manifold* of dimension n^2. Actually, matrix multiplication and inversion are smooth functions (rational functions of the entries with nonzero denominators) and so $\mathbf{U}_n$ is a *Lie group*. //

Example 2.3.6. **1.** $\mathbf{U}_1 = S^1 = \{z \in \mathbb{C} : \ |z| = 1\}$ and can be parametrized by a single real parameter ϕ as $\mathrm{e}^{\mathrm{i}\phi}$.

2. $\mathbf{U}_2$ consists of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $|a|^2 + |b|^2 = 1 = |c|^2 + |d|^2$ and $a\bar{c} + b\bar{d} = 0$. Thus any matrix in $\mathbf{U}_2$ can be represented as

$$\begin{pmatrix} \mathrm{e}^{\mathrm{i}(\psi_1-\psi_2)}\cos\theta & \mathrm{e}^{\mathrm{i}\psi_3}\sin\theta \\ -\mathrm{e}^{\mathrm{i}(\psi_1-\psi_3)}\sin\theta & \mathrm{e}^{\mathrm{i}\psi_2}\cos\theta \end{pmatrix} \tag{2.3.8}$$

in terms of $4 = 2^2$ real parameters $\psi_1, \psi_2, \psi_3, \theta$. //

Invariant measure on $\mathbf{U}_n$.

Definition 2.3.7. Given an Hermitian matrix $M \in \mathbf{H}_n$, $M = X + \mathrm{i}Y$ for real $n \times n$ symmetric and skew-symmetric matrices X and Y, respectively, we denote

$$((M)) = (X_{1,1}, \ldots, X_{n,n}, X_{1,2}, Y_{1,2}, \ldots, X_{1,n}, Y_{1,n}, X_{2,3}, Y_{2,3}, \ldots, X_{n-1,n}, Y_{n-1,n}) \in \mathbb{R}^{n^2}, \tag{2.3.9}$$

the real-valued row vector of size n^2 obtained by 'unrolling' the Hermitian matrix M (and omitting repetitions). //

Note that the standard metric on $\mathbf{H}_n$, see (2.2.2), can be written as (again, $M = X + \mathrm{i}Y$)

$$\langle M, M\rangle = \sum_{i=1}^{n} X_{ii}^2 + 2\sum_{1\le i<j\le n} (X_{ij}^2 + Y_{ij}^2) = ((M))\, D\, ((M))^\top, \tag{2.3.10}$$

for the $n^2 \times n^2$ diagonal matrix $D = (\underbrace{1, \ldots, 1}_{n \text{ times}}, \underbrace{2, \ldots, 2}_{n(n-1) \text{ times}})$.

Now, for any $U \in \mathbf{U}_n$ there exists $\widehat{U} \in \mathbb{R}^{n^2\times n^2}$ such that

$$((UMU^\dagger)) = ((M))\widehat{U}. \tag{2.3.11}$$

Lemma 2.3.8. *For any $U \in \mathbf{U}_n$, the matrix $\widehat{U} \in \mathbb{R}^{n^2\times n^2}$ has $\det \widehat{U} = 1$. In particular, for any list of n^2 Hermitian matrices $M_1, \dots, M_{n^2} \in \mathbf{H}_n$ and any $U \in \mathbf{U}_n$ we have*

$$\det\begin{pmatrix} ((U^\dagger M_1 U)) \\ \hline \vdots \\ \hline ((U^\dagger M_{n^2} U)) \end{pmatrix} = \det\begin{pmatrix} ((M_1)) \\ \hline \vdots \\ \hline ((M_{n^2})) \end{pmatrix} \tag{2.3.12}$$

Proof. From (2.3.10) and the fact that $\mathrm{Ad}_U : M \mapsto UMU^\dagger$ is an isometry we obtain

$$D = \widehat{U}\, D\, \widehat{U}^\top. \tag{2.3.13}$$

Taking determinants we obtain $\det \widehat{U} = \pm 1$. Since $\mathbf{U}_n$ is connected and $\widehat{\mathbf{1}_n} = \mathbf{1}_{n^2}$, necessarily $\det \widehat{U} = 1$. ◻

Proposition 2.3.9 (Invariant volume-form in local coordinates)**.** *Let $p = (p_1, \dots, p_{n^2})$ be any local coordinates on $\mathbf{U}_n$, and denote $p \mapsto U(p)$ the parametrization of $\mathbf{U}_n$. The differential form*

$$\Omega_{\mathbf{U}_n} = g(p_1, \dots, p_{n^2})\mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_{n^2}, \qquad g(p_1, \dots, p_{n^2}) := \det\begin{pmatrix} ((\mathrm{i}U^\dagger(p)\frac{\partial U(p)}{\partial p_1})) \\ \hline \vdots \\ \hline ((\mathrm{i}U^\dagger(p)\frac{\partial U(p)}{\partial p_{n^2}})) \end{pmatrix} \tag{2.3.14}$$

gives an invariant measure on $\mathbf{U}_n$.

By the general uniqueness of invariant measures, any invariant measure on $\mathbf{U}_n$ is a multiple of $\Omega_{\mathbf{U}_n}$.

Proof. We have two sanity checks to perform before delving into the proof proper.

First, the matrix $\mathrm{i}U^\dagger \frac{\partial U}{\partial p_i}$ is Hermitian: this is left as an **exercise**.

Second, the differential form $\Omega_{\mathbf{U}_n}$ does not depend on the choice of coordinates p_i. Namely, suppose we have two different parametrizations $p \mapsto U(p)$ and $\widetilde{p} \mapsto \widetilde{U}(\widetilde{p})$, and the change of coordinates (defined by $U(p) = \widetilde{U}(\widetilde{p})$) is $\widetilde{p}_k = \widetilde{p}_k(p_1, \dots, p_{n^2})$ $(1 \le k \le n^2)$. Then, as for a general transformation, we have

$$\mathrm{d}\widetilde{p}_1 \wedge \cdots \wedge \mathrm{d}\widetilde{p}_{n^2} = \det_{1\le i,j\le n^2}\left(\frac{\partial \widetilde{p}_i}{\partial p_j}\right) \mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_{n^2}. \tag{2.3.15}$$

On the other hand, by the chain rule,

$$\mathrm{i}\widetilde{U}^\dagger(\widetilde{p})\frac{\partial \widetilde{U}(\widetilde{p})}{\partial \widetilde{p}_j} = \sum_{k=1}^{n^2} \mathrm{i}\frac{\partial p_k}{\partial \widetilde{p}_j} U^\dagger(p)\frac{\partial U(p)}{\partial p_k}, \qquad \text{for all } 1 \le j \le n^2 \tag{2.3.16}$$

and this relation can be re-written as the matrix identity

$$\begin{pmatrix} ((\mathrm{i}\widetilde{U}^\dagger(\widetilde{p})\frac{\partial \widetilde{U}(\widetilde{p})}{\partial \widetilde{p}_1})) \\ \hline \vdots \\ \hline ((\mathrm{i}\widetilde{U}^\dagger(\widetilde{p})\frac{\partial \widetilde{U}(\widetilde{p})}{\partial \widetilde{p}_{n^2}})) \end{pmatrix} = \begin{pmatrix} \frac{\partial p_1}{\partial \widetilde{p}_1} & \cdots & \frac{\partial p_{n^2}}{\partial \widetilde{p}_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial p_1}{\partial \widetilde{p}_{n^2}} & \cdots & \frac{\partial p_{n^2}}{\partial \widetilde{p}_{n^2}} \end{pmatrix} \cdot \begin{pmatrix} ((\mathrm{i}U^\dagger(p)\frac{\partial U(p)}{\partial p_1})) \\ \hline \vdots \\ \hline ((\mathrm{i}U^\dagger(p)\frac{\partial U(p)}{\partial p_{n^2}})) \end{pmatrix}. \tag{2.3.17}$$

Taking determinants we see that

$$g(\widetilde{p}_1, \dots, \widetilde{p}_{n^2}) = \det_{1\le i,j\le n^2}\left(\frac{\partial p_i}{\partial \widetilde{p}_j}\right) g(p_1, \dots, p_{n^2}) \tag{2.3.18}$$

and combining (2.3.15) with (2.3.18) we obtain

$$g(\widetilde{p}_1, \dots, \widetilde{p}_{n^2})\mathrm{d}\widetilde{p}_1 \wedge \cdots \wedge \mathrm{d}\widetilde{p}_{n^2} = g(p_1, \dots, p_{n^2})\mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_{n^2}, \tag{2.3.19}$$

namely, the definition of $\Omega_{\mathbf{U}_n}$ does not depend on the choice of coordinates.

Now for the proof of invariance; we need to show that the transformation $U \mapsto VU$ (for a fixed $V \in \mathbf{U}_n$) preserves $\Omega_{\mathbf{U}_n}$. For a given parametrization $p \mapsto U(p)$, this transformation corresponds to a diffeomorphism $p = (p_1, \ldots, p_{n^2}) \mapsto \widetilde{p} = (\widetilde{p}_1, \ldots, \widetilde{p}_{n^2})$ between domains of $\mathbb{R}^{n^2}$, defined by the property that $U(\widetilde{p}) = VU(p)$. Again, the measure $\mathrm{d}p_1 \wedge \cdots \wedge \mathrm{d}p_{n^2}$ transforms according to the general rule (2.3.15); on the other hand

$$\mathrm{i}U^\dagger(\widetilde{p})\frac{\partial U(\widetilde{p})}{\partial \widetilde{p}_j} = \sum_{k=1}^{n^2} \mathrm{i}\frac{\partial p_k}{\partial \widetilde{p}_j}(VU^\dagger(p))\frac{\partial (VU(p))}{\partial p_k} = \sum_{k=1}^{n^2} \mathrm{i}\frac{\partial p_k}{\partial \widetilde{p}_j}U^\dagger(p)\frac{\partial U(p)}{\partial p_k}. \tag{2.3.20}$$

Then the proof continues as the proof of invariance of $\Omega_{\mathbf{U}_n}$ under coordinate changes.

In general for compact Lie groups, left-invariance (which we have proved) implies right-invariance; to see it directly, the transformation $U \mapsto UV$ (for a fixed $V \in \mathbf{U}_n$), in terms of a parametrization $p \mapsto U(p)$, corresponds to a diffeomorphism $p = (p_1, \ldots, p_{n^2}) \mapsto \widetilde{p} = (\widetilde{p}_1, \ldots, \widetilde{p}_{n^2})$ between domains of $\mathbb{R}^{n^2}$, defined by the property that $U(\widetilde{p}) = U(p)V$. Now we have

$$\mathrm{i}U^\dagger(\widetilde{p})\frac{\partial U(\widetilde{p})}{\partial \widetilde{p}_j} = \sum_{k=1}^{n^2} \mathrm{i}\frac{\partial p_k}{\partial \widetilde{p}_j}(U^\dagger(p)V)\frac{\partial (U(p)V)}{\partial p_k} = V^\dagger\left(\sum_{k=1}^{n^2} \mathrm{i}\frac{\partial p_k}{\partial \widetilde{p}_j}U^\dagger(p)\frac{\partial U(p)}{\partial p_k}\right)V. \tag{2.3.21}$$

Taking into account Lemma 2.3.8, the proof of right-invariance proceeds parallel to that for left-invariance. □

Remark 2.3.10. In general, left (resp. right) invariant measures for Lie groups G can be obtained similarly by taking the wedge product of the independent entries of the one-form $g^{-1}\mathrm{d}g$ (resp. $\mathrm{d}gg^{-1}$), $g \in G$. //

Example 2.3.11. For $\mathbf{U}_1$, whose elements are just phases $\mathrm{e}^{\mathrm{i}\phi}$, we have $\Omega_{\mathbf{U}_1} = \mathrm{i}\mathrm{e}^{-\mathrm{i}\phi}\frac{\partial \mathrm{e}^{\mathrm{i}\phi}}{\partial \phi}\mathrm{d}\phi = -\mathrm{d}\phi$, which is clearly invariant. For $\mathbf{U}_2$, with the notations of Example 2.3.6, with some effort one can compute $\Omega_{\mathbf{U}_2} = \frac{1}{2}\sin(2\theta)\mathrm{d}\theta \wedge \mathrm{d}\psi_1 \wedge \mathrm{d}\psi_2 \wedge \mathrm{d}\psi_3$. //

Remark 2.3.12. Let us recall the general procudure by which a volume-form (in this case, $\Omega_{\mathbf{U}_n}$) allows one to integrate functions on a manifold (in this case, $\mathbf{U}_n$), by a partition of unity argument. By compactness and orientability of $\mathbf{U}_n$, there exists a finite number (say, K) of local parametrizations $U_a : \mathcal{D}_a \to \mathbf{U}_n : p \mapsto U_a(p)$, $a = 1, \ldots, K$, which are diffeomorphisms from open domains $\mathcal{D}_a \subset \mathbb{R}^{n^2}$ onto open domains $U_a(\mathcal{D}_a) \subset \mathbf{U}_n$ and such that the change of coordinates between any two parametrizations is orientation-preserving. It is possible to construct smooth functions $\rho_a : \mathbf{U}_n \to \mathbb{R}$ such that $\sum_{a=1}^K \rho_a(U) = 1$ for all $U \in \mathbf{U}_n$ and $\rho_a(U) = 0$ for all $U \in \mathbf{U}_n \setminus U_a(\mathcal{D}_a)$ (a "partition of unity"). Then, for any (sufficiently regular) function $\psi : \mathbf{U}_n \to \mathbb{C}$,

$$\int_{\mathbf{U}_n} \psi(U)\,\Omega_{\mathbf{U}_n}(\mathrm{d}U) := \sum_{a=1}^K \int_{\mathcal{D}_a} \rho_a(U_a(p))\psi(U_a(p))\det\begin{pmatrix} ((\mathrm{i}U_a^\dagger(p)\frac{\partial U_a(p)}{\partial p_1})) \\ \hline \vdots \\ \hline ((\mathrm{i}U_a^\dagger(p)\frac{\partial U_a(p)}{\partial p_{n^2}})) \end{pmatrix}\mathrm{d}p_1 \cdots \mathrm{d}p_{n^2}. \tag{2.3.22}$$

where on the right we have integrals over domains in $\mathbb{R}^{n^2}$. The construction does not depend on the choice of local parametrization and of partition of unity. In particular one can define the invariant measure on Borel subsets of $\mathbf{U}_n$ by integrating their indicator functions. //

Since $\mathbf{U}_n$ is compact, the total measure $\int_{\mathbf{U}_n} \Omega_{\mathbf{U}_n}$ is a finite number. This allows us to give the following definition.

Definition 2.3.13. The **Circular Unitary Ensemble** (CUE) is $\mathbf{U}_n$ with the unique invariant probability measure. //

Note that in this case there is no need to add a weight function in front of the measure $\Omega_{\mathbf{U}_n}$ to ensure that the total measure is finite (and so that we can define a probability measure), as we had to do in Definition 2.2.5. One could of course add a general weight of the form $\exp\operatorname{tr}(V(U))$ in front of $\Omega_{\mathbf{U}_n}$ to obtain more general models.

Remark 2.3.14. For both GUE and CUE, their probability densities $p(M)\Omega_{\mathcal{M}}(\mathrm{d}M)$ maximize the *information entropy*

$$S[p] := -\int_{\mathcal{M}} p(M)\log(p(M))\,\Omega_{\mathcal{M}}(\mathrm{d}M), \qquad \mathcal{M} = \mathbf{U}_n, \mathbf{H}_n \tag{2.3.23}$$

(in the Hermitian case with the constraint that $\mathbb{E}[M^2]$ is kept constant, otherwise there is no maximum). This is a natural assumption from the point of view of Statistical Mechanics.
To make it more clear, let us consider the scalar case only.

- Among all probability measures $p(\phi)\mathrm{d}\phi$ on $S^1 = \mathbf{U}_1$, the maximum entropy is attained by the uniform measure $\mathrm{d}\phi/2\pi$.
- Among all measures $p(x)\mathrm{d}x$ on $\mathbb{R} = \mathbf{H}_1$ with fixed variance σ^2, the maximum entropy is attained by Gaussian distributions of variance σ^2.[2]

//

2.3.3 More about the invariant measure on $\mathbf{U}_n$

Gaussian distribution in $\mathbb{C}^{n\times n}$ and Gram–Schmidt algorithm. Let $\mathrm{GL}_n(\mathbb{C}) \subseteq \mathbb{C}^{n\times n}$ the set of invertible matrices, i.e. $M \in \mathrm{GL}_n(\mathbb{C})$ if and only if $\det M \neq 0$. $\mathrm{GL}_n(\mathbb{C})$ is an open set of full measure in $\mathbb{C}^{n\times n}$ (see also Remark 2.4.4 below for general facts about full measure sets in $\mathbb{C}^N$). We regard the Gram–Schmidt algorithm as a map

$$\mathsf{GS} : \mathrm{GL}_n(\mathbb{C}) \to \mathbf{U}_n. \tag{2.3.24}$$

More precisely the output of the function GS on the matrix $M = (m_1|\cdots|m_n)$ (where $m_j \in \mathbb{C}^{n\times 1}$ are the columns of M) is given by the unitary matrix $U = (u_1|\cdots|u_n)$ where the columns are $u_j := \widetilde{u}_j/\|\widetilde{u}_j\|$ with

$$\widetilde{u}_1 := m_1, \quad \widetilde{u}_2 := m_2 - (m_2, u_1)u_1, \ \ldots, \widetilde{u}_j := m_j - \sum_{\ell=1}^{j-1}(m_j, u_\ell)u_\ell, \ \ldots \tag{2.3.25}$$

Lemma 2.3.15. *For any $U \in \mathbf{U}_n$ and $M \in \mathrm{GL}_n(\mathbb{C})$, we have $\mathsf{GS}(UM) = U\mathsf{GS}(M)$.*

Proof. Exercise (it may be useful to note that $\mathsf{GS}(M)$ is the unique unitary matrix whose first k columns generate the same space as the first k columns of M, for all $k = 1, \ldots, n$; otherwise use the explicit formula (2.3.25)). □

Proposition 2.3.16. *Let μ be the probability measure on $\mathrm{GL}_n(\mathbb{C})$, obtained by restriction of the Gaussian measure*

$$\frac{1}{(2\pi)^{n^2}}\mathrm{e}^{-\mathrm{tr}\,(M^\dagger M)/2}\prod_{1\le i,j\le n}\mathrm{d}X_{ij}\mathrm{d}Y_{ij}, \qquad M = X + \mathrm{i}Y, \tag{2.3.26}$$

on $\mathbb{C}^{n\times n}$. The pushforward measure $\mathsf{GS}_*\mu$ is the invariant probability measure on $\mathbf{U}_n$.

This provides a convenient method for sampling the invariant probability measure on $\mathbf{U}_n$.

Before the proof, let us recall that if (X, Σ) and (X', Σ') are measurable spaces, $f : X \to X'$ is a measurable map (i.e., $f^{-1}(E) \in \Sigma$ for any $E \in \Sigma'$), and μ is a measure on (X, Σ), then $f_*\mu$ (the *pushforward measure*) is the measure on (X', Σ') defined by $(f_*\mu)(E) = \mu(f^{-1}(E))$ for any $E \in \Sigma'$. Observe that the push-forward of a probability measure is also a probability measure.

Proof. Let $E \subseteq \mathbf{U}_n$ be a Borel set, and $U \in \mathbf{U}_n$. Then

$$\mathsf{GS}_*\mu(UE) \overset{(1)}{=} \mu(\mathsf{GS}^{-1}(UE)) \overset{(2)}{=} \mu(U\mathsf{GS}^{-1}(E)) \overset{(3)}{=} \mu(\mathsf{GS}^{-1}(E)) \overset{(4)}{=} \mathsf{GS}_*\mu(E).$$

Let us prove each equality.

[2] There is no maximum of the entropy among all probability measures on $\mathbb{R}$. For instance, the uniform distribution supported on the interval $[-L/2, L/2]$ has entropy $\log L$, which can be arbitrarily large as we send $L \to +\infty$; the first moment of these distributions is always 0, however the variance is $L^2/12$ which diverges to $+\infty$ when $L \to +\infty$, so fixing the variance is a natural constraint.

(1) This is the definition of pushforward measure.

(2) It is enough to prove that $\mathsf{GS}^{-1}(UE) = U\mathsf{GS}^{-1}(E)$, which follows from

$$\begin{aligned}\mathsf{GS}^{-1}(UE) &= \{M \in \mathrm{GL}_n(\mathbb{C}) : \ \mathsf{GS}(M) \in UE\} \\ &= \left\{M \in \mathrm{GL}_n(\mathbb{C}) : \ U^\dagger \mathsf{GS}(M) \in E\right\} \\ &\overset{(*)}{=} \left\{M \in \mathrm{GL}_n(\mathbb{C}) : \ \mathsf{GS}(U^\dagger M) \in E\right\} \\ &= U \ \{M \in \mathrm{GL}_n(\mathbb{C}) : \ \mathsf{GS}(M) \in E\} = U\mathsf{GS}^{-1}(E),\end{aligned} \tag{2.3.27}$$

where the equality $(*)$ is a consequence of Lemma 2.3.15

(3) We have in general that $\mu(UX) = \mu(X)$ for any $X \subseteq \mathrm{GL}_n(\mathbb{C})$ Borel subset by the definition (2.3.26) of μ.

(4) This is the definition of pushforward measure.

□

As a sequence of sampling of unit vectors. Another interpretation of the invariant probability measure on $\mathbf{U}_n$ is given by the following way of sampling; take a vector u_1 at random from $S^{2n-1} = \{v \in \mathbb{C}^n : (v, v) = 1\}$, where the probability measure is given by the standard volume form on S^{2n-1} (i.e., the uniform distribution on the sphere).

Then pick another vector u_2 at random, this time from the sphere $S^{2n-3} = \{v \in \mathbb{C}^n : (v, v) = 1, \ (v, u_1) = 0\}$, again with respect to the uniform distribution on the sphere.

Proceed similarly by picking the vector u_j at random from the sphere $S^{2(n-j)+1} = \{v \in \mathbb{C}^n : (v, v) = 1, \ (v, u_1) = \cdots = (v, u_{j-1}) = 0\}$, for $j = 1, \ldots, n$.

The matrix $U = (u_1|\ldots|u_n)$ is a random unitary matrix and it is possible to prove that it is distributed according to the invariant probability measure on $\mathbf{U}_n$ (essentially because $\mathbf{U}_n$ preserves the area form on the spheres).

2.4 Weyl integration formula(s)

2.4.1 The Hermitian case

The idea of Weyl integration formula is to consider (thanks to the spectral theorems) eigenvalues and eigenvectors, instead of the entries, to parametrize a matrix.

For example, for an Hermitian matrix M we would like to consider a change of variables $M \mapsto (U, (x_1, \ldots, x_n))$ where U is unitary and $x_i \in \mathbb{R}$ are such that $M = U\mathrm{diag}(x_1, \ldots, x_n)U^\dagger$. However, such a map $M \mapsto (U, (x_1, \ldots, x_n))$, as it stands, is not well defined for the following reasons:

- the vector of eigenvalues $(x_1, \ldots, x_n)$ is defined only up to permutations, and
- the matrix $U \in \mathbf{U}_n$ which puts M in diagonal form is also not unique.

A way around is to introduce

$$\mathbf{Q}_n := \{U \in \mathbf{U}_n : \ U_{ii} > 0, \ i = 1, \ldots, n\}, \quad \mathfrak{X}_n := \{(x_1, \ldots, x_n) \in \mathbb{R}^n : \ x_1 < \cdots < x_n\}. \tag{2.4.1}$$

Lemma 2.4.1. *$\mathbf{Q}_n$ is a smooth manifold of dimension $m = n(n-1)$, i.e. it is locally parametrized by m real coordinates $q_1, \ldots, q_m$.*

Proof. The proof is similar to the argument used for $\mathbf{U}_n$; first, the subset $\mathbf{U}_n^0$ of unitary matrices U with nonzero diagonal entries is open in $\mathbf{U}_n$, and hence a smooth manifold itself of dimension n^2. Next, $\mathbf{Q}_n$ is the preimage of $(1, \ldots, 1)$ under the map $v : \mathbf{U}_n^0 \to (S^1)^n : U \mapsto (\arg U_{11}, \ldots, \arg U_{nn})$ (here $\arg z = z/|z|$ for any nonzero complex number z) whose differential is surjective. □

Example 2.4.2. Using the parametrization of $\mathbf{U}_2$ in Example 2.3.6 we deduce that we can parametrize (an open set in) $\mathbf{Q}_2$ as

$$\begin{pmatrix} \cos\theta & e^{i\psi}\sin\theta \\ -e^{-i\psi}\sin\theta & \cos\theta \end{pmatrix} \tag{2.4.2}$$

in terms of 2 real parameters ψ, θ. //

Lemma 2.4.3. *The map $\phi : \mathbf{Q}_n \times \mathcal{X}_n \to \mathbf{H}_n$ defined by*

$$(U, (x_1, \dots, x_n)) \mapsto U \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix} U^\dagger \tag{2.4.3}$$

is smooth, injective, and its image has full measure (that is to say, $\mathbf{H}_n \setminus \phi(\mathbf{Q}_n \times \mathcal{X}_n)$ has measure zero with respect to the Lebesgue measure $\Omega_{\mathbf{H}_n}$).

Remark 2.4.4. Before the proof let us recall some useful facts.

1. Any subset of $\mathbb{R}^d$ contained in the zero locus of a nonzero polynomial (or more generally, analytic) function on $\mathbb{R}^d$ has Lebesgue measure zero. This is most easily shown by induction on $d \geq 1$ with the help of Fubini Theorem.

2. Given two polynomials $P(\lambda) = \sum_{i=0}^{d} p_i \lambda^{d-i}, Q(\lambda) = \sum_{j=0}^{e} q_j \lambda^{e-j}$, their resultant is defined as

$$\mathrm{Res}(P, Q) := \det \begin{pmatrix} p_0 & p_1 & \cdots & p_d & 0 & \cdots & 0 \\ 0 & p_0 & \cdots & p_{d-1} & p_d & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p_0 & \cdots & \cdots & p_d \\ q_0 & q_1 & \cdots & \cdots & q_e & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & q_0 & \cdots & \cdots & \cdots & q_e \end{pmatrix}. \tag{2.4.4}$$

   (The matrix in this definition is a square matrix of size $d + e$.) $\mathrm{Res}(P, Q)$ is a polynomial function of the coefficients p_i, q_j, $0 \leq i \leq d$, $0 \leq j \leq e$; moreover, P, Q share a root if and only if $\mathrm{Res}(P, Q) = 0$. To prove it, let $\mathbb{C}[\lambda]_{<d}$ be the vector space of polynomials in x of degree at $< d$ and consider the map

$$\begin{array}{rccc} R: & \mathbb{C}[\lambda]_{<e} \oplus \mathbb{C}[\lambda]_{<d} & \to & \mathbb{C}[\lambda]_{<d+e} \\ & (U, V) & \mapsto & PU + QV. \end{array} \tag{2.4.5}$$

   It is then a simple exercise to show that R is represented in the monomial basis by the matrix in (2.4.4) whose determinant defines the resultant, and that R has a nontrivial kernel if and only if P, Q share a common root.

3. As a particular case of the above construction, a (wlog monic) polynomial P of degree $d \geq 2$ has at least a root of multiplicity ≥ 2 if and only if its "discriminant" $\Delta(P) := (-1)^{d(d-1)/2}\mathrm{Res}(P, P')$ vanishes. For example for $P(\lambda) = \lambda^2 + a\lambda + b$

$$\Delta(P) = -\mathrm{Res}(P, P') = -\det \begin{pmatrix} 1 & a & b \\ 2 & a & 0 \\ 0 & 2 & a \end{pmatrix} = a^2 - 4b \tag{2.4.6}$$

   is the familiar discriminant of a quadratic polynomial. Similarly, for $P = x^3 + ax^2 + bx + c$ we have $\Delta(P) = a^2b^2 - 4a^3c - 4b^3 + 18abc - 27c^2$. It is an interesting exercise to prove that $\Delta(P) = \prod_{1 \leq i < j \leq d}(r_i - r_j)^2$ where $r_1, \dots, r_d$ are the roots of P, in any order. (This explains the pre-factor $(-1)^{d(d-1)/2}$.)

//

Proof of lemma. To see that ϕ is injective, assume $U, U' \in \mathbf{Q}_n$ and $x_1 < \cdots < x_n, x_1' < \cdots < x_n'$ are such that

$$U \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix} U^\dagger = U' \begin{pmatrix} x_1' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n' \end{pmatrix} (U')^\dagger. \tag{2.4.7}$$

By taking the characteristic polynomial of both sides of this identity we find that $x_j = x_j'$ for all $j = 1, \ldots, n$ (recall that the x_j's and x_j''s are ordered). Then setting $V = U^\dagger U'$ we get

$$\begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix} V = V \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix}, \tag{2.4.8}$$

i.e. $x_i V_{ij} = V_{ij} x_j$ and so $V_{ij} = 0$ unless $i = j$, namely V is diagonal. V is a unitary matrix, whence $|V_{ii}| = 1$; finally $U_{ii}' = V_{ii} U_{ii}$ implies, as $U, U' \in \mathbf{Q}_n$, that $V_{ii} = 1$. Thus $V = \mathbf{1}$ and so $U = U'$.

It remains to prove that the image of ϕ has full measure. For, $\mathbf{H}_n \setminus \phi(\mathbf{Q}_n \times \mathcal{X}_n) \subseteq \mathcal{Z}_1 \cup \mathcal{Z}_2$ where $\mathcal{Z}_1$ is the set of hermitian matrices with at least one eigenvalue of multiplicity ≥ 2, and $\mathcal{Z}_2$ is the set of hermitian matrices that can be written as $UDU^\dagger$ with D (real) diagonal and $U_{ii} = 0$ for some $1 \leq i \leq n$. Using the facts summarized before this proof we can easily show that $\mathcal{Z}_1, \mathcal{Z}_2$ both have measure zero, concluding the proof of the lemma.

Setting $P(\lambda) := \det(M - \lambda \mathbf{1}_n)$ we see that $\mathcal{Z}_1$ is the zero locus of the discriminant $\Delta(P)$, which is a nonzero polynomial function in the entries of $M \in \mathbf{H}_n$. Hence $\mathcal{Z}_1$ has measure zero.

To see that $\mathcal{Z}_2$ has measure zero too we first consider the set $\mathcal{Z}_2'$ of Hermitian matrices $M = UDU^\dagger$ with $U \in \mathbf{U}_n$, D diagonal and $U_{11} = 0$. Equivalently, $\mathcal{Z}_2'$ is the set of Hermitian matrices M with an eigenvector orthogonal to e_1, so that writing M in block-form

$$M = \left(\begin{array}{c|c} a & b^\dagger \\ \hline b & C \end{array} \right) \tag{2.4.9}$$

for $a \in \mathbb{R}$, $b \in \mathbb{C}^{n-1}$, $C \in \mathbf{H}_{n-1}$, the condition is equivalent to existence of an eigenvector of C which is orthogonal to b. Setting $P(\lambda) := \det(C - \lambda \mathbf{1}_{n-1})$, $Q(\lambda) := \det(C + bb^\dagger - \lambda \mathbf{1}_{n-1})$, we conclude that $\mathcal{Z}_2'$ is contained in the zero locus of the resultant $\mathrm{Res}(P, Q)$, which is a nonzero polynomial function of the entries of M. Thus $\mathcal{Z}_2'$ has measure zero. In general if $M = UDU^\dagger$ with D diagonal, $U \in \mathbf{U}_n$ with $U_{ii} = 0$ then $\Pi H \Pi \in \mathcal{Z}_2'$, where Π is the permutation matrix

$$\Pi_{r,s} = \begin{cases} 1 & \text{if } r = 1, s = i \text{ or } s = 1, r = i \text{ or } 1 \neq r = s \neq j, \\ 0 & \text{otherwise,} \end{cases} \tag{2.4.10}$$

and we can reason in the same way to show that this set is contained in the zero locus of a polynomial function of the entries of M. Thus we conclude that $\mathcal{Z}_2$ is contained in the union of n measure zero sets, therefore it has measure zero. □

We are ready to state and prove Weyl integration formula.

Definition 2.4.5. Given n variables $\xi_1, \ldots, \xi_n$, the **Vandermonde determinant** $\Delta(\xi_1, \ldots, \xi_n)$ is defined by

$$\Delta(\xi_1, \ldots, \xi_n) := \prod_{1 \leq i < j \leq n} (\xi_j - \xi_i). \tag{2.4.11}$$

The term "determinant" is clarified in Lemma 2.5.1. //

Theorem 2.4.6. *Let $f \in L^1(\mathbf{H}_n, \Omega_{\mathbf{H}_n})$ be a class-function, i.e. for all $U \in \mathbf{U}_n$ we have $f(UMU^\dagger) = f(M)$ for almost all $M \in \mathbf{H}_n$. Then*

$$\int_{\mathbf{H}_n} f(M)\, \Omega_{\mathbf{H}_n}(\mathrm{d}M) = c_n \int_{\mathbb{R}^n} \widetilde{f}(x_1, \ldots, x_n)\, \Delta^2(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \tag{2.4.12}$$

where $\widetilde{f}(x_1,\dots,x_n) := f(\begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix})$ *and* c_n *is a constant depending on* n *only.*

Proof. We need to compute the Jacobian of the change of coordinates given by ϕ (which is defined on a full measure set by Lemma 2.4.3): using coordinates $q = (q_1,\dots,q_m)$ $(m = n(n-1))$ on $\mathbf{Q}_n$, we have

$$\frac{\partial \phi}{\partial x_j} = U(q)E_{jj}U^\dagger(q), \qquad j = 1,\dots,n \tag{2.4.13}$$

and, denoting from now on

$$D := \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_n \end{pmatrix}, \tag{2.4.14}$$

we have (**exercise**)

$$\frac{\partial \phi}{\partial q_s} = U(q)\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_s}, D\right]U^\dagger(q), \qquad s = 1,\dots,m. \tag{2.4.15}$$

where $[A,B] := AB - BA$ is the commutator. Therefore the Jacobian determinant is

$$\det\begin{pmatrix} ((\,U(q)E_{11}U^\dagger(q)\,)) \\ \hline \vdots \\ \hline ((\,U(q)E_{nn}U^\dagger(q)\,)) \\ \hline ((\,U(q)\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_1}, D\right]U^\dagger(q)\,)) \\ \hline \vdots \\ \hline ((\,U(q)\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_m}, D\right]U^\dagger(q)\,)) \end{pmatrix} = \det\begin{pmatrix} ((\,E_{11}\,)) \\ \hline \vdots \\ \hline ((\,E_{nn}\,)) \\ \hline ((\,\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_1}, D\right]\,)) \\ \hline \vdots \\ \hline ((\,\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_m}, D\right]\,)) \end{pmatrix} \tag{2.4.16}$$

where we use Lemma 2.3.8. Denoting $U^\dagger(q)\frac{\partial U(q)}{\partial q_s} =: X^{(s)} + \mathrm{i}Y^{(s)}$ (for $X^{(s)}, Y^{(s)} \in \mathbb{R}^{m\times m}$), the entry i,j of

$$\left[U^\dagger(q)\frac{\partial U(q)}{\partial q_s}, D\right] \tag{2.4.17}$$

is $(x_j - x_i)(X^{(s)}_{ij} + \mathrm{i}Y^{(s)}_{ij})$ (see Exercise 2.4.8) and so the previous determinant is equal to

$$\det\left(\begin{array}{c|c} \mathbf{1}_n & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{J} \end{array}\right) = \det \mathbf{J}, \tag{2.4.18}$$

where the $m\times m$ matrix $\mathbf{J}$ is

$$\begin{pmatrix} X^{(1)}_{12}(x_2-x_1) & Y^{(1)}_{12}(x_2-x_1) & X^{(1)}_{13}(x_3-x_1) & Y^{(1)}_{13}(x_3-x_1) & \cdots & X^{(1)}_{n-1,n}(x_n-x_{n-1}) & Y^{(1)}_{n-1,n}(x_n-x_{n-1}) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X^{(m)}_{12}(x_2-x_1) & Y^{(m)}_{12}(x_2-x_1) & X^{(m)}_{13}(x_3-x_1) & Y^{(m)}_{13}(x_3-x_1) & \cdots & X^{(m)}_{n-1,n}(x_n-x_{n-1}) & Y^{(m)}_{n-1,n}(x_n-x_{n-1}) \end{pmatrix}$$

Denoting $\Omega_{\mathbf{Q}_n} := h(q_1,\dots,q_m)\mathrm{d}q_1\cdots\mathrm{d}q_m$, where

$$h := \det\begin{pmatrix} X^{(1)}_{12} & Y^{(1)}_{12} & X^{(1)}_{13} & Y^{(1)}_{13} & \cdots & X^{(1)}_{n-1,n} & Y^{(1)}_{n-1,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ X^{(m)}_{12} & Y^{(m)}_{12} & X^{(m)}_{13} & Y^{(m)}_{13} & \cdots & X^{(m)}_{n-1,n} & Y^{(m)}_{n-1,n} \end{pmatrix}, \tag{2.4.19}$$

we have proven that (recall formula (2.2.3))

$$\phi^*\Omega_{\mathbf{H}_n} = 2^{n(n-1)/2}\Delta^2(x_1,\dots,x_n)\,\mathrm{d}x_1\cdots\mathrm{d}x_n\,\Omega_{\mathbf{Q}_n}. \tag{2.4.20}$$

Finally, we have the following chain of equalities

$$\begin{aligned}\int_{\mathbf{H}_n} f(M)\,\Omega_{\mathbf{H}_n}(\mathrm{d}M) &= \int_{\phi(\mathbf{Q}_n\times\mathcal{X}_n)} f(M)\,\Omega_{\mathbf{H}_n}(\mathrm{d}M)\\ &= 2^{n(n-1)/2}\int_{\mathbf{Q}_n\times\mathcal{X}_n}\widetilde{f}(x_1,\dots,x_n)\Delta^2(x_1,\dots,x_n)\Omega_{\mathbf{Q}_n}\mathrm{d}x_1\dots\mathrm{d}x_n\\ &= \frac{2^{n(n-1)/2}}{n!}\left(\int_{\mathbf{Q}_n}\Omega_{\mathbf{Q}_n}\right)\int_{\mathbb{R}^n}\widetilde{f}(x_1,\dots,x_n)\mathrm{d}x_1\dots\mathrm{d}x_n,\end{aligned} \tag{2.4.21}$$

In the last step we have noted that $\widetilde{f}$ is a symmetric function of the x_i's (because permutation matrices are unitary matrices). The proof is complete by setting $c_n := \frac{2^{n(n-1)/2}}{n!}\left(\int_{\mathbf{Q}_n}\Omega_{\mathbf{Q}_n}\right)$. □

Remark 2.4.7. $\mathbf{Q}_n$ is diffeomomorphic to an open dense set in the homogeneous space $\mathbf{U}_n/\mathbf{U}_1^n$ (where $\mathbf{U}_1^n\subset\mathbf{U}_n$ is the subgroup of diagonal unitary matrices and acts on $\mathbf{U}_n$ by multiplication on the right). The measure $\Omega_{\mathbf{Q}_n} := h(q_1,\dots,q_m)\mathrm{d}q_1\cdots\mathrm{d}q_m$ defined in this proof is an invariant measure on this homogeneous space. //

Exercise 2.4.8. Let D be a diagonal $k\times k$ matrix with diagonal entries $d_1,\dots,d_k$ and let $A\in\mathbb{C}^{k\times k}$. Prove that the entries of $[A,D]=AD-DA$ are $[A,D]_{ij}=A_{ij}(d_j-d_i)$ //

Exercise 2.4.9. Prove by a direct computation the case $n=2$ of Theorem 2.4.6, using the parametrization (2.4.2) of $\mathbf{Q}_2$. //

Exercise 2.4.10. Compute the marginal eigenvalue distribution for an Unitary Invariant Ensemble of Hermitian matrices. //

Exercise 2.4.11. Compute the marginal eigenvalue distribution for an ensemble of positive-definite Hermitian matrices with joint probability distribution

$$\frac{1}{C_n}\exp(-\mathrm{tr}\,(M^2)-(\mathrm{tr}\,(M))^2)\,\Omega_{\mathbf{H}_n}(\mathrm{d}M), \tag{2.4.22}$$

where C_n is the normalizing constant. Is this a Unitary-Invariant Ensemble in the sense of Definition 2.2.5? //

2.4.2 The unitary case

Let

$$\mathcal{Y}_n := \{(\varphi_1,\dots,\varphi_n)\in\mathbb{R}^n : 0<\varphi_1<\dots<\varphi_n<2\pi\}. \tag{2.4.23}$$

We have the following analogue of Lemma 2.4.3, which is proved similarly.

Lemma 2.4.12. *The map* $\rho:\mathbf{Q}_n\times\mathcal{Y}_n\to\mathbf{U}_n$ *defined by*

$$(U,(\varphi_1,\dots,\varphi_n))\mapsto U\begin{pmatrix}\mathrm{e}^{\mathrm{i}\varphi_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \mathrm{e}^{\mathrm{i}\varphi_n}\end{pmatrix}U^\dagger \tag{2.4.24}$$

is smooth, injective, and its image has full measure (that is to say, $\mathbf{U}_n\setminus\rho(\mathbf{Q}_n\times\mathcal{Y}_n)$ *has measure zero with respect to the invariant measure* $\Omega_{\mathbf{U}_n}$*).*

The unitary version of Weyl integration formula follows.

Theorem 2.4.13. *Let $f \in L^1(\mathbf{U}_n, \Omega_{\mathbf{U}_n})$ be a class-function, i.e. for all $U \in \mathbf{U}_n$ we have $f(UVU^\dagger) = f(V)$ for almost all $V \in \mathbf{U}_n$. Then*

$$\int_{\mathbf{U}_n} f(U)\,\Omega_{\mathbf{U}_n}(\mathrm{d}U) = \widetilde{c}_n \int_{(0,2\pi)^n} \widetilde{f}(\mathrm{e}^{\mathrm{i}\varphi_1},\dots,\mathrm{e}^{\mathrm{i}\varphi_n})\,|\Delta|^2(\mathrm{e}^{\mathrm{i}\varphi_1},\dots,\mathrm{e}^{\mathrm{i}\varphi_n})\,\mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_n, \tag{2.4.25}$$

where $\widetilde{f}(\mathrm{e}^{\mathrm{i}\varphi_1},\dots,\mathrm{e}^{\mathrm{i}\varphi_n}) := f(\begin{pmatrix} \mathrm{e}^{\mathrm{i}\varphi_1} & \cdots & 0 \\ \dots & \ddots & \vdots \\ 0 & \cdots & \mathrm{e}^{\mathrm{i}\varphi_n}\end{pmatrix})$ *and $\widetilde{c}_n$ is a constant depending on n only.*

Proof. Let us use coordinates $q = (q_1,\dots,q_m)$ on $\mathbf{Q}_n$ $(m = n(n-1))$ and denote $q \mapsto V(q)$ the local parametrization of $\mathbf{Q}_n$. Hence we can locally parametrize $\mathbf{U}_n$ in terms of the coordinates q and of n angles $\varphi = (\varphi_1,\cdots,\varphi_n) \in \mathcal{Y}$ by $(\varphi,q) \mapsto U(\varphi,q) := V(q)\mathrm{diag}(\mathrm{e}^{\mathrm{i}\varphi_1},\dots,\mathrm{e}^{\mathrm{i}\varphi_n})V^\dagger(q)$. We want to use the formula (2.3.14) for the invariant measure on $\mathbf{U}_n$, keeping in mind that now the coordinates $p = (p_1,\dots,p_{n^2})$ on $\mathbf{U}_n$ are $p = (\varphi,q)$. We have

$$\mathrm{i}U^\dagger(\varphi,q)\frac{\partial U(\varphi,q)}{\partial\varphi_j} = -V(q)E_{jj}V^\dagger(q), \qquad j = 1,\dots,n \tag{2.4.26}$$

and, denoting from now on

$$D := \begin{pmatrix} \mathrm{e}^{\mathrm{i}\varphi_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{e}^{\mathrm{i}\varphi_n}\end{pmatrix}, \tag{2.4.27}$$

we have (**exercise**)

$$\mathrm{i}U^\dagger(\varphi,q)\frac{\partial U(\varphi,q)}{\partial q_s} = -\mathrm{i}V(q)\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_s}D, D^\dagger\right]V^\dagger(q), \qquad s = 1,\dots,m, \tag{2.4.28}$$

where $[A,B] := AB - BA$ is the commutator. Using (2.3.14) and Lemma 2.3.8 we deduce that the measure $\Omega_{\mathbf{U}_n} = g(\varphi,q)\mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_n\mathrm{d}q_1\cdots\mathrm{d}q_m$ (as above, $m = n(n-1)$) can be written as

$$g(\varphi,q) = (-1)^{n^2}\det\begin{pmatrix} ((\,E_{11}\,)) \\ \hline \vdots \\ \hline ((\,E_{nn}\,)) \\ \hline ((\,\mathrm{i}\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_1}D, D^\dagger\right]\,)) \\ \hline \vdots \\ \hline ((\,\mathrm{i}\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_m}D, D^\dagger\right]\,)) \end{pmatrix} = (-1)^{n^2}\det\begin{pmatrix} ((\,\mathrm{i}\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_1}D, D^\dagger\right]\,))' \\ \hline \vdots \\ \hline ((\,\mathrm{i}\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_m}D, D^\dagger\right]\,))' \end{pmatrix}, \tag{2.4.29}$$

where $((M))'$ for an Hermitian matrix M denotes the m-dimensional row vector obtained by $((M))$ by omission of the first n components, i.e. by omission of the diagonal entries of M, see (2.3.9).

Next, let

$$\mathrm{i}V^\dagger(q)\frac{\partial V(q)}{\partial q_s} =: X^{(s)} + \mathrm{i}Y^{(s)} \tag{2.4.30}$$

and

$$\mathrm{i}\left[V^\dagger(q)\frac{\partial V(q)}{\partial q_s}D, D^\dagger\right] =: \widetilde{X}^{(s)} + \mathrm{i}\widetilde{Y}^{(s)} \tag{2.4.31}$$

for $X^{(s)}, Y^{(s)}, \widetilde{X}^{(s)}, \widetilde{Y}^{(s)} \in \mathbb{R}^{m\times m}$ (and $s = 1,\dots,m$). Now (**exercise**)

$$\widetilde{X}^{(s)}_{ij} + \mathrm{i}\widetilde{Y}^{(s)}_{ij} = (X^{(s)}_{ij} + \mathrm{i}Y^{(s)}_{ij})(1 - \mathrm{e}^{\mathrm{i}(\varphi_i-\varphi_j)}) \tag{2.4.32}$$

and so

$$(\widetilde{X}^{(s)}_{ij}, \widetilde{Y}^{(s)}_{ij}) = (X^{(s)}_{ij}, Y^{(s)}_{ij})\Phi_{ij}, \qquad \Phi_{ij} := \begin{pmatrix} 1-\cos(\varphi_i-\varphi_j) & \sin(\varphi_i-\varphi_j) \\ -\sin(\varphi_i-\varphi_j) & 1-\cos(\varphi_i-\varphi_j) \end{pmatrix} \tag{2.4.33}$$

and finally

$$((\,\widetilde{X}^{(s)} + \mathrm{i}\widetilde{Y}^{(s)}\,))' = ((\,X^{(s)} + \mathrm{i}Y^{(s)}\,))' \left(\begin{array}{c|c|c} \Phi_{12} & \cdots & \mathbf{0}_{2\times 2} \\ \hline \vdots & \ddots & \vdots \\ \hline \mathbf{0}_{2\times 2} & \cdots & \Phi_{n-1,n} \end{array}\right). \tag{2.4.34}$$

Note that[3] $\det \Phi_{ij} = |1-\mathrm{e}^{\mathrm{i}(\varphi_i-\varphi_j)}|^2$. Hence with the same definition of $\Omega_{\mathbf{Q}_n} = h(q)\mathrm{d}q_1 \dots \mathrm{d}q_m$ as in (2.4.19) we obtain

$$\begin{aligned} g(\varphi, q) = (-1)^{n^2} h(q) \prod_{1\le i<j\le n} |1-\mathrm{e}^{\mathrm{i}(\varphi_i-\varphi_j)}|^2 &= (-1)^{n^2} h(q) \prod_{1\le i<j\le n} |\mathrm{e}^{\mathrm{i}\varphi_j} - \mathrm{e}^{\mathrm{i}\varphi_i}|^2 \\ &= (-1)^{n^2} h(q) |\Delta|^2(\mathrm{e}^{\mathrm{i}\varphi_1}, \dots, \mathrm{e}^{\mathrm{i}\varphi_n}) \end{aligned} \tag{2.4.35}$$

and the proof is completed similarly to the Hermitian case (now $\widetilde{c}_n = \frac{(-1)^{n^2}}{n!} \int_{\mathbf{Q}_n} \Omega_{\mathbf{Q}_n}$). □

Exercise 2.4.14. Prove by a direct computation the case $n = 2$ of Theorem 2.4.13, using the parametrization (2.4.2) of $\mathbf{Q}_2$. //

Exercise 2.4.15. Compute the marginal eigenvalue distribution for the CUE. //

2.4.3 Comments

Weyl integration formula shows that the Lebesgue measure on $\mathbf{H}_n$ and the the invariant measure on $\mathbf{U}_n$ both split into a product measure over eigenvalues and over eigenvectors.

Moreover, it explains the phenomenon of *eigenvalue repulsion*: configurations where the eigenvalues are close to each other are very unlikely due to the presence of the Vandermonde squared, as the probability density near such configurations vanishes to second order.

2.5 The Vandermonde determinant

Lemma 2.5.1. *The quantity* $\Delta(\xi_1, \dots, \xi_n) := \prod_{1\le i<j\le n}(\xi_j - \xi_i)$ *is the "Vandermonde determinant":*

$$\Delta(\xi_1, \dots, \xi_n) = \det \begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-1} \\ 1 & \xi_3 & \xi_3^2 & \cdots & \xi_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-1} \end{pmatrix}. \tag{2.5.1}$$

Proof. Let us call $\widetilde{\Delta}$ the RHS of (2.5.1), so that we aim at proving $\Delta = \widetilde{\Delta}$. We start by noting that $\widetilde{\Delta}$ is a polynomial of degree $n-1$ in ξ_n; one way to see it is to consider the Laplace expansion of the determinant with respect to the last row. Moreover, this polynomial vanishes whenever $\xi_n = \xi_j$ for some $1 \le j \le n-1$ (because then the jst and nth rows in the matrix in the RHS of (2.5.1) coincide). It follows that

$$\det \begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-1} \\ 1 & \xi_3 & \xi_3^2 & \cdots & \xi_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-1} \end{pmatrix} = C(\xi_1, \dots, \xi_{n-1}) \prod_{j=1}^{n-1} (\xi_n - \xi_j), \tag{2.5.2}$$

[3] In general multiplication by the complex number $z = a + \mathrm{i}b$, seen as a linear transformation (of row vectors) $\mathbb{R}^2 \to \mathbb{R}^2$, is represented by the 2×2 matrix $Z = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, whose determinant is $|z|^2 = a^2 + b^2$.

for some normalizing constant $C(\xi_2,\dots,\xi_n)$. By looking at the RHS of (2.5.2), $C(\xi_1,\dots,\xi_{n-1})$ is the coefficient in front of ξ_1^{n-1} in the polynomial $\widetilde{\Delta}$. Next, by looking at the LHS of (2.5.2), with the help of Laplace expansion with respect to the first row, this coefficient is

$$C(\xi_1,\dots,\xi_{n-1}) = \det\begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-2} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_{n-1} & \xi_{n-1}^2 & \cdots & \xi_{n-1}^{n-2} \end{pmatrix} = \widetilde{\Delta}(\xi_1,\dots,\xi_{n-1}). \tag{2.5.3}$$

Summarizing:

$$\widetilde{\Delta}(\xi_1,\dots,\xi_n) = \widetilde{\Delta}(\xi_1,\dots,\xi_{n-1})\prod_{j=1}^{n-1}(\xi_n-\xi_j), \tag{2.5.4}$$

and it is now easy to complete the proof that $\Delta = \widetilde{\Delta}$ by induction on n. □

Exercise 2.5.2. Give an alternative proof of Lemma 2.5.1 as follows. First, subtract the first row from the other rows:

$$\det\begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-1} \\ 1 & \xi_3 & \xi_3^2 & \cdots & \xi_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-1} \end{pmatrix} = \det\begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-1} \\ 0 & \xi_2-\xi_1 & \xi_2^2-\xi_1^2 & \cdots & \xi_2^{n-1}-\xi_1^{n-1} \\ 0 & \xi_3-\xi_1 & \xi_3^2-\xi_1^2 & \cdots & \xi_3^{n-1}-\xi_1^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \xi_n-\xi_1 & \xi_n^2-\xi_1^2 & \cdots & \xi_n^{n-1}-\xi_1^{n-1} \end{pmatrix}$$
$$= (\xi_2-\xi_1)\cdots(\xi_n-\xi_1)\det\begin{pmatrix} 1 & \frac{\xi_2^2-\xi_1^2}{\xi_2-\xi_1} & \cdots & \frac{\xi_2^{n-1}-\xi_1^{n-1}}{\xi_2-\xi_1} \\ 1 & \frac{\xi_3^2-\xi_1^2}{\xi_3-\xi_1} & \cdots & \frac{\xi_3^{n-1}-\xi_1^{n-1}}{\xi_3-\xi_1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{\xi_n^2-\xi_1^2}{\xi_n-\xi_1} & \cdots & \frac{\xi_n^{n-1}-\xi_1^{n-1}}{\xi_n-\xi_1} \end{pmatrix}. \tag{2.5.5}$$

Then use the formula $\frac{a^{k+1}-b^{k+1}}{a-b} = a^k + a^{k-1}b + \cdots + ab^{k-1} + b^k$ to prove that

$$\begin{pmatrix} 1 & \frac{\xi_2^2-\xi_1^2}{\xi_2-\xi_1} & \cdots & \frac{\xi_2^{n-1}-\xi_1^{n-1}}{\xi_2-\xi_1} \\ 1 & \frac{\xi_3^2-\xi_1^2}{\xi_3-\xi_1} & \cdots & \frac{\xi_3^{n-1}-\xi_1^{n-1}}{\xi_3-\xi_1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \frac{\xi_n^2-\xi_1^2}{\xi_n-\xi_1} & \cdots & \frac{\xi_n^{n-1}-\xi_1^{n-1}}{\xi_n-\xi_1} \end{pmatrix} = \begin{pmatrix} 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-2} \\ 1 & \xi_3 & \xi_3^2 & \cdots & \xi_3^{n-2} \\ 1 & \xi_4 & \xi_4^2 & \cdots & \xi_4^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-2} \end{pmatrix}\begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-2} \\ 0 & 1 & \xi_1 & \cdots & \xi_1^{n-3} \\ 0 & 0 & 1 & \cdots & \xi_1^{n-4} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{2.5.6}$$

Take determinant and use induction on n. //

Corollary 2.5.3. *For any collection of polynomials $p_0(\xi),\dots,p_{n-1}(\xi)$ of polynomials in the variable ξ such that p_j has degree j and leading coefficient κ_j, we have*

$$\Delta(\xi_1,\dots,\xi_n) = \frac{1}{\kappa_0\cdots\kappa_{n-1}}\det\begin{pmatrix} p_0(\xi_1) & p_1(\xi_1) & p_2(\xi_1) & \cdots & p_{n-1}(\xi_1) \\ p_0(\xi_2) & p_1(\xi_2) & p_2(\xi_2) & \cdots & p_{n-1}(\xi_2) \\ p_0(\xi_3) & p_1(\xi_3) & p_2(\xi_3) & \cdots & p_{n-1}(\xi_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_0(\xi_n) & p_1(\xi_n) & p_2(\xi_n) & \cdots & p_{n-1}(\xi_n) \end{pmatrix}. \tag{2.5.7}$$

Proof. Let $p_j(x) = \kappa_j x^j + a_{j-1,j}x^{j-1} + \cdots + a_{1,j}x + a_{0,j}$. Then

$$\begin{pmatrix} p_0(\xi_1) & p_1(\xi_1) & p_2(\xi_1) & \cdots & p_{n-1}(\xi_1) \\ p_0(\xi_2) & p_1(\xi_2) & p_2(\xi_2) & \cdots & p_{n-1}(\xi_2) \\ p_0(\xi_3) & p_1(\xi_3) & p_2(\xi_3) & \cdots & p_{n-1}(\xi_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_0(\xi_n) & p_1(\xi_n) & p_2(\xi_n) & \cdots & p_{n-1}(\xi_n) \end{pmatrix} = \begin{pmatrix} 1 & \xi_1 & \xi_1^2 & \cdots & \xi_1^{n-1} \\ 1 & \xi_2 & \xi_2^2 & \cdots & \xi_2^{n-1} \\ 1 & \xi_3 & \xi_3^2 & \cdots & \xi_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_n & \xi_n^2 & \cdots & \xi_n^{n-1} \end{pmatrix}\cdot\begin{pmatrix} \kappa_1 & a_{0,1} & a_{0,2} & \cdots & a_{0,n-1} \\ 0 & \kappa_2 & a_{1,2} & \cdots & a_{1,n-1} \\ 0 & 0 & \kappa_3 & \cdots & a_{2,n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \kappa_{n-1} \end{pmatrix} \tag{2.5.8}$$

and the proof is immediate by taking the determinant. □

Chapter 3

Orthogonal Polynomials

3.1 Andréief formula

Lemma 3.1.1 (Andréief, 1886)**.** *Let* $(X, \mathcal{F}, \mu)$ *be a measure space and let* $f_1, g_1, \cdots, f_n, g_n : X \to \mathbb{C}$ *measurable functions (defined almost everywhere) such that* $f_i g_j \in L^1(X, \mu)$ *for any* $1 \le i, j \le n$*. Then we have the identity*

$$\int_{X^n} \det_{1\le i,j\le n} (f_i(x_j)) \det_{1\le i,j\le n} (g_i(x_j))\, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n) \;=\; n! \det_{1\le i,j\le n} \left(\int_X f_i(x) g_j(x) \mu(\mathrm{d}x) \right). \tag{3.1.1}$$

Proof. Recall that in general for a matrix $A \in \mathbb{C}^{n\times n}$ with entries A_{ij} we have

$$\det_{1\le i,j\le n}(A_{ij}) = \sum_{\pi \in \mathfrak{S}_n} (-1)^{|\pi|} A_{1,\pi(1)} \cdots A_{n,\pi(n)}. \tag{3.1.2}$$

Here $\mathfrak{S}_n$ is the group of permutations of $\{1, \ldots, n\}$. For any $\pi \in \mathfrak{S}_n$, its parity (i.e. the parity of the number of transpositions in any factorization of π into transpositions) is denoted $|\pi|$ (and can be 0 or 1) and $(-1)^{|\pi|}$ is the sign of π (i.e. $+1$ when the permutation π is a product of an even number of transpositions, and -1 when it is the product of an odd number of transpositions). Note that $|\pi^{-1}| = |\pi|$ and $(-1)^{|\pi\pi'|} = (-1)^{|\pi|}(-1)^{|\pi'|}$.

Then the proof is given by the following chain of equalities.

$$\int_{X^n} \det_{1\le i,j\le n} (f_i(x_j)) \det_{1\le i,j\le n} (g_i(x_j))\, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n) \tag{3.1.3}$$

(definition of determinant)

$$= \sum_{\pi,\rho \in \mathfrak{S}_n} (-1)^{|\pi|}(-1)^{|\rho|} \int_{X^n} f_1(x_{\pi(1)}) \cdots f_n(x_{\pi(n)}) g_1(x_{\rho(1)}) \cdots g_n(x_{\rho(n)}) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n) \tag{3.1.4}$$

(change of variables $\widetilde{x}_i = x_{\pi(i)}$ in the integrals)

$$\sum_{\pi,\rho \in \mathfrak{S}_n} (-1)^{|\pi^{-1}\rho|} \int_{X^n} f_1(\widetilde{x}_1) \cdots f_n(\widetilde{x}_n) g_1(\widetilde{x}_{\pi^{-1}\rho(1)}) \cdots g_n(\widetilde{x}_{\pi^{-1}\rho(n)}) \mu(\mathrm{d}\widetilde{x}_1) \cdots \mu(\mathrm{d}\widetilde{x}_n) \tag{3.1.5}$$

(rename $\sigma := \pi^{-1}\rho$ and observe that the terms in the sum, now with indices $\pi, \sigma \in \mathfrak{S}_n$, do not depend on π)

$$= n! \sum_{\sigma \in \mathfrak{S}_n} (-1)^{|\sigma|} \int_{X^n} f_1(\widetilde{x}_1) \cdots f_n(\widetilde{x}_n) g_1(\widetilde{x}_{\sigma(1)}) \cdots g_n(\widetilde{x}_{\sigma(n)}) \mu(\mathrm{d}\widetilde{x}_1) \cdots \mu(\mathrm{d}\widetilde{x}_n) \tag{3.1.6}$$

(rename $\sigma \mapsto \sigma^{-1}$)

$$= n! \sum_{\sigma \in \mathfrak{S}_n} (-1)^{|\sigma|} \int_{X^n} f_1(\widetilde{x}_1) \cdots f_n(\widetilde{x}_n) g_{\sigma(1)}(\widetilde{x}_1) \cdots g_{\sigma(n)}(\widetilde{x}_n) \mu(\mathrm{d}\widetilde{x}_1) \cdots \mu(\mathrm{d}\widetilde{x}_n) \tag{3.1.7}$$

(the integrand is factorized)

$$= n! \sum_{\sigma \in \mathfrak{S}_n} (-1)^{|\sigma|} \int_X f_1(x) g_{\sigma(1)}(x) \mu(\mathrm{d}x) \cdots \int_X f_n(x) g_{\sigma(n)}(x) \mu(\mathrm{d}x) \tag{3.1.8}$$

(definition of determinant)

$$= n! \det_{1 \le i,j \le n} \left(\int_X f_i(x) g_j(x) \mu(\mathrm{d}x) \right). \tag{3.1.9}$$

□

Remark 3.1.2. Let $A, B \in \mathbb{C}^{n \times m}$ with $m \geq n$. The Binet–Cauchy identity is

$$\sum_I \det A_I \det B_I \; = \; \det(AB^\top)\,, \tag{3.1.10}$$

where the sum in the left-hand side runs over the sets $I \subset \{1, \ldots, m\}$ of cardinality n and A_I (respectively, B_I) is the square matrix of size n obtained from A (respectively, B) by removing the columns whose index are not in I. The Andréief identity can be considered as a generalization of the Binet–Cauchy identity; more concretely, to derive (3.1.10) from (3.1.1), given the matrices $A, B \in \mathbb{C}^{n \times m}$ it suffices to take $X := \{1, \ldots, m\}$ and μ the counting measure on X and define functions $f_1, g_1, \ldots, f_n, g_n$ on X by $f_i(j) := A_{ij}$ and $g_i(j) = B_{ij}$, for $1 \le i \le n$ and $1 \le j \le m$. (In this way, one obtains $n! \times$(3.1.10) from (3.1.1).) //

3.1.1 CUE

We learn from Theorem 2.4.13 that the joint eigenvalue distribution of the CUE is given by

$$\frac{1}{\widehat{Z}_n^{\mathsf{CUE}}} \prod_{1 \le i < j \le n} |\mathrm{e}^{\mathrm{i}\varphi_i} - \mathrm{e}^{\mathrm{i}\varphi_j}|^2 \prod_{i=1}^n \mathrm{d}\varphi_i, \qquad \varphi_i \in (0, 2\pi). \tag{3.1.11}$$

Let us note that

$$\prod_{1 \le i < j \le n} |\mathrm{e}^{\mathrm{i}\varphi_i} - \mathrm{e}^{\mathrm{i}\varphi_j}|^2 = |\Delta|^2(\mathrm{e}^{\mathrm{i}\varphi_i}, \ldots, \mathrm{e}^{\mathrm{i}\varphi_n}) = \det_{1 \le \ell, m \le n} (\mathrm{e}^{\mathrm{i}(\ell-1)\varphi_m}) \det_{1 \le \ell, m \le n} (\mathrm{e}^{-\mathrm{i}(\ell-1)\varphi_m}). \tag{3.1.12}$$

Armed with Andréief formula we can compute the CUE normalizing constant;

$$\begin{aligned} \widehat{Z}_n^{\mathsf{CUE}} &= \int_{(0,2\pi)^n} |\Delta|^2 (\mathrm{e}^{\mathrm{i}\varphi_1}, \ldots, \mathrm{e}^{\mathrm{i}\varphi_n}) \mathrm{d}\varphi_1 \cdots \mathrm{d}\varphi_n \\ &= \int_{(0,2\pi)^n} \det_{1 \le \ell, m \le n} (\mathrm{e}^{\mathrm{i}(\ell-1)\varphi_m}) \det_{1 \le \ell, m \le n} (\mathrm{e}^{-\mathrm{i}(\ell-1)\varphi_m}) \mathrm{d}\varphi_1 \cdots \mathrm{d}\varphi_n \\ &= n! \det_{1 \le \ell, m \le n} \left(\int_0^{2\pi} \mathrm{e}^{\mathrm{i}(\ell - m)\varphi} \mathrm{d}\varphi \right) \\ &= n! \det_{1 \le \ell, m \le n} (2\pi\, \delta_{\ell,m}) \\ &= n!\,(2\pi)^n. \end{aligned} \tag{3.1.13}$$

Note two important ingredients in this computation.

- The form of the eigenvalue interaction $|\Delta|^2$ allows us to use Andréief identity. (This is not the case for any other β-ensemble.)
- The orthogonality of $\{\mathrm{e}^{\mathrm{i}\ell\varphi}\}_{\ell \in \mathbb{Z}}$ on the unit circle simplifies the computation.

3.1.2 GUE

By Corollary 2.5.3 we have

$$\Delta(\xi_1,\dots,\xi_n) = \det_{1\le i,j\le n}(\xi_i^{j-1}) = \det_{1\le i,j\le n}(p_{j-1}(\xi_i)) \tag{3.1.14}$$

for any family $p_j(\xi)$ of monic polynomials of degree j, for integers $j \ge 0$.

Hence we try to repeat the computation (3.1.13) for the GUE:

$$\begin{aligned}\widehat{Z}_n^{\mathsf{GUE}} &= \int_{\mathbb{R}^n} \Delta^2(x_1,\dots,x_n)\mathrm{e}^{-\sum_{i=1}^n \frac{x_i^2}{2}}\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &= \int_{\mathbb{R}^n} \det_{1\le\ell,m\le n}(p_{\ell-1}(x_m)) \det_{1\le\ell,m\le n}(p_{\ell-1}(x_m))\mathrm{e}^{-\sum_{i=1}^n \frac{x_i^2}{2}}\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &= n! \det_{1\le\ell,m\le n}\left(\int_{-\infty}^{+\infty} p_{\ell-1}(x)p_{m-1}(x)\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right).\end{aligned} \tag{3.1.15}$$

Therefore if we choose polynomials with the orthogonality property

$$\int_{-\infty}^{+\infty} p_a(x)p_b(x)\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x = 0 \text{ unless } a=b \tag{3.1.16}$$

the computation simplifies. But this is the characterizing property of the Hermite polynomials!

Definition 3.1.3. The **Hermite polynomials** are defined by the formula

$$P_k(x) := \mathrm{e}^{\frac{x^2}{2}}\left[\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^k \mathrm{e}^{-\frac{x^2}{2}}\right]. \qquad \textit{(Rodrigues' formula.)} \tag{3.1.17}$$

The first few of them are

$$P_0(x)=1,\ P_1(x)=x,\ P_2(x)=x^2-1,\ P_3(x)=x^3-3x,\ P_4(x)=x^4-6x^2+3. \tag{3.1.18}$$

//

Exercise 3.1.4. Use the definition (3.1.17) to prove the following facts.

1. Prove that $P_k(-x) = (-1)^k P_k(x)$.
2. Prove the recurrence relation $xP_k(x) = P_{k+1}(x) + kP_{k-1}(x)$ for all $k \ge 1$. Deduce that P_k is a monic polynomial of degree k.
3. Prove the relation $P_k'(x) = kP_{k-1}(x)$.

//

Proposition 3.1.5. *For all $k,\ell \ge 0$ we have*

$$\int_{-\infty}^{+\infty} P_k(x)P_\ell(x)\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x = k!\sqrt{2\pi}\,\delta_{k,\ell}. \tag{3.1.19}$$

Proof. It is enough to prove that for all $\ell \le k$ we have

$$\int_{-\infty}^{+\infty} P_k(x)x^\ell\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x = \begin{cases}0, & \text{when } \ell<k,\\ k!\sqrt{2\pi}, & \text{when } \ell=k.\end{cases} \tag{3.1.20}$$

This is proved by integration by parts using the definition (3.1.17);

$$\int_{-\infty}^{+\infty} P_k(x)x^\ell\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x = \int_{-\infty}^{+\infty}\left[\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^k\mathrm{e}^{-\frac{x^2}{2}}\right]x^\ell\mathrm{d}x = \int_{-\infty}^{+\infty}\mathrm{e}^{-\frac{x^2}{2}}\left[\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k x^\ell\right]\mathrm{d}x \tag{3.1.21}$$

and now $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k x^\ell = 0$ if $\ell<k$ and $\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^k x^\ell = k!$ if $\ell = k$. □

Resuming our computation (3.1.15), with P_k the Hermite polynomials just introduced, we have

$$\begin{aligned}\widehat{Z}_n^{\mathsf{GUE}} &= n! \det_{1\le \ell,m\le n}\left(\int_{-\infty}^{+\infty} P_{\ell-1}(x)P_{m-1}(x)\mathrm{e}^{-\frac{x^2}{2}}\mathrm{d}x\right)\\ &= n! \det_{1\le \ell,m\le n}\left((\ell-1)!\sqrt{2\pi}\,\delta_{\ell,m}\right)\\ &= \left(\prod_{i=1}^{n} i!\right)(2\pi)^{n/2}. \end{aligned} \tag{3.1.22}$$

3.2 Orthogonal Polynomials on the Real Line (OPRL)

3.2.1 Orthogonality with respect to a moment functional

Let $\mathbb{R}[x]$ be the algebra of polynomial functions in the variable x. Let $\mathcal{S} : \mathbb{R}[x] \to \mathbb{R}$ be a linear functional.

Definition 3.2.1. The (monic) **Orthogonal Polynomials on the Real Line** (OPRL) (associated with the functional $\mathcal{S}$) are a family of monic polynomials $P_k(x)$, for all integers $k \ge 0$, of degree k such that $\mathcal{S}[P_kP_\ell] = 0$ unless $k = \ell$, in which case we instead require $\mathcal{S}[P_k^2] \ne 0$. //

To study the existence and uniqueness of OPRL, let $m_j := \mathcal{S}[x^j]$ and let $\mathcal{M}$ be the infinite matrix

$$\mathcal{M} := \begin{pmatrix} m_0 & m_1 & m_2 & \cdots \\ m_1 & m_2 & m_3 & \cdots \\ m_2 & m_3 & m_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{3.2.1}$$

i.e. $\mathcal{M} = (\mathcal{M}_{i,j})_{i,j=1}^{+\infty}$ with $\mathcal{M}_{i,j} := m_{i+j-2}$. Furthermore, introduce

$$D_k := \det_{1\le i,j\le k}(\mathcal{M}_{i,j}). \tag{3.2.2}$$

Remark 3.2.2. Matrices whose entries depend only on the sum of their indexes are called *Hankel matrices*. The matrix $\mathcal{M}$ is a Hankel matrix; sometimes determinants of Hankel matrices are referred to as *Hankel determinants*. //

Proposition 3.2.3. *The OPRL exist and are unique if and only if $D_k \ne 0$ for all $k \ge 0$, in which case they are given by*

$$P_k(x) = \frac{1}{D_k}\det\begin{pmatrix} m_0 & m_1 & m_2 & \cdots & m_k \\ m_1 & m_2 & m_3 & \cdots & m_{k+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{k-1} & m_\ell & m_{k+1} & \cdots & m_{2k-1} \\ 1 & x & x^2 & \cdots & x^k \end{pmatrix} \tag{3.2.3}$$

and the constants $h_k := \mathcal{S}[P_k(x)^2] = \mathcal{S}[x^kP_k(x)]$ are given by

$$h_k = \frac{D_{k+1}}{D_k}. \tag{3.2.4}$$

Proof. The main observation is the the orthogonal polynomials $P_k(x) = a_{k,0}+a_{k,1}x+\cdots+a_{k,k-1}x^{k-1}+x^k$ are uniquely determined by the linear system $\mathcal{S}[P_k(x)x^\ell] = 0$ for $\ell = 0,\dots,k-1$. Spelling out this linear system for the unkown coefficients $a_{k,j}$ we get

$$\begin{cases} a_{k,0}m_0 + a_{k,1}m_1 + \cdots + a_{k,k-1}m_{k-1} = -m_k, \\ a_{k,0}m_1 + a_{k,1}m_2 + \cdots + a_{k,k-1}m_k = -m_{k+1}, \\ \vdots \\ a_{k,0}m_{k-1} + a_{k,1}m_k + \cdots + a_{k,k-1}m_{2k-2} = -m_{2k-1}. \end{cases} \tag{3.2.5}$$

This linear system has one and only one solution for $(a_{k,0},\dots,a_{k,k-1})$ if and only if the coefficient matrix $(m_{i+j-2})_{1\le i,j\le k}$ has non-vanishing determinant, i.e. if and only if $D_k\neq 0$. Next, supposing $D_k\neq 0$ we can solve this system by Cramer's rule as

$$a_{k,j}=-\frac{D_k^{[j]}}{D_k},\qquad j=0,\dots,k-1, \tag{3.2.6}$$

where $D_k^{[j]}$ is the determinant of the matrix obtained from the coefficient matrix by replacing the jth row with the vector $(m_k,m_{k+1},\dots,m_{2k-1})$. The proof of (3.2.3) follows by Laplace expanding the determinant in the numerator of (3.2.3). Finally, (3.2.4) follows by a direct computation using (3.2.3);

$$\begin{aligned} h_k=\mathcal{S}[x^kP_k(x)]&=\frac{1}{D_k}\mathcal{S}\left[\det\begin{pmatrix} m_0 & m_1 & m_2 & \cdots & m_k\\ m_1 & m_2 & m_3 & \cdots & m_{k+1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ m_{k-1} & m_\ell & m_{k+1} & \cdots & m_{2k-1}\\ x^k & x^{k+1} & x^{k+2} & \cdots & x^{2k}\end{pmatrix}\right]\\ &=\frac{1}{D_k}\det\begin{pmatrix} m_0 & m_1 & m_2 & \cdots & m_k\\ m_1 & m_2 & m_3 & \cdots & m_{k+1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ m_{k-1} & m_\ell & m_{k+1} & \cdots & m_{2k-1}\\ m_k & m_{k+1} & m_{k+2} & \cdots & m_{2k}\end{pmatrix}=\frac{D_{k+1}}{D_k}. \end{aligned} \tag{3.2.7}$$

□

Exercise 3.2.4. Prove by a direct computation (i.e., without resorting to the linear system in the proof and to Cramer's rule) that the polynomials defined by (3.2.3) satisfy the orthogonality $\mathcal{S}[x^\ell P_k]=0$ for all $0\le\ell<k$. //

Exercise 3.2.5. Suppose $D_k\neq 0$ for all $k\ge 0$. Define the infinite lower-triangular unipotent matrix[1] $\mathcal{L}=(\mathcal{L}_{i,j})_{i,j=1}^{+\infty}$ by

$$\mathcal{L}^{-1}\begin{pmatrix}1\\ x\\ x^2\\ \vdots\end{pmatrix}=\begin{pmatrix}P_0(x)\\ P_1(x)\\ P_2(x)\\ \vdots\end{pmatrix}. \tag{3.2.8}$$

Prove that $\mathcal{L}\mathcal{D}\mathcal{L}^\top=\mathcal{M}$, where $\mathcal{D}=\operatorname{diag}(h_0,h_1,h_2,\dots)$. Moreover, prove that for any $k\ge 1$ the truncated matrices $\mathcal{L}_{[k]}=(\mathcal{L}_{i,j})_{1\le i,j\le k}, \mathcal{D}_{[k]}=(\mathcal{D}_{i,j})_{1\le i,j\le k}, \mathcal{M}_{[k]}=(\mathcal{M}_{i,j})_{1\le i,j\le k}$ also satisfy $\mathcal{L}_{[k]}\mathcal{D}_{[k]}\mathcal{L}_{[k]}^\top=\mathcal{M}_{[k]}$. //

According to this exercise, to compute the first k OPRL (as well as h_i for $0\le i\le k-1$) it is enough to compute the Choleski factorization of the truncated Hankel matrix of moments $\mathcal{M}_{[k]}=(m_{i+j-2})_{1\le i,j\le k}$; this is a numerically very effective way of computing the OPRL, whereas the determinantal formula (3.2.3) is of theoretical importance but it is numerically much less effective.

3.2.2 Orthogonality with respect to a Borel measure

For simplicity we now specialize to the case of interest, although many of the results below could be formulated more generally. Namely we shall now restrict to the case where $\mathcal{S}[f]=\int_\mathbb{R} f(x)\mu(\mathrm{d}x)$ for a Borel measure[2] μ on $\mathbb{R}$ with finite moments of all orders (namely, $\int_\mathbb{R}|x|^k\mu(\mathrm{d}x)<+\infty$ for all integers $k\ge 0$, or equivalently, $\mathbb{R}[x]\subseteq L^1(\mathbb{R},\mu)$). In this case, Definition 3.2.1 retrieves the usual notion of orthogonality.

Note that in this case the function $\mathcal{S}$ is positive-semidefinite: for all polynomials $f\in\mathbb{R}[x]$ which are non-negative on the real line we have $\mathcal{S}[f]\ge 0$.

[1]Namely, $\mathcal{L}_{i,j}=0$ for $i<j$ and $\mathcal{L}_{i,i}=1$.
[2]Without further mention, all measures in these notes are considered to be positive.

Proposition 3.2.6. *Suppose μ is a Borel measure with $\int_{\mathbb{R}} |x|^k \mu(\mathrm{d}x) < +\infty$ for all integers $k \geq 0$. The OPRL exist if and only if the support*[3] *of μ is infinite; in this case we further have $D_k > 0$ and $h_k > 0$.*

Proof. The $k \times k$ matrix $(\mathcal{M}_{i,j})_{i,j=1}^k$, where $\mathcal{M}_{i,j} = m_{i+j-2}$, $m_j = \int_{\mathbb{R}} x^k \mu(\mathrm{d}x)$, represents the quadratic form

$$P \mapsto \int_{\mathbb{R}} P(x)^2 \mu(\mathrm{d}x) \tag{3.2.9}$$

on the space of polynomials P of degree at most $k-1$, with respect to the monomial basis. This quadratic form is positive-definite: indeed, if $\int_{\mathbb{R}} P(x)^2 \mu(\mathrm{d}x) = 0$ then $P(x)$ must be zero on the support of μ, which is infinite by assumption, and then $P(x) = 0$ identically in x. In particular, $D_k = \det_{1 \leq i,j \leq k}(m_{i+j-2}) > 0$ for all $k \geq 0$. □

Exercise 3.2.7. Let $\mu = c_1 \delta_{x_1} + \cdots + c_r \delta_{x_r}$, where $x_1, \ldots, x_r \in \mathbb{R}$ are distinct points, $c_i > 0$, and δ_x denotes the Dirac delta measure

$$\delta_x(E) := \begin{cases} 1, & x \in E, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2.10}$$

Prove that monic polynomials $P_0, P_1, \ldots, P_r$, with P_ℓ of degree ℓ, satisfying $\int_{\mathbb{R}} P_\ell(x) P_m(x) \mu(\mathrm{d}x) = 0$ if $\ell \neq m$ for all $0 \leq \ell, m \leq r$ always exist and are unique. Prove that $P_r(x) = \prod_{j=1}^r (x - x_j)$. //

Exercise 3.2.8. Assume that μ is a Borel measure on $\mathbb{R}$ with finite moments of all orders and infinite support.

1. Prove that

$$\int_{\mathbb{R}^n} \Delta^2(x_1, \ldots, x_n) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n) = n! \prod_{i=0}^{n-1} h_i = n!\, D_n. \tag{3.2.11}$$

2. Prove that (*Heine formula*)

$$\frac{1}{n!\, D_n} \int_{\mathbb{R}^n} \Delta^2(x_1, \ldots, x_n) \left(\prod_{i=1}^n (y - x_i) \right) \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n) = P_n(y). \tag{3.2.12}$$

3. Let V be a potential function as in Def. 2.2.5. Deduce that

$$\frac{\int_{\mathbf{H}_n} \det(y \mathbf{1}_n - M) \exp(-\operatorname{tr} V(M)) \Omega_{\mathbf{H}_n}(\mathrm{d}M)}{\int_{\mathbf{H}_n} \exp(-\operatorname{tr} V(M)) \Omega_{\mathbf{H}_n}(\mathrm{d}M)} = P_n(y), \tag{3.2.13}$$

   in words: the expectation of the characteristic polynomials in a unitary-invariant ensemble are the associated OPRL.

(Hint: in 1. use Andréief identity, Lemma 3.1.1, writing $\Delta(x_1, \ldots, x_n) = \det_{1 \leq i,j \leq n}(P_{j-1}(x_i))$; in 2. use Andréief identity, writing $\Delta^2(x_1, \ldots, x_n) \prod_{i=1}^n (y - x_i) = \det_{1 \leq i,j \leq n}(x_i^{j-1}) \det_{1 \leq i,j \leq n}(x_i^{j-1}(y - x_i))$, and use (3.2.3) with elementary column operation; in 3. use Weyl integration formula, Theorem 2.4.6.) //

There is in this case another more-symmetric normalization for the OPRL, no longer monic but *orthonormal*: $p_k(x) := P_k(x)/\sqrt{h_k}$ satisfy $\mathcal{S}[p_k p_\ell] = \delta_{k,\ell}$.

3.2.3 Three-term recurrence, Jacobi operators, Favard Theorem

The following property characterizes orthogonality on the real line.

[3] The support of a measure is the smallest closed subset of $\mathbb{R}$ whose complement has zero measure. It always exists: it is the intersection of the nonempty family of closed subsets of $\mathbb{R}$ whose complement has zero measure. The support of a measure is a finite set if and only if it is a finite linear combination (with positive coefficients) of delta measures.

Proposition 3.2.9 (Three-term recurrence). *Let $(P_k(x))_{k=0}^{+\infty}$ be a family of monic OPRL (with respect to a linear functional $\mathcal{S}$). Then there exist sequences of real numbers $(b_k)_{k=0}^{+\infty}$ and $(w_k)_{k=1}^{+\infty}$ such that*

$$xP_k(x) = P_{k+1}(x) + b_k P_k(x) + w_k P_{k-1}(x), \qquad k \geq 0. \tag{3.2.14}$$

(With the understanding that for $k = 0$ we set $P_{-1} := 0$.) Moreover, $w_k = h_k/h_{k-1} \neq 0$ for $k \geq 1$.

Proof. Since $\{P_l(x) : 0 \leq l \leq k+1\}$ form a basis of the space of polynomials of degree at most $k+1$ we have

$$xP_k(x) = \sum_{j=0}^{k+1} c_{k,j} P_j(x), \tag{3.2.15}$$

for some $c_{k,j} \in \mathbb{R}$. Since $xP_k(x)$ is monic, $c_{k,k+1} = 1$. We now show that $c_{k,r} = 0$ for $r = 0, \ldots, k-2$. Indeed, multiply (3.2.15) by $P_r(x)$ and apply $\mathcal{S}$:

$$\mathcal{S}[xP_r(x)P_k(x)] = \sum_{j=0}^{k+1} c_{k,j} \mathcal{S}[P_r(x)P_j(x)] \tag{3.2.16}$$

and so, if $0 \leq r \leq k-2$ we obtain

$$0 = h_r c_{k,r} \tag{3.2.17}$$

and since $h_r \neq 0$, see (3.2.4), $c_{k,r} = 0$ for all $r = 0, \ldots, k-2$. Then suffices to set $b_k := c_{k,k}$ and $w_k := c_{k,k-1}$. Finally, multiply (3.2.15) by $P_{k-1}(x)$ and apply $\mathcal{S}$ to obtain $h_k = w_k h_{k-1}$, and the proof is complete. □

Exercise 3.2.10. Let $P_k(x)$ be a family of monic OPRL. Express the three-term recurrence of the polynomials $P_k(ax+b)$ in terms of that for $P_k(x)$. //

We shall assume for simplicity from now on that $w_k > 0$; this is always the case for orthogonality with respect to a measure.

In terms of the orthonormal polynomials $p_k(x) := P_k(x)/\sqrt{h_k}$ we have

$$xp_k(x) = a_k p_{k+1}(x) + b_k p_k(x) + a_{k-1} p_{k-1}(x), \qquad a_k = \sqrt{h_{k+1}/h_k} = \sqrt{w_{k+1}}. \tag{3.2.18}$$

Introducing the infinite tridiagonal matrix

$$\mathcal{J} = \begin{pmatrix} b_0 & a_0 & 0 & 0 & \cdots \\ a_0 & b_1 & a_1 & 0 & \ddots \\ 0 & a_1 & b_2 & a_2 & \ddots \\ 0 & 0 & a_2 & b_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{3.2.19}$$

we can rewrite the three-term recursion as an infinite-vector identity

$$x \begin{pmatrix} p_0(x) \\ p_1(x) \\ p_2(x) \\ \vdots \end{pmatrix} = \mathcal{J} \begin{pmatrix} p_0(x) \\ p_1(x) \\ p_2(x) \\ \vdots \end{pmatrix} \tag{3.2.20}$$

Definition 3.2.11. A real tridiagonal matrix J (i.e. its entries $J_{i,k}$ vanish unless $i - k = 0, \pm 1$) is called a **Jacobi matrix** (or, if it is infinite as in (3.2.19), a **Jacobi operator**) if it is symmetric and the off-diagonal entries are positive. //

Let us see how the Jacobi operator (3.2.19) encodes the moments.

Lemma 3.2.12. *The moments* $m_k = \mathcal{S}[x^k]$ *for* $k \geq 1$ *can be retrieved from the Jacobi matrix* $\mathcal{J}$ *in* (3.2.19) *by*

$$\frac{m_k}{m_0} = (\mathcal{J}^k)_{1,1} \tag{3.2.21}$$

namely m_k/m_0 *is the* $1,1$*-entry in* $\mathcal{J}^k$*.*

Note that a scalar rescaling of the measure does not affect the family of OPRL, hence we can only recover the moments up to a common factor from the OPRL.

Moreover, note that in general the multiplication of infinite matrices might not be defined; however powers of the Jacobi infinite matrix (3.2.19) are well defined.

Proof. From (3.2.20)

$$x^k \begin{pmatrix} p_0(x) \\ p_1(x) \\ p_2(x) \\ \vdots \end{pmatrix} = \mathcal{J}^k \begin{pmatrix} p_0(x) \\ p_1(x) \\ p_2(x) \\ \vdots \end{pmatrix}. \tag{3.2.22}$$

Noting that $p_0 = 1/\sqrt{m_0}$ (because $\mathcal{S}[p_0^2] = 1$) we can write

$$x^k = (\sqrt{m_0}, 0, 0, \dots)\mathcal{J}^k \begin{pmatrix} p_0(x) \\ p_1(x) \\ p_2(x) \\ \vdots \end{pmatrix}. \tag{3.2.23}$$

Apply $\mathcal{S}$ to obtain

$$m_k = (\sqrt{m_0}, 0, 0, \dots)\mathcal{J}^k \begin{pmatrix} \mathcal{S}[p_0(x)] \\ \mathcal{S}[p_1(x)] \\ \mathcal{S}[p_2(x)] \\ \vdots \end{pmatrix}. \tag{3.2.24}$$

Finally, $\mathcal{S}[p_0] = \sqrt{m_0}$ and $\mathcal{S}[p_l] = 0$ for all $l \geq 1$ and the proof is complete. □

We now aim at a converse of Proposition 3.2.9. We first have a lemma of independent interest.

Lemma 3.2.13. *Let* $(b_k)_{k=0}^{+\infty}$ *and* $(w_k)_{k=1}^{+\infty}$ *be sequences of real numbers and define monic polynomials* $P_k(x)$ *of degree* k *by* $P_{-1} := 0$, $P_0(x) = 1$ *and*

$$P_{k+1} = (x - b_k)P_k(x) - w_k P_{k-1}, \qquad k \geq 0. \tag{3.2.25}$$

Let $\mathcal{J}$ *the corresponding Jacobi operator (with* $a_k = \sqrt{w_{k+1}}$ *for* $k \geq 0$*). Then* $P_k(x)$ *for* $k \geq 1$ *is the characteristic polynomial of the truncated Jacobi matrix* $\mathcal{J}_{[k]} = (\mathcal{J}_{i,j})_{1 \leq i,j \leq k}$*, i.e.*

$$P_k(x) = \det(x\mathbf{1}_k - \mathcal{J}_{[k]}). \tag{3.2.26}$$

Proof. The first few cases can be computed directly, $P_1(x) = x - b_0$, $P_2(x) = (x - b_1)(x - b_0) - a_0^2 = \det \begin{pmatrix} x - b_0 & -a_0 \\ -a_0 & x - b_1 \end{pmatrix}$. We can then proceed by induction in k. We compute

$$\det(x\mathbf{1}_{k+1} - \mathcal{J}_{[k+1]}) = \det \left(\begin{array}{ccc|c|c} & & & 0 & 0 \\ & x\mathbf{1}_{k-1} - \mathcal{J}_{[k-1]} & & \vdots & \vdots \\ & & & -a_{k-2} & 0 \\ \hline 0 & \cdots & -a_{k-2} & x - b_{k-1} & -a_{k-1} \\ \hline 0 & \cdots & 0 & -a_{k-1} & x - b_k \end{array}\right) \tag{3.2.27}$$

by Laplace expansion with respect to the last row:

$$\det(x\mathbf{1}_{k+1} - \mathcal{J}_{[k+1]}) = (x-b_k)\det(x\mathbf{1}_k - \mathcal{J}_{[k]}) + a_{k-1}\det\left(\begin{array}{ccc|c} & x\mathbf{1}_{k-1} - \mathcal{J}_{[k-1]} & & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 & \cdots & -a_{k-2} & -a_{k-1} \end{array}\right)$$
$$= (x-b_k)\det(x\mathbf{1}_k - \mathcal{J}_{[k]}) - a_{k-1}^2 \det(x\mathbf{1}_{k-1} - \mathcal{J}_{[k-1]}) \quad (3.2.28)$$

and so by the induction hypothesis

$$\det(x\mathbf{1}_{k+1} - \mathcal{J}_{[k+1]}) = (x-b_k)P_k(x) - a_{k-1}^2 P_k(x) = P_{k+1}(x) \quad (3.2.29)$$

and the proof is complete. □

Remark 3.2.14. It follows that the zeros of orthogonal polynomials are the eigenvalues of the truncated Jacobi matrices. //

Exercise 3.2.15. Prove that for a finite-size $n \times n$ Jacobi matrix, the eigenvalues are distinct and the eigenvectors $(v_1, \dots, v_n)^\top$ satisfy $v_1 \neq 0 \neq v_n$. (Hint: first show that for an eigenvector $(v_1, \dots, v_n)$ with eigenvalue λ, the pair (v_1, λ) completely determines all v_k's with $k \geq 2$, and similarly (v_n, λ) determines all v_k with $k \leq n-1$; deduce that no eigenvalue can have multiplicity larger than one.) Deduce that orthogonal polynomials have real and simple zeros. //

Theorem 3.2.16 (Formal Favard theorem)**.** *Let $(b_k)_{k=0}^{+\infty}$ and $(w_k)_{k=1}^{+\infty}$ be sequences of real numbers, with $w_k > 0$ for all $k \geq 1$, and define monic polynomials $P_k(x)$ of degree k by $P_{-1} := 0$, $P_0(x) = 1$ and*

$$P_{k+1} = (x-b_k)P_k(x) - w_k P_{k-1}, \qquad k \geq 0. \quad (3.2.30)$$

Let $\mathcal{J}$ the corresponding Jacobi operator (with $a_k = \sqrt{w_{k+1}}$ for $k \geq 0$). Then the P_k's are the monic OPRL with respect to the functional $\mathcal{S} : \mathbb{R}[x] \to \mathbb{R}$ defined by $\mathcal{S}[x^k] := (\mathcal{J}^k)_{1,1}$, and they satisfy $\mathcal{S}_k[P_k^2] = \prod_{i=1}^k w_i > 0$.

Proof. We first show that $\mathcal{S}[P_k(x)] = 0$ for all $k \geq 1$, where $\mathcal{S}$ is the functional defined in the statement. Indeed, by the definition $\mathcal{S}[x^k] := (\mathcal{J}^k)_{1,1}$ we get $\mathcal{S}[P_k(x)] = (P_k(\mathcal{J}))_{1,1}$. Next, for any polynomial $\Phi(x) \in \mathbb{R}[x]$ of degree $\leq k$ we have (**exercise**)

$$(\Phi(\mathcal{J}))_{1,1} = (\Phi(\mathcal{J}_{[k]}))_{1,1} \quad (3.2.31)$$

Therefore $\mathcal{S}[P_k(x)] = (P_k(\mathcal{J}_{[k]}))_{1,1} = 0$ for all $k \geq 1$ by Lemma 3.2.13 and the Cayley–Hamilton theorem.

Next, we need to prove that for all $\ell \geq 0$ we have $\mathcal{S}[x^\ell P_k(x)] = 0$ for all $k > \ell$. We prove this by induction on ℓ; we have already proved the base case $\ell = 0$, so let us assume this is true for some $\ell \geq 0$; then

$$\mathcal{S}[x^{\ell+1}P_k(x)] = \mathcal{S}[x^\ell P_{k+1}(x)] + b_k \mathcal{S}[x^\ell P_{k+1}] + w_k \mathcal{S}[x^\ell P_{k-1}] \quad (3.2.32)$$

which vanishes provided $\ell < k-1$, i.e. provided $\ell+1 < k$, which is the statement for $\ell+1$. Finally, it is also easy to establish that $\mathcal{S}[x^k P_k(x)] = w_k\, \mathcal{S}[x^{k-1}P_{k-1}(x)]$ for all $k \geq 1$, and so $\mathcal{S}[x^k P_k(x)] = \prod_{i=1}^k w_i$ (note that $\mathcal{S}[1] = 1$). □

3.2.4 Connection with the moment problem

The formal notion of orthogonality relative to a linear functional $\mathcal{S}$ and the more familiar notion of orthogonality with respect to a Borel measure with finite moments of all orders are related by the famous *(Hamburg) moment problem*: given a sequence of numbers $(m_k)_{k=0}^{+\infty}$, with $m_0 = 1$, find, if any exists, all Borel probability measures μ on $\mathbb{R}$ such that $m_k = \int_{\mathbb{R}} x^k \mu(\mathrm{d}x)$. An obvious necessary condition is that the truncated Hankel matrices $\mathcal{M}_{[k]} = (m_{i+j-2})_{1 \le i,j \le k}$ are positive-definite for all $k \ge 1$. We have seen that this condition holds true in the formal Favard theorem (Theorem 3.2.16).

It is a classical result that, under the positive-definiteness condition on $\mathcal{M}_{[k]}$, there is always at least one Borel probability measure μ fulfilling the required condition $m_k = \int_{\mathbb{R}} x^k \mu(\mathrm{d}x)$; however, uniqueness is only guaranteed if the m_k's do not grow too fast as $k \to +\infty$. This is for instance ensured if the sequences of b_k's and w_k's are bounded uniformly in k, in which case the support of μ is also bounded. (This is not a necessary condition however.)

There are at least two approaches to the moment problem that we would like to mention here for their connection to OPRL.

Stieltjes transform. Consider a Borel probability measure μ on $\mathbb{R}$ and assume it is absolutely continuous with respect to the Lebesgue measure, $\mu(\mathrm{d}x) = \nu(x)\mathrm{d}x$ and assume it has compact support (i.e. for some $R > 0$ we have $\nu(x) = 0$ for all $|x| > R$). The Steltjes transform of μ is the function $F(z) := \int_{\mathbb{R}} \frac{\mu(\mathrm{d}x)}{z-x} = \int_{-R}^{R} \frac{\nu(x)}{z-x}\mathrm{d}x$. We will come back to this function later, but two main properties can be understood immediately.

- It is a *generating function* of moments, in the sense that it is analytic at least within $|z| > R$, and admits a Taylor expansion at $z = \infty$ whose coefficients are the moments. Indeed, by the geometric series
$$\frac{1}{z-x} = \sum_{k=0}^{+\infty} \frac{x^k}{z^{k+1}}, \qquad \text{for all } |z| > |x|, \tag{3.2.33}$$
we obtain (**exercise**: prove that we can interchange series and integral)
$$F(z) = \int_{-R}^{R} \nu(x) \sum_{k=0}^{+\infty} \frac{x^k}{z^{k+1}} \mathrm{d}x = \sum_{k=0}^{+\infty} \frac{m_k}{z^{k+1}}, \qquad |z| > R. \tag{3.2.34}$$
- On the other hand, $F(z)$ is not defined on the support of ν; it can be shown however under mild analytic assumptions on ν that $F(z)$ is continuous up to the support of ν from above and below the real axis and that the *jump* across the real axis recovers ν:
$$\nu(x) = \frac{\mathrm{i}}{2\pi} \lim_{\epsilon \to 0_+} \left[F(x + \mathrm{i}\epsilon) - F(x - \mathrm{i}\epsilon) \right]. \tag{3.2.35}$$

Exercise 3.2.17. (Hard.) Let the polynomials $P_k(x)$ be defined by $P_{-1}(x) = 0$, $P_0(x) = 1$, and $P_{k+1} = xP_k(x) - P_{k-1}(x)$ for $k \ge 0$. Find the orthogonality measure of these polynomials. (Hint: find the moments $m_k = (\mathcal{J}^k)_{1,1}$ to obtain the formal generating function $F(z) = \sum_{k=0}^{+\infty} \frac{m_k}{z^{k+1}}$ and compute the jump of F across the real axis.) //

Not surprisingly OPRL associated to a measure μ enter the picture in this approach via the theory of Padé approximation (that in general is useful, among other things, to detect poles outside the region of convergence of the Taylor series of an analytic function) applied to the Stieltjes transform $F(z)$.

Spectral theory of Jacobi operators. We consider for simplicity the finite dimensional case as an example first. Namely, in Exercise 3.2.7 the OPRL $P_k(x)$ for a measure with finite support, say of cardinality r, are proved to exist unique with $h_k \neq 0$ up to $k \le r-1$. Let $p_k(x) = P_k(x)/\sqrt{h_k}$ be their orthonormal version, $0 \le k \le r-1$. It can be checked that the three-term recursion $xp_k(x) = a_k p_{k+1} + b_k p_k(x) + a_{k-1} p_{k-1}(x)$ still holds true, as long as $k \le r-1$ and so we can consider the $r \times r$ Jacobi matrix
$$J = \begin{pmatrix} b_0 & a_0 & 0 & \cdots & 0 \\ a_0 & b_1 & a_1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a_{r-3} & b_{r-2} & a_{r-2} \\ 0 & \cdots & 0 & a_{r-2} & b_{r-1} \end{pmatrix}. \tag{3.2.36}$$

We can retrieve the orthogonality measure from J: first, J is real symmetric so

$$J = O \begin{pmatrix} x_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_r \end{pmatrix} O^\top \tag{3.2.37}$$

for some orthogonal $r \times r$ matrix O. We can use Lemma (3.2.12) to compute

$$m_k = (e_1, J^k e_1) = (e_1, O \begin{pmatrix} x_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_r^k \end{pmatrix} O^\top e_1) = (O^\top e_1, \begin{pmatrix} x_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_r^k \end{pmatrix} O^\top e_1) = q_1^2 x_1^k + \cdots + q_r^2 x_r^k, \tag{3.2.38}$$

where we set $(q_1, \ldots, q_n) = e_1^\top O$ (the first row of O). Therefore we recover the orthogonality measure $\mu = q_1^2 \delta_{x_1} + \cdots + q_r^2 \delta_{x_r}$. (Note that $q_i \neq 0$ by Exercise 3.2.15.)

An approach to the moment problem in general consists of the following steps. First, given the sequence of moments one constructs the associated OPRL and compute the Jacobi operator $\mathcal{J}$. Then one performs the construction above for all truncations $\mathcal{J}_{[k]}$, obtaining a sequence of measures μ_k, supported at k points. Extracting a convergent subsequence of measures as $k \to +\infty$ can be done via classical (but advanced) theorems in analysis, like Helly's selection principle. Then the limiting measure provides an answer to the moment problem.

Another approach is to consider directly the infinite-dimensional version of the spectral theorem. Suppose $\mathcal{J}$ defines a bounded operator on $\ell^2(\mathbb{Z}_{\geq 0})$ (this happens if and only if the sequences a_k, b_k are bounded). Then the spectral theory of bounded self-adjoint operators provides a projection-valued measure E such that

$$\mathcal{J} = \int_{\sigma(\mathcal{J})} x \, E(\mathrm{d}x). \qquad (\sigma(\mathcal{J}) := \text{spectrum of } \mathcal{J}.) \tag{3.2.39}$$

Therefore $\mu := (e_1, E e_1)$ gives the orthogonality measure:

$$m_k = (e_1, \mathcal{J}^k e_1) = \int_{\sigma(\mathcal{J})} x^k \, \mu(\mathrm{d}x). \tag{3.2.40}$$

Not surprisingly then, the general case can be tackled by the spectral theory of unbounded self-adjoint operators. A major result is that the solution to the moment problem is unique if and only if the Jacobi operator is essentially self-adjoint. For more details, see Chapter 16 in: Schmudgen, "Unbounded Self-Adjoint Operators on Hilbert Space", GTM, 265. Springer, 2012.

3.2.5 Christoffel–Darboux identity

Proposition 3.2.18 (Christoffel–Darboux identity)**.** *Let $(P_k(x))_{k=0}^{+\infty}$ be a family of monic OPRL. Then, for all integers $n \geq 1$ we have*

$$\sum_{\ell=0}^{n-1} \frac{P_\ell(x) P_\ell(y)}{h_\ell} = \frac{1}{h_{n-1}} \frac{P_n(x) P_{n-1}(x) - P_{n-1}(x) P_n(y)}{x - y} \tag{3.2.41}$$

Proof. It is equivalent to prove the formula

$$\sum_{\ell=0}^{n-1} p_\ell(x) p_\ell(y) = \sqrt{\frac{h_n}{h_{n-1}}} \frac{p_n(x) p_{n-1}(y) - p_{n-1}(x) p_n(y)}{x - y}. \tag{3.2.42}$$

for the orthonormal polynomials $p_k(x) := P_k(x)/\sqrt{h_k}$. The three-term recurrence, see (3.2.14),

$$x p_\ell(x) = a_\ell p_{\ell+1}(x) + b_\ell p_\ell(x) + a_{\ell-1} p_{\ell-1}(x), \quad \ell \geq 1, \qquad x p_0(x) = a_0 p_1(x) + b_0 p_0(x) \tag{3.2.43}$$

can be used to simplify the expression

$$\begin{aligned}(x-y)\sum_{\ell=0}^{n-1}p_\ell(x)p_\ell(y) &= \sum_{\ell=0}^{n-1}(xp_\ell(x))p_\ell(y) - \sum_{\ell=0}^{n-1}p_\ell(x)(yp_\ell(y)) \\ &= \sum_{\ell=0}^{n-1}a_\ell p_{\ell+1}(x)p_\ell(y) + \sum_{\ell=1}^{n-1}a_{\ell-1}p_{\ell-1}(x)p_\ell(y) \\ &\quad - \sum_{\ell=0}^{n-1}a_\ell p_\ell(x)p_{\ell+1}(y) - \sum_{\ell=1}^{n-1}a_{\ell-1}p_\ell(x)p_{\ell-1}(y) \end{aligned} \tag{3.2.44}$$

where the terms with b_ℓ's cancel each other. In the last expression, the first and fourth sums cancel each other but for the term $a_{n-1}p_n(x)p_{n-1}(y)$ and the second and third sum cancel each other but for the term $-a_{n-1}p_{n-1}(x)p_n(y)$ □

Exercise 3.2.19. Let $(p_k(x))_{k=0}^{+\infty}$ be a family of orthonormal polynomials on the real line. Prove the identity (*confluent* Christoffel–Darboux identity)

$$\sum_{\ell=0}^{n-1}p_\ell^2(x) = \sqrt{\frac{h_n}{h_{n-1}}}(p_n'(x)p_{n-1}(x) - p_n(x)p_{n-1}'(x)), \qquad p_k'(x) := \frac{\mathrm{d}}{\mathrm{d}x}p_k(x). \tag{3.2.45}$$

Write the corresponding formula for monic OPRL. (Hint: let $y \to x$ in (3.2.42).) //

3.2.6 Further properties of the OPRL

Proposition 3.2.20 (zeros of OPRL)**.** *Let $(P_k(x))_{k=0}^{+\infty}$ be a family of monic OPRL associated with a Borel measure μ (with infinite support) on $\mathbb{R}$. Let $[a,b]$ be the convex hull of the support of μ. Then $P_k(x)$ has k distinct real zeros lying in $[a,b]$. Moreover, let $x_1^{[k]} < \cdots < x_k^{[k]}$ be the zeros of $P_k(x)$, we have*

$$x_1^{[k+1]} < x_1^{[k]} < x_2^{[k+1]} < \cdots < x_i^{[k+1]} < x_i^{[k]} < x_{i+1}^{[k+1]} < \cdots < x_k^{[k]} < x_{k+1}^{[k+1]}. \tag{3.2.46}$$

The inequalities (3.2.46) go under the name of *interlacing* property for the zeros of OPRL.

Proof. Let $x_1 < \cdots < x_r$ be the zeros of $P_k(x)$ such that: 1) they lie in the convex hull of the support of μ and 2) they have odd multiplicity. Set $Q(x) := \prod_{i=1}^r(x-x_i)$. Note that $Q(x)P(x)$ never changes sign on the support of μ (all the zeros of $Q(x)P(x)$ have even multiplicity). Hence $\int_\mathbb{R} Q(x)P(x)\mu(\mathrm{d}x) \neq 0$; however $\deg_x Q = r \leq k$ and to avoid a contradiction with the orthogonality property we need to have $k = r$. Hence $x_1, \ldots, x_r$ are all the zeros of P_k.

For the second part, by the confluent Christoffel–Darboux formula (3.2.45) we have

$$P_{k+1}'(x)P_k(x) - P_k'(x)P_{k+1}(x) \geq \frac{h_k}{h_0} > 0, \qquad x \in \mathbb{R}. \tag{3.2.47}$$

Hence for any zero x_* of P_{k+1} we have

$$P_{k+1}'(x_*)P_k(x_*) > 0. \tag{3.2.48}$$

Finally, at two consecutive zeros of P_{k+1} the derivative P_{k+1}' must take values with different sign (because the zeros are simple), and so by the inequality (3.2.48) P_k must take values with different sign, hence P_k must have a zero between any two consecutive zeros of P_{k+1}. □

3.2.7 Classical OPLR

Hermite polynomials

The Hermite polynomials correspond to the Gaussian measure $\mu(E) = \int_E \exp(-x^2/2)\mathrm{d}x$. The support of the measure is $\mathbb{R}$. They can be defined in terms of the Rodrigues' formula (3.1.17) above. Their three-term recurrence is given by $b_k = 0$, $w_k = k$ (see Exercise 3.1.4).

Laguerre polynomials

The Laguerre polynomials correspond to the Laguerre measure $\mu(E) = \int_{E\cap(0,+\infty)} \exp(-x)x^{\alpha}\mathrm{d}x$, for a parameter $\alpha > -1$. The support of the measure is $[0, +\infty)$. They can be defined in terms of the Rodrigues' formula

$$P_k(x) = \mathrm{e}^{x}x^{-\alpha}\left[\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^{k}(\mathrm{e}^{-x}x^{\alpha+k})\right] \tag{3.2.49}$$

Exercise 3.2.21. Use (3.2.49) to prove the orthogonality relation with respect to the Laguerre measure. Compute the constants h_k. Use again (3.2.49) to compute the three-term recurrence for the Laguerre polynomials, i.e. compute b_k for $k \geq 0$ and w_k for $k \geq 1$. //

Jacobi polynomials

The Jacobi polynomials correspond to the Jacobi measure $\mu(E) = \int_{E\cap(0,1)} x^{\alpha}(1-x)^{\beta}\mathrm{d}x$, for parameters $\alpha, \beta > -1$. The support of the measure is $[0, 1]$. They can be defined in terms of the Rodrigues' formula

$$P_k(x) = \frac{1}{\prod_{j=1}^{k}(\alpha+\beta+k+j)}x^{-\alpha}(1-x)^{-\beta}\left[\left(-\frac{\mathrm{d}}{\mathrm{d}x}\right)^{k}(x^{\alpha+k}(1-x)^{\beta+k})\right]. \tag{3.2.50}$$

Exercise 3.2.22. Use (3.2.50) to prove that $P_k(x)$ are monic polynomials of degree k and the orthogonality relation with respect to the Jacobi measure. Use again (3.2.50) to compute the three-term recurrence for the Jacobi polynomials, i.e. compute b_k for $k \geq 0$ and w_k for $k \geq 1$. //

3.3 Determinantal structure of correlation functions in unitary invariant models

3.3.1 Christoffel–Darboux kernel

Let V be a potential function as in Definition 2.2.5, so that $\mathbb{R}[x]\mathrm{e}^{-\frac{V(x)}{2}} \subset L^2(\mathbb{R}, \mathrm{d}x)$. Let $\mathcal{K}_n^V$ be the rank n orthogonal projector of the Hilbert space $L^2(\mathbb{R}, \mathrm{d}x)$ onto the subspace $\mathbb{R}[x]_{\leq n-1}\mathrm{e}^{-\frac{V(x)}{2}}$, where $\mathbb{R}[x]_{\leq d}$ is the space of polynomials of degree $\leq d$. In other words, for all $f \in L^2(\mathbb{R}, \mathrm{d}x)$ we have

$$(\mathcal{K}_n^V f)(x) = \int_{\mathbb{R}} K_n^V(x,y)f(y)\mathrm{d}y \tag{3.3.1}$$

for the kernel

$$K_n^V(x,y) = \sum_{\ell=0}^{n-1}\psi_\ell(x)\psi_\ell(y), \qquad \psi_\ell(x) := \frac{P_\ell(x)}{\sqrt{h_\ell}}\mathrm{e}^{-\frac{V(x)}{2}} = p_\ell(x)\mathrm{e}^{-\frac{V(x)}{2}}. \tag{3.3.2}$$

Here $P_\ell(x), p_\ell(x)$ are the OPRL (respectively monic and orthonormal) for the measure $\mu(\mathrm{d}x) = \mathrm{e}^{-V(x)}\mathrm{d}x$, whose existence and properties have been studied in general above. The orthogonality property can be expressed by

$$\int_{\mathbb{R}} \psi_k(x)\psi_\ell(x)\mathrm{d}x = \delta_{k,\ell}. \tag{3.3.3}$$

Definition 3.3.1. $\mathcal{K}_n^V$ and $K_n^V(\cdot,\cdot)$ are called Christoffel–Darboux (CD) projector and CD kernel, respectively. //

The CD operator $\mathcal{K}_n^V$ is a rank n orthogonal projector, a fact which can be equivalently expressed by the two conditions

$$\operatorname{tr}\mathcal{K}_n^V = n, \qquad (\mathcal{K}_n^V)^2 = \mathcal{K}_n. \tag{3.3.4}$$

These two conditions can be equivalently expressed in terms of the CD kernel as

$$\int_{\mathbb{R}} K_n(x,x)\mathrm{d}x = n, \qquad \int_{\mathbb{R}} K_n(x,y)K_n(y,z)\mathrm{d}y = K_n(x,z). \tag{3.3.5}$$

Exercise 3.3.2. Prove (3.3.5) from (3.3.2) directly using the orthogonality (3.3.3). //

Note that by the Christoffel–Darboux identity (Proposition 3.2.18) we have

$$K_n^V(x,y) = \sqrt{\frac{h_n}{h_{n-1}}} \exp\left(-\frac{V(x)+V(y)}{2}\right) \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y} \tag{3.3.6}$$

and by (3.2.45)

$$K_n^V(x,x) = \sqrt{\frac{h_n}{h_{n-1}}} \exp\left(-V(x)\right)\left(p_n'(x)p_{n-1}(x) - p_n(x)p_{n-1}'(x)\right). \tag{3.3.7}$$

The relevance of the CD kernel for Hermitian random matrices with unitary symmetry is manifest from the following proposition.

Proposition 3.3.3. *The joint eigenvalue density of an unitary-invariant ensemble of Hermitian matrices can be written as*

$$\frac{1}{\widehat{Z}_n^V}\Delta^2(x_1,\dots,x_n)\mathrm{e}^{-V(x_1)}\cdots\mathrm{e}^{-V(x_n)} = \frac{1}{n!}\det_{1\le i,j\le n}\left(K_n^V(x_i,x_j)\right). \tag{3.3.8}$$

Proof. First, recall that $\widehat{Z}_n^V = n! D_n$, see Exercise 3.2.8. Moreover, we have (using also Corollary 2.5.3)

$$\Delta^2(x_1,\cdots,x_n)\mathrm{e}^{-V(x_1)}\cdots\mathrm{e}^{-V(x_n)} = h_0\cdots h_{n-1}\left(\det_{1\le i,j\le n}\left(\psi_{j-1}(x_i)\right)\right)^2. \tag{3.3.9}$$

Next, $h_0\cdots h_{n-1} = D_n$ and

$$\begin{aligned}
\left(\det_{1\le i,j\le n}\left(\psi_{j-1}(x_i)\right)\right)^2 &= \det\begin{pmatrix}\psi_0(x_1) & \cdots & \psi_{n-1}(x_1)\\ \vdots & \ddots & \vdots \\ \psi_0(x_n) & \cdots & \psi_{n-1}(x_n)\end{pmatrix}\cdot\det\begin{pmatrix}\psi_0(x_1) & \cdots & \psi_0(x_n)\\ \vdots & \ddots & \vdots \\ \psi_{n-1}(x_1) & \cdots & \psi_{n-1}(x_n)\end{pmatrix}\\
&= \det\left(\begin{pmatrix}\psi_0(x_1) & \cdots & \psi_{n-1}(x_1)\\ \vdots & \ddots & \vdots \\ \psi_0(x_n) & \cdots & \psi_{n-1}(x_n)\end{pmatrix}\cdot\begin{pmatrix}\psi_0(x_1) & \cdots & \psi_0(x_n)\\ \vdots & \ddots & \vdots \\ \psi_{n-1}(x_1) & \cdots & \psi_{n-1}(x_n)\end{pmatrix}\right) &(3.3.10)\\
&= \det\begin{pmatrix}K_n^V(x_1,x_1) & \cdots & K_n^V(x_1,x_n)\\ \vdots & \ddots & \vdots \\ K_n^V(x_n,x_1) & \cdots & K_n^V(x_n,x_n)\end{pmatrix} &(3.3.11)
\end{aligned}$$

and the proof is complete. □

This expression for the joint eigenvalue distribution is particularly useful to study the marginal eigenvalue distributions because of the following "*integrating-out*" formula.

Lemma 3.3.4 (Dyson, 1970)**.** *Let $(X,\mathcal{F},\mu)$ be a measure space and let $K: X^2\to\mathbb{C}$ be a measurable function satisfying*

$$\int_X K(x,y)K(y,z)\mu(\mathrm{d}y) = K(x,z). \tag{3.3.12}$$

Then for all $r\ge 1$ we have the identity

$$\int_X \det_{1\le i,j\le r} K(x_i,x_j)\mu(\mathrm{d}x_r) = \left(\int_X K(x,x)\mu(\mathrm{d}x) - r + 1\right)\det_{1\le i,j\le r-1} K(x_i,x_j). \tag{3.3.13}$$

Proof. Let us expand the determinant in the left side[4]

$$\int_X \det_{1\le i,j\le r} K(x_i,x_j)\mu(\mathrm{d}x_r) = \sum_{\pi\in\mathfrak{S}_r}(-1)^{|\pi|}\int_X K(x_1,x_{\pi(1)})\cdots K(x_r,x_{\pi(r)})\mu(\mathrm{d}x_r). \tag{3.3.14}$$

For each term in the sum over permutations $\pi\in\mathfrak{S}_r$ we have two cases.

[4]See the proof of Lemma 3.1.1 for the notations regarding permutations.

- $\boxed{\pi(r) = r.}$ In this case we regard π as a permutation of $\{1, \cdots, r-1\}$ only, i.e. $\pi \in \mathfrak{S}_{r-1}$ and note that the sign of the permutation π does not change whether we consider it as a permutation in $\mathfrak{S}_r$ or in $\mathfrak{S}_{r-1}$. The corresponding terms in the RHS of (3.3.14) add up to

$$\sum_{\pi \in \mathfrak{S}_{r-1}} (-1)^{|\pi|} K(x_1, x_{\pi(1)}) \cdots K(x_{r-1}, x_{\pi(r-1)}) \int_X K(x_r, x_r)\mu(\mathrm{d}x_r)$$
$$= \left(\int_X K(x,x)\mu(\mathrm{d}x) \right) \det_{1 \le i,j \le r-1} K(x_i, x_j). \tag{3.3.15}$$

- $\boxed{\pi(r) \neq r.}$ In this case there exists necessarily $j \in \{1, \ldots, r-1\}$ such that[5] $\pi = (j,r) \circ \widetilde{\pi}$, where $\widetilde{\pi} \in \mathfrak{S}_r$ is such that $\widetilde{\pi}(r) = r$, and so we view it as $\widetilde{\pi} \in \mathfrak{S}_{r-1}$, while (j,r) is the transposition exchanging j and r and fixing everything else. Note that $(-1)^{|\pi|} = -(-1)^{|\widetilde{\pi}|}$ and in the product $\prod_{i=1}^r K(x_i, x_{\pi(i)})$ the only two factors that involve x_r are $K(x_r, x_{\pi(r)}) = K(x_r, x_j)$ and $K(x_{\pi^{-1}(r)}, x_r) = K(x_{\widetilde{\pi}^{-1}(j)}, x_r)$. Hence, the corresponding terms in the RHS of (3.3.14) add up to

$$\begin{aligned}
&-\sum_{j=1}^{r-1} \sum_{\widetilde{\pi} \in \mathfrak{S}_{r-1}} (-1)^{|\widetilde{\pi}|} \left(\prod_{\substack{1 \le i \le r-1 \\ \widetilde{\pi}(i) \neq j}} K(x_i, x_{\widetilde{\pi}(i)}) \right) \int_X K(x_{\widetilde{\pi}^{-1}(j)}, x_r) K(x_r, x_j)\mu(\mathrm{d}x_r) \\
&= -\sum_{j=1}^{r-1} \sum_{\widetilde{\pi} \in \mathfrak{S}_{r-1}} (-1)^{|\widetilde{\pi}|} \left(\prod_{\substack{1 \le i \le r-1 \\ \widetilde{\pi}(i) \neq j}} K(x_i, x_{\widetilde{\pi}(i)}) \right) K(x_{\widetilde{\pi}^{-1}(j)}, x_j) \\
&= -\sum_{j=1}^{r-1} \sum_{\widetilde{\pi} \in \mathfrak{S}_{r-1}} (-1)^{|\widetilde{\pi}|} \prod_{1 \le i \le r-1} K(x_i, x_{\widetilde{\pi}(i)}) \\
&= -(r-1) \det_{1 \le i,j \le r-1} K(x_i, x_j).
\end{aligned} \tag{3.3.16}$$

The proof is complete by combining (3.3.15) and (3.3.16). □

Corollary 3.3.5. *With the assumptions of Lemma 3.3.4 and denoting*

$$d := \int_X K(x,x)\mu(\mathrm{d}x), \tag{3.3.17}$$

for any integers r, s satisfying $r \ge s \ge 1$ we have

$$\int_{X^s} \det_{1 \le i,j \le r} K(x_i, x_j)\mu(\mathrm{d}x_{r-s+1}) \cdots \mu(\mathrm{d}x_r) = \left(\prod_{j=1}^s (d - r + j) \right) \det_{1 \le i,j \le r-s} K(x_i, x_j). \tag{3.3.18}$$

Note that in view of (3.3.5) we deduce from the above corollary that for the Christoffel–Darboux kernel K_n^V in (3.3.2), we have

$$\boxed{\int_{\mathbb{R}^s} \det_{1 \le i,j \le n} K_n^V(x_i, x_j)\, \mathrm{d}x_{n-s+1} \cdots \mathrm{d}x_n = s! \det_{1 \le i,j \le n-s} K^V(x_i, x_j), \quad \text{for all } 1 \le s \le n.} \tag{3.3.19}$$

[5] We compose from right to left as usual for composition of functions, namely we first perform $\widetilde{\pi}$ and then exchange j with r.

3.3.2 Correlation functions

We have seen (Proposition 3.3.3) that the eigenvalues $x_1, \ldots, x_n$ of a random Hermitian matrix distributed according to a unitary-invariant model (Definition 2.2.5) have the joint probability distribution

$$\frac{1}{n!} \det_{1 \le i,j \le n} (K_n^V(x_i, x_j))\, \mathrm{d}x_1 \cdots \mathrm{d}x_n, \tag{3.3.20}$$

which is symmetric in the x_i's.

One-point correlation function

A first natural question is to describe the eigenvalue density[6]

$$\rho_1(x) := \lim_{\epsilon \to 0_+} \frac{1}{\epsilon} \mathbb{E}\left[\text{number of eigenvalues in } (x - \frac{\epsilon}{2}, x + \frac{\epsilon}{2})\right]. \tag{3.3.21}$$

This function would represent the average histogram of eigenvalues.

In more precise terms, introduce the random variable (for any Borel $A \subseteq \mathbb{R}$)

$$\#_A := \text{number of eigenvalues in } A. \tag{3.3.22}$$

Then the function $\rho_1(x)$ is defined in terms of the property

$$\mathbb{E}\left[\#_A\right] = \int_A \rho_1(x)\mathrm{d}x, \quad \text{for all Borel sets } A \subseteq \mathbb{R} \tag{3.3.23}$$

and is called **one-point correlation function**. We now show that it exists by a direct computation:[7]

$$\begin{aligned}
\mathbb{E}[\#_A] &= \frac{1}{n!} \int_{\mathbb{R}^n} (1_A(x_1) + \cdots + 1_A(x_n)) \det_{1 \le i,j \le n} (K_n^V(x_i, x_j))\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \frac{1}{(n-1)!} \int_A \left(\int_{\mathbb{R}^{n-1}} \det_{1 \le i,j \le n} (K_n^V(x_i, x_j))\mathrm{d}x_2 \cdots \mathrm{d}x_n \right) \mathrm{d}x_1 \\
&= \int_A K_n^V(x_1, x_1)\, \mathrm{d}x_1
\end{aligned} \tag{3.3.24}$$

where in the last step we use (3.3.19) (with $s = n - 1$). Therefore

$$\rho_1(x) = K_n^V(x, x). \tag{3.3.25}$$

Remark 3.3.6. Sanity check: obviously, $\#_{\mathbb{R}} = n$ surely, and indeed $\mathbb{E}[\#_{\mathbb{R}}] = \int_{\mathbb{R}} K_n^V(x, x)\mathrm{d}x = n$. //

Exercise 3.3.7. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function with polynomial growth as $x \to \pm\infty$. Prove that

$$\mathbb{E}[\mathrm{tr}\,(f(M))] = \int_{\mathbb{R}} f(x) K_n^V(x, x)\mathrm{d}x. \tag{3.3.26}$$

where the expectation on the left is taked with respect to an unitary-invariant ensemble of Hermitian matrices with potential V (Definition 2.2.5). Deduce the following identity for the moments of $\rho_1(x) = K_n^V(x, x)$:

$$\int_{\mathbb{R}} x^\ell \rho_1(x)\mathrm{d}x = \mathbb{E}[\mathrm{tr}\,(M^\ell)]. \tag{3.3.27}$$

//

[6] Hereafter we use $\mathbb{E}$ to denote the probabilistic expectation.

[7] We denote $1_S(x) := \begin{cases} 1, & x \in S, \\ 0, & \text{else}, \end{cases}$ the characteristic function of a set S.

Two-point correlation function

Since the joint distribution of eigenvalues is not factorized we cannot expect that $\mathbb{E}[\#_{A_1} \cdot \#_{A_2}] = \mathbb{E}[\#_{A_1}]\mathbb{E}[\#_{A_2}]$, not even for disjoint Borel sets $A_1, A_2 \subseteq \mathbb{R}$, $A_1 \cap A_2 = \emptyset$. In other words, the random variables $\#_{A_1}, \#_{A_2}$ are not independent (even when $A_1 \cap A_2 = \emptyset$).

Let us compute, for $n \geq 2$ and disjoint Borel sets $A_1, A_2 \subseteq \mathbb{R}$, $A_1 \cap A_2 = \emptyset$,

$$\begin{aligned}\mathbb{E}[\#_{A_1} \cdot \#_{A_2}] &= \frac{1}{n!}\int_{\mathbb{R}^n}\left(\sum_{l=1}^n 1_{A_1}(x_l)\right)\left(\sum_{m=1}^n 1_{A_2}(x_m)\right)\det_{1\leq i,j\leq n}(K_n^V(x_i,x_j))\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &= \frac{1}{n!}\sum_{l,m=1}^n\int_{\mathbb{R}^n} 1_{A_1}(x_l)1_{A_2}(x_m)\det_{1\leq i,j\leq n}(K_n^V(x_i,x_j))\mathrm{d}x_1\cdots\mathrm{d}x_n. \qquad (3.3.28)\end{aligned}$$

Here the summands corresponding to $l = m$ do not contribute because $A_1 \cap A_2 = \emptyset$, and so by symmetry we have

$$\begin{aligned}\mathbb{E}[\#_{A_1} \cdot \#_{A_2}] &= \frac{1}{(n-2)!}\int_{A_1\times A_2}\left(\int_{\mathbb{R}^{n-2}}\det(K_n^V(x_i,x_j))\mathrm{d}x_3\cdots\mathrm{d}x_n\right)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int_{A_1\times A_2}\rho_2(x_1,x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \qquad (3.3.29)\end{aligned}$$

where we use (3.3.19) with $s = n-2$ and we set

$$\rho_2(x_1,x_2) := \det\begin{pmatrix}K_n^V(x_1,x_1) & K_n^V(x_1,x_2)\\ K_n^V(x_2,x_1) & K_n^V(x_2,x_2)\end{pmatrix} = K_n^V(x_1,x_1)K_n^V(x_2,x_2) - K_n^V(x_1,x_2)K_n^V(x_2,x_1). \qquad (3.3.30)$$

The function $\rho_2(x_1,x_2)$ is called **two-point correlation function**.

Let us repeat the same computation for $A_1 = A_2 = A$ instead; we get

$$\begin{aligned}\mathbb{E}[(\#_A)^2] &= \mathbb{E}\left[\left(\sum_{l=1}^n 1_A(x_l)\right)\left(\sum_{m=1}^n 1_A(x_m)\right)\right] \\ &= \frac{1}{n!}\sum_{l,m=1}^n\int_{\mathbb{R}^n}1_A(x_l)1_A(x_m)\det_{1\leq i,j\leq n}(K_n^V(x_i,x_j))\mathrm{d}x_1\cdots\mathrm{d}x_n. \qquad (3.3.31)\end{aligned}$$

Now divide the sum into terms with $l = m$ and $l \neq m$, and by symmetry we have

$$\begin{aligned}\mathbb{E}[(\#_A)^2] &= \frac{1}{(n-1)!}\int_A\left(\int_{\mathbb{R}^{n-1}}\det_{1\leq i,j\leq n}(K_n^V(x_i,x_j))\mathrm{d}x_2\cdots\mathrm{d}x_n\right)\mathrm{d}x_1 \\ &\qquad + \frac{1}{(n-2)!}\int_{A^2}\left(\int_{\mathbb{R}^{n-2}}\det_{1\leq i,j\leq n}(K_n^V(x_i,x_j))\mathrm{d}x_3\cdots\mathrm{d}x_n\right)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int_A\rho_1(x_1)\mathrm{d}x_1 + \int_{A^2}\rho_2(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \mathbb{E}[\#_A] + \int_{A^2}\rho_2(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2. \qquad (3.3.32)\end{aligned}$$

Summarizing:

- For $A_1, A_2 \subseteq \mathbb{R}$ disjoint Borel sets $A_1 \cap A_2 = \emptyset$

$$\mathbb{E}[\#_{A_1} \cdot \#_{A_2}] = \int_{A_1\times A_2}\rho_2(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2. \qquad (3.3.33)$$

- For $A \subseteq \mathbb{R}$ Borel set we have

$$\mathbb{E}[\#_A \cdot (\#_A - 1)] = \int_{A^2}\rho_2(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2. \qquad (3.3.34)$$

Note that we can express the result uniformly for both cases: $\mathbb{E}[\#_{A_1} \cdot \#_{A_2}]$ for A_1, A_2 disjoint is the expected number of ordered pairs (λ_1, λ_2) of distinct eigenvalues with $\lambda_1 \in A_1$ and $\lambda_2 \in A_2$; in the same way, also $\mathbb{E}[\#_A \cdot (\#_A - 1)]$ is the number of ordered pairs (λ_1, λ_2) of distinct eigenvalues with $\lambda_1 \in A$ and $\lambda_2 \in A$.

Therefore, by additivity we obtain the following uniform interpretation for the quantity

$$m_2(A \times B) = \int_{A\times B} \rho_2(x_1, x_2)\mathrm{d}x \tag{3.3.35}$$

for arbitrary Borel sets $A, B \subseteq \mathbb{R}$: it is the expected number of ordered pairs of distinct eigenvalues (λ_1, λ_2) with $\lambda_1 \in A$ and $\lambda_2 \in B$. By standard theorems about measures on product spaces (here, $\mathbb{R}\times\mathbb{R}$), m_2 defines a measure on $\mathbb{R}^2$, called **two-point correlation measure** and its density with respect to the Lebesgue measure is the two-point correlation function ρ_2.

Exercise 3.3.8. Let $f, g : \mathbb{R} \to \mathbb{R}$ be two continuous functions with polynomial growth at $x \to \pm\infty$. Prove that

$$\mathbb{E}[\operatorname{tr} f(M) \cdot \operatorname{tr} g(M)] = \int_{\mathbb{R}^2} f(x)g(y)\rho_2(x, y)\mathrm{d}x\mathrm{d}y + \int_{\mathbb{R}} f(x)g(x)\rho_1(x)\mathrm{d}x. \tag{3.3.36}$$

//

Correlation functions in general

For $k \geq 1$ and Borel sets $A_1, \cdots, A_k$ define (assuming the size n of the random matrix satisfies $n \geq k$)

$$m_k(A_1 \times \cdots \times A_k) \tag{3.3.37}$$

to be the expected number of ordered $k$tuples of eigenvalues $(\lambda_1, \ldots, \lambda_k)$ with $\lambda_i \in A_i$.

Exercise 3.3.9.
1. Prove that for any Borel sets $A_1, \ldots, A_k \subseteq \mathbb{R}$ which are pairwise disjoint, $A_i \cap A_j = \emptyset$ for all $i \neq j$, we have

$$m_k(A_1 \times \cdots \times A_k) = \mathbb{E}[\#_{A_1} \cdots \#_{A_k}] \tag{3.3.38}$$

2. Prove that for all Borel sets $A \subseteq \mathbb{R}$

$$m_k(A^k) = k!\, \mathbb{E}\left[\binom{\#_A}{k}\right]. \tag{3.3.39}$$

where $\binom{\#_A}{k} := \frac{\#_A \cdot (\#_A - 1) \cdots (\#_A - k + 1)}{k!}$.

3. Finally, prove that for any $1 \leq s \leq k$, any collection of integers $k_1, \ldots, k_s > 0$ satisfying $k_1 + \cdots + k_s = k$, and any disjoint Borel subsets $A_1, \ldots, A_s \subseteq \mathbb{R}$ ($A_i \cap A_j = \emptyset$ for all $i \neq j$) we have

$$m_k(A_1^{k_1} \times \cdots \times A_s^{k_s}) = k_1! \cdots k_s!\, \mathbb{E}\left[\binom{\#_{A_1}}{k_1} \cdots \binom{\#_{A_s}}{k_s}\right]. \tag{3.3.40}$$

//

Theorem 3.3.10. *m_k in (3.3.37) defines a measure on $\mathbb{R}^k$ which is absolutely continuous with respect to the Lebesgue measure and with density*

$$\rho_k(x_1, \ldots, x_k) = \det_{1\leq i,j\leq k} K_n^V(x_i, x_j). \tag{3.3.41}$$

Definition 3.3.11. m_k is called k-point **correlation measure** and ρ_k is called k-point **correlation function**. //

Proof. By standard theorems in measure theory, m_k defined in (3.3.37) extends to a measure on $\mathbb{R}^k$. Note that any "rectangle" $B_1 \times \cdots \times B_k$, for arbitrary Borel sets $B_i \subseteq \mathbb{R}$ can be decomposed into a finite disjoint union of sets of the form $A_1^{k_1} \times \cdots \times A_s^{k_s}$ for disjoint Borel $A_i \subseteq \mathbb{R}$, with $k_1 + \cdots + k_s = k$. Therefore we only need to prove the identity

$$\mathbb{E}\left[\binom{\#_{A_1}}{k_1} \cdots \binom{\#_{A_s}}{k_s}\right] = \frac{1}{k_1! \cdots k_s!} \int_{A_1^{k_1} \times \cdots \times A_s^{k_s}} \det_{1 \le i,j \le k} K_n^V(x_i, x_j) \mathrm{d}x_1 \cdots \mathrm{d}x_k. \tag{3.3.42}$$

To prove it, first note that

$$\binom{\#_A}{k} = \sum_{\substack{J \subseteq \{1,\dots,n\} \\ |J| = k}} \prod_{j \in J} 1_A(x_j), \tag{3.3.43}$$

because both sides count the number of unordered k-tuples of eigenvalues lying in A. Hence

$$\begin{aligned} \mathbb{E}\left[\binom{\#_{A_1}}{k_1} \cdots \binom{\#_{A_s}}{k_s}\right] &= \frac{1}{n!} \int_{\mathbb{R}^n} \det_{1 \le p,q \le n} K_n^V(x_p, x_q) \sum_{\substack{J_1,\dots,J_s \subseteq \{1,\dots,n\} \\ |J_i| = k_i}} \prod_{i=1}^s \prod_{j \in J_i} 1_{A_i}(x_j) \\ &= \frac{1}{n!} \int_{\mathbb{R}^n} \det_{1 \le p,q \le n} K_n^V(x_p, x_q) \sum_{\substack{J_1,\dots,J_s \subseteq \{1,\dots,n\} \\ |J_i| = k_i \\ J_a \cap J_b = \emptyset}} \prod_{i=1}^s \prod_{j \in J_i} 1_{A_i}(x_j), \end{aligned} \tag{3.3.44}$$

where in the last step we use that the A_i's are pairwise disjoint. Due to the symmetry in $x_1, \dots, x_n$ of the integrand, it is enough to compute the contribution from $J_1 = \{1, \dots, k_1\}$, $J_2 = \{k_1{+}1, \dots, k_1{+}k_2\}$, $\dots$, $J_s = \{k_1 + \cdots + k_{s-1} + 1, \dots, k_1 + \cdots + k_s\}$ and multiply it for the number of ways in which we can choose $J_1, \dots, J_s \subseteq \{1, \dots, n\}$ with $|J_i| = k_i$ and pairwise disjoint; the latter factor is the multinomial coefficient

$$\binom{n}{k_1 \cdots k_s} = \frac{n!}{k_1! \cdots k_s! (n - \sum_{i=1}^s k_i)!} \tag{3.3.45}$$

hence we get

$$\begin{aligned} \mathbb{E}\left[\binom{\#_{A_1}}{k_1} \cdots \binom{\#_{A_s}}{k_s}\right] &= \frac{1}{k! \cdots k_s! (n - \sum_{i=1}^s k_i)!} \int_{A_1^{k_1} \times \cdots \times A_s^{k_s}} \left(\int_{\mathbb{R}^{n-k}} \det_{1 \le p,q \le n} K_n^V(x_p, x_q) \right) \mathrm{d}x_1 \cdots \mathrm{d}x_k \\ &= \frac{1}{k! \cdots k_s!} \int_{A_1^{k_1} \times \cdots \times A_s^{k_s}} \det_{1 \le p,q \le k} K_n^V(x_p, x_q) \, \mathrm{d}x_1 \cdots \mathrm{d}x_k \end{aligned} \tag{3.3.46}$$

where in the last step we use (3.3.19) with $s = n - k$. □

Remark 3.3.12. Note that m_k is not in general a probability measure on $\mathbb{R}^k$:

$$m_k(\mathbb{R}^k) = \int_{\mathbb{R}^k} \rho_k(x_1, \dots, x_k) \mathrm{d}x_1 \cdots \mathrm{d}x_k = \frac{n!}{(n-k)!}. \tag{3.3.47}$$

//

3.3.3 CUE and a first encounter with the Sine-kernel

The situation for the CUE can be analyzed in a similar way. First, the joint eigenvalue distribution can be expressed as (**exercise**, mimic the Hermitian case, Proposition 3.3.3)

$$\frac{1}{n!(2\pi)^n} |\Delta|^2 (\mathrm{e}^{\mathrm{i}\varphi_1}, \dots, \mathrm{e}^{\mathrm{i}\varphi_n}) = \frac{1}{n!} \det_{1 \le i,j \le n} K_n^{\mathsf{CUE}}(\varphi_i, \varphi_j) \tag{3.3.48}$$

where[8]

$$K_n^{\mathsf{CUE}}(\varphi_1, \varphi_2) = \sum_{\ell=0}^{n-1} \frac{\mathrm{e}^{\mathrm{i}(\varphi_1 - \varphi_2)\ell}}{2\pi} = \frac{1}{2\pi} \frac{\mathrm{e}^{\mathrm{i}(\varphi_1 - \varphi_2)n} - 1}{\mathrm{e}^{\mathrm{i}(\varphi_1 - \varphi_2)} - 1}. \tag{3.3.49}$$

[8] The last equality is also true when $\varphi_1 = \varphi_2$ in the sense of the limit $\varphi_2 \to \varphi_1$. We shall omit this remark in similar situations in the following.

K_n^{CUE} is the kernel of the orthogonal projector of $L^2(S^1, \mathrm{d}\varphi)$ onto the span of $\mathrm{e}^{\mathrm{i}\ell\varphi}$ for $0 \le \ell \le n-1$. As such, we have

$$\int_0^{2\pi} K_n^{\mathsf{CUE}}(\varphi,\varphi)\mathrm{d}\varphi = n, \qquad \int_0^{2\pi} K_n^{\mathsf{CUE}}(\varphi_1,\varphi_2)K_n^{\mathsf{CUE}}(\varphi_2,\varphi_3)\mathrm{d}\varphi_2 = K_n^{\mathsf{CUE}}(\varphi_1,\varphi_3), \tag{3.3.50}$$

which are the analogue of (3.3.5). Note however the kernel is not symmetric, but rather self-adjoint:

$$K_n^{\mathsf{CUE}}(\varphi_2,\varphi_1) = \overline{K_n^{\mathsf{CUE}}(\varphi_1,\varphi_2)}. \tag{3.3.51}$$

(Again, $\bar{z} = x - \mathrm{i}y$ is the complex conjugate of $z = x + \mathrm{i}y$.)

In particular Lemma 3.3.4 and Corollary 3.3.5 apply to K_n^{CUE} as well.

The correlation functions $\rho_k(\varphi_1,\dots,\varphi_k)$ are defined by the property

$$\mathbb{E}\left[\binom{\#_{A_1}}{k_1}\cdots\binom{\#_{A_s}}{k_s}\right] = \frac{1}{k_1!\cdots k_s!}\int_{A_1^{k_1}\times\cdots\times A_s^{k_s}} \rho_k(\varphi_1,\dots,\varphi_k)\mathrm{d}\varphi_1\cdots\mathrm{d}\varphi_k, \tag{3.3.52}$$

for $k_1+\cdots+k_s = k$ and $A_1,\dots,A_s$ disjoint Borel sets in $[0,2\pi)$. The computation of the correlation functions proceeds similarly to the Hermitian case: explicitly (**exercise**)

$$\rho_k(\varphi_1,\dots,\varphi_k) = \det_{1\le i,j\le k} K_n^{\mathsf{CUE}}(\varphi_i,\varphi_j). \tag{3.3.53}$$

An equivalent formulation of the CUE kernel

It is customary to rewrite the CUE kernel in a more symmetric way as follows

$$K_n^{\mathsf{CUE}}(\varphi_1,\varphi_2) = \frac{1}{2\pi}\frac{\mathrm{e}^{\mathrm{i}(\varphi_1-\varphi_2)n}-1}{\mathrm{e}^{\mathrm{i}(\varphi_1-\varphi_2)}-1} = \frac{1}{2\pi}\frac{\mathrm{e}^{\frac{\mathrm{i}}{2}n(\varphi_1-\varphi_2)}}{\mathrm{e}^{\frac{\mathrm{i}}{2}(\varphi_1-\varphi_2)}}\frac{\sin\left(\frac{n}{2}(\varphi_1-\varphi_2)\right)}{\sin\left(\frac{1}{2}(\varphi_1-\varphi_2)\right)} = \frac{q_n(\varphi_1)}{q_n(\varphi_2)}\widehat{K}_n^{\mathsf{CUE}} \tag{3.3.54}$$

where we set

$$q_n(\varphi) := \mathrm{e}^{\mathrm{i}\frac{n-1}{2}\varphi}, \qquad \widehat{K}_n^{\mathsf{CUE}}(\varphi_1,\varphi_2) := \frac{1}{2\pi}\frac{\sin\left(\frac{n}{2}(\varphi_1-\varphi_2)\right)}{\sin\left(\frac{1}{2}(\varphi_1-\varphi_2)\right)}. \tag{3.3.55}$$

Note that the CUE eigenvalue distribution can be equivalently expressed as

$$\frac{1}{n!}\det_{1\le i,j\le n} K_n^{\mathsf{CUE}}(\varphi_i,\varphi_j) = \frac{1}{n!}\det_{1\le i,j\le n} \widehat{K}_n^{\mathsf{CUE}}(\varphi_i,\varphi_j) \tag{3.3.56}$$

and similarly for the correlation functions

$$\rho_k(\varphi_1,\dots,\varphi_k) = \det_{1\le i,j\le k} \widehat{K}_n^{\mathsf{CUE}}(\varphi_i,\varphi_j). \tag{3.3.57}$$

Scaling and Sine kernel

The one-point correlation function ρ_1, i.e. the eigenvalue density, is uniform:

$$\rho_1^{\mathsf{CUE}}(\varphi) = K_n^{\mathsf{CUE}}(\varphi,\varphi) = \widehat{K}_n^{\mathsf{CUE}}(\varphi,\varphi) = \frac{n}{2\pi}. \tag{3.3.58}$$

It is therefore natural to rescale the eigenvalues according to

$$(-\pi,\pi) \ni \varphi_i = \frac{2\pi}{n}x_i, \qquad x_i \in \mathbb{R}. \tag{3.3.59}$$

Proposition 3.3.13. *For all $\varphi \in [-\pi,\pi)$ and all $x,y \in \mathbb{R}$ we have*

$$\lim_{n\to+\infty}\frac{2\pi}{n}\widehat{K}_n^{\mathsf{CUE}}\left(\varphi+\frac{2\pi}{n}x, \varphi+\frac{2\pi}{n}y\right) = K^{\mathsf{sine}}(x,y) := \frac{\sin(\pi(x-y))}{\pi(x-y)}. \tag{3.3.60}$$

Note that the statement can be equivalently formulated as

$$\lim_{n\to+\infty} \frac{1}{\rho_1(\varphi)} \widehat{K}_n^{\mathsf{CUE}} \left(\varphi + \frac{2\pi}{n}x, \varphi + \frac{2\pi}{n}y\right) = K^{\mathsf{sine}}(x, y). \tag{3.3.61}$$

It follows that for large size n, the microscopic eigenvalue correlations $(2\pi/n)^k \rho_k(\varphi + \frac{2\pi}{n}x_1, \ldots, \varphi + \frac{2\pi}{n}x_k)$ converge to

$$\rho_k^{\mathsf{sine}}(x_1, \ldots, x_k) = \det_{1\le i,j\le k} \left(\frac{\sin(\pi(x_i - x_j))}{\pi(x_i - x_j)}\right). \tag{3.3.62}$$

This type of correlation functions is *universal*, in a sense that we will explore further in the rest of the course. However, first we want to describe a more precise notion of convergence of eigenvalue distributions as $n \to +\infty$ which can be given by the theory of point processes.

Remark 3.3.14. The universality of the Sine kernel goes well beyond the realm of random matrices. One outstanding example is the discovery by Montgomery and Dyson that the same kernel describes the two-point correlation of zeros of the Riemann zeta function (generalized to the k-point correlation function by Hejhal and Rudnick–Zarnack). //

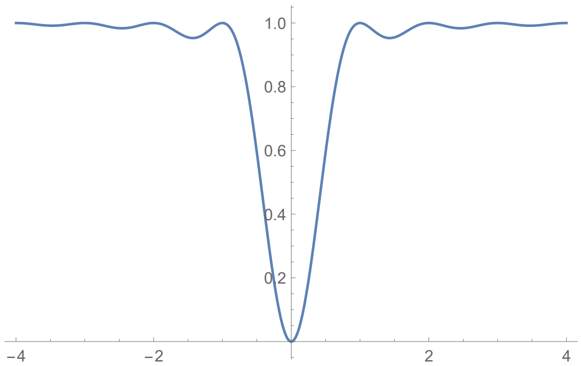


Figure 3.1: Two-point correlation function $\rho_2^{\mathsf{sine}}(x_1, x_2) = 1 - K^{\mathsf{sine}}(x_1, x_2)K^{\mathsf{sine}}(x_2, x_1) = 1 - \frac{\sin^2(\pi y)}{\pi^2 y^2}$ as a function of $y := x_1 - x_2$.

Chapter 4

Determinantal Point Processes

4.1 Point Processes and correlation functions

Let E be a locally compact[1] separable[2] complete[3] metric space. In applications usually E is $\mathbb{R}^d, S^d, \mathbb{Z}^d$. Let $\mathcal{B}(E)$ be the sigma-algebra of Borel subsets of E (i.e. the smallest sigma-algebra of subsets of E containing all closed and open subsets of E).

Definition 4.1.1. A subset $X \subseteq E$ is said to be **locally finite** if and only if $X \cap B$ is a finite set for all bounded $B \subseteq E$. Let $\mathrm{Conf}\,(E)$ be the set of locally finite subsets of E.

For all bounded subsets $B \subseteq E$, let the **counting function of** B be[4]

$$\#_B : \mathrm{Conf}\,(E) \to \mathbb{Z}_{\geq 0} : X \mapsto \#(X \cap B). \tag{4.1.1}$$

Namely, they assign to any locally finite subset of E the number of points which are in B.

Let $\mathcal{B}(\mathrm{Conf}\,(E))$ be the smallest sigma-algebra of subsets of $\mathrm{Conf}\,(E)$ that makes the counting functions measurable.

A **(simple)**[5] **point process** on E is a probability measure $\mathbb{P}$ on $(\mathrm{Conf}\,(E), \mathcal{B}(\mathrm{Conf}\,(E)))$. //

A convenient way to characterize point processes is given by correlation functions, which are a direct generalization of the notion we have introduced for the eigenvalues of random matrices.

Definition 4.1.2. Let ν be a fixed reference measure on E and let a point process on E be given. A function $\rho_k \in L^1_{loc}(E^k, \nu^{\otimes k})$ is called k-**point function** if and only if for any $1 \leq s \leq k$, any collection of integers $k_1, \dots, k_s > 0$ satisfying $k_1 + \cdots + k_s = k$, and any disjoint Borel subsets $A_1, \dots, A_s \in \mathcal{B}(E)$ ($A_i \cap A_j = \emptyset$ for all $i \neq j$) we have

$$k_1! \cdots k_s!\, \mathbb{E}\left[\binom{\#_{A_1}}{k_1} \cdots \binom{\#_{A_s}}{k_s}\right] = \int_{A_1^{k_1} \times \cdots \times A_s^{k_s}} \rho_k(x_1, \dots, x_k)\nu(\mathrm{d}x_1) \cdots \nu(\mathrm{d}x_k). \tag{4.1.2}$$

In all our examples, $E = \mathbb{R}, S^1, \mathbb{Z}$ a reference measure is given (Lebesgue measure, arc-length measure, counting measure) and the correlation functions, if they exists, are tacitly assumed to be defined with respect to these measures. //

Example 4.1.3.
1. Eigenvalue process (on $E = \mathbb{R}$ for unitary-invariant ensembles of Hermitian matrices, on $E = S^1$ for the CUE).
2. The previous example is contained in the class of processes with a deterministic finite number n of particles. An easy construction is the following. Let $n \geq 1$ be a fixed number and let $p(x_1, \dots, x_n)$ be a symmetric probability density on $\mathbb{R}^n$. It defines a point process by taking

[1] i.e., for all $x \in E$ there exists an open $U \subseteq E$ and a compact $K \subseteq E$ such that $x \in U \subseteq K$.

[2] i.e., there exists a countable dense subset of E.

[3] i.e., all Cauchy sequences converge.

[4] Here $\#(S)$ denotes the cardinality of the set S.

[5] More complicated point processes can account for particles with multiplicity.

the random configuration of points to be the set $\{x_1, \ldots, x_n\}$ where $(x_1, \ldots, x_n)$ is a random vector valued random variable with joint probability density function given by p (and so, $x_i \neq x_j$ almost surely so that almost surely there are n particles in the configuration). The correlation functions exist and are given by (**exercise**)

$$\rho_k(x_1, \ldots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \ldots, x_n) \mathrm{d}x_{k+1} \cdots \mathrm{d}x_n. \tag{4.1.3}$$

(**Exercise**: what are the correlation functions for $k > n$?)

3. Let X be a random variable with range the non-negative integers n; we say that X is a Poisson random variable of intensity $\lambda > 0$, and write $X \sim \mathrm{Pois}(\lambda)$, if and only if

$$\mathbb{P}(X = n) = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}. \tag{4.1.4}$$

The Poisson process of *intensity* $\lambda > 0$ on $\mathbb{R}$ is the unique point process such that $\#_{[a,b)} \sim \mathrm{Pois}(\lambda(b-a))$ for all $a < b$ and such that $\#_{[a,b)}$ and $\#_{[c,d)}$ are independent for all $a < b \leq c < d$. The correlation functions exist and are given by (**exercise**)

$$\rho_k(x_1, \ldots, x_k) = \lambda^k. \tag{4.1.5}$$

Note in particular that the Poisson point process has infinitely many particles. The example can be generalized naturally by taking λ no longer constant, but rather $\lambda \in L^1_{loc}$ and in this case $\rho_k(x_1, \ldots, x_k) = \lambda(x_1) \cdots \lambda(x_k)$.

//

The following important theorem, which we do not prove here, allows one to define point processes just by assigning correlation functions.

Theorem 4.1.4 (Lenard, 1973)**.** *Let E be a locally compact, separable, complete metric space, and let ν be a fixed measure on E. Let symmetric functions $\rho_k \in L^1_{loc}(E^k, \nu^{\otimes k})$ be given for all $k \geq 1$. The following two assertions are equivalent.*

1. *$\rho_k(x_1, \ldots, x_k)$ are the correlation functions of a point process on E.*

2. *For any collection $\phi_0, \phi_1, \ldots, \phi_N$ of compactly supported measurable functions $\phi_k : E^k \to \mathbb{R}$ satisfying*

$$\phi_0 + \sum_{k=1}^{N} \sum_{\substack{1 \leq i_1, \cdots, i_k \leq N \\ i_a \neq i_b}} \phi_k(x_{i_1}, \ldots, x_{i_k}) \geq 0 \tag{4.1.6}$$

we have

$$\phi_0 + \sum_{k=1}^{N} \int_{E^k} \phi_k(x_1, \ldots, x_k) \rho_k(x_1, \ldots, x_k) \mathrm{d}x_1 \cdots \mathrm{d}x_k \geq 0. \tag{4.1.7}$$

Moreover, if the ρ_k's satisfy these conditions, the point process with correlation functions ρ_k is unique if and only if $\#_B$ are determined by their moments for all Borel $B \subseteq E$.

The criterion can be quite hard to apply in general. However, we can restrict to a subclass of point processes for which the situation is simpler and which covers all examples that we need from Random Matrix Theory (and, actually, many more).

4.2 Determinantal Point Processes

The following notion goes back to Macchi (1975).

Definition 4.2.1. A point process on E is called **determinantal** if and only if it has all correlation functions ρ_k and moreover

$$\rho_k(x_1, \dots, x_k) = \det_{1 \le i,j \le k} K(x_i, x_j) \tag{4.2.1}$$

for some function $K : E \times E \to \mathbb{C}$ called **correlation kernel** satisfying

$$\overline{K(x,y)} = K(y,x). \tag{4.2.2}$$

//

Remark 4.2.2.
1. Note that as long as the determinantal point process is concerned, another kernel $\widetilde{K}(x,y) = q(x)K(x,y)/q(y)$ (for a nonzero function $q : E \to \mathbb{C} \setminus \{0\}$) defines the same correlation functions, hence the same point process. Therefore, the correlation kernel for a determinantal process is not unique.
2. The matrices $(K(x_i, x_j))_{i,j=1}^k$ need to be semipositive-definite[6] (for almost all $(x_1, \dots, x_k) \in \mathbb{R}^k$) by Sylvester's criterion because $\rho_k(x_1, \dots, x_k) \ge 0$ for almost all $(x_1, \dots, x_k) \in \mathbb{R}$ and all $k \ge 1$.

//

There is a close connection to the theory of operators on the Hilbert space $L^2(E, \nu)$ which is convenient for the general theory of determinantal point processes. For our purposes and to keep technicalities to a minimum we shall assume $E = \mathbb{R}$ and ν =Lebesgue measure.

Let us denote $\langle f, g \rangle := \int_{\mathbb{R}} f(x)\overline{g(x)}\mathrm{d}x$ the standard scalar product on $L^2(\mathbb{R})$. The correlation kernel $K(x,y)$ corresponds to an operator $\mathcal{K}$ on $L^2(\mathbb{R})$ via

$$\mathcal{K} : f \mapsto \int_{\mathbb{R}} K(x,y)f(y)\mathrm{d}y, \qquad f \in L^2(\mathbb{R}). \tag{4.2.3}$$

The composition of two operators $\mathcal{K}_1, \mathcal{K}_2$ of this form is again of this form, with kernel

$$\int_{\mathbb{R}} K_1(x,t)K_2(t,y)\mathrm{d}t. \tag{4.2.4}$$

Let us recall the following notions which apply to an operator of the form (4.2.3).

- The condition (4.2.2) implies that the operator $\mathcal{K}$ is **self-adjoint** in the sense that

$$\langle f, \mathcal{K}g \rangle = \langle \mathcal{K}f, g \rangle, \qquad \forall f, g \in L^2(\mathbb{R}). \tag{4.2.5}$$

- For a self-adjoint operator $\mathcal{K}$ on $L^2(\mathbb{R})$ we write $\mathcal{K} \ge 0$ if and only if $\mathcal{K}$ is semipositive-definite, i.e. $\langle f, \mathcal{K}f \rangle \ge 0$ for all $f \in L^2(\mathbb{R})$. We write $\mathcal{K} \ge \lambda$ (for $\lambda \in \mathbb{R}$) if and only if $\mathcal{K} - \lambda I \ge 0$, and $\mathcal{K} \le \lambda$ if and only if $\lambda I - \mathcal{K} \ge 0$.
- An operator $\mathcal{K}$ in the form (4.2.3) is **Hilbert–Schmidt** if and only if

$$\int_{\mathbb{R}^2} |K(x,y)|^2 \mathrm{d}x\mathrm{d}y < +\infty. \tag{4.2.6}$$

- An operator on $L^2(\mathbb{R})$ is **trace-class** if and only if it is the composition of two Hilbert–Schmidt operators.
- An operator on $L^2(\mathbb{R})$ is **locally trace-class** if and only if its restriction $1_B \mathcal{K} 1_B$ to $L^2(B)$ is a trace-class operator for all bounded Borel sets $B \subset \mathbb{R}$.

Theorem 4.2.3 (Soshnikov)**.** *A kernel $K : \mathbb{R}^2 \to \mathbb{C}$ satisfying (4.2.2) defines uniquely a point process if and only if the associated self-adjoint operator $\mathcal{K}$ on $L^2(\mathbb{R})$ is locally trace-class and $0 \le \mathcal{K} \le 1$.*

[6] An Hermitian matrix $P \in \mathbf{H}_n$ is semipositive-definite if and only if $v^\dagger P v \ge 0$ for all $v \in \mathbb{C}^n$.

We omit the proof, which can be found, e.g., in the survey article by Soshnikov.

Remark 4.2.4. In particular, a locally trace-class orthogonal projector $\mathcal{K}$ (i.e., $\mathcal{K} = \mathcal{K}^\dagger = \mathcal{K}^2$, and $\mathcal{K}$ locally trace-class) always satisfies the conditions of Theorem 4.2.3. Indeed, any orthogonal projector $\mathcal{K} = \mathcal{K}^\dagger = \mathcal{K}^2$ is ≥ 0,

$$\langle f, \mathcal{K} f\rangle = \langle f, \mathcal{K}^2 f\rangle = \langle \mathcal{K}^\dagger f, \mathcal{K} f\rangle = \langle \mathcal{K} f, \mathcal{K} f\rangle \geq 0, \tag{4.2.7}$$

and $1 - \mathcal{K}$ is also an orthogonal projector, hence semi-positive definite as well.

Not all determinantal point processes come from orthogonal projectors (even though in these notes we will be mostly concerned with this case). //

Remark 4.2.5. The condition that $\mathcal{K}$ is locally trace-class ensures the necessary condition that $\#_B$ is almost surely finite when B is a bounded Borel set. This is indeed the case if and only if its expectation $\mathbb{E}[\#_B] = \int_B K(x,x)\mathrm{d}x$ is a finite number. Indeed, if $1_B \mathcal{K} 1_B = \mathcal{K}_1\mathcal{K}_2$ with Hilbert–Schmidt operators $\mathcal{K}_1, \mathcal{K}_2$ then

$$\begin{aligned}\int_B K(x,x)\mathrm{d}x &= \int_{\mathbb{R}} 1_B(x) K(x,x)\mathrm{d}x \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} K_1(x,y)K_2(y,x)\mathrm{d}y\mathrm{d}x \\ &\leq \int_{\mathbb{R}^2} |K_1(x,y)K_2(y,x)|\mathrm{d}x\mathrm{d}y \\ &\leq \sqrt{\int_{\mathbb{R}^2} |K_1(x,y)|^2\mathrm{d}x\mathrm{d}y}\sqrt{\int_{\mathbb{R}^2} |K_2(x,y)|^2\mathrm{d}x\mathrm{d}y} < +\infty, \end{aligned} \tag{4.2.8}$$

where in the last step we use Cauchy–Schwarz. //

Sine process.

We aim at showing, using Theorem 4.2.3, that the sine kernel

$$K^{\mathsf{sine}}(x,y) = \frac{\sin(\pi(x-y))}{\pi(x-y)} = \overline{K^{\mathsf{sine}}(y,x)} \tag{4.2.9}$$

(which we introduced above) defines uniquely a point process; namely, there exists a unique point process on $\mathbb{R}$ with correlation functions $\rho_k^{\mathsf{sine}}(x_1,\dots,x_k) = \det_{1\leq i,j\leq k} K^{\mathsf{sine}}(x_i,x_j)$.

The main idea is to express the sine kernel as

$$K^{\mathsf{sine}}(x,y) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(x-y)t}\frac{\mathrm{d}t}{2\pi}. \tag{4.2.10}$$

(Informally this representation could be regarded as the large n limit of (3.3.49).) We can rewrite this expression as

$$K^{\mathsf{sine}}(x,y) = \int_{-\pi}^{\pi} \frac{\mathrm{e}^{\mathrm{i}xt}}{\sqrt{2\pi}}\frac{\mathrm{e}^{-\mathrm{i}ty}}{\sqrt{2\pi}}\mathrm{d}t \tag{4.2.11}$$

i.e. the operator $\mathcal{K}^{\mathsf{sine}} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is the composition $\mathcal{K}^{\mathsf{sine}} = \mathcal{K}_1\mathcal{K}_2$ of the operators $\mathcal{K}_1, \mathcal{K}_2$ on $L^2(\mathbb{R})$ defined by

$$(\mathcal{K}_j f)(x) = \int_{\mathbb{R}} K_j(x,y)f(y)\mathrm{d}y \qquad (j=1,2) \tag{4.2.12}$$

with

$$K_1(x,t) := \frac{\mathrm{e}^{\mathrm{i}xt}}{\sqrt{2\pi}} 1_{(-\pi,\pi)}(t), \qquad K_2(t,y) := 1_{(-\pi,\pi)}(t)\frac{\mathrm{e}^{-\mathrm{i}ty}}{\sqrt{2\pi}}. \tag{4.2.13}$$

We are ready to check the conditions of Theorem 4.2.3.

1. $\boxed{\mathcal{K} \text{ is locally trace-class.}}$ Take any bounded Borel set $B \subseteq \mathbb{R}$, then $1_B\mathcal{K}^{\mathsf{sine}}1_B = 1_B\mathcal{K}_1\mathcal{K}_2 1_B$. The operator $1_B\mathcal{K}_1$ has kernel $1_B(x)K_1(x,t)$ and hence is Hilbert–Schmidt; the operator $\mathcal{K}_2 1_B$ has kernel $K_2(t,y)1_B(y)$ and hence is Hilbert–Schmidt. Therefore $1_B\mathcal{K}1_B$ is the composition of two Hilbert–Schmidt operators and so is trace-class.

2. $\boxed{0 \leq \mathcal{K} \leq 1.}$ By Remark 4.2.4, it is enough to show that $\mathcal{K}^{\mathsf{sine}}$ is an orthogonal projector. Indeed, recall the Fourier and inverse Fourier transforms;

$$(\mathcal{F}f)(x) = \int_{\mathbb{R}} \frac{\mathrm{e}^{-\mathrm{i}xy}}{\sqrt{2\pi}} f(y)\mathrm{d}y, \qquad (\mathcal{F}^{-1}f)(x) = \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}xy}}{\sqrt{2\pi}} f(y)\mathrm{d}y. \tag{4.2.14}$$

$\mathcal{F}$ and $\mathcal{F}^{-1} = \mathcal{F}^{\dagger}$ are unitary operators on $L^2(\mathbb{R})$ and by the above calculations we also have

$$\mathcal{K}^{\mathsf{sine}} = \mathcal{F}^{\dagger} 1_{(-\pi,\pi)} \mathcal{F}. \tag{4.2.15}$$

Hence $\mathcal{K}^{\mathsf{sine}}$ is the orthogonal projector onto the subspace in $L^2(\mathbb{R})$ of functions whose Fourier transform is supported in $[-\pi, \pi]$.

4.3 Convergence of point processes

Let $\mathbb{P}^{(n)}$ be a sequence of point processes on E. A natural notion of convergence for point processes is convergence in probability to a point process $\mathbb{P}$ on E: namely

$$\mathbb{P}^{(n)}(\mathcal{X}) \to \mathbb{P}(\mathcal{X}) \tag{4.3.1}$$

for all $\mathcal{X} \in \mathcal{B}(\mathrm{Conf}\,(E))$.

- $\mathbb{P}^{(n)}$ converges in probability to $\mathbb{P}$ if and only if, for any $B_1, \ldots, B_s \in \mathcal{B}(E)$, the vector-valued random variables $(\#_{B_1}^{(n)}, \ldots, \#_{B_s}^{(n)})$ converge in law to $(\#_{B_1}, \ldots, \#_{B_s})$, or
- $\mathbb{P}^{(n)}$ converges in probability to $\mathbb{P}$ provided that all processes have correlation functions of all orders and $\rho_k^{(n)}(x_1, \ldots, x_k)$ converge locally uniformly in the variables x_i to $\rho_k(x_1, \ldots, x_k)$. In the determinantal case, it is enough that $K^{(n)}(x, y)$ converges locally uniformly in the variables x, y to $K(x, y)$.

For example, we have seen that the re-scaled eigenvalue process of the CUE converges, when the size tends to infinity, to the sine process. Hence, local statistics of eigenvalues of the CUE can be computed in the large size limit using the correlation functions of the sine process.

The *universality* phenomenon in random matrix theory refers to the existence of few point processes representing limits of many diverse eigenvalue processes; the sine process is one of these universal processes.

4.4 Generating functions, gap probabilities

Let a point process $\mathbb{P}$ on E be given and let B be a bounded Borel set. The series of z given by

$$\mathbb{E}[z^{\#_B}] = \sum_{n \geq 0} z^n \, \mathbb{P}(\#_B = n) \tag{4.4.1}$$

is useful in computations and is termed **generating function**. For example,

$$\mathbb{E}[z^{\#_B}]\Big|_{z=0} = \mathbb{P}(\#_B = 0), \tag{4.4.2}$$

(a **gap probability**) or, more generally,

$$\frac{1}{n!}\frac{\mathrm{d}^n}{\mathrm{d}z^n}\mathbb{E}[z^{\#_B}]\Big|_{z=0} = \mathbb{P}(\#_B = n). \tag{4.4.3}$$

Theorem 4.4.1. *If $\mathbb{P}$ is determinantal with correlation kernel K, we have*

$$\mathbb{E}[z^{\#_B}] = 1 + \sum_{k \geq 1} \frac{(z-1)^k}{k!} \int_{B^k} \det_{1 \leq i,j \leq k} K(x_i, x_j)\mathrm{d}x_1 \ldots \mathrm{d}x_k \tag{4.4.4}$$

and the series is absolutely convergent for all $z \in \mathbb{C}$, so that $\mathbb{E}[z^{\#_B}]$ is an entire function of z.

Proof. First, we compute

$$\begin{aligned}\mathbb{E}[z^{\#_B}] &= \sum_{n\geq 0} z^n\, \mathbb{P}(\#_B = n) \\ &= \sum_{n\geq 0} (z-1+1)^n\, \mathbb{P}(\#_B = n) \\ &= \sum_{n\geq 0}\sum_{k=0}^{n} \binom{n}{k} (z-1)^k\, \mathbb{P}(\#_B = n) \\ &\overset{*}{=} \sum_{k\geq 0} (z-1)^k \sum_{n\geq k} \binom{n}{k}\, \mathbb{P}(\#_B = n) \\ &= \sum_{k\geq 0} (z-1)^k\, \mathbb{E}\left[\binom{\#_B}{k}\right] \\ &= 1 + \sum_{k\geq 1} \frac{(z-1)^k}{k!} \int_{B^k} \rho_k(x_1,\dots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \\ &= 1 + \sum_{k\geq 1} \frac{(z-1)^k}{k!} \int_{B^k} \det_{1\leq i,j\leq k} K(x_i,x_j)\mathrm{d}x_1\dots\mathrm{d}x_k \end{aligned} \tag{4.4.5}$$

In the step marked with $*$ we exchange order of summation; this can be done because the double series in n, k is absolutely convergent, as we now show. Indeed, for a semipositive-definite matrix the determinant is smaller than the product of the diagonal entries (Lemma 4.4.2 below), hence

$$\int_{B^k} \det_{1\leq i,j\leq k} K(x_i,x_j)\mathrm{d}x_1\dots\mathrm{d}x_k \leq \int_{B^k} \prod_{i=1}^{k} K(x_i,x_i)\mathrm{d}x_1\cdots\mathrm{d}x_k = \left(\int_B K(x,x)\mathrm{d}x\right)^k. \tag{4.4.6}$$

However, $C_B := \mathbb{E}[\#_B] = \int_B |K(x,x)|\mathrm{d}x < +\infty$ and so

$$1 + \sum_{k\geq 1} \left| \frac{(z-1)^k}{k!} \int_{B^k} \det_{1\leq i,j\leq k} K(x_i,x_j)\mathrm{d}x_1\dots\mathrm{d}x_k \right| \leq \sum_{k\geq 0} \frac{|z-1|^k C_B^k}{k!} = \mathrm{e}^{C_B|z-1|} \tag{4.4.7}$$

and the series is absolutely convergent. □

Lemma 4.4.2. *Let $P \in \mathbf{H}_n$ be a semi-positive definite Hermitian matrix of size n, i.e. $v^\dagger P v \geq 0$ for all $v \in \mathbb{C}^n$. Then*

$$\det P \leq \prod_{i=1}^{n} P_{ii}. \tag{4.4.8}$$

Proof. First, if $\det P = 0$ the inequality is ensured by the fact that $P_{ii} = e_i^\dagger P e_i \geq 0$; therefore, let us assume that P is non-singular, $\det P > 0$. The case $n = 1$ is obvious, and so we reason by induction. For $n \geq 2$ factorize P as follows:

$$P = \left(\begin{array}{c|c} \widetilde{P} & v \\ \hline v^\dagger & P_{nn} \end{array}\right) = \left(\begin{array}{c|c} \mathbf{1}_{n-1} & 0 \\ \hline b^\dagger & 1 \end{array}\right) \left(\begin{array}{c|c} \widetilde{P} & v \\ \hline 0 & c \end{array}\right) \tag{4.4.9}$$

where $\widetilde{P}$ is the $(n-1)\times(n-1)$ submatrix of P obtained removing the last row and column, and $b \in \mathbb{C}^{n-1}$, $c \in \mathbb{C}$ are determined by

$$b^\dagger \widetilde{P} = v^\dagger \Rightarrow b^\dagger = v^\dagger \widetilde{P}^{-1}, \qquad b^\dagger v + c = P_{nn} \Rightarrow c = P_{nn} - v^\dagger \widetilde{P}^{-1} v \leq P_{nn}, \tag{4.4.10}$$

because $\widetilde{P}^{-1}$ is positive-definite provided $\widetilde{P}$ is, and $\widetilde{P}$ is positive-definite because P is (by Sylvester's criterion). Finally:

$$\det P = c \det \widetilde{P} \leq P_{nn} \det \widetilde{P} \underset{\text{induction hypothesis}}{\leq} P_{nn} \prod_{i=1}^{n-1} \widetilde{P}_{ii} = \prod_{i=1}^{n} P_{ii}, \tag{4.4.11}$$

and the proof is complete. □

Exercise 4.4.3. Prove that for any matrix $M = (c_1|\cdots|c_n)$, where $c_i \in \mathbb{C}^n$ are its columns, we have $|\det M| \le \prod_{i=1}^n \|c_i\|$, which is known as "Hadamard inequality". (Hint: apply Lemma 4.4.2 to $P := M^\dagger M$.) //

Exercise 4.4.4. Deduce from Lemma 4.4.2 that $\rho_k(x_1,\dots,x_k) \le \rho_1(x_1)\cdots\rho_1(x_k)$, and so that

$$\mathbb{E}[\#_{B_1}\cdots\#_{B_k}] \le \mathbb{E}[\#_{B_1}]\cdots\mathbb{E}[\#_{B_k}] \quad \text{for disjoint Borel sets } B_1,\dots,B_k. \tag{4.4.12}$$

(This is another manifestation of repulsion in determinantal point processes). //

Exercise 4.4.5. Let $B_1,\dots,B_s$ be a family of bounded disjoint Borel sets in E. Under the same assumptions of the previous theorem, prove that

$$\mathbb{E}[z_1^{\#_{B_1}}\cdots z_s^{\#_{B_s}}] = \sum_{k_1,\cdots,k_s\ge 0} \frac{(z_1-1)^{k_1}\cdots(z_s-1)^{k_s}}{k_1!\cdots k_s!}\int_{B_1^{k_1}\times\cdots\times B_s^{k_s}} \det_{1\le i,j\le k} K(x_i,x_j)\mathrm{d}x_1\cdots\mathrm{d}x_k \tag{4.4.13}$$

where in the right-hand side we denote $k := k_1+\cdots+k_s$. //

Fredholm determinants

It is appropriate here to make a little detour in the theory of infinite-dimensional determinants.

Let us first recall that an operator $\mathcal{K}$ on $L^2(\mathbb{R})$ (or, more generally, on a Hilbert space) is compact if and only if there exists an orthonormal system $(e_i)_{i\ge1}$ in $L^2(\mathbb{R})$ and complex numbers $(\lambda_i)_{i\ge1}$ such that $\lambda_i \to 0$ as $i\to+\infty$, such that

$$\mathcal{K}f = \sum_{i=1}^{+\infty}\lambda_i e_i(e_i,f), \qquad f\in L^2(\mathbb{R}). \tag{4.4.14}$$

Theorem 4.4.6. *Hilbert–Schmidt and trace-class operators are compact.*

Theorem 4.4.7 (Lidskii)**.** *A compact operator $\mathcal{K}$ is trace-class if and only if the series $\sum_{i\ge1}|\lambda_i|$, where λ_i are the eigenvalues of $\mathcal{K}$, converges.*

A motivation for why trace-class operators are important to define infinite-dimensional determinants is provided by the following classical criterion for the convergence of an infinite product.

Exercise 4.4.8. Let $(a_i)_{i\ge1}$ be a sequence of nonnegative numbers, $a_i \ge 0$. Then $\prod_{i\ge1}(1+a_i)$ converges if and only if the series $\sum_{i\ge1} a_i$ converges. //

Definition 4.4.9. Given a trace-class operator $\mathcal{L}$ on $L^2(\mathbb{R})$, with kernel L, the **Fredholm determinant** of $1+\mathcal{L}$ is defined to be

$$\det(1+\mathcal{L}) := 1 + \sum_{k\ge1}\frac{1}{k!}\int_{\mathbb{R}^k}\det_{1\le i,j\le k} L(x_i,x_j)\mathrm{d}x_1\cdots\mathrm{d}x_k \tag{4.4.15}$$

//

Theorem 4.4.10. *Given a trace-class operator $\mathcal{L}$ on $L^2(\mathbb{R})$ with eigenvalues λ_i, as in (4.4.14), we have*

$$\det(1+\mathcal{L}) = \prod_{i\ge1}(1+\lambda_i). \tag{4.4.16}$$

Applying this theory, from (4.4.4), we have

$$\mathbb{E}[z^{\#_B}] = \det(1+(z-1)1_B\mathcal{K}1_B) = \prod_{i\ge1}(1+(z-1)\lambda_i(B)) \tag{4.4.17}$$

where we denote $\lambda_i(B)$ the eigenvalues of the trace-class, hence compact, operator $1_B\mathcal{K}1_B$.

We can at this point explain the necessity of the condition $\mathcal{K} \leq 1$ in Theorem 4.2.3. Indeed, if $\mathcal{K} \nleq 1$ it means that there exists a sufficiently large bounded interval $B \subseteq \mathbb{R}$ and an eigenvalue $\lambda > 1$ of $1_B \mathcal{K} 1_B$. Solving the equation $1 + (z-1)\lambda = 0$ gives $z_0 = \frac{\lambda - 1}{\lambda} \in (0,1)$; hence

$$0 = \prod_{i \geq 1} (1 + (z_0 - 1)\lambda_i(B)) = \mathbb{E}[z_0^{\#_B}] = \sum_{n \geq 1} z_0^n \, \mathbb{P}(\#_B = n) \tag{4.4.18}$$

where the first equality holds because one of the factors of the infinite product is zero. Since in the right-hand side we have a sum of non-negative terms, we must have $\mathbb{P}(\#_B = n) = 0$ for all $n \geq 0$, a contradiction (because B is bounded hence $\#_B$ is almost surely finite and so $1 = \sum_{n \geq 0} \mathbb{P}(\#_B = n)$).

4.5 Total number of particles in a determinantal point process

Proposition 4.5.1. *Let a point process on $\mathbb{R}$ be given and let, as usual, $\#_{\mathbb{R}}$ be the total number of particles and $\mathcal{K}$ the associated operator on $L^2(\mathbb{R})$.*

1. *$\#_{\mathbb{R}}$ is either almost surely[7] finite or almost surely infinite; the first situation happens if and only if $\int_{\mathbb{R}} K(x,x)\mathrm{d}x < +\infty$, the second if and only if $\int_{\mathbb{R}} K(x,x)\mathrm{d}x = +\infty$.*
2. *$\#_{\mathbb{R}}$ is almost surely $\leq n$ if and only if $\mathcal{K}$ is a finite-rank operator of rank $\leq n$.*
3. *$\#_{\mathbb{R}}$ is almost surely n if and only if $\mathcal{K}$ is an orthogonal projector onto a space of dimension n.*

The proof is omitted and can be found in Soshnikov's article.

[7] An even $\mathcal{E}$ occurs almost surely if and only if $\mathbb{P}(\mathcal{E}) = 1$.

Chapter 5

Steepest descent method, Airy functions

5.1 Bachmann–Landau notations

Definition 5.1.1. Let X be a topological space, $A \subseteq X$ a subspace, and $x_0 \in \overline{A}$ (the closure of A in X). Consider two functions $f, g : A \to \mathbb{C}$ and assume that there is a neighborhood $\mathcal{U}$ of x_0 in X such that

$$g(x) \neq 0 \text{ for all } x \in \mathcal{U} \cap A. \tag{5.1.1}$$

Then we write:

$$f = O(g) \iff |f/g| \text{ is bounded in a neighborhood of } x_0 \text{ in } A; \tag{5.1.2}$$

$$f = o(g) \iff f(x)/g(x) \to 0 \text{ as } x \to x_0 \text{ in } A; \tag{5.1.3}$$

$$f \sim g \iff f(x)/g(x) \to 1 \text{ as } x \to x_0 \text{ in } A. \tag{5.1.4}$$

//

The relation $f = O(g)$ is read "f is big-O of g"; the relation $f = o(g)$ is read "f is little-o of g"; The relation $f \sim g$ is read "f is asymptotically equivalent to g", or "f and g are asymptotically equivalent", or just "f is asymptotic to g".

Remark 5.1.2. This notion is very convenient, although the notation is potentially misleading; the equality sign is by no means a standard equality of functions, but rather these notations should be regarded as (set-theoretic) relations between functions. In this respect it is also worth-noting that that the $O, o, \sim$-relations are transitive, the $O, \sim$-relations are reflexive, and the $\sim$-relation is also symmetric (so that the latter is an equivalence relation). Sometimes we shall write

$$f = g(1 + o(1)) \tag{5.1.5}$$

to mean $f/g \to 1$. //

We shall use these notations essentially in two cases.

1. $X = \mathbb{C} \cup \{\infty\}$, the one-point compactification of the complex plane; a basis of open neighborhoods of ∞ is given by $\{z : |z| > R\} \cup \{\infty\}$, for $R > 0$, so that limits at ∞ coincide with the usual notion.

2. $X = \mathbb{N} \cup \{\infty\}$, the one-point compactification of $\mathbb{N}$; in this case the definition is non-trivial only for $x_0 = \infty$.

It is important to always stress the data of A and x_0 in writing such relations; e.g.

- $f(z) = 1/z$ is $O(1)$ in $A = \mathbb{C} \setminus \{0\}$ as $z \to \infty$, but not as $z \to 0$;
- $f(z) = \mathrm{e}^{-z}$ is $o(1)$ as $z \to \infty$ in $A = \{z \in \mathbb{C} : \operatorname{Re} z > 0\}$, but not in $A = \mathbb{C}$.

5.2 Laplace's method

Consider the integral

$$I(M) = \int_a^b f(x)\mathrm{e}^{M\varphi(x)}\mathrm{d}x, \tag{5.2.1}$$

depending parametrically on M. We want to study the asymptotics of $I(M)$ for $M \to +\infty$; one of the main tools for this task is *Laplace's method*, which relies on the observation that the leading contribution to the integral $I(M)$ comes from the maxima of the function φ only. A formal statement can be given as follows.

Theorem 5.2.1 ("Laplace's method"). *Suppose $-\infty \le a < b \le +\infty$, $f \in \mathcal{C}^0(a,b)$ and $\varphi \in \mathcal{C}^2(a,b)$ such that φ has a unique global maximum at $x_0 \in (a,b)$ satisfying $\varphi''(x_0) \lneq 0$. Moreover*[1] *we assume that there exist $\delta, \eta > 0$ such that*

$$\varphi(x) - \varphi(x_0) \le -\delta \text{ for all } x \in (a,b) \setminus (x_0-\eta, x_0+\eta). \tag{5.2.2}$$

Then

$$I(M) = \int_a^b f(x)\mathrm{e}^{M\varphi(x)}\mathrm{d}x = f(x_0)\mathrm{e}^{M\varphi(x_0)}\sqrt{-\frac{2\pi}{M\varphi''(x_0)}}(1+o(1)), \qquad M \to +\infty, \tag{5.2.3}$$

provided $I(M)$ is convergent for $M > M_0$ for some $-\infty \le M_0 < \infty$.

Proof. First, for later convenience, let us assume without loss of generality that

$$\varphi'(x) \neq 0 \text{ for all } x \in (x_0-\eta, x_0+\eta) \setminus \{x_0\}. \tag{5.2.4}$$

Indeed, this can always be achieved, possibly at cost of choosing smaller $\eta, \delta > 0$ in (5.2.2). Then, let us write $I(M) = \mathrm{e}^{M\varphi(x_0)}(J_1(M) + J_2(M))$, where

$$J_1(M) := \int_{(a,b)\setminus(x_0-\eta,x_0+\eta)} f(x)\mathrm{e}^{M(\varphi(x)-\varphi(x_0))}\mathrm{d}x, \qquad J_2(M) := \int_{x_0-\eta}^{x_0+\eta} f(x)\mathrm{e}^{M(\varphi(x)-\varphi(x_0))}\mathrm{d}x. \tag{5.2.5}$$

The contribution from $J_1(M)$ is exponentially small as $M \to +\infty$. To see this, write

$$|J_1(M)| \le \mathrm{e}^{-M\delta}\int_{(a,b)\setminus(x_0-\eta,x_0+\eta)} |f(x)|\mathrm{e}^{M(\varphi(x)-\varphi(x_0)+\delta)}\mathrm{d}x. \tag{5.2.6}$$

Then observe that $\int_{(a,b)\setminus(x_0-\eta,x_0+\eta)} |f(x)|\mathrm{e}^{M(\varphi(x)-\varphi(x_0)+\delta)}\mathrm{d}x$ is a decreasing function of M (because $\varphi(x) - \varphi(x_0) + \delta \le 0$ in the integration domain) and thus we obtain the claim $J_1(M) = O(\mathrm{e}^{-M\delta})$ as $M \to +\infty$, because

$$|J_1(M)| \le C\mathrm{e}^{-M\delta}, \qquad C := \int_a^b |f(x)|\mathrm{e}^{-M_1(\varphi(x)-\varphi(x_0)+\delta)}\mathrm{d}x \tag{5.2.7}$$

for all $M \ge M_1$ where $M_1 > M_0$ can be arbitrarily fixed. To compute the contribution from $J_1(M)$ let us introduce the function $y : (x_0-\eta, x_0+\eta) \to \mathbb{R}$ by

$$y(x) := \begin{cases} -\sqrt{\varphi(x_0)-\varphi(x)} & x_0-\eta < x < x_0, \\ 0 & x = x_0, \\ \sqrt{\varphi(x_0)-\varphi(x)} & x_0 < x < x_0+\eta. \end{cases} \tag{5.2.8}$$

We compute the derivative y' as follows;

$$y'(x) = \begin{cases} \frac{\varphi'(x)}{2\sqrt{\varphi(x_0)-\varphi(x)}} & x_0 - \eta < x < x_0, \\ \sqrt{-\frac{\varphi''(x_0)}{2}} & x = x_0, \\ -\frac{\varphi'(x)}{2\sqrt{\varphi(x)-\varphi(x_0)}} & x_0 < x < x_0 + \eta. \end{cases} \tag{5.2.9}$$

We conclude that $y \in \mathcal{C}^1(x_0 - \eta, x_0 + \eta)$ and $y'(x) > 0$ for all $x \in (x_0 - \eta, x_0 + \eta)$; hence we can change variable in the integral

$$J_1(M) = \int_{y_-}^{y_+} g(y) \mathrm{e}^{-My^2} \mathrm{d}y = \frac{1}{\sqrt{M}} \int_{y_-\sqrt{M}}^{y_+\sqrt{M}} g\left(\frac{w}{\sqrt{M}}\right) \mathrm{e}^{-w^2} \mathrm{d}w, \tag{5.2.10}$$

where we set

$$g(y) := \frac{f(x(y))}{y'(x(y))}, \qquad y_\pm = \pm\sqrt{\varphi(x_0) - \varphi(x_0 \pm \eta)}, \tag{5.2.11}$$

denoting $x(y)$ the inverse of the function y defined in (5.2.8). Since for all $w \in \mathbb{R}$ we have

$$\left| g\left(\frac{w}{\sqrt{M}}\right) \mathrm{e}^{-w^2} 1_{(y_-\sqrt{M}, y_+\sqrt{M})}(w) \right| \le \left(\sup_{y \in [y_-, y_+]} |g(y)| \right) \mathrm{e}^{-w^2} \in L^1(\mathbb{R}, \mathrm{d}w), \tag{5.2.12}$$

we can apply Lebesgue's Dominated Convergence Theorem to get

$$\int_{y_-\sqrt{M}}^{y_+\sqrt{M}} g\left(\frac{w}{\sqrt{M}}\right) \mathrm{e}^{-w^2} \mathrm{d}w = \int_{\mathbb{R}} g\left(\frac{w}{\sqrt{M}}\right) \mathrm{e}^{-w^2} 1_{(y_-\sqrt{M}, y_+\sqrt{M})}(w) \mathrm{d}w \to g(0) \int_{\mathbb{R}} \mathrm{e}^{-w^2} \mathrm{d}w = g(0)\sqrt{\pi}, \tag{5.2.13}$$

as $M \to +\infty$. Since $y(x_0) = 0$ we obtain

$$g(0) = f(x_0)\sqrt{-\frac{2}{\varphi''(x_0)}}. \tag{5.2.14}$$

Putting everything together, we have shown that

$$I(M)\sqrt{M}\mathrm{e}^{-M\varphi(x_0)} = O(\sqrt{M}\mathrm{e}^{-M\delta}) + \sqrt{-\frac{2\pi}{\varphi''(x_0)}} f(x_0) + o(1) = \sqrt{-\frac{2\pi}{\varphi''(x_0)}} f(x_0) + o(1), \tag{5.2.15}$$

which completes the proof. □

Remark 5.2.2. 1. The condition that there is exactly one maximum can be lifted; if there are several points in (a, b) where φ which are non-degenerate maxima (in particular, φ attains the *same* maximum value at these points), it is enough to split the integral over intervals with disjoint interiors each of them containing one of these global maxima and apply the result of the theorem above to each interval. The final asymptotics is a linear combination of pieces of the form (5.2.3).

2. The condition $\varphi''(x_0) \neq 0$ (the maximum is "non-degenerate") is clearly necessary, as the resulting estimate for $I(M)$ contains $\varphi''(x_0)$ in the denominator. However, one can mimic the proof above to show that under similar assumptions, the only difference being that for some even integer $p \ge 2$ we have $\varphi \in \mathcal{C}^p(a, b)$, $\varphi^{(i)}(x_0) = 0$ for $i = 1, \ldots, p-1$, and $\varphi^{(p)} \lneq 0$, then

$$I(M) = c_p f(x_0) \left(-\frac{p!}{M\varphi^{(p)}(x_0)}\right)^{1/p} \mathrm{e}^{M\varphi(x_0)}(1 + o(1)). \tag{5.2.16}$$

where $c_p := \int_{\mathbb{R}} \exp(-w^p) \mathrm{d}w$, which can also be expressed in terms of the Euler Gamma function as $c_p = 2p^{-1}\Gamma(p^{-1})$. Indeed, the estimate for of $J_2(M)$ can be extended directly, and for $J_1(M)$ one can proceed as in the case $p = 2$ with the function $y(x) := \operatorname{sign}(x - x_0)(\varphi(x_0) - \varphi(x))^{1/p}$.

3. If f, φ have more regularity we can provide more terms in the asymptotics for $I(M)$. Indeed, assume in addition to the hypothesis of Theorem 5.2.1 that for some integer $k \geq 0$ we have $\varphi \in \mathcal{C}^{2k+2}(a,b)$ and $f \in \mathcal{C}^{2k}(a,b)$. Then (5.2.3) is improved to

$$I(M) = \frac{\mathrm{e}^{M\varphi(x_0)}}{\sqrt{M}} \left(\sum_{j=0}^{k} a_j M^{-j} + o(M^{-k}) \right), \qquad M \to +\infty, \tag{5.2.17}$$

where

$$a_j = \frac{g^{(2j)}(0)}{(2j)!} \int_{\mathbb{R}} w^{2j} \mathrm{e}^{-w^2} \mathrm{d}w = \sqrt{\pi} \frac{(2j-1)!!}{2^j (2j)!} g^{(2j)}(0) = \frac{\sqrt{\pi}}{j! 4^j} g^{(2j)}(0), \tag{5.2.18}$$

the function g being defined as in the proof, cf. (5.2.11). In particular, if $f, \varphi \in \mathcal{C}^{\infty}(a,b)$ then we have an infinite asymptotic series

$$I(M) \sim \frac{\mathrm{e}^{M\varphi(x_0)}}{\sqrt{M}} \sum_{j=0}^{+\infty} a_j M^{-j}, \qquad M \to +\infty. \tag{5.2.19}$$

In practice, there is a simpler way to compute the coefficients a_j in these expansions, rather than using formula (5.2.18). The mnemonic rule is to remove from φ the terms up to the quadratic one in the Taylor series at x_0 by defining

$$\widetilde{\varphi}(x) := \varphi(x) - \varphi(x_0) - \frac{\varphi''(x_0)}{2}(x - x_0)^2, \tag{5.2.20}$$

and then to change variable $x \to \xi$ by rescaling in the vicinity of the maximum x_0 by $x = x_0 + \frac{\xi}{\sqrt{M}}$ and then expand

$$f\left(x_0 + \frac{\xi}{\sqrt{M}}\right) \mathrm{e}^{M\widetilde{\varphi}(x_0 + \frac{\xi}{\sqrt{M}})} = \sum_{j \geq 0} \frac{b_j(\xi)}{M^{j/2}}, \tag{5.2.21}$$

so that the b_j's are polynomial in ξ of degree $3j$ satisfying $b_j(-\xi) = (-1)^j b_j(\xi)$; finally, the coefficients a_j are given by the Gaussian integrals (recall $\varphi''(x_0) < 0$)

$$a_j = \int_{\mathbb{R}} b_{2j}(\xi) \mathrm{e}^{\frac{\varphi''(x_0)}{2}\xi^2} \mathrm{d}\xi. \tag{5.2.22}$$

4. Laplace's method only applies in case the maxima of φ are *inside* the integration domain; in case the maxima are at the boundary of the integration domain one should resort to other methods (e.g. integration by parts, Watson's lemma, which will not be discussed here).

5. Laplace's method shares some similarities with the "stationary phase method", but there are important differences between the two. The latter applies to oscillating integrals $\int_a^b f(x)\mathrm{e}^{\mathrm{i}M\varphi(x)}\mathrm{d}x$ (the difference is that the exponential part is now purely imaginary instead of purely real); again, the leading contributions are localized, however this time they come from *all stationary points* of φ (i.e. all points at which $\varphi' = 0$). Moreover, the localization mechanism is completely different in the two cases: for the Laplace method the integrand $f(x)\mathrm{e}^{M\varphi(x)}$ is simply exponentially larger at x_0, for the stationary phase method the rapid oscillations of $\mathrm{e}^{\mathrm{i}M\varphi(x)}$ tend to cancel out in the integral and the leading contribution occur when φ varies slower, that is at stationary points, where $\varphi' = 0$.

//

Exercise 5.2.3 (Stirling approximation). 1. Prove the identity $n! = \int_0^{+\infty} x^n \mathrm{e}^{-x} \mathrm{d}x$ for all nonnegative integers n. (Hint: integrate by parts and use induction on n.)

2. Prove Stirling approximation

$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}(1 + O(n^{-1})), \qquad n \to +\infty. \tag{5.2.23}$$

(Hint: by an appropriate change of variable, rewrite the result in the previous step as $n! = n^{n+1} \int_0^{+\infty} \mathrm{e}^{n(\log x - x)} \mathrm{d}x$ and apply Laplace's method.)

3. Prove Stirling approximation to the first sub-leading order
$$n! = \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}\left(1 + \frac{1}{12n} + O(n^{-2})\right), \qquad n \to +\infty. \tag{5.2.24}$$
(Hint: expand the function $\exp\left[n\left(\log\left(\frac{y}{\sqrt{n}}+1\right) - \left(\frac{y}{\sqrt{n}}+1\right)\right) + n + \frac{y^2}{2}\right] = \sum_{j\geq 0} b_j(y) n^{-j/2}$ as $n \to +\infty$ to compute the Gaussian integral $a_1 = \int_{\mathbb{R}} b_2(y)\mathrm{e}^{-y^2/2}\mathrm{d}y$, cf. (5.2.22).)

//

5.3 The steepest descent method

Let us consider now a more general situation. We want to study large-M asymptotics for a contour integral of the form
$$I(M) = \int_\gamma f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z, \tag{5.3.1}$$
where γ is a piece-wise $\mathcal{C}^1$ contour in $\mathbb{C}$ and f, φ are analytic in an open set $\mathcal{U} \subseteq \mathbb{C}$, with $\gamma \subset \mathcal{U}$.

The idea to tackle this problem is to exploit the freedom allowed by the Cauchy Theorem in order to deform the contour γ within the domain of analyticity $\mathcal{U}$ of f, φ in such a way that the integral can be analyzed by the Laplace's method studied in the previous section.

Let us first study the structure of critical points z_0 ($\varphi'(z_0) = 0$) of analytic functions φ. Given the important role played by these points, this asymptotic method is also known as *saddle point method*.

Lemma 5.3.1. *Let φ be an analytic function on the open set $\mathcal{U}$. Any point $z_0 \in \mathcal{U}$ such that $\varphi'(z_0) = 0$ is a saddle point. If $\varphi'(z_0) = 0$ and $\varphi''(z_0) \neq 0$, the locus $\operatorname{Im}\varphi(z) = \operatorname{Im}\varphi(z_0)$ restricted to a sufficiently small neighborhood of z_0 is the union of two smooth curves Γ_+ and Γ_-, which meet at z_0 only, where they meet perpendicularly. Moreover, the function $\operatorname{Re}\varphi(z)$ restricted to Γ_+ (resp. Γ_-) has a unique global maximum (resp. minimum) at z_0.*

Definition 5.3.2. The contour Γ_+ (resp. Γ_-) is called the **steepest descent contour** (resp. **steepest ascent contour**) at z_0. //

Proof. By the maximum modulus principle, no point of $\mathcal{U}$ can be a maximum or minimum of φ, hence points where $\varphi'(z_0)$ must be saddle points.

Next, under the hypothesis $\varphi''(z_0) \neq 0$, we claim that there exists a sufficiently small neighborhood $\mathcal{N} \subseteq \mathcal{U}$ of z_0 and an analytic function ψ on $\mathcal{N}$ which is nowhere vanishing and such that
$$\varphi(z) - \varphi(z_0) = (\psi(z)(z - z_0))^2. \tag{5.3.2}$$
To prove it, note that the Taylor expansion of φ at $z \to z_0$
$$\varphi(z) - \varphi(z_0) = (z - z_0)^2 \sum_{j\geq 0} \frac{\varphi^{(j+2)}(z_0)}{(j+2)!}(z - z_0)^j \tag{5.3.3}$$
converges in a disk around z_0; then set $\widetilde{\psi}(z) := \sum_{j\geq 0} \frac{\varphi^{(j+2)}(z_0)}{(j+2)!}(z - z_0)^j$, which is analytic in the same disk and (possibly up to shrinking the disk) never vanishing (using the hypothesis $\varphi''(z_0) \neq 0$). Thus we can introduce a square root
$$\psi(z) := \sqrt{\widetilde{\psi}(z)}, \tag{5.3.4}$$
possibly defined in a smaller disk, and analytic there. This proves the claim.

Then $\{z \in \mathcal{N} : \operatorname{Im}\varphi(z) = \operatorname{Im}\varphi(z_0)\} = \Gamma_+ \cup \Gamma_-$ where
$$\Gamma_+ := \{z \in \mathcal{N} : \operatorname{Re}(\psi(z)(z - z_0)) = 0\}, \qquad \Gamma_- := \{z \in \mathcal{N} : \operatorname{Im}(\psi(z)(z - z_0)) = 0\}. \tag{5.3.5}$$
Since $\psi(z) \neq 0$ in $\mathcal{N}$, these curves only meet at $z = z_0$. The perpendicularity at z_0 follows from the Cauchy–Riemann equations.

Finally, by (5.3.2) we see that $\operatorname{Re}\varphi(z)$ along Γ_+ (resp. Γ_-) has a unique maximum (resp. minimum) at z_0. □

For a rather general setting (enough for our purposes) the strategy of the steepest descent method is described as follows.

- First, we need to find a piece-wise $\mathcal{C}^1$ contour γ_1 in $\mathcal{U}$ which is homotopic to γ in $\mathcal{U}$ relative to its endpoints, and satisfies some further conditions detailed below. The homotopy relation means the following; assume $\gamma : [a,b] \to \mathcal{U}$ is a contour, then another contour $\gamma_1 : [a,b] \to \mathcal{U}$ is homotopic to γ in $\mathcal{U}$ relative to its endpoints if and only if there exists $F : [a,b] \times [0,1] \to \mathcal{U}$ (jointly) continuous such that

$$F(t,0) = \gamma(t),\ F(t,1) = \gamma_1(t), \qquad \text{for all } t \in [a,b], \tag{5.3.6}$$
$$F(a,s) = \gamma(a),\ F(b,s) = \gamma(b), \qquad \text{for all } s \in [0,1]. \tag{5.3.7}$$

  (In particular $\gamma(a) = \gamma_1(a)$, $\gamma(b) = \gamma_1(b)$.) Let us denote $\gamma_s := F(\cdot, s)$; we assume also that γ_s is piecewise $\mathcal{C}^1$ for all $s \in [0,1]$, and that there exists $-\infty \le M_0 < +\infty$ such that $\int_{\gamma_s} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z$ converges for all $M > M_0$ and for all $s \in [0,1]$.

- γ_1 passes through $k \ge 1$ saddle points $z_1, \ldots, z_k \in \mathcal{U}$ of φ (which are not necessarily *all* the saddle points of φ in $\mathcal{U}$); moreover, $\varphi''(z_i) \ne 0$ for all $1 \le i \le k$, and there exist neighborhoods $\mathcal{N}_i$ of z_i, for all $1 \le i \le k$, such that
  - $\gamma_1 \cap \mathcal{N}_i$ coincides with the steepest descent contour Γ_+ at z_i, and
  - there exists $\delta > 0$ such that $\operatorname{Re}(\varphi(z_0) - \varphi(z)) \ge \delta$ for all $z \in \gamma_1 \setminus (\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_k)$. (This last condition should be compared with (5.2.2).)

Assuming the first condition is met, Cauchy's Theorem[2] implies that the value of the integral $\int_{\gamma_s} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z$ is independent of s, hence

$$I(M) = \int_{\gamma_1} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z. \tag{5.3.8}$$

Using now the second condition we can write

$$I(M) = \sum_{i=1}^{k} \int_{\gamma_1 \cap \mathcal{N}_i} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z + \int_{\gamma \setminus (\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_k)} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z, \tag{5.3.9}$$

and the first k integrals can be estimated when M is large by Laplace's method (Theorem 5.2.1) while the last integral is $O(\mathrm{e}^{-M\delta})$ as $M \to +\infty$. More precisely, the first k integrals are

$$\int_{\gamma_1 \cap \mathcal{N}_i} f(z)\mathrm{e}^{M\varphi(z)}\mathrm{d}z = \mathrm{e}^{M\operatorname{Im}\varphi(z_i)} \int_{\gamma_1 \cap \mathcal{N}_i} f(z)\mathrm{e}^{M\operatorname{Re}\varphi(z)}\mathrm{d}z = \mathrm{e}^{M\operatorname{Im}\varphi(z_i)} \int_{a_i}^{b_i} f(\gamma_1(t))\gamma_1'(t)\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t))}\mathrm{d}t \tag{5.3.10}$$

where $\gamma_1 \cap \mathcal{N}_i : (a_i, b_i) \to \mathcal{U} : t \mapsto \gamma_1(t)$ $(a < a_i < b_i < b)$, and real and imaginary parts of the integral

$$\int_{a_i}^{b_i} f(\gamma_1(t))\gamma_1'(t)\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t))}\mathrm{d}t \tag{5.3.11}$$

satisfy the conditions of Theorem 5.2.1 so that

$$\boxed{I(M) = \sum_{i=1}^{k} \mathrm{e}^{M\varphi(z_i)} f(z_i)\mathrm{e}^{\mathrm{i}\theta_i}\sqrt{\frac{2\pi}{M|\varphi''(z_i)|}}(1 + O(M^{-1})), \qquad M \to +\infty,} \tag{5.3.12}$$

[2]Or, equivalently, Stokes' theorem, because $F(z)$ is *holomorphic* if and only if $F(z)\mathrm{d}z = F(x+\mathrm{i}y)(\mathrm{d}x + \mathrm{i}\mathrm{d}y)$ is a ($\mathbb{C}$-valued) *closed* one-form on $\mathbb{R}^2 \ni (x,y)$.

where θ_i is the angle formed by the steepest descent curve Γ_+ at z_i with a line passing through z_i and parallel to the real axis. The angle can be computed simply by solving $\varphi''(z_i)\mathrm{e}^{2\mathrm{i}\theta_i} < 0$: hence

$$\theta_i = \frac{\pm\pi - \arg\varphi''(z_i)}{2} \tag{5.3.13}$$

and more precisely the sign $\pm$ is obtained by looking in which orientation the steepest descent path is traversed (this is part of the data γ, which is an oriented contour).

To prove (5.3.12), it suffices to apply the general formula (5.2.3) to (5.3.11); denoting $t_i \in (a_i, b_i)$ the point for which $\gamma(t_i) = z_i$, we have

$$\int_{a_i}^{b_i} f(\gamma_1(t))\gamma_1'(t)\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t))}\mathrm{d}t = f(\gamma(t_i))\gamma'(t_i)\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t_i))}\sqrt{-\frac{2\pi}{\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}\operatorname{Re}\varphi(\gamma_1(t))\right)\Big|_{t=t_i}}} \tag{5.3.14}$$

It remains to compute $\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}\operatorname{Re}\varphi(\gamma_1(t))\right)\Big|_{t=t_i}$; we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\operatorname{Re}\varphi(\gamma_1(t)) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi(\gamma_1(t)) = \varphi''(\gamma_1(t))(\gamma_1'(t))^2 + \varphi'(\gamma_1(t))\gamma_1''(t) \tag{5.3.15}$$

where the first equality is because $\gamma_1(t)$ coincides with the steepest descent contour Γ_+ where $\operatorname{Im}\varphi$ is constant. Evaluating at $t = t_1$ we obtain

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2}\operatorname{Re}\varphi(\gamma_1(t))\right)\Big|_{t=t_i} = \varphi''(t_1)(\gamma_1'(t_1))^2. \tag{5.3.16}$$

Next, this quantity is < 0, again because $\gamma_1(t)$ coincides with the steepest descent contour; hence

$$-\varphi''(t_1)(\gamma_1'(t_1))^2 > 0 \tag{5.3.17}$$

and so

$$\int_{a_i}^{b_i} f(\gamma_1(t))\gamma_1'(t)\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t))}\mathrm{d}t = f(\gamma(t_i))\frac{\gamma'(t_i)}{|\gamma'(t_i)|}\mathrm{e}^{M\operatorname{Re}\varphi(\gamma_1(t_i))}\sqrt{\frac{2\pi}{|\varphi''(z_i)|}} \tag{5.3.18}$$

and (5.3.12) follows because $\gamma'(t_i)/|\gamma'(t_i)| = \mathrm{e}^{\mathrm{i}\theta_i}$.

Finally, note that the sum in (5.3.12) can be restricted to the saddles z_i such that $\operatorname{Re}\varphi(z_i)$ is maximal, as the other terms are exponentially smaller as $M \to +\infty$.

5.4 Airy functions

5.4.1 Airy equation

The Airy equation is the linear second-order ODE

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x) = xf(x). \tag{5.4.1}$$

Its solutions provide a simple model of functions interpolating between oscillatory and exponential behaviors, see Figure 5.1. This model is actually quite universal to describe the transition between oscillatory and exponential behaviors (cf. Section 5.4.3).

5.4.2 Solution by contour integral

The solutions of (5.4.1) are transcendental, in the sense that they cannot be expressed through algebraic expressions of elementary functions. However, the solutions of (5.4.1) can be effectively constructed in terms of contour integrals of elementary functions (more precisely, of exponentials). This

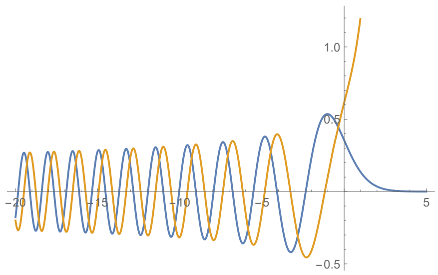


Figure 5.1: Plot of the functions $\mathrm{Ai}(x)$ and $\mathrm{Bi}(x)$, which form a basis of the solution space of (5.4.1), in blue and yellow respectively. They oscillate (with interlacing zeros and amplitude decaying as $|x|^{-1/4}$) when $x \to -\infty$, and they are exponentially decaying (Ai) and growing (Bi) when $x \to +\infty$. The precise asymptotics as $x \to \pm\infty$ are proven below in Proposition 5.4.8.

has the important consequence that we can analyze the asymptotics of the solutions by means of the steepest descent method (as we will do in Section 5.4.4).[3]

Let us first introduce the following notations:

$$\Omega := \left\{z \in \mathbb{C} : \mathrm{Re}\,(z^3) < 0\right\} = \left\{z \in \mathbb{C} :\ \arg z \in \left(\tfrac{\pi}{6}, \tfrac{\pi}{2}\right) \cup \left(\tfrac{5\pi}{6}, \tfrac{7\pi}{6}\right) \cup \left(\tfrac{3\pi}{2}, \tfrac{11\pi}{6}\right)\right\}, \tag{5.4.2}$$

$$\Omega_\epsilon := \{z \in \Omega :\ |z - w| \geq \epsilon \text{ for all } w \in \mathbb{C} \setminus \Omega\}. \tag{5.4.3}$$

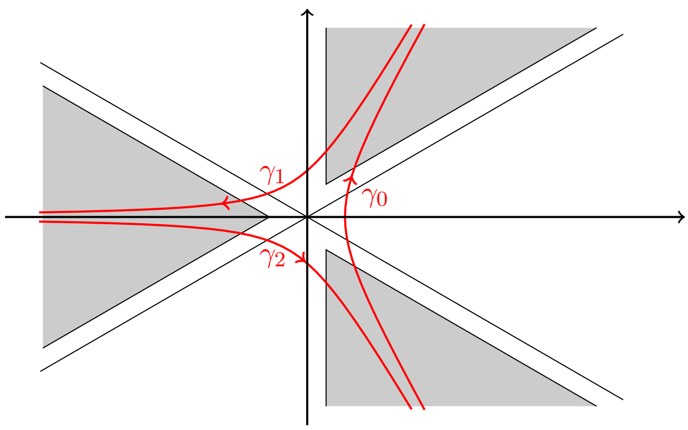


Figure 5.2: The set Ω_ϵ is in gray. The contour γ in Lemma 5.4.1 has go reach infinity within the gray regions. Ultimately, up to homotopy deformation and orientation reversal, the contours can only be the red ones (denoted $\gamma_0, \gamma_1, \gamma_2$) shown in the picture.

Lemma 5.4.1. *Let $\gamma : (-\infty, +\infty) \to \mathbb{C}$ be any oriented smooth contour in the complex plane such that $\lim_{t\to\pm\infty} |\gamma(t)| = +\infty$ and there exist $\epsilon, t_0 > 0$ such that $\gamma(t) \in \Omega_\epsilon$ for all $|t| > t_0$. Then*

$$f(x) := \int_\gamma \mathrm{e}^{\frac{s^3}{3} - xs} \mathrm{d}s \tag{5.4.4}$$

defines an entire function of $x \in \mathbb{C}$ which solves (5.4.1).

Proof. To check that the contour integral is well-defined we only need to check that it converges. To this end, fix $M > 0$ and assume that $|x| < M$ and $s = R\mathrm{e}^{\mathrm{i}\theta}$ for $R > 0, \theta \in \mathbb{R}/2\pi\mathbb{Z}$. Then the integrand can be estimated as follows:

$$\left|\mathrm{e}^{\frac{s^3}{3} - xs}\right| = \mathrm{e}^{\mathrm{Re}\left(\frac{s^3}{3} - xs\right)} \leq \mathrm{e}^{\frac{R^3}{3}\cos(3\theta) + RM}. \tag{5.4.5}$$

[3] The solutions could also be constructed by the power series method, namely plugging the ansatz $f(x) = \sum_{i\geq 0} f_i x_i$ into $\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) = x f(x)$ and solving for $f_2, f_3, \cdots$ in terms of two arbitrary constants f_0, f_1. This method can be used to show that the solutions are entire functions: once the solution for f_i is plugged into $f(x) = \sum_{i\geq 0} f_i x_i$, the series is easily shown to have infinite radius of convergence. On the other hand, asymptotics of the solutions $f(x)$ for large $|x|$ do not easily follow from the series representation.

An elementary geometric inspection of Ω_ϵ implies that if $Re^{i\theta} \in \Omega_\epsilon$ then $\cos\theta < -\frac{\kappa}{R}$ for R sufficiently large and for some $\kappa > 0$ depending on ϵ only. Since the tails of γ at ∞ lie entirely in Ω_ϵ, the integrand along these tails is $O(\mathrm{e}^{-\frac{\kappa}{3}R^2+MR}) = O(e^{-\kappa R^2/4})$ as $|s| = R \to +\infty$. Therefore, the integral defining $f(x)$ is absolutely convergent, with a bound uniform for x in compact subsets of $\mathbb{C}$. Formally differentiating in x under the integral sign we get

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = -\int_\gamma s\,\mathrm{e}^{\frac{s^3}{3}-xs}\mathrm{d}s. \tag{5.4.6}$$

By completely similar estimates, the integral in the right hand side is absolutely convergent with a uniform bound for x in compact subsets of $\mathbb{C}$, ensuring that we can indeed differentiate under integral sign. It follows that $f(x)$ is an entire function.

Finally, by the same argument, we can show that it is possible to differentiate once more under integral sign, which gives

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) = \int_\gamma s^2 \mathrm{e}^{\frac{s^3}{3}-xs}\mathrm{d}s = \int_\gamma \left(\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{e}^{\frac{s^3}{3}}\right)\mathrm{e}^{-xs}\mathrm{d}s = -\int_\gamma \mathrm{e}^{\frac{s^3}{3}}\left(\frac{\mathrm{d}}{\mathrm{d}s}\mathrm{e}^{-xs}\right)\mathrm{d}s = \int_\gamma \mathrm{e}^{\frac{s^3}{3}} x\mathrm{e}^{-xs}\mathrm{d}s = xf(x),$$

where we integrate by parts, the boundary terms vanishing thanks to our assumption on the tails of γ at ∞. □

Remark 5.4.2. Linear ODEs of the form $\left[p(\frac{\mathrm{d}}{\mathrm{d}x}) + xq(\frac{\mathrm{d}}{\mathrm{d}x})\right] f(x) = 0$ for p, q polynomials may be considered as the simplest linear ODEs beyond the constant coefficients case (i.e., $q = 0$). The Airy equation belongs to this class (with $p(s) = s^2$, $q(s) = 1$). The solutions to these equations can in general be expressed by contour integrals. For, writing the solution in the form $\int_\gamma \exp(v(s) + xs)\mathrm{d}s$ and formally differentiating under integral sign we find that v is determined by the first order ODE $[p(s) - q'(s) - q(s)\partial_s]\, v(s) = 0$, namely $v'(s) = \frac{p(s)-q'(s)}{q(s)}$. This ODE for v amounts to finding the primitive of a rational function and can therefore be explicitly integrated in terms of elementary functions. Once v has been computed, it is then possible to choose wisely a set of contours γ in the complex s-plane for which the integral $\int_\gamma \exp(v(s)+xs)\mathrm{d}s$ converges absolutely, thus providing a basis of solutions to the given ODE. //

By Cauchy theorem, the only possibilities for the integration contour γ are the three contours $\gamma_0, \gamma_1, \gamma_2$ depicted in Figure 5.2, distinguished by the sectors in which they approach ∞. Note that if γ approaches ∞ in the same sector, the resulting integral vanishes by Cauchy theorem. Therefore let us set

$$w_j(x) = \frac{1}{2\pi\mathrm{i}}\int_{\gamma_j} \mathrm{e}^{\frac{s^3}{3}-xs}\mathrm{d}s, \qquad j = 0, 1, 2. \tag{5.4.7}$$

The solution space of $\mathrm{d}^2/\mathrm{d}x^2 - x$ is two-dimensional, indicating some redundancy in the set of solutions w_0, w_1, w_2. Indeed, the sum (in homology) of the three contours is zero, so that by Cauchy theorem $w_0 + w_1 + w_2 = 0$. Any two out of w_0, w_1, w_2 form a basis of the solution space of $\mathrm{d}^2/\mathrm{d}x^2 - x$ (see Exercise 5.4.5 below).

A canonical basis of solutions to the Airy equation is defined as follows:

$$\mathrm{Ai}(z) := w_0(z), \qquad \mathrm{Bi}(z) := \mathrm{i}(w_2(z) - w_1(z)). \tag{5.4.8}$$

Ai is called **Airy function**, and Bi **Airy function of the second kind**, or, more colloquially, Bairy function.

Some basic properties of the Airy functions are left as exercise below.

Exercise 5.4.3. Show that $\mathrm{Ai}(x), \mathrm{Bi}(x)$ are real for real x. (Hint: use the general formula $\overline{\int_\gamma f(s)\mathrm{d}s} = \int_{\overline{\gamma}} \overline{f(\overline{s})}\mathrm{d}s$, where $\overline{\gamma}$ is the oriented contour obtained by conjugation of the oriented contour γ.) //

Exercise 5.4.4. Let $\omega := \mathrm{e}^{2\pi\mathrm{i}/3}$.

1. If $f(z)$ solves $\frac{\mathrm{d}^2}{\mathrm{d}z^2} f(z) = zf(z)$, show that $g(z) := f(\omega z)$ solves the same equation, $\frac{\mathrm{d}^2}{\mathrm{d}z^2} g(z) = zg(z)$.

2. Show that $w_1(z) = \omega w_0(\omega z)$ and $w_2(z) = \omega^2 w_0(\omega^2 z)$.

//

Exercise 5.4.5. Prove that $w_0(0) = \mathrm{Ai}(0) = \frac{1}{3^{2/3}\Gamma(2/3)}$ and $w_0'(0) = \mathrm{Ai}'(0) = -\frac{1}{3^{1/3}\Gamma(1/3)}$, where Γ is the Euler Gamma function. Use the fact that $w_0(0) \neq 0 \neq w_0'(0)$ and the second point of Exercise 5.4.4 to show that any two out of w_0, w_1, w_2 are linearly independent. (Hint: deform the integration contour γ_0 into $\mathrm{e}^{i\pi/3}\mathbb{R}_+ \cup \mathrm{e}^{-\mathrm{i}\pi/3}\mathbb{R}_+$ (with appropriate orientation) and apply the definition $\Gamma(z) = \int_0^{+\infty} \mathrm{e}^{-s} s^{z-1} \mathrm{d}s$.) //

Exercise 5.4.6. Prove the following improper Riemann integral representation for the Airy functions, valid for $x \in \mathbb{R}$:

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^{+\infty} \cos\left(\frac{s^3}{3} + xs\right)\mathrm{d}s, \tag{5.4.9}$$

$$\mathrm{Bi}(x) = \frac{1}{\pi}\int_0^{+\infty} \left[\exp\left(-\frac{s^3}{3} + xs\right) + \sin\left(\frac{s^3}{3} + xs\right)\right]\mathrm{d}s, \tag{5.4.10}$$

where $\int_0^{+\infty} := \lim_{L\to+\infty}\int_0^L$. (Hint: deform the integration contour γ_0 into $\mathrm{i}[-L, L]$ and rays from $\pm\mathrm{i}L$ to ∞ in the appropriate sectors and integrate by parts to show that the integrals along these rays vanish as $L \to +\infty$.) //

Exercise 5.4.7. For any $j, k \in \{0, 1, 2\}$, consider the Wronskian

$$W_{jk} := w_j(z)w_k'(z) - w_j'(z)w_k(z). \tag{5.4.11}$$

Prove that W_{jk} does not depend on z. Compute W_{jk} using Exercises 5.4.5 and 5.4.4. Deduce that

$$\mathrm{Ai}(z)\mathrm{Bi}'(z) - \mathrm{Ai}'(z)\mathrm{Bi}(z) = \frac{1}{\pi}. \tag{5.4.12}$$

Use this identity to show that the real zeros of Ai, Bi are

- "simple": if $x \in \mathbb{R}$ is such that $\mathrm{Ai}(x) = 0$ then $\mathrm{Ai}'(x) \neq 0$, and the same for Bi in place of Ai, and
- "interlacing": if there are real numbers $x_1 < x_2$ such that $\mathrm{Ai}(x_1) = 0 = \mathrm{Ai}(x_2)$, there exists $y \in (x_1, x_2)$ such that $\mathrm{Bi}(y) = 0$, and, conversely, the same with the roles of Ai, Bi interchanged.

(Hint: for the last part, mimic the proof of Proposition 3.2.20.) //

5.4.3 Applications of Airy functions

All appearances of the Airy functions are related, one way or another, to the transition between oscillatory and exponential (decaying or growing) behaviors. A brief (and certainly not exhaustive) list of places where Airy functions appear is given below.

- Sir George Biddell Airy first introduced the Airy function in his 1838 paper "On the intensity of light in the neighbourhood of a caustic". The Airy function in this case describes the transition between dark and light regions of space.
- In Quantum Mechanics, the stationary Schrödinger equation

  $$\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V(x)\right)\psi(x) = E\psi(x) \qquad (E \in \mathbb{R}) \tag{5.4.13}$$

  near a "turning point" $x_0 \in \mathbb{R}$, $V(x_0) = E$, is approximated as

  $$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) = (V(x) - E)\psi(x) = (x - x_0)(V'(x_0) + O(x - x_0))\psi(x) \tag{5.4.14}$$

so that formally neglecting the higher order terms in $x - x_0$ we get the approximated equation

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\psi(x) = (x - x_0)V'(x_0)\psi(x) \tag{5.4.15}$$

which is solved in terms of the Airy functions, as writing $\psi(x) = f\left((x - x_0)\sqrt[3]{V'(x_0)}\right)$ implies that f satisfies the Airy equation (5.4.1). It can be rigorously proved that the Airy function indeed governs the transition between the oscillatory behavior of the stationary wave function ψ with energy E in the classically allowed region $V(x) < E$ and the decaying behavior of ψ in the classically prohibited region $V(x) > E$.

- We shall see below that the Airy equation also describes the transition of the behavior of orthogonal polynomials of large degree at the edge between the regions where they have zeros and where they diverge to ∞.

5.4.4 Asymptotics

As anticipated, the contour integral representation of solutions to the Airy equation is important because it allows us to describe very precisely the asymptotic properties of its solutions.

Proposition 5.4.8. *We have*

$$\mathrm{Ai}(x) = \begin{cases} \frac{1}{2\sqrt{\pi}}x^{-\frac{1}{4}}\exp\left(-\frac{2}{3}x^{3/2}\right)(1 + O(x^{-3/2})), & x \to +\infty \\ \frac{1}{\sqrt{\pi}}|x|^{-\frac{1}{4}}\sin\left(\frac{\pi}{4} + \frac{2}{3}|x|^{3/2}\right)(1 + O(|x|^{-3/2})), & x \to -\infty. \end{cases} \tag{5.4.16}$$

Proof. Let us first consider the case $x \to +\infty$. The first step is to make the change of variables $s = x^{1/2}t$ so that

$$\mathrm{Ai}(x) = \frac{1}{2\pi\mathrm{i}}\int_{\gamma_0} \mathrm{e}^{\frac{s^3}{3} - xs}\mathrm{d}s = \frac{x^{1/2}}{2\pi\mathrm{i}}\int_{\gamma_0} \mathrm{e}^{x^{3/2}\left(\frac{t^3}{3} - t\right)}\mathrm{d}t. \tag{5.4.17}$$

(We use the Cauchy theorem to deform the contour $x^{-1/2}\gamma_0$ back to γ_0.) The function $\varphi(t) := \frac{t^3}{3} - t$ satisfies $\varphi'(t) = t^2 - 1$ hence it has two saddle points $t_\pm = \pm 1$. The steepest descent direction at $t = 1$ is vertical, the steepest descent direction at $t = -1$ is horizontal. The locus $\mathrm{Im}\,\varphi(t) = \mathrm{Im}\,\varphi(+1) = 0$, writing $t = u + \mathrm{i}v$ reduces to $v(1 - u^2 + \frac{1}{3}v^2) = 0$ which consists of the two curves $v = 0$ (steepest ascent direction, as $t = +1$ is a minimum of $\varphi(t)$ along $t \in \mathbb{R}$) and a branch of the hyperbola $1 - u^2 + \frac{1}{3}v^2 = 0$ (steepest descent direction). It is then possible to globally deform γ_0 into the branch of the hyperbola passing through $u = 1, v = 0$ (see Figure 5.3). By the general formula (5.3.12) we get

$$\mathrm{Ai}(x) = \frac{x^{1/2}}{2\pi\mathrm{i}}\mathrm{e}^{x^{3/2}\varphi(1)}\mathrm{e}^{\mathrm{i}\frac{\pi}{2}}\sqrt{\frac{2\pi}{x^{3/2}|\varphi''(1)|}}(1 + O(x^{-3/2})) \tag{5.4.18}$$

and by $\varphi(1) = -2/3$, $\varphi''(1) = 2$ and elementary simplification we get the first formula in (5.4.16).

Let us not consider the case $x \to -\infty$. Again, the first step is to make the change of variables $s = |x|^{1/2}t$

$$\mathrm{Ai}(x) = \frac{1}{2\pi\mathrm{i}}\int_{\gamma_0} \mathrm{e}^{\frac{s^3}{3} - xs}\mathrm{d}s = \frac{|x|^{1/2}}{2\pi\mathrm{i}}\int_{\gamma_0} \mathrm{e}^{|x|^{3/2}\left(\frac{t^3}{3} + t\right)}\mathrm{d}t \tag{5.4.19}$$

so that here $\varphi(t) = \frac{t^3}{3} + t$, with $\varphi'(t) = t^2 + 1$ and saddle points are $t_1 = \mathrm{i}, t_2 = -\mathrm{i}$. The directions $\theta_{1,2}$ of steepest descent at the saddles are computed by the general formula (5.3.13) as

$$\theta_1 = \frac{\pm\pi - \arg\varphi''(\mathrm{i})}{2} = \pm\frac{\pi}{2} - \frac{\pi}{4}, \qquad \theta_2 = \frac{\pm\pi - \arg\varphi''(-\mathrm{i})}{2} = \pm\frac{\pi}{2} + \frac{\pi}{4}, \tag{5.4.20}$$

By Cauchy theorem, we can deform the contour γ_0 into the union of the two steepest descent curves, see Figure 5.3 and this forces us to take $\theta_1 = \frac{\pi}{4}$ and $\theta_2 = \frac{3}{4}\pi$. By the general formula (5.3.12), we obtain

$$\mathrm{Ai}(x) = \sum_{i=1,2} \frac{|x|^{1/2}}{2\pi \mathrm{i}} \mathrm{e}^{|x|^{3/2}\varphi(t_i)} \mathrm{e}^{\mathrm{i}\theta_i} \sqrt{\frac{2\pi}{|x|^{3/2}|\varphi''(t_i)|}} (1 + O(|x|^{-3/2})) \tag{5.4.21}$$

and by an elementary simplification we get the second formula in (5.4.16). □

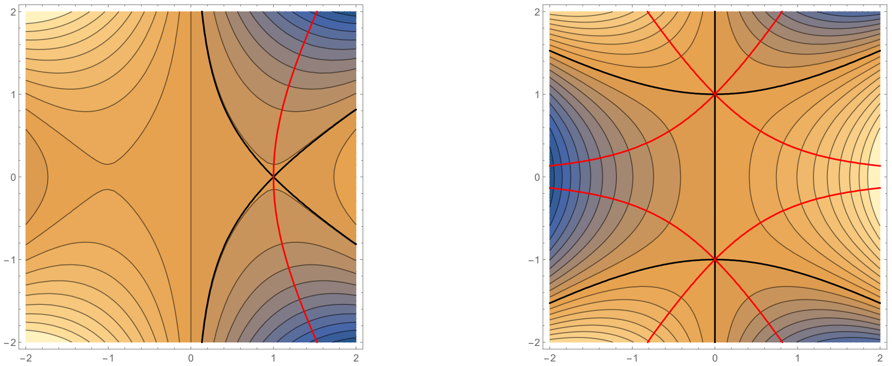


Figure 5.3: *On the left*, the steepest descent contour (red) when $x > 0$ through $t = 1$; in black the level curves of the real part of $\frac{t^3}{3} - t$. Blue colors correspond to large negative values of the real part of $\frac{t^3}{3} - t$, and orange colors correspond to large positive values of the real part of $\frac{t^3}{3} - t$. We can deform γ_0 globally into this steepest descent contour. *On the right*, the steepest descent and ascent contours (red) when $x < 0$ passing through the saddle points $t = \pm \mathrm{i}$; in black the level curves of the real part of $\frac{t^3}{3} + t$. Blue colors correspond to large negative values of the real part of $\frac{t^3}{3} + t$, and orange colors correspond to large positive values of the real part of $\frac{t^3}{3} + t$. The steepest descent curves have ends in the blue regions. We can deform γ_0 into the union of the two steepest descent contours.

Exercise 5.4.9. Use the steepest descent method to prove

$$\mathrm{Bi}(x) = \begin{cases} \frac{1}{\sqrt{\pi}} x^{-\frac{1}{4}} \exp\left(\frac{2}{3}x^{3/2}\right)(1 + O(x^{-3/2})), & x \to +\infty \\ \frac{1}{\sqrt{\pi}} |x|^{-\frac{1}{4}} \cos\left(\frac{\pi}{4} + \frac{2}{3}|x|^{3/2}\right)(1 + O(|x|^{-3/2})), & x \to -\infty. \end{cases} \tag{5.4.22}$$

and (using the formula $\mathrm{Ai}'(x) = -\frac{1}{2\pi \mathrm{i}} \int_{\gamma_0} s\, \mathrm{e}^{\frac{s^3}{3} - xs} \mathrm{d}s$)

$$\mathrm{Ai}'(x) = \begin{cases} -\frac{x^{\frac{1}{4}}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right)(1 + O(x^{-3/2})), & x \to +\infty \\ -\frac{1}{\sqrt{\pi}} |x|^{\frac{1}{4}} \cos\left(\frac{\pi}{4} + \frac{2}{3}|x|^{3/2}\right)(1 + O(|x|^{-3/2})), & x \to -\infty. \end{cases} \tag{5.4.23}$$

//

5.4.5 Stokes' phenomenon

The Stokes' phenomenon refers to the fact that the Airy functions have radically different asymptotic behaviors at $x = \pm\infty$. Namely, if we analytically continue the functions $\frac{\exp(-\frac{2}{3}x^{3/2})}{2\sqrt{\pi}x^{1/4}}$ and $\frac{\sin(\frac{\pi}{4} + \frac{2}{3}(-x)^{3/2})}{2\sqrt{\pi}(-x)^{1/4}}$, appearing in the $x \to \pm\infty$ asymptotics for the Airy function, to the complex x-plane we get two different functions. In the paper "On the numerical calculation of a class of definite integrals and infinite series" of 1847 (almost a decade after the introduction of the Airy function), Sir George Gabriel Stokes first addressed the study of this behavior of Airy functions.

An indication of the appearance of the Stokes' phenomenon comes from the fact that the Airy function is entire while its asymptotic expansion at $+\infty$, once analytically continued to the complex

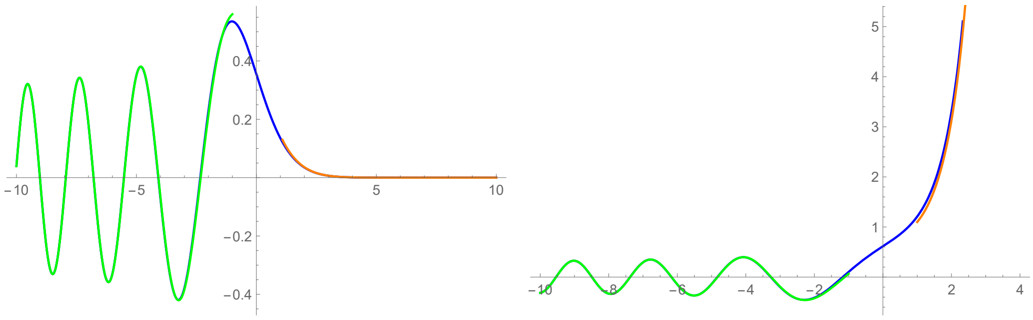


Figure 5.4: Ai (left) and Bi (right) functions (in blue) plotted together with their asymptotic behaviors for $x \to \pm\infty$.

plane has a nontrivial monodromy and so we cannot expect it to hold uniformly as $x \to \infty$ in the whole complex plane.

A more precise explanation of the Stokes' phenomenon of the Airy function is the following. Along the lines of the proof of Proposition 5.4.8, we first change variable $s = |x|^{1/2}t$ to obtain

$$\mathrm{Ai}(x) = \frac{1}{2\pi\mathrm{i}} \int_{\gamma_0} \mathrm{e}^{\frac{s^3}{3} - xs} \mathrm{d}s = \frac{|x|^{1/2}}{2\pi\mathrm{i}} \int_{\gamma_0} \mathrm{e}^{|x|^{3/2}\left(\frac{t^3}{3} - \mathrm{e}^{\mathrm{i}\nu} t\right)} \mathrm{d}t, \tag{5.4.24}$$

where we denote $x = |x|\mathrm{e}^{\mathrm{i}\nu}$. The last expression can be studied with the steepest descent method; however, the saddle points $\pm\mathrm{e}^{\mathrm{i}\nu/2}$ move around the unit circle as the argument ν of x varies, as well as the steepest descent contours also change in dependence of ν. On the other hand, the contour γ_0 of integration must remain fixed (at least up to homotopy deformations admissible by the Cauchy theorem). As a consequence, the type of contour deformation of γ_0 into steepest descent contour(s) which are allowed changes when we vary the argument of x, and it is this mechanism at the origin of the Stokes' phenomenon.

More precisely, we have already seen in the proof of Proposition 5.4.8 that the situation is drastically different in the cases $\nu = 0, \pi$. In general, it can be shown (see Figure 5.5) that the contour γ_0 can be deformed into a steepest descent contour passing through the saddle $\mathrm{e}^{\mathrm{i}\nu/2}$ only, provided that $-\frac{2}{3}\pi < \nu < \frac{2}{3}\pi$. When ν is not in this region, both saddle points contribute. The critical lines $\arg x = \nu = \pm\frac{2}{3}\pi$ are termed Stokes' lines.

By computing explicitly the contribution from both saddles we obtain the complete asymptotics as $x \to \infty$ in the complex plane is[4]

$$\mathrm{Ai}(x) = \begin{cases} \frac{x^{-\frac{1}{4}}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{\frac{3}{2}}\right)(1 + O(x^{-\frac{3}{2}})), & |\arg x| < \frac{2}{3}\pi \\ \frac{x^{-\frac{1}{4}}}{2\sqrt{\pi}} \left(\exp\left(-\frac{2}{3}x^{\frac{3}{2}}\right) + \mathrm{i}\exp\left(\frac{2}{3}x^{\frac{3}{2}}\right)\right)(1 + O(x^{-\frac{3}{2}})), & -\pi < \arg x < -\frac{2}{3}\pi \text{ or } \frac{2}{3}\pi < \arg x \le \pi. \end{cases} \tag{5.4.25}$$

The change of the asymptotics across the Stokes' lines $\arg x = \pm\frac{2}{3}\pi$ do not imply an abrupt change in the asymptotic behavior of Airy functions. Indeed, the additional term that starts contributing when $|\arg \nu| > \frac{2}{3}\pi$ is exponentially smaller than the other term along the Stokes' lines, and so practically invisible along a Stokes' line. The effects of this additional contribution are then only manifest when we approach the negative real line, where now the contributions of both terms are of the same order and produce the oscillatory behavior, see Figure 5.6.

[4]These asymptotics can be rewritten more symmetrically in the region $|\arg(-x)| < \frac{1}{3}\pi$ as $\mathrm{Ai}(x) = \frac{\sin(\frac{\pi}{4} + \frac{2}{3}(-x)^{3/2})}{\sqrt{\pi}(-x)^{1/4}}(1 + O(x^{3/2}))$.

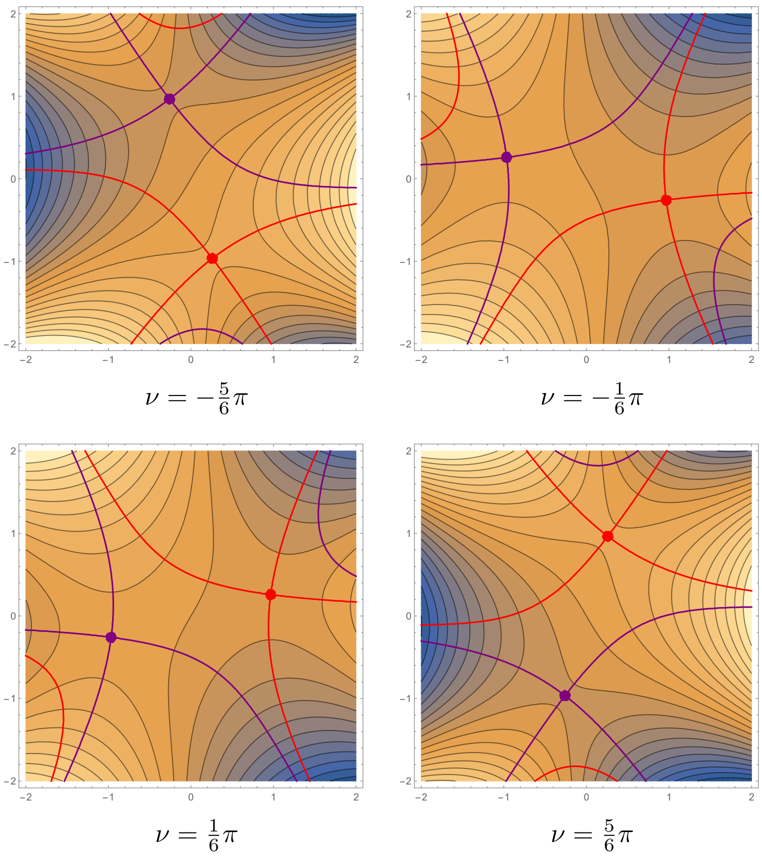


Figure 5.5: Geometry of steepest ascent/descent contours for various values of $\nu = \arg x \in (-\pi, \pi)$. The saddle $\exp(\mathrm{i}\nu/2)$ is in red, the saddle $-\exp(\mathrm{i}\nu/2)$ is in purple. Also plotted, with the same colors, the curves where the imaginary part of $t^3/3 - \mathrm{e}^{\mathrm{i}\nu}t$ is constant (steepest ascent/descent curves). The Stokes' phenomenon of Airy functions is clearly explained as follows: when $|\nu| < \frac{2}{3}\pi$ (as in the second and third pictures) the contour γ_0 can be deformed into the steepest descent path through the red saddle only; when $|\nu| \geq \frac{2}{3}\pi$ (as in the first and last pictures) the contour γ_0 can be deformed into the union of the two steepest descent paths through the red and through the purple saddles. Incidentally, note that after a full turn $\arg x \mapsto \arg x + 2\pi$, the two saddle points swap.

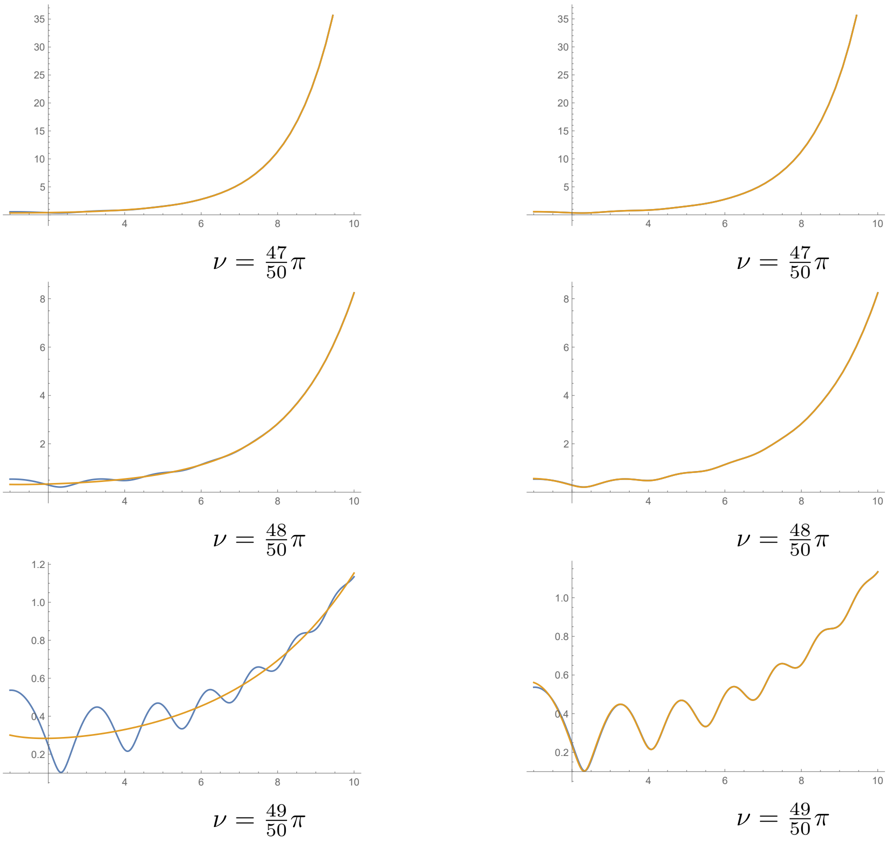


Figure 5.6: *In blue:* $|\mathrm{Ai}(x\mathrm{e}^{\mathrm{i}\nu})|$ for $1 < x < 10$ and values of ν approaching π from below. *In yellow:* in the first column, the approximation given in the first line of (5.4.25); in the second column, the full approximation given in the second line of (5.4.25).

Chapter 6

Applications to the GUE

6.1 Large n scaling of random Hermitian matrices

What is the scale of eigenvalues of a large-size Hermitian matrix? If M is a (random) matrix of size n with entries $M_{ij} = O(1)$, then

$$\mathrm{tr}\,(M^2) = \sum_{i,j=1}^{n} |M_{ij}|^2 = O(n^2). \tag{6.1.1}$$

Therefore, if $x_1, \dots, x_n$ are the eigenvalues of M,

$$\sum_{i=1}^{n} x_i^2 = O(n^2) \tag{6.1.2}$$

i.e., in average we expect $\lambda_i = O(\sqrt{n})$. Therefore, it is meaningful to rescale

$$M \mapsto \frac{M}{\sqrt{n}} \tag{6.1.3}$$

(with eigenvalues which are $O(1)$ in average) in order to study the eigenvalues in the large size limit $n \to +\infty$. Therefore, from now one we shall consider the GUE as the measure

$$\frac{1}{\pi^{\frac{n^2}{2}} 2^{\frac{n(n-1)}{2}}} \exp\left(-n\,\mathrm{tr}\,(M^2)\right) \prod_{i=1}^{n} \mathrm{d}M_{ii} \prod_{1\le i<j\le n} \mathrm{dRe}\,M_{ij}\mathrm{dIm}\,M_{ij} \tag{6.1.4}$$

on $\mathbf{H}_n$. (**Exercise**: prove that this distribution is normalized to 1; this is similar to Exercise 2.2.7.)

Another way to motivate the rescaling (6.1.3) comes from the following observation; the distribution of eigenvalues is proportional to[1]

$$\Delta^2(x_1, \dots, x_n) \mathrm{e}^{-\frac{x_1^2}{2}} \cdots \mathrm{e}^{-\frac{x_n^2}{2}} = \exp\left(2 \sum_{1\le i<j\le n} \log|x_i - x_j| - \frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \tag{6.1.5}$$

and hence it is subject to two competing effects: the eigenvalues repel each other and, at the same time, they cannot escape too far to infinity as they are confined by the Gaussian potential. The appropriate scale is then the one where the two effects are of the same order in n as $n \to +\infty$; scaling $x_i = \sqrt{n} y_i$ for $i = 1, \dots, n$, their distribution is proportional to

$$\exp\left(2 \sum_{1\le i<j\le n} \log|y_i - y_j| - \frac{n}{2}\sum_{i=1}^{n} y_i^2\right) \tag{6.1.6}$$

[1]For the physically-inclined reader, note that this is closely related to the Boltzmann distribution $\exp\left(-\frac{1}{k_B T} U(x_1, \dots x_n)\right)$ for a canonical ensemble of n identical particles with electrical charge q constrained to lie on a line, trapped by a Gaussian potential, and interacting as a 2D Coulomb gas, namely their potential energy is $U(x_1, \dots, x_n) = q \sum_{1\le i<j\le n} \log|x_i - x_j| - \sum_{i=1}^{n} \frac{x_i^2}{2}$).

so that now both terms in the exponent are of the same order $O(n^2)$ provided $y_i = O(1)$ and not too close to each other, therefore this is the regime where we see the interesting balance of the competing effects of repulsion and confining.

That this is the correct scaling as $n \to +\infty$ is also manifest from the numerical examples in Section 1.1.1.

6.2 Plancherel–Rotach asymptotics for Hermite polynomials

6.2.1 Contour integral representation for Hermite polynomials

Introduce

$$\widetilde{P}_k(x;n) = \frac{1}{\sqrt{n}^k} P_k(\sqrt{n}\, x) \tag{6.2.1}$$

where P_k are the Hermite polynomials defined in (3.1.17). The normalization ensures that they are monic.

Exercise 6.2.1. Prove that $\int_{\mathbb{R}} \widetilde{P}_k(x;n)\widetilde{P}_\ell(x;n)\mathrm{e}^{-n\frac{x^2}{2}}\mathrm{d}x = k! n^{-k-\frac{1}{2}}\sqrt{2\pi}\,\delta_{k,\ell}$. (Hint: use Proposition 3.1.5.) //

Therefore, $\widetilde{P}_k(x;n)$ are the monic orthogonal polynomials with respect to the measure $\mathrm{e}^{-n\frac{x^2}{2}}\mathrm{d}x$, which we need in order to study the GUE as in (6.1.4).

To find large n asymptotics for (6.2.1) we will apply the steepest descent method. To start with, we need a contour integral representation.

Lemma 6.2.2. *Let $P_k(x)$ be the Hermite polynomials. For all $x, t \in \mathbb{C}$ we have*

$$\sum_{j\geq 0} P_j(x)\frac{t^j}{j!} = \mathrm{e}^{xt-\frac{t^2}{2}}. \tag{6.2.2}$$

Proof. By the defining relation (3.1.17) of Hermite polynomials we have

$$\sum_{j\geq 0} P_j(x)\frac{t^j}{j!} = \mathrm{e}^{\frac{x^2}{2}}\sum_{j\geq 0}\frac{(-t\partial_x)^j}{j!}\mathrm{e}^{-\frac{x^2}{2}} = \mathrm{e}^{\frac{x^2}{2}}\mathrm{e}^{-t\partial_x}\mathrm{e}^{-\frac{x^2}{2}} = \mathrm{e}^{\frac{x^2}{2}-\frac{(x-t)^2}{2}}. \tag{6.2.3}$$

□

Remark 6.2.3. In the above proof we have used that for all entire functions f and all $x, s \in \mathbb{C}$ we have

$$(\exp(s\partial_x)f)(x) := \sum_{j\geq 0}\frac{s^j(\partial_x^j f)(x)}{j!} = f(x+s), \tag{6.2.4}$$

because the Taylor series of f at any point is convergent. //

As a consequence we have the following.

Lemma 6.2.4. *Let $P_k(x)$ be the Hermite polynomials as defined in (3.1.17). For all $k \geq 0$ we have*

$$P_k(x) = \frac{k!}{2\pi\mathrm{i}}\int_\gamma \mathrm{e}^{xt-\frac{t^2}{2}}\frac{\mathrm{d}t}{t^{k+1}} \tag{6.2.5}$$

where γ is any piece-wise $\mathcal{C}^1$ closed contour in the complex t-plane with $\mathrm{Ind}\,(\gamma; t=0) = 1$.

Proof. By the Cauchy residue theorem

$$\int_\gamma \mathrm{e}^{xt-\frac{t^2}{2}}\frac{\mathrm{d}t}{t^{k+1}} = 2\pi\mathrm{i}\,\mathrm{Res}\left(\frac{\mathrm{e}^{xt-\frac{t^2}{2}}}{t^{k+1}}; t=0\right) = 2\pi\mathrm{i}\frac{P_k(x)}{k!} \tag{6.2.6}$$

where to compute the residue we apply Lemma 6.2.2 to obtain the Laurent expansion at $t = 0$

$$\frac{\mathrm{e}^{xt-\frac{t^2}{2}}}{t^{k+1}} = \sum_{j\geq 0} \frac{P_j(x)}{j!} t^{j-k-1} \tag{6.2.7}$$

so that the coefficient in front of t^{-1} is $P_k(x)/k!$. ☐

Corollary 6.2.5. *For all $n \geq 0$ and all $j \in \mathbb{Z}$ such that $n + j \geq 0$ we have*

$$\widetilde{P}_{n+j}(x; n) = \frac{(n+j)!}{2\pi\mathrm{i}\, n^{n+j}} \int_\gamma \mathrm{e}^{n(xz-\frac{z^2}{2}-\log z)} \frac{\mathrm{d}z}{z^{j+1}} \tag{6.2.8}$$

where γ is any piece-wise $\mathcal{C}^1$ closed contour in the complex z-plane with $\mathrm{Ind}\,(\gamma; z = 0) = 1$.

Proof. Change variable $t = \sqrt{n}\, z$ in the result of the lemma. ☐

We consider $\widetilde{P}_k(x; n)$ of order $k = n+j$ because we would like to compute the large n asymptotics for the correlation functions in the GUE, which can be computed in terms of the Christoffel–Darboux kernel

$$K_n^{\mathsf{GUE}}(x, y) := \mathrm{e}^{-n\frac{x^2+y^2}{2}} \sum_{\ell=0}^{n-1} \frac{\widetilde{P}_\ell(x; n)\widetilde{P}_\ell(y; n)}{h_\ell(n)} = \frac{\mathrm{e}^{-n\frac{x^2+y^2}{2}}}{h_{n-1}(n)} \frac{\widetilde{P}_n(x; n)\widetilde{P}_{n-1}(y; n) - \widetilde{P}_{n-1}(x; n)\widetilde{P}_n(y; n)}{x - y}, \tag{6.2.9}$$

where $h_\ell(n) = \ell! n^{-\ell-\frac{1}{2}}\sqrt{2\pi}$ by Exercise 6.2.1. So we will need $\widetilde{P}_k(x; n)$ of order $k = n+j$ in particular for $j = 0, -1$ as $n \to +\infty$.

6.2.2 Steepest descent analysis

The expression (6.2.8) makes the study of large n asymptotics amenable by the steepest descent method introduced above. The functions φ, f are

$$\varphi(z) = xz - \frac{z^2}{2} - \log z, \qquad f(z) = 1/z^{j+1}, \tag{6.2.10}$$

which are analytic in the open set $\mathcal{U} = \mathbb{C} \setminus (-\infty, 0]$. (The branch of the log is dictated by $\log z \in \mathbb{R}$ for $z > 0$.) Therefore $\varphi'(z) = x - z - \frac{1}{z}$ and the saddle-points are obtained by $\varphi'(z) = 0$ which yields

$$z^2 - xz + 1 = 0 \Rightarrow z = z_\pm(x) := \frac{x \pm \sqrt{x^2 - 4}}{2}. \tag{6.2.11}$$

Exterior: $|x| > 2$

When $|x| > 2$ the two saddle points $z_\pm(x)$ are both real. By studying the function $\varphi = -\frac{z^2}{2} + xz - \log z$ for $z > 0$ (Figure 6.1) we see that $z_-(x)$ is a minimum and $z_+(x)$ is a maximum; we deduce that the contour of steepest descent is vertical at $z_-(x)$ and horizontal at $z_+(x)$. We can deform the contour of integration as we indicate in Figure 6.2.

Applying the general formula (5.3.12) with $M = n$ to (6.2.8) (here $\theta = \pi/2$ and $\varphi''(z_-(x)) < 0$) we get

$$\begin{aligned}\widetilde{P}_{n+j}(x; n) &= \frac{(n+j)!}{2\pi \cancel{\mathrm{i}}\, n^{n+j}} \cancel{\mathrm{e}^{\mathrm{i}\frac{\pi}{2}}} \sqrt{-\frac{2\pi}{n\varphi''(z_-(x))}} \frac{\exp\left(n\varphi(z_-(x))\right)}{z_-(x)^{j+1}} (1 + O(n^{-1})) \\ &= \frac{2^{-j-n+\frac{1}{2}} e^{\frac{n}{4}\left(x^2-2-x\sqrt{x^2-4}\right)} \left(\sqrt{x^2-4}+x\right)^{j+n}}{\sqrt{4 - x^2 + x\sqrt{x^2-4}}} (1 + O(n^{-1})), \qquad x > 2, \end{aligned} \tag{6.2.12}$$

where we use the explicit formulas $z_-(x) = \frac{x-\sqrt{x^2-4}}{2}$, $\varphi(z) = xz - \frac{z^2}{2} - \log z$, and the following consequence of Stirling's asymptotics (5.2.23)

$$\frac{(n+j)!}{n^{n+j}} = \sqrt{2\pi(n+j)}\mathrm{e}^{-n}(1+O(n^{-1})) = \sqrt{2\pi n}\,\mathrm{e}^{-n}(1+O(n^{-1})), \qquad n\to+\infty,\ j \text{ fixed.} \tag{6.2.13}$$

We have assumed $x > 2$ in the last step of (6.2.12); for $x < -2$ we can either use a similar simplification, or the parity of Hermite polynomials.

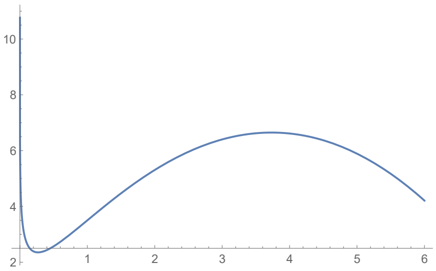


Figure 6.1: Function $\varphi(z) = xz - \frac{z^2}{2} - \log z$ for $z > 0$ (and $x = 4$); the points $z_\pm(x) = \frac{x-\sqrt{x^2-4}}{2}$ are the local minimum (z_-) and local maximum (z_+) for $z > 0$.

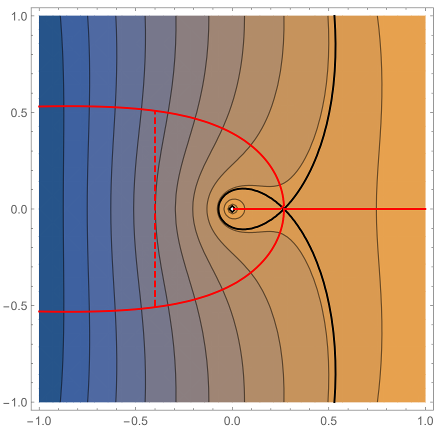


Figure 6.2: Contour deformation for the Hermite polynomials when $|x| > 2$ (here $x = 4$). The plot shows the values of $\operatorname{Re}\varphi(z)$ for z in the complex plane (for $-1 < \operatorname{Re} z, \operatorname{Im} z < 1$): blue = small values, orange = large values, and the thick black curve corresponds to the level curve $\operatorname{Re}\varphi(z) = \operatorname{Re}\varphi(z_-(x))$. The level curves $\operatorname{Im}\varphi(z) = \operatorname{Im}\varphi(z_-(x))$ are shown in red; the steepest descent one is the one that crosses the real axis perpendicularly. The contour of integration around $z = 0$ can be chosen as a part of this steepest descent curve closed by means of the dashed vertical line, whose exact shape is inessential as long as it lies entirely in a region where $\operatorname{Re}\varphi$ is strictly smaller than $\operatorname{Re}\varphi(z_-(x))$. The other saddle point $z_+(x) > z_-(x)$ is not shown in the picture.

Bulk, oscillatory regime: $|x| < 2$

In this case, the saddle points are complex conjugate, $z_\pm(x) = \frac{x\pm \mathrm{i}\sqrt{4-x^2}}{2}$; moreover, it is easily checked that $|z_\pm(x)| = 1$ hence it is convenient to write

$$z_\pm(x) = \mathrm{e}^{\pm\mathrm{i}\omega(x)}, \qquad \omega(x) = \arccos(x/2) \in (0, 2\pi), \qquad |x| < 2. \tag{6.2.14}$$

We have

$$\varphi(z_\pm(x)) = \frac{1}{4}\left(2 + x^2 \pm \mathrm{i}\left(x\sqrt{4-x^2} - 4\omega(x)\right)\right) \tag{6.2.15}$$

and in particular $\operatorname{Re}\varphi(z_-(x)) = \operatorname{Re}\varphi(z_+(x))$ hence in (5.3.12) we will need to take into account the contributions of both saddles; indeed, we can deform the contour of integration around $z = 0$ as depicted in Figure 6.4 so that it passes through both saddles.

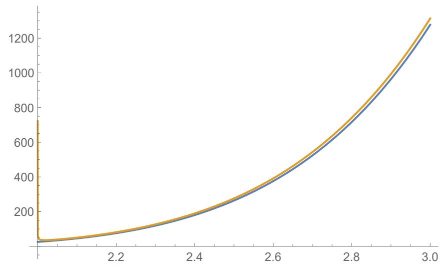


Figure 6.3: Approximation (in yellow) of $\widetilde{P}_{n+j}(x;n)$ (in blue) for $x > 2$, $n = 5$, $j = 2$. We see that the steepest descent method provides an extremely accurate approximation even for small n.

The directions $\theta_\pm(x)$ (that now depend on x) of steepest descent at the saddle points $z_\pm(x)$ are computed by

$$\varphi''(z_\pm(x))\mathrm{e}^{2\mathrm{i}\theta_\pm(x)} < 0 \tag{6.2.16}$$

and since

$$\varphi''(z_\pm(x)) = -1 + \frac{1}{z_\pm(x)^2} = -1 + \mathrm{e}^{\mp 2\mathrm{i}\omega(x)} = 2\sin(\omega(x))\mathrm{e}^{\mp\mathrm{i}\left(\frac{\pi}{2}+\omega(x)\right)} \tag{6.2.17}$$

we get (taking into account that we are encircling $z = 0$ in the positive direction)

$$\theta_-(x) = \frac{\pi}{4} - \frac{\omega(x)}{2}, \qquad \theta_+ = \pi - \theta_-. \tag{6.2.18}$$

We have all the ingredients to apply (5.3.12): using again (6.2.13) we obtain

$$\begin{aligned}
\widetilde{P}_{n+j}(x;n) &= \frac{\mathrm{e}^{-n}}{\mathrm{i}}\left(\mathrm{e}^{\mathrm{i}\theta_-(x)}\frac{\mathrm{e}^{n\varphi(z_-(x))}}{z_-(x)^{j+1}\sqrt{|\varphi''(z_-(x))|}} + \mathrm{e}^{\mathrm{i}\theta_+(x)}\frac{\mathrm{e}^{n\varphi(z_+(x))}}{z_+(x)^{j+1}\sqrt{|\varphi''(z_+(x))|}}\right)(1+O(n^{-1})) \\
&= 2\,\mathrm{Im}\left(\frac{\mathrm{e}^{\mathrm{i}\theta_-(x)}\mathrm{e}^{n(\varphi(z_-(x))-1)}}{z_-(x)^{j+1}\sqrt{|\varphi''(z_-(x))|}}\right)(1+O(n^{-1})) \\
&= \frac{\sqrt{2}}{\sqrt{\sin(\omega(x))}}\,\mathrm{Im}\left(\mathrm{e}^{\mathrm{i}(\frac{\pi}{4}+(j+\frac{1}{2})\omega(x))}\mathrm{e}^{\frac{n}{4}\left(2+x^2-\mathrm{i}x\sqrt{4-x^2}+4\mathrm{i}\omega(x)\right)}\right)(1+O(n^{-1})) \\
&= \frac{2}{(4-x^2)^{1/4}}\mathrm{e}^{\frac{n}{4}(x^2-2)}\sin\left(\frac{\pi}{4}+\left(n+j+\frac{1}{2}\right)\omega(x) - \frac{n}{4}x\sqrt{4-x^2}\right)(1+O(n^{-1})).
\end{aligned} \tag{6.2.19}$$

Summary

Theorem 6.2.6 (Plancherel–Rotach asymptotics for Hermite polynomials)**.** *As $n \to +\infty$, we have*

$$\widetilde{P}_{n+j}(x;n) = (1+O(n^{-1}))\times\begin{cases}\dfrac{2^{-j-n+\frac{1}{2}}e^{\frac{n}{4}\left(x^2-2-x\sqrt{x^2-4}\right)}\left(\sqrt{x^2-4}+x\right)^{j+n}}{\sqrt{4-x^2+x\sqrt{x^2-4}}}, & x > 2, \\ \dfrac{2\mathrm{e}^{\frac{n}{4}(x^2-2)}}{(4-x^2)^{1/4}}\sin\left(\frac{\pi}{4}+\left(n+j+\frac{1}{2}\right)\arccos\left(\frac{x}{2}\right)-\frac{n}{4}x\sqrt{4-x^2}\right), & |x| < 2.\end{cases} \tag{6.2.20}$$

Exercise 6.2.7. Show that

$$\frac{2^{-j-n+\frac{1}{2}}e^{\frac{n}{4}\left(x^2-2-x\sqrt{x^2-4}\right)}\left(\sqrt{x^2-4}+x\right)^{j+n}}{\sqrt{4-x^2+x\sqrt{x^2-4}}} \sim x^{n+j}, \qquad x\to\infty,\ n,j \text{ fixed.} \tag{6.2.21}$$

//

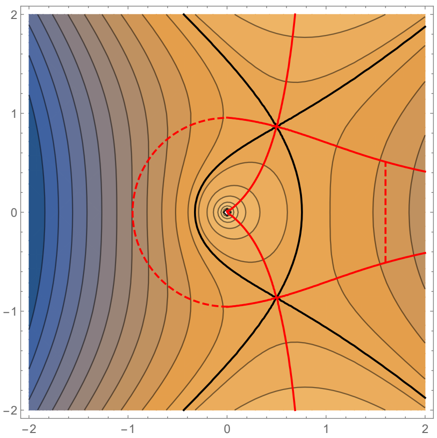


Figure 6.4: Contour deformation for the Hermite polynomials when $|x| < 2$ (here $x = 1$). The plot shows the values of $\operatorname{Re}\varphi(z)$ for z in the complex plane (for $-2 < \operatorname{Re} z, \operatorname{Im} z < 2$): blue = small values, orange = large values, and the thick black curves corresponds to the two level curves $\operatorname{Re}\varphi(z) = \operatorname{Re}\varphi(z_\pm(x))$. The level curves $\operatorname{Im}\varphi(z) = \operatorname{Im}\varphi(z_\pm(x))$ are shown in red; the steepest descent ones can be recognized by how they traverse the various level curves of the real part. The contour of integration around $z = 0$ can be chosen as parts of the steepest descent curves closed by means of the dashed curves, whose exact shape is inessential as long as they lie entirely in a region where $\operatorname{Re}\varphi$ is strictly smaller than $\operatorname{Re}\varphi(z_\pm(x))$.

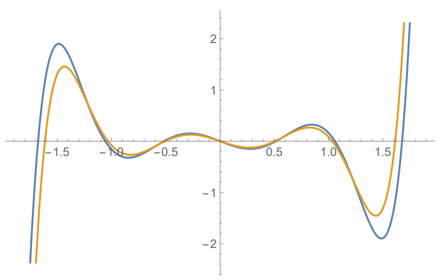


Figure 6.5: Approximation (in yellow) of $\widetilde{P}_{n+j}(x;n)$ (in blue) for $|x| < 2$, $n = 5$, $j = 2$.

Edge regime: $|x| = 2$

This regime is more complicated and hence omitted. When $x = \pm 2$, the two saddle points $z_\pm = \frac{x \pm \sqrt{x^2-4}}{2}$ coalesce into a single saddle point $z = 1$ which is degenerate, namely $\varphi''(1) = 0$ for $x = \pm 2$. The local structure of the level lines near the saddle point presents three principal directions; it is a so-called "monkey saddle" (Figure 6.6).

We report without proof the asymptotics for Hermite polynomials in this critical regime, which involve the Airy function:

$$\widetilde{P}_n(x;n)\mathrm{e}^{-n\frac{x^2}{4}}\Big|_{x=2+\frac{y}{n^{2/3}}} = \sqrt{2\pi}n^{1/6}\mathrm{e}^{-\frac{n}{2}}(\mathrm{Ai}(y) + O(n^{-2/3})). \tag{6.2.22}$$

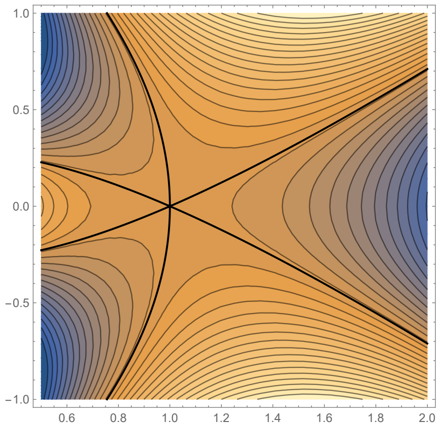


Figure 6.6: Level curves of $\mathrm{Re}\,\varphi(z)$ for $x = 2$ near the degenerate saddle point $z = 1$.

6.3 Applications to the GUE

6.3.1 Wigner's semicircle law

Theorem 6.3.1 (Wigner's semicircle law)**.** *As $n \to +\infty$ with $j \in \mathbb{Z}$ fixed, we have*

$$\lim_{n\to+\infty} \frac{1}{n}\rho_1(x) = \frac{\sqrt{4-x^2}}{2\pi}1_{(-2,2)}(x). \tag{6.3.1}$$

Proof. We have ($' := \partial_x$)

$$\frac{1}{n}\rho_1(x) = \frac{\mathrm{e}^{-n\frac{x^2}{2}}}{n\,h_{n-1}(n)}\left(\widetilde{P}_n'(x;n)\widetilde{P}_{n-1}(x;n) - \widetilde{P}_n(x;n)\widetilde{P}_{n-1}'(x;n)\right) \tag{6.3.2}$$

$$h_{n+j}(n) = (n+j)!n^{-n-j-\frac{1}{2}}\sqrt{2\pi} = 2\pi\mathrm{e}^{-n}(1+O(n^{-1})), \qquad n\to+\infty,\ j \text{ fixed}, \tag{6.3.3}$$

where the last asymptotic relation is obtained by (5.2.23). Since differentiating asymptotic relations poses some issues, we use the relation

$$\widetilde{P}_k'(x;n) = k\widetilde{P}_{k-1}(x;n), \tag{6.3.4}$$

which follows from Exercise 3.1.4, to write, using also (6.3.3),

$$\frac{1}{n}\rho_1(x) = \frac{\mathrm{e}^{-\frac{n}{2}(x^2-2)}}{2\pi}\left(\widetilde{P}_{n-1}(x;n)^2 - \frac{n-1}{n}\widetilde{P}_n(x;n)\widetilde{P}_{n-2}(x;n)\right)(1+O(n^{-1})). \tag{6.3.5}$$

- $x > 2$: using the simple estimate

$$\widetilde{P}_{n+j}(x;n) = O\left(\exp\left(\frac{n}{4}\left(x^2 - 2 - x\sqrt{x^2-4} + 4\log\frac{x+\sqrt{x^2-4}}{2}\right)\right)\right) \tag{6.3.6}$$

as $n \to +\infty$ with $j \in \mathbb{Z}$ and $x > 2$ fixed, which follows directly from (6.2.20), we obtain

$$\frac{1}{n}\rho_1(x) = O\left(\frac{1}{n}\exp\left(\frac{n}{2}\left(-x\sqrt{x^2-4} + 4\log\frac{x+\sqrt{x^2-4}}{2}\right)\right)\right), \qquad n \to +\infty, \tag{6.3.7}$$

and since $-x\sqrt{x^2-4} + 4\log\frac{x+\sqrt{x^2-4}}{2} < 0$ for all $x > 2$ we obtain $\frac{1}{n}\rho_1(x) \to 0$ as $n \to +\infty$.

- $x < -2$: by parity of Hermite polynomials we have $\rho_1(-x) = \rho_1(x)$ so it follows from the previous case that $\frac{1}{n}\rho_1(x) \to 0$ as $n \to +\infty$ for all $x < -2$ as well.

- $|x| < 2$: using (6.2.20) in the form

$$\mathrm{e}^{-\frac{n}{4}(x^2-2)}\widetilde{P}_{n+j}(x;n) = \frac{2}{(4-x^2)^{1/4}}\sin(n\alpha(x) + j\beta(x) + \gamma(x)) \tag{6.3.8}$$

with

$$\alpha(x) := \arccos\frac{x}{2} - \frac{x}{4}\sqrt{4-x^2}, \quad \beta(x) := \arccos\frac{x}{2}, \qquad \gamma(x) := \frac{\pi}{4} + \frac{1}{2}\arccos\frac{x}{2}, \tag{6.3.9}$$

we get

$$\begin{aligned}
&\mathrm{e}^{-\frac{n}{2}(x^2-2)}\widetilde{P}_{n+j_1}(x;n)\widetilde{P}_{n+j_2}(x;n) \\
&\quad = \frac{4}{\sqrt{4-x^2}}\sin\Big(n\alpha(x) + j_1\beta(x) + \gamma(x)\Big)\sin\Big(n\alpha(x) + j_2\beta(x) + \gamma(x)\Big)(1 + O(n^{-1})) \\
&\quad = \frac{2}{\sqrt{4-x^2}}\left[\cos\Big((j_1 - j_2)\beta(x)\Big) - \cos\Big(2n\alpha(x) + (j_1+j_2)\beta(x) + 2\gamma(x)\Big)\right](1 + O(n^{-1}))
\end{aligned} \tag{6.3.10}$$

using the prosthaphaeresis identity $\sin(u)\sin(v) = \frac{1}{2}(\cos(u-v) - \cos(u+v))$. Next, we observe that $\mathrm{e}^{-\frac{n}{2}(x^2-2)}\widetilde{P}_{n+j_1}(x;n)\widetilde{P}_{n+j_2}(x;n) = O(1)$ as $n \to +\infty$ so that we can replace the factor $\frac{n-1}{n}$ in (6.3.5) by 1, absorbing the difference in the $O(n^{-1})$ error, to get

$$\begin{aligned}
\frac{1}{n}\rho_1(x) &= \frac{2}{\sqrt{4-x^2}}\Big[\cos(0) - \cancel{\cos\Big(2n\alpha(x) - 2\beta(x) + 2\gamma(x)\Big)} \\
&\qquad\qquad - \cos(2\beta(x)) + \cancel{\cos\Big(2n\alpha(x) - 2\beta(x) + 2\gamma(x)\Big)}\Big](1 + O(n^{-1})) \\
&= \frac{1 - \cos(2\beta(x))}{\pi\sqrt{4-x^2}}(1 + O(n^{-1})) \\
&= \frac{\sqrt{4-x^2}}{2\pi}(1 + O(n^{-1}))
\end{aligned} \tag{6.3.11}$$

using the trigonometric identity $\cos(2\beta(x)) = 2\cos(\beta(x))^2 - 1 = \frac{x^2}{2} - 1$.

□

Corollary 6.3.2. *For any continuous $f : \mathbb{R} \to \mathbb{R}$ we have*

$$\lim_{n \to +\infty} \mathbb{E}\left[\frac{1}{n}\mathrm{tr}\,(f(M))\right] = \int_{-2}^{2} f(s)\frac{\sqrt{4-s^2}}{2\pi}\mathrm{d}s. \tag{6.3.12}$$

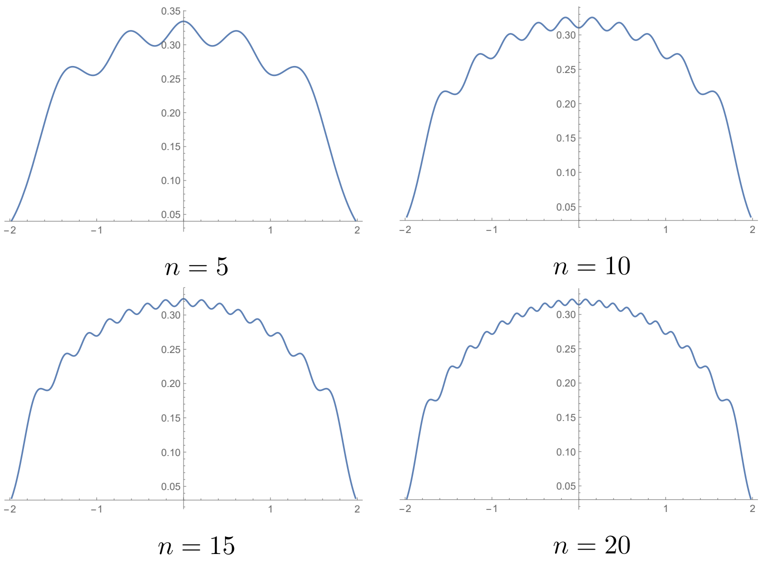


Figure 6.7: Plot of $\frac{1}{n}K_n^{\mathsf{GUE}}(x,x)$ for various values of n.

6.3.2 Some remarks on the semicircle law for Wigner matrices

A different model of random matrix theory is that of Wigner matrices. Informally, invariant ensembles (which are the primary focus of these notes) are characterized by invariance with respect to the unitary adjoint action, whereas Wigner ensembles are characterized by independence of the entries. It is not too hard to show that the Gaussian ensembles are the only random matrix ensembles which are simultaneously Wigner and invariant (but we will not prove it here).

Definition 6.3.3. Let Y be a real-valued random variable, and let Z be a complex-valued random variable. We assume that

$$\mathbb{E}[Y] = 0 = \mathbb{E}[Z], \qquad \mathbb{E}[Z^2] = 1, \tag{6.3.13}$$

and moreover that

$$\mathbb{E}[|Y|^k] \text{ and } \mathbb{E}[|Z|^k] \text{ are finite for all } k. \tag{6.3.14}$$

A(n Hermitian) **Wigner matrix** is an Hermitian random matrix M with upper triangular and diagonal entries independent and distributed as

$$M_{ii} \sim Y,\ 1 \le i \le n, \qquad M_{ij} = \overline{M_{ji}} \sim Z,\ 1 \le i < j \le n. \tag{6.3.15}$$

For an Hermitian Wigner matrix M of size n, the **empirical spectral measure** is

$$\mu_n := \frac{1}{n} \sum_{\lambda \text{ eigenvalue of } M} \delta_{\lambda/\sqrt{n}}, \tag{6.3.16}$$

which is a random probability measure on $\mathbb{R}$. //

These conditions could be relaxed in several directions, but we will stick to this simple setting here to formulate the next result. Note that a GUE matrix is an Hermitian Wigner matrix.

Theorem 6.3.4 (Wigner)**.** *Let M be an Hermitian Wigner matrix of size n, in the sense of the above definition and let μ_n be the associated empirical spectral measure. Then we have the weak convergence in probability*

$$\mu_n \to \frac{\sqrt{4-x^2}}{2\pi} 1_{(-2,2)}(x)\mathrm{d}x \tag{6.3.17}$$

which more explicitly means that for all $f:\mathbb{R}\to\mathbb{R}$ bounded continuous function and for all $\epsilon>0$ we have

$$\mathbb{P}\left(\left|\int_{\mathbb{R}} f(x)\mu_n(\mathrm{d}x) - \int_{-2}^{2} f(x)\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x\right| > \epsilon\right) \to 0, \text{ as } n\to+\infty. \tag{6.3.18}$$

6.3.3 Sine kernel in the bulk

Theorem 6.3.5. *Fix $x_0\in(-2,2)$ and denote*

$$\psi(x_0) := \frac{\sqrt{4-x_0^2}}{2\pi}, \tag{6.3.19}$$

which is the limiting eigenvalue density by Theorem 6.3.1. Then the GUE correlation kernel defined in (6.2.9) satisfies

$$\lim_{n\to+\infty}\frac{1}{n\psi(x_0)}K_n^{\mathsf{GUE}}\left(x_0+\frac{\xi}{n\psi(x_0)}, x_0+\frac{\eta}{n\psi(x_0)}\right) = K^{\mathsf{sine}}(\xi,\eta) := \frac{\sin(\pi(\xi-\eta))}{\pi(\xi-\eta)}. \tag{6.3.20}$$

Proof. Throughout this proof, $x := x_0+\frac{\xi}{n\psi(x_0)}$, $y := x_0+\frac{\eta}{n\psi(x_0)}$ and the O-notation is always taken in the limit $n\to+\infty$. In particular

$$\frac{1}{n\psi(x_0)(x-y)} = \frac{1}{\xi-\eta}, \tag{6.3.21}$$

and x,y satisfy $|x|<2, |y|<2$, provided n is sufficiently large. Therefore, with $\alpha(x),\beta(x),\gamma(x)$ as in (6.3.8)–(6.3.9), using also (6.3.3),

$$\begin{aligned}\frac{K_n^{\mathsf{GUE}}(x,y)}{n\psi(x_0)} &= \frac{\mathrm{e}^{-\frac{n}{4}(x^2+y^2)}}{h_{n-1}(n)\,(\xi-\eta)\,(4-x)^{1/4}\,(4-y)^{1/4}}\left[\widetilde{P}_n(x;n)\widetilde{P}_{n-1}(y;n)-\widetilde{P}_{n-1}(x;n)\widetilde{P}_n(y;n)\right]\\ &= \frac{2}{\pi(\xi-\eta)\sqrt{4-x_0^2}}\Bigg[\sin\Big(n\alpha(x)+\gamma(x)\Big)\sin\Big(n\alpha(y)-\beta(y)+\gamma(y)\Big)\\ &\qquad -\sin\Big(n\alpha(x)-\beta(x)+\gamma(x)\Big)\sin\Big(n\alpha(y)+\gamma(y)\Big)\Bigg](1+O(n^{-1}))\\ &= \frac{1}{\pi(\xi-\eta)\sqrt{4-x_0^2}}\Bigg[\cos\Big(n(\alpha(x)-\alpha(y))+\beta(y)+\gamma(x)-\gamma(y)\Big)\\ &\qquad -\cos\Big(n(\alpha(x)+\alpha(y))-\beta(y)+\gamma(x)+\gamma(y)\Big)\\ &\qquad -\cos\Big(n(\alpha(y)-\alpha(x))+\beta(x)+\gamma(y)-\gamma(x)\Big)\\ &\qquad +\cos\Big(n(\alpha(x)+\alpha(y))-\beta(x)+\gamma(x)+\gamma(y)\Big)\Bigg](1+O(n^{-1})),\end{aligned} \tag{6.3.22}$$

where we use again the prosthaphaeresis identity $\sin(u)\sin(v) = \frac{1}{2}(\cos(u-v)-\cos(u+v))$. In the last expression, the arguments of cos in the second and fourth lines differ by $\beta(x)-\beta(y) = O(n^{-1})$ so we can absorb these terms into the error $O(n^{-1})$; for the remaining two terms, we compute

$$\begin{aligned}\cos\Big(n(\alpha(x)-\alpha(y))+\beta(x)+\gamma(x)-\gamma(y)\Big)-\cos\Big(n(\alpha(y)-\alpha(x))+\beta(y)+\gamma(y)-\gamma(x)\Big)\\ =2\sin\left(\frac{\beta(x)+\beta(y)}{2}\right)\sin\left(n(\alpha(y)-\alpha(x))+\frac{\beta(y)-\beta(x)}{2}+\gamma(y)-\gamma(x)\right)\end{aligned} \tag{6.3.23}$$

using the inverse prosthaphaeresis identity $\cos(u)-\cos(v)=2\sin(\frac{u+v}{2})\sin(\frac{v-u}{2})$. Finally, the first factor is

$$\sin(\beta(x_0)+O(n^{-1}))=\sin\left(\arccos\left(\frac{x_0}{2}\right)\right)(1+O(n^{-1}))=\sqrt{1-\left(\frac{x_0}{2}\right)^2}\,(1+O(n^{-1})) \tag{6.3.24}$$

the second one is

$$\sin\left(\alpha'(x_0)\frac{\eta-\xi}{\psi(x_0)}+O(n^{-1})\right)=\sin(\pi(\xi-\eta))(1+O(n^{-1})), \tag{6.3.25}$$

where we use the elementary computation $\alpha'(x)=-\frac{1}{2}\sqrt{4-x^2}$, for $|x|<2$. By assembling all these terms together, we finally obtain

$$\frac{K_n^{\mathsf{GUE}}(x,y)}{n\psi(x_0)}=\frac{\sin(\pi(\xi-\eta))}{\pi(\xi-\eta)}(1+O(n^{-1})), \tag{6.3.26}$$

which completes the proof. □

6.3.4 Airy kernel at the edge

Theorem 6.3.6. *The GUE correlation kernel defined in* (6.2.9) *satisfies*

$$\lim_{n\to+\infty} n^{-2/3}K_n^{\mathsf{GUE}}\left(2+\frac{\xi}{n^{2/3}},2+\frac{\eta}{n^{2/3}}\right)=K^{\mathsf{Airy}}(\xi,\eta):=\frac{\mathrm{Ai}(\xi)\mathrm{Ai}'(\eta)-\mathrm{Ai}'(\xi)\mathrm{Ai}(\eta)}{\xi-\eta}. \tag{6.3.27}$$

The proof relies on (6.2.22) and is omitted.

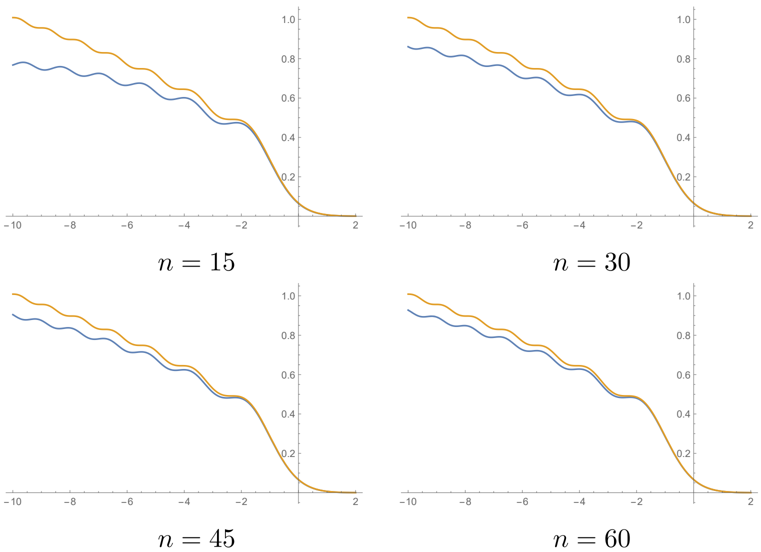


Figure 6.8: Plot of $n^{-2/3}K_n^{\mathsf{GUE}}(2+n^{-2/3}x,2+n^{-2/3}x)$ (blue) for various values of n and of the one-point correlation function of the Airy point process $K^{\mathsf{Airy}}(x,x)=\mathrm{Ai}'(x)^2-x\mathrm{Ai}(x)^2$ (yellow).

6.3.5 Airy process

Exercise 6.3.7. Prove that

$$K^{\mathsf{Airy}}(\xi,\eta) := \frac{\mathrm{Ai}(\xi)\mathrm{Ai}'(\eta)-\mathrm{Ai}'(\xi)\mathrm{Ai}(\eta)}{\xi-\eta} = \int_0^{+\infty}\mathrm{Ai}(\xi+t)\mathrm{Ai}(\eta+t)\mathrm{d}t. \tag{6.3.28}$$

Note that the integral in the right side is absolutely convergent by Proposition 5.4.8. (Hint: use the Airy equation and integrate by parts the expression $(\xi-\eta)\int_0^{+\infty}\mathrm{Ai}(\xi+t)\mathrm{Ai}(\eta+t)\mathrm{d}t$.) //

As a matter of fact (we will not prove it here) the operator $\mathcal{A}$ defined for compactly supported smooth functions $f:\mathbb{R}\to\mathbb{R}$ as

$$(\mathcal{A}f)(x) = \int_{\mathbb{R}}\mathrm{Ai}(x+t)f(t)\mathrm{d}t \tag{6.3.29}$$

extends to a unitary involution of $L^2(\mathbb{R})$:

$$\mathcal{A}=\mathcal{A}^\dagger, \qquad \mathcal{A}^2 = Id_{L^2(\mathbb{R})}. \tag{6.3.30}$$

By the above exercise, the operator $\mathcal{K}^{\mathsf{Airy}}$ acting on $L^2(\mathbb{R})$ as

$$(\mathcal{K}^{\mathsf{Airy}}f)(x) = \int_{\mathbb{R}} K^{\mathrm{Ai}}(x,y)f(y)\mathrm{d}y \tag{6.3.31}$$

can be expressed as

$$\mathcal{K}^{\mathsf{Airy}} = \mathcal{A}1_{(0,+\infty)}\mathcal{A}. \tag{6.3.32}$$

It follows that $\mathcal{K}^{\mathsf{Airy}}$ is an orthogonal projector, $(\mathcal{K}^{\mathsf{Airy}})^2=\mathcal{K}^{\mathsf{Airy}}$; it is also verified that $1_B\mathcal{A}1_{(0,+\infty)}$ is Hilbert–Schmidt for all bounded Borel subsets B of $\mathbb{R}$ (this follows from the asymptotics of $\mathrm{Ai}(x)$ as $x\to+\infty$, proven in Proposition 5.4.8), hence $\mathcal{K}^{\mathsf{Airy}}$ is locally trace-class. By Theorem 4.2.3 and Remark 4.2.4 we conclude that there exists a determinantal point process with correlation kernel $K^{\mathsf{Airy}}(\cdot,\cdot)$.

Definition 6.3.8. This determinantal point process is called **Airy point process**. //

6.3.6 Largest eigenvalue fluctuations and Tracy–Widom distribution

Let $\lambda_{\mathsf{max}}^{(n)}$ be the largest eigenvalue of a GUE matrix of size n.

Theorem 6.3.9. *The following limit exists for all $s\in\mathbb{R}$*

$$F(s) = \lim_{n\to+\infty}\mathbb{P}((\lambda_{\mathsf{max}}^{(n)}-2)n^{2/3}\le s) \tag{6.3.33}$$

and is given by the Fredholm series

$$F(s) = 1+\sum_{\ell\ge1}\frac{(-1)^\ell}{\ell!}\int_{(s,+\infty)^\ell}\det_{1\le i,j\le\ell}\left(K^{\mathsf{Airy}}(x_i,x_j)\right)\mathrm{d}x_1\cdots\mathrm{d}x_\ell. \tag{6.3.34}$$

Proof. Using notations and results from Section 4.4, in particular Theorem 4.4.1, we know that

$$\begin{aligned}
\mathbb{P}((\lambda_{\mathsf{max}}^{(n)}-2)n^{2/3}\le s) &= \mathbb{P}(\lambda_{\mathsf{max}}^{(n)}\le 2+\frac{s}{n^{2/3}}) \\
&= \mathbb{P}(\#_{(2+\frac{s}{n^{2/3}},+\infty)}=0) \\
&= 1+\sum_{\ell\ge1}\frac{(-1)^\ell}{\ell!}\int_{(2+\frac{s}{n^{2/3}},+\infty)^\ell}\det_{1\le i,j\le\ell}\left(K_n^{\mathsf{GUE}}(x_i,x_j)\right)\mathrm{d}x_1\cdots\mathrm{d}x_\ell \\
&= 1+\sum_{\ell\ge1}\frac{(-1)^\ell}{\ell!}\int_{(s,+\infty)^\ell}\det_{1\le i,j\le\ell}\left(\frac{1}{n^{2/3}}K_n^{\mathsf{GUE}}\left(2+\frac{y_i}{n^{2/3}},2+\frac{y_j}{n^{2/3}}\right)\right)\mathrm{d}y_1\cdots\mathrm{d}y_\ell
\end{aligned} \tag{6.3.35}$$

where in the last step we change variables $x_i = 2 + \frac{y_i}{n^{2/3}}$. It finally suffices to apply Theorem 6.3.6 (with some details that we omit in passing the limit $n \to +\infty$ inside the integral). $\square$

See Figure 6.9. An important result on this distribution is the following caracterization

Theorem 6.3.10 (Tracy & Widom, 1994). *We have*

$$F(s) = \exp\left(-\int_s^{+\infty}(x-s)q(x)^2\mathrm{d}x\right) \tag{6.3.36}$$

where $q(s)$ is the unique solution to the boundary value problem

$$\begin{cases} \frac{\mathrm{d}^2}{\mathrm{d}s^2}q(s) = sq(s) + 2q(s)^3 & \textit{(“Painlevé II equation”)} \\ q(s) \sim \mathrm{Ai}(s), \quad \textit{as } s \to +\infty. \end{cases} \tag{6.3.37}$$

It is also relevant to state the following important result that was proven independently before the Tracy–Widom Theorem.

Theorem 6.3.11 (Hastings & McLeod, 1980). *There exists a unique solution to the boundary value problem* (6.3.37)*.*

The relevance of Theorem 6.3.10 stems from the “universality” of the Tracy–Widom distribution $F(s)$. It indeed appears to be a universal distribution describing many different models at the transition between weakly and strongly coupled phases of a systems[2]. It appears for example (beyond many instances in Random Matrix Theory) in the study of

- the longest increasing subsequence of a random permutation,
- asymmetric simple exclusion process,
- the Kardar–Parisi–Zhang equation.

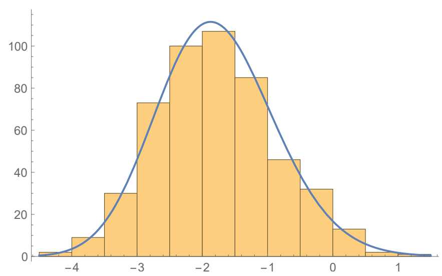


Figure 6.9: Histogram of 500 samples of the random variable $(\lambda_{\mathsf{max}}^{(n)} - 2)n^{2/3}$ for $n = 20$ plotted against the Tracy–Widom probability distribution function $F'(s)$ (suitably rescaled so to have the same total mass as the histogram). The probability distribution function $F'(s)$ has a bell-shaped curve, with a negative mean value approximately equal to -1.77109 and a fatter tail to the right-hand side of the mean than to the left-hand side (in statistical jargon, it has positive *skewness*).

[2]For a divulgative introduction see: Wolchover, “At the Far Ends of a New Universal Law”, in *Quanta Magazine*, October 2014, publicly available at `https://www.quantamagazine.org/beyond-the-bell-curve-a-new-universal-law-20141015/`.

Chapter 7

Riemann–Hilbert problems

7.1 Introduction and basic theory

Riemann–Hilbert (from now on, RH) problems are an analytic way of characterizing interesting quantities: here we shall be concerned only with Orthogonal Polynomials on the real line and their applications to Random Matrix Theory, but we should mention also

- more general types of orthogonality (e.g. different contours, multi-orthogonality)
- Hankel and Töplitz determinants,
- Fredholm determinants,
- "integrable" partial differential equations (one of the the prototypical examples being the Korteweg–de Vries equation, which models waves in shallow water).

In a sense (that we shall explore further below), RH problems could be seen as a 'non-abelian' generalization of contour integral representations. It should be noted from the beginning that one is rarely able to solve explicitly a RH problem; rather, the approach is to exploit the analytic characterization of certain quantities (for us, the OPRL) in terms of the unique solution to a RH problem to extract information about these quantities. This is parallel to how we generally use contour integral representations, which are often not explicitly computable in terms of elementary functions, but still useful to extract information (e.g. asymptotics, symmetries) — the study of the Airy and Hermite functions carried out in previous chapters should be a guidance here.

We do not aim at a general exhaustive discussion (which would greatly exceed the goals of the course) but rather at giving an introduction with focus on applications to OPRL and RMT.

7.1.1 Scalar additive RH problems: simple closed contour case

A RH problem requires to find solutions to certain conditions:

RH1 Analyticity;

RH2 Jump condition;

RH3 Normalization.

A first simple (but important and pedagogical) example is given by *scalar additive* RH problems. The data of the problem are the following.

- A smooth oriented contour $\gamma \subset \mathbb{C}$, which for now we assume to be simple (i.e. it has no self-intersections) and closed. We introduce the convention (which is customary and will be used throughout all this chapter) to denote the two sides of γ with signs $+$ (left) and $-$ (right) (see Figure 7.1).
- A function $m : \gamma \to \mathbb{C}$ which is sufficiently regular (more details below when we construct the solution).

The solution is a function $y = y(z)$ of a single variable satisfying the following conditions.

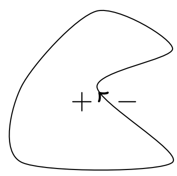


Figure 7.1: The simple closed smooth loop γ in the complex plane. The orientation of γ induces a choice of $\pm$ sides of γ by the rule: $+$ =left, $-$ =right.

Scalar additive RH problem

RH1 $y(z)$ is analytic for $z \in \mathbb{C} \setminus \gamma$.

RH2 The **boundary values** $y_\pm(z) := \lim_{w \to z_\pm} y(w)$ exist for all $z \in \gamma$, are continuous on γ, and satisfy the relation

$$y_+(z) = y_-(z) + m(z), \quad z \in \gamma. \tag{7.1.1}$$

Here $\lim_{w \to z_\pm}$ for a point $z \in \gamma$ means that the limit is taken along any (smooth) curve ending in z and lying entirely to the $\pm$ side of γ in a neighborhood of z and should be independent of the curve (and thus depend only on the side $\pm$).

RH3 As $z \to \infty$ in the complex plane, we have $y(z) \to 0$.

The following result could be formulated in more generality, for instance relaxing the regularity of γ and of m (and, consequently, also of meaning of the boundary values for the solution y); the most notable cases are when m is only assumed to be Hölder continuous, or even just in $L^2(\gamma)$.

Proposition 7.1.1. *Assume, in addition to the previous conditions, that there exists an open neighborhood $\mathcal{N}_\gamma$ of $\gamma \subset \mathbb{C}$ such that m admits an analytic extension to $\mathcal{N}_\gamma$. Then, the solution $y(z)$ is unique and is given by the formula*

$$y(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma m(s) \frac{\mathrm{d}s}{s - z}. \tag{7.1.2}$$

Definition 7.1.2. An integral of the form (7.1.2) is called a **Cauchy integral**. //

Proof. RH1 means that $y(z)$ is holomorphic for all $z \notin \gamma$; this holds because $y(z)$ is complex-differentiable:

$$y'(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma m(s) \frac{\mathrm{d}s}{(s - z)^2}. \tag{7.1.3}$$

(Differentiation under integral sign is justified because the integral (7.1.3) converges uniformly for z at a finite distance from γ.)

The jump condition RH2 is proved by a contour deformation argument. Fix a point $z \in \gamma$ and refer to Figure 7.2; introduce the contours $\widetilde{\gamma}_\pm$ which coincide with γ outside a sufficiently small neighborhood of $z \in \gamma$, and within such a small neighborhood of z they are given by a deformation which leaves z inside the $\pm$ region and which anyway is small enough so that we never exit the neighborhood $\mathcal{N}_\gamma$. By Cauchy Theorem we can compute the boundary values as

$$2\pi \mathrm{i}\, y_\pm(z) = \lim_{w \to z_\pm} \int_\gamma m(s) \frac{\mathrm{d}s}{s - w} = \lim_{w \to z_\pm} \int_{\widetilde{\gamma}_\pm} m(s) \frac{\mathrm{d}s}{s - w} = \int_{\widetilde{\gamma}_\pm} m(s) \frac{\mathrm{d}s}{s - z}. \tag{7.1.4}$$

Since $\widetilde{\gamma}_+ - \widetilde{\gamma}_-$ is homotopically equivalent to a small circle positively oriented around z (Figure 7.3), by Cauchy theorem we get

$$y_+(z) - y_-(z) = \frac{1}{2\pi \mathrm{i}} \int_{\widetilde{\gamma}_+ - \widetilde{\gamma}_-} m(s) \frac{\mathrm{d}s}{s - z} = \operatorname{Res}\left(m(s) \frac{\mathrm{d}s}{s - z}, s = z \right) = m(z). \tag{7.1.5}$$

The normalization condition RH3 follows by the simple estimate

$$|y(z)| \leq \frac{\max_\gamma |m|}{\text{dist}\,(z,\gamma)}\,, \tag{7.1.6}$$

and the last expression tends to 0 as $z \to \infty$.

Finally, uniqueness is proven as follows; let $\widetilde{y} = \widetilde{y}(z)$ be another solution and consider $\epsilon(z) := y(z) - \widetilde{y}(z)$. Since $\delta_\pm(z)$ exist and $\delta_+(z) = \delta_-(z)$ for all $z \in \gamma$ and since m is analytic in $\mathcal{N}_\gamma$, we conclude that $\epsilon(z)$ is an entire function of z; since $\epsilon(z) \to 0$ as $z \to \infty$, we conclude by Liouville Theorem that $\epsilon(z) = 0$ identically in z. □

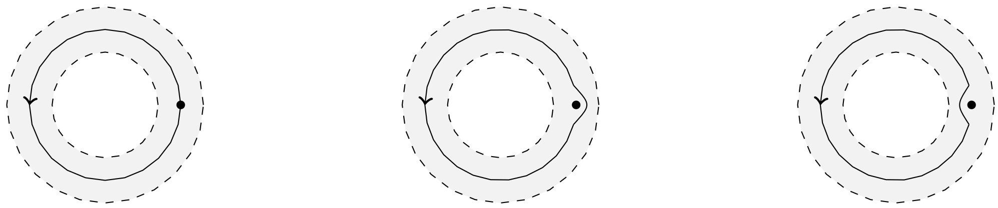


Figure 7.2: In grey the neighborhood $\mathcal{N}_\gamma$. In the first figure, the original contour γ and some point $z \in \gamma$. In the second figure, the deformation $\widetilde{\gamma}_+$ used to compute the boundary value $y_+(z)$. In the third figure, the deformation $\widetilde{\gamma}_-$ used to compute the boundary value $y_-(z)$.

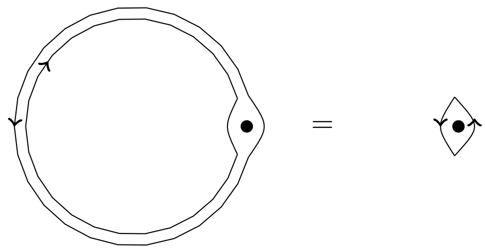


Figure 7.3: The integration contour $\widetilde{\gamma}_+ - \widetilde{\gamma}_-$ is homotopically equivalent to a small circle around z (with index +1).

The following exercise can be of help in understanding Cauchy integrals.

Exercise 7.1.3. In this setting, γ is a smooth simple closed loop, and hence it divides the plane into two disjoint regions Ω_+, Ω_- (interior and exterior, respectively). More precisely: Ω_+ and Ω_- are open, $\overline{\Omega_+}$ is compact, $\Omega_- \cap \Omega_+ = \emptyset$, $\overline{\Omega_+} \cap \overline{\Omega_-} = \gamma$, and $\Omega_+ \cup \Omega_- = \mathbb{C} \setminus \gamma$.

1. Let m admit an analytic extension to an open neighborhood of $\overline{\Omega_+}$. Find the solution to the additive RH problem.
2. Assume m admits a meromorphic extension to an open neighborhood of $\overline{\Omega_+}$, whose only singularity is a simple pole at $z_0 \in \Omega_+$. Find the solution to the additive RH problem.
3. Generalize to the case of an arbitrary meromorphic function m in an open neighborhood of Ω_+.

(Hint: use (7.1.2) and Cauchy Theorem.) //

7.1.2 Scalar additive RH problems: open contour case

Now, let γ be a smooth simple contour which is not closed (rather it has endpoints, which may be finite or infinite); see Figure 7.4. We stipulate that the endpoints are part of γ, and denote by γ° the interior of γ. The jump function m can be defined just on γ°; however, for simplicity we assume that $m(z) \to 0$ as $z \to a$ faster than any power of $(z - a)$, for any endpoint a of γ.

The scalar additive RH problem is in this case formulated as

RH1 $y(z)$ is analytic for $z \in \mathbb{C} \setminus \gamma$.

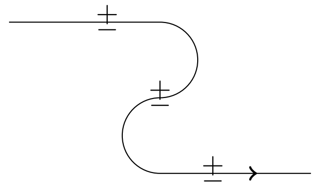


Figure 7.4: The simple smooth curve γ in the complex plane. The orientation of γ induces a choice of $\pm$ sides of γ by the rule: $+$ =left, $-$ =right. Contrarily to the closed loop case, the curve γ does not divide the complex plane into disjoint open components and $\pm$ acquire meaning only in a sufficiently close neighborhood of γ.

RH2 The limits $y_\pm(z) := \lim_{w\to z_\pm} y(w)$ exist for all $z \in \gamma^\circ$, are continuous on γ, and satisfy the relation

$$y_+(z) = y_-(z) + m(z), \quad z \in \gamma^\circ. \tag{7.1.7}$$

RH3 As $z \to \infty$ in the complex plane, we have $y(z) \to 0$.

It can be checked exactly as in Proposition 7.1.1, that a solution is again given by the formula (7.1.2). However, now the solution is not unique: for example, we are free to add to any solution $y(z)$ a multiple of $(z-a)^{-p}$ for any integer $p \geq 1$ and any endpoint a. (**Exercise**: check that this does not spoil any of the conditions RH1–RH3). To restore uniqueness of the solution, one has to amend RH3:

RH3' As $z \to \infty$ in the complex plane, we have $y(z) \to 0$, and as $z \to$ any endpoint of γ, we have $y(z) = O(1)$.

It can be checked that $y(z)$ given by the formula (7.1.2) is then the unique solution of RH1, RH2, RH3'; the arguments in the proof of Proposition 7.1.1 must be complemented by the analysis of the Cauchy integral as z approaches the endpoints of γ. We omit the details.

Exercise 7.1.4. Solve the scalar additive RH problem on $\gamma = \mathbb{R}$ with $m(z) = \frac{1}{z^2+1}$. (Hint: close the contour of integration $\mathbb{R}$ with a large semicircle in either half-plane.) //

7.1.3 Scalar multiplicative case

We discuss the closed loop case only for simplicity (generalizing to the open contour case could be done essentially as in the previous paragraph).

A scalar multiplicative RH problem is given by the following conditions.

RH1 $y(z)$ is analytic for $z \in \mathbb{C} \setminus \gamma$.

RH2 The limits $y_\pm(z) := \lim_{w\to z_\pm} y(w)$ exist for all $z \in \gamma$, are continuous on γ, and satisfy the relation

$$y_+(z) = y_-(z)m(z), \quad z \in \gamma. \tag{7.1.8}$$

RH3 As $z \to \infty$ in the complex plane, we have $y(z) \to 1$.

Proposition 7.1.5. *Assume, in addition to the previous conditions, that there exists an open neighborhood $\mathcal{N}_\gamma$ of $\gamma \subset \mathbb{C}$ such that m admits an analytic extension to $\mathcal{N}_\gamma$ and also that there exists a logarithm $\log m(z)$ defined for $z \in \gamma$ in such a way that it is* continuous *on γ. Then the scalar multiplicative RH problem admits a unique solution*

$$y(z) = \exp\left(\frac{1}{2\pi \mathrm{i}} \int_\gamma \log(m(s)) \frac{\mathrm{d}s}{s-z}\right). \tag{7.1.9}$$

Proof. We first check that $y(z)$ defined in (7.1.9) is a solution. Indeed,

$$\frac{y_+(z)}{y_-(z)} = \exp\Big(\varphi_+(z) - \varphi_-(z)\Big) \tag{7.1.10}$$

where

$$\varphi(z) = \frac{1}{2\pi \mathrm{i}} \int_\gamma \log(m(s)) \frac{\mathrm{d}s}{s - z}. \tag{7.1.11}$$

We know from Proposition 7.1.1 that

$$\varphi_+(z) - \varphi_-(z) = \log(m(z)) \tag{7.1.12}$$

and so (7.1.9) solves RH2. Since $y(z) = \exp(\varphi(z))$, conditions RH1 and RH3 are also easily checked (the former follows from the fact that φ is analytic for $z \in \mathbb{C} \setminus \gamma$, and the latter follows from $\varphi(z) \to 0$ as $z \to \infty$).

Uniqueness can be proven as follows; let $\widetilde{y}(z)$ be another solution and consider the ratio $\epsilon(z) := \widetilde{y}(z)/y(z)$. Since $\epsilon_\pm(z)$ exist for all $z \in \gamma$ and $\epsilon_+(z) = \epsilon_-(z)$, and since m has an analytic extension to a neighborhood of γ (as well as $\log m$), we deduce that $\epsilon(z)$ is an entire function of z. (Note that, by construction, $y(z)$ has no zeros.) Since $\epsilon(z) \to 1$ as $z \to \infty$, by Liouville Theorem we have $\epsilon(z) = 1$ identically in z. □

The condition that $m(z)$ admits a continuous logarithm is essential here; if the contour is closed, it is equivalent to the *index* condition

$$0 = \oint_\gamma \mathrm{d} \log m(s) \mathrm{d}s = \oint_\gamma \frac{m'(s)}{m(s)} \mathrm{d}s. \tag{7.1.13}$$

Exercise 7.1.6. Suppose, as in the previous proposition, that m admits a continuous logarithm on γ, and let y solve the above scalar multiplicative RH problem. Prove uniqueness and give an integral representation of the solution to the *in-homogeneous* scalar RH problem (for a given function f, also analytic in a neighborhood of γ):

RH1 $\phi(z)$ analytic for $z \in \mathbb{C} \setminus \gamma$;

RH2 boundary values $\phi_\pm(z)$ exist and $\phi_+(z) = \phi_-(z)m(z) + f(z)$, for all $z \in \gamma$;

RH3 $\phi(z) \to 0$ as $z \to \infty$.

(Hint: about uniqueness, given two solutions $\phi, \widetilde{\phi}$ consider the function $(\phi(z) - \widetilde{\phi}(z))/y(z)$ and show that it is entire and use Liouville Theorem; about the integral representation, write RH2 as $\frac{\phi_+(z)}{y_+(z)} = \frac{\phi_-(z)}{y_-(z)} + \frac{f(z)}{y_+(z)}$ and apply the theory relative to scalar additive RH problems.) //

7.2 General theory of matrix RH problems

7.2.1 Formulation of the RH problem

In the simplified setting which is enough for these notes, a ($k \times k$-)matrix RH problem consists of the following data.

- A smooth simple contour γ in the complex plane. It might be closed or open; in the former case note that $\gamma^\circ = \gamma$.
- A matrix function $M : \gamma \to \mathrm{SL}_k(\mathbb{C})$, where $\mathrm{SL}_k(\mathbb{C})$ is the *special linear group* consisting of $k \times k$ matrix with complex entries and unit determinant. If γ has endpoints, we require that $M(z) \to \mathbf{1}_k$ as $z \to a$ faster than any power of $z - a$, for any endpoint $a \in \gamma$. We shall also always assume (for simplicity) that $M(z)$ admits an analytic extension to a neighborhood of γ.

The solution to the RH problem specified by the above data is a $k \times k$ matrix function $Y = Y(z)$ satisfying the following conditions.

RH1 $Y(z)$ is analytic (entry-wise) for $z \in \mathbb{C} \setminus \gamma$.

RH2 The limits $Y_\pm(z) := \lim_{w \to z_\pm} Y(w)$ exist for all $z \in \gamma^\circ$, are continuous on γ°, and satisfy the relation

$$Y_+(z) = Y_-(z)M(z), \quad z \in \gamma^\circ. \tag{7.2.1}$$

RH3 As $z \to \infty$ in the complex plane, we have $Y(z) \to \mathbf{1}_k$; as $z \to a$, where a is any endpoint of γ, we have $Y(z) = O(1)$.

7.2.2 Uniqueness of the solution

Contrarily to the scalar case, we do not have a general formula for the solution to the matrix RH problem. However, we have the following important result concerning uniqueness.

Proposition 7.2.1. *If a solution $Y(z)$ to the above matrix RH problem exists, it is unique.*

Proof. Assuming the existence of a solution $Y(z)$, $\det Y(z)$ is an entire function (because by assumption, $M(z)$ has unit determinant and has an analytic extension to a neighborhood of γ). Since $\det Y(z) \to 1$ as $z \to \infty$, Liouville Theorem implies that $\det Y(z) = 1$ identically in z. It follows that the inverse $Y(z)^{-1}$ exists.

Let now $\widetilde{Y}(z)$ be another solution and define the matrix $R(z) := Y(z)\widetilde{Y}(z)^{-1}$. Since $R_\pm(z)$ exist and $R_+(z) = R_-(z)$ for all $z \in \gamma$, it follows that $R(z)$ is an entire (matrix-valued) function of z. Since $R(z) \to \mathbf{1}_k$ as $z \to \infty$, it follows by Liouville Theorem that $R(z) = \mathbf{1}_k$ identically in z, hence $\widetilde{Y}(z) = Y(z)$. □

7.2.3 Cauchy operator formulation and small-norm theorem

The goal of this section is to obtain a perturbative version of the simple fact that if $M = \mathbf{1}_k$ then the unique solution exists and is $Y(z) = \mathbf{1}_k$. Namely, we aim at showing that if the jump matrix M is close to $\mathbf{1}_k$ (in an appropriate sense) then also the solution Y is close to $\mathbf{1}_k$. The relation of RH problems to integral equations is the key to understand such result.

Definition 7.2.2. The **Cauchy operator** associated with the smooth oriented contour γ is defined by the formula

$$(\mathcal{C}f)(z) := \frac{1}{2\pi \mathrm{i}} \int_\gamma f(s) \frac{\mathrm{d}s}{s - z}, \qquad z \in \mathbb{C} \setminus \gamma. \tag{7.2.2}$$

Moreover, we introduce

$$(\mathcal{C}_\pm f)(z) := \lim_{w \to z_\pm} (\mathcal{C}f)(w). \tag{7.2.3}$$

//

Definition 7.2.3. For a (matrix-valued) function $f : \gamma \to \mathbb{C}^{k \times k}$ and $p \in [1, +\infty]$ we denote

$$\|f\|_{L^p(\gamma)} := \begin{cases} \int_\gamma |f(s)|^p |\mathrm{d}s|, & p \neq \infty, \\ \sup_\gamma |f|, & p = \infty. \end{cases} \tag{7.2.4}$$

Here, $|f|$ denotes any matrix norm (e.g., the maximum of the absolute value of the entries) and $\int \cdot |\mathrm{d}s|$ means the integral with respect to the arclength measure on γ.

By $L^p(\gamma)$ (omitting the size k for simplicity) we mean the Banach space of (equivalence classes of) functions $f : \gamma \to \mathbb{C}^{k \times k}$ with $\|f\|_{L^p(\gamma)} < +\infty$ (the equivalence relation being equality almost everywhere with respect to the arclength measure on γ). //

The following result will not be proved here.

Theorem 7.2.4. *If $f \in L^2(\gamma)$ then $(\mathcal{C}f)(z)$ is an analytic function of $z \in \mathbb{C} \setminus \gamma$ and $(\mathcal{C}_\pm f)(z)$ exist for almost all $z \in \gamma$ and $\mathcal{C}_\pm f \in L^2(\gamma)$. Moreover, there exists a constant $c(\gamma)$ depending on γ only such that*

$$\|(\mathcal{C}_\pm f)\|_{L^2(\gamma)} \leq c(\gamma)\|f\|_{L^2(\gamma)}. \tag{7.2.5}$$

Moreover,

$$\mathcal{C}_+ - \mathcal{C}_- = Id_{L^2(\gamma)}. \tag{7.2.6}$$

Remark 7.2.5. The identity (7.2.6) is a generalization of the argument in the proof of Proposition 7.1.1. It implies that the solution to the additive RH problem $X_+ - X_- = V$ (with $X \to 0$ at ∞) can be expressed as $X = \mathcal{C}V$ in terms of the Cauchy operator on γ. //

Let us explain how the general matrix RH problem is related to Cauchy operators. Keeping in mind our objective of proving that if M is close to $\mathbf{1}$ then Y is close to $\mathbf{1}$ it is convenient to rewrite $Y = \mathbf{1} + X$, $M = \mathbf{1} + \Delta$ (let us also assume that X, Δ belong to $L^p(\gamma)$ for $p = 1, 2, \infty$). By simple algebra, the jump condition $Y_+ = Y_- M$ becomes

$$X_+ - X_- = (\mathbf{1} + X_-)\Delta. \tag{7.2.7}$$

This can be understood as an additive RH problem, for which we know how to write the solution (see the last remark):

$$X = \mathcal{C}((\mathbf{1} + X_-)\Delta) = \mathcal{C}(\Delta) + \mathcal{C}(X_-\Delta). \tag{7.2.8}$$

Taking the $-$ boundary value we obtain

$$X_- = \mathcal{C}_-(\Delta) + \mathcal{C}_-(X_-\Delta), \tag{7.2.9}$$

or, equivalently,

$$(Id - \mathcal{L})X_- = \mathcal{C}_-(\Delta), \qquad \mathcal{L}f := \mathcal{C}_-(f\Delta). \tag{7.2.10}$$

Therefore we would like to invert $Id - \mathcal{L}$. Note that, by (7.2.5),

$$\|\mathcal{L}f\|_{L^2(\gamma)} \leq c(\gamma)\|f\Delta\|_{L^2(\gamma)} \leq c(\gamma)\|\Delta\|_{L^\infty(\gamma)}\|f\|_{L^2(\gamma)}. \tag{7.2.11}$$

Provided that $c(\gamma)\|\Delta\|_{L^\infty(\gamma)} < 1$ we can invert the operator $\mathcal{L}$. This is a general fact recalled in the following exercise.

Exercise 7.2.6. Let $\mathcal{L}$ be a linear operator on a Banach space V such that $\|\mathcal{L}v\|_V \leq k\|v\|_V$ for all $v \in V$ for some $0 \leq k < 1$. Prove that the series $\mathcal{T}v := \sum_{j\geq 0}\mathcal{L}^j v$ converges in norm for all v, and more precisely that we have

$$\|\mathcal{T}v\|_V \leq \frac{1}{1-k}\|v\|_V, \qquad \text{for all } v \in V. \tag{7.2.12}$$

Conclude that $\mathcal{T}$ is a linear operator on V which is inverse to $Id_V - \mathcal{L}$, i.e. $\mathcal{T}(Id_V - \mathcal{L}) = (Id_V - \mathcal{L})\mathcal{T} = Id_V$. //

Therefore, we can solve the equation (7.2.10) as

$$X_- = (Id - \mathcal{L})^{-1}\mathcal{C}_-(\Delta) \tag{7.2.13}$$

and so (see previous exercise) we have the L^2-norm estimate

$$\|X_-\|_{L^2(\gamma)} \leq \frac{c(\gamma)\|\Delta\|_{L^2(\gamma)}}{1 - c(\gamma)\|\Delta\|_{L^\infty(\gamma)}}. \tag{7.2.14}$$

Finally, spelling out (7.2.8) we have

$$X(z) = \frac{1}{2\pi \mathrm{i}}\int_\gamma \frac{\Delta(s)}{s-z}\mathrm{d}s + \frac{1}{2\pi \mathrm{i}}\int_\gamma \frac{X_-(s)\Delta(s)}{s-z}\mathrm{d}s \tag{7.2.15}$$

and so we get a uniform estimate for any $z \in \mathbb{C} \setminus \gamma$:

$$\begin{aligned} 2\pi|X(z)| &= \left| \int_\gamma \frac{\Delta(s)}{s-z} \mathrm{d}s + \int_\gamma \frac{X_-(s)\Delta(s)}{s-z} \mathrm{d}s \right| \\ &\le \left(\int_\gamma \left| \frac{\Delta(s)}{s-z} \mathrm{d}s \right| + \int_\gamma \left| \frac{X_-(s)\Delta(s)}{s-z} \mathrm{d}s \right| \right) \\ &\le \frac{1}{\mathrm{dist}\,(z,\gamma)} \left(\|\Delta\|_{L^1(\gamma)} + \|X_-\|_{L^2(\gamma)} \|\Delta\|_{L^2(\gamma)} \right) \\ &= \frac{1}{\mathrm{dist}\,(z,\gamma)} \left(\|\Delta\|_{L^1(\gamma)} + \frac{c(\gamma)\|\Delta\|^2_{L^2(\gamma)}}{1 - c(\gamma)\|\Delta\|_{L^\infty(\gamma)}} \right) \end{aligned} \tag{7.2.16}$$

where we used Cauchy–Schwarz inequality.

Recalling that $Y = \mathbf{1} + X$ and $M = \mathbf{1} + \Delta$ we have proved the following result.

Theorem 7.2.7 (Small-norm theorem)**.** *Suppose $M(z;t)$ also depends on a real parameter $t \ge t_0$ and define*

$$\delta(t) := \max\{\|M(\cdot;t) - \mathbf{1}\|_{L^1(\gamma)}, \|M(\cdot;t) - \mathbf{1}\|_{L^2(\gamma)}, \|M(\cdot;t) - \mathbf{1}\|_{L^\infty(\gamma)}\}. \tag{7.2.17}$$

Then, if $\delta(t) \to 0$ as $t \to +\infty$, there exists $t_1 \ge t_0$ such that the solution $Y(z;t)$ exists for $t \ge t_1$. Moreover, there exists $C > 0$ independent of t, z such that

$$|Y(z;t) - \mathbf{1}| \le \frac{C\delta(t)}{\mathrm{dist}\,(z,\gamma)}, \qquad \textit{for } t \ge t_1,\ z \in \mathbb{C} \setminus \gamma. \tag{7.2.18}$$

Remark 7.2.8. Assume that $M(z;t)$ admits an analytic extension to a neighborhood $\mathcal{N}_\gamma$ of γ (independent of t) and also that there is a function $\delta(t)$ for $t \ge t_0$ satisfying $\delta(t) \to 0$ as $t \to +\infty$, and

$$\delta(t) \ge \max\{\|M(\cdot;t) - \mathbf{1}\|_{L^1(\widetilde{\gamma})}, \|M(\cdot;t) - \mathbf{1}\|_{L^2(\widetilde{\gamma})}, \|M(\cdot;t) - \mathbf{1}\|_{L^\infty(\widetilde{\gamma})}\}, \tag{7.2.19}$$

whenever $\widetilde{\gamma}$ is a small deformation of γ (homomotopy equivalent to γ) that never exits $\mathcal{N}_\gamma$. By a contour deformation argument we can now avoid the singularities in the Cauchy integrals and, consequently, the estimate (7.2.18) can be refined to

$$|Y(z;t) - \mathbf{1}| \le \frac{C\delta(t)}{1 + \mathrm{dist}\,(z,\gamma)}, \qquad \text{for } t \ge t_1,\ z \in \mathbb{C} \setminus \gamma, \tag{7.2.20}$$

for $t_1 \ge t_0$ and some constant $C > 0$ independent of t, z. This inequality now applies also to $z \in \gamma$ in the sense of boundary values, i.e. $|Y_\pm(z;t) - \mathbf{1}| \le C\delta(t)$. //

7.3 The RH problem for OPRL

We shall consider the following assumptions on the measure of orthogonality $\mu(\mathrm{d}x) = w(x)\mathrm{d}x$; these assumptions are quite far from being minimal, but they are convenient in order to work with RH problems in the easy analytic sense discussed above.

Throughout this section, the weight function $w : \mathbb{R} \to \mathbb{R}_{\ge 0}$ is such that for some $\epsilon > 0$, w admits an analytic extension to the domain

$$\Omega_\epsilon := \{z \in \mathbb{C} : |\arg z| < \epsilon \text{ or } |\pi + \arg z| < \epsilon \text{ or } |\mathrm{Im}\, z| < \epsilon\} \tag{7.3.1}$$

such that $\int_{\mathbb{R}e^{\mathrm{i}\psi}} |w(s)s^k \mathrm{d}s| < +\infty$ for all $-\epsilon < \psi < \epsilon$ and all integers $k \ge 0$.

(In particular, for $\psi = 0$, all the moments are finite, $\int_{\mathbb{R}} w(s)|s|^k \mathrm{d}s < +\infty$ for all integers $k \ge 0$.)

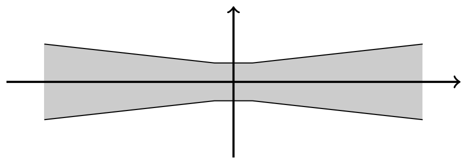


Figure 7.5: Analyticity domain for w.

Example 7.3.1. An important example for us is the class of measures $\mu(\mathrm{d}x) = \exp(-V(x))\mathrm{d}x$ where $V(x)$ is a polynomial of x of even degree. This is a generalization of the Gaussian measure. //

The RH problem for OPRL is the problem of finding a 2×2-matrix function $Y_n = Y_n(z)$, for a fixed integer $n \geq 1$, satisfying the following conditions.

RH1 $Y_n(z)$ is analytic for $z \in \mathbb{C} \setminus \mathbb{R}$,

RH2 $Y_{n,\pm}(z)$ exist for all $z \in \mathbb{R}$ and are related as

$$Y_{n,+}(z) = Y_{n,-}(z) \begin{pmatrix} 1 & w(z) \\ 0 & 1 \end{pmatrix}. \tag{7.3.2}$$

RH3 $Y_n(z) \begin{pmatrix} z^{-n} & 0 \\ 0 & z^n \end{pmatrix} \to \mathbf{1}_2$ as $z \to \infty$.

Straightforward variations of this RH problem can be made so that it applies to orthogonality on a finite interval rather than on the real line, but let us stick to this formulation.

Remark 7.3.2. It will be convenient to use the notation

$$\sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{7.3.3}$$

so that in particular

$$a^{\sigma_3} := \exp(\log(a)\sigma_3) = \begin{pmatrix} a & 0 \\ 0 & \frac{1}{a} \end{pmatrix}, \qquad \text{for all } a \neq 0. \tag{7.3.4}$$

The condition RH3 can be written as $Y_n(z)z^{-n\sigma_3} \to \mathbf{1}_2$ as $z \to \infty$. //

Theorem 7.3.3 (Fokas–Its–Kitaev, 1992)**.** *The RH problem above admits a unique solution $Y_n(z)$ for all integers $n \geq 1$ which is given explicitly by*

$$Y_n(z) = \begin{pmatrix} P_n(z) & \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} P_n(s) w(s) \frac{\mathrm{d}s}{s-z} \\ -\frac{2\pi \mathrm{i}}{h_{n-1}} P_{n-1}(z) & -\frac{1}{h_{n-1}} \int_{\mathbb{R}} P_{n-1}(s) w(s) \frac{\mathrm{d}s}{s-z} \end{pmatrix}, \qquad z \in \mathbb{C} \setminus \mathbb{R}. \tag{7.3.5}$$

Here, $P_\ell = P_\ell(x)$ are the monic orthogonal polynomials on the real line, satisfying

$$\int_{\mathbb{R}} P_\ell(x) P_k(x) w(x) \mathrm{d}x = h_\ell \delta_{\ell,k}. \tag{7.3.6}$$

Proof. First, the uniqueness, which is proven along the lines of Proposition 7.2.1. Indeed, the scalar function $\det Y(z)$ satisfies $(\det Y(z))_+ = (\det Y(z))_-$ for all $z \in \mathbb{R}$, hence $\det Y(z)$ extends to an entire function of z; by RH3 we have $\det Y(z) = \det(Y(z)z^{-n\sigma_3}) \to 1$ as $z \to \infty$. By Liouville Theorem,

$\det Y(z) = 1$ identically in z. Hence, every solution is invertible; let $Y, \widetilde{Y}$ be two solutions to the the above RH problem and set $R(z) := \widetilde{Y}(z)Y(z)^{-1}$, so that $R_+(z) = R_-(z)$ and so $R(z)$ is a matrix of entire functions. Since

$$R(z) = \widetilde{Y}(z)Y(z)^{-1} = \widetilde{Y}(z)z^{-n\sigma_3}(Y(z)z^{-n\sigma_3})^{-1} \to \mathbf{1}_2 \tag{7.3.7}$$

we conclude, once again thanks to Liouville Theorem that $R(z) = \mathbf{1}_2$ identically in z. Hence, the solution Y is unique.

Actually, that the solution is unique also follows from the following part of the proof, where we explicitly show that the RH conditions RH1–RH3 uniquely determine the solution in the form (7.3.5); it is however instructive to see the general argument based on Liouville Theorem too.

To derive the structure of the solution, let us denote

$$Y(z) = \begin{pmatrix} a(z) & b(z) \\ c(z) & d(z) \end{pmatrix}, \tag{7.3.8}$$

so that RH1 implies a, b, c, d are analytic in $\mathbb{C} \setminus \mathbb{R}$, RH3 implies

$$a(z)z^{-n} \to 1, \quad b(z)z^n \to 0, \quad c(z)z^{-n} \to 0, \quad d(z)z^n \to 1, \tag{7.3.9}$$

and RH2 implies

$$\begin{pmatrix} a_+(z) & b_+(z) \\ c_+(z) & d_+(z) \end{pmatrix} = \begin{pmatrix} a_-(z) & b_-(z) \\ c_-(z) & d_-(z) \end{pmatrix} \begin{pmatrix} 1 & w(z) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_-(z) & b_-(z) + a_-(z)w(z) \\ c_-(z) & d_-(z) + c_-(z)w(z) \end{pmatrix} \tag{7.3.10}$$

in particular

$$a_+ = a_-, \qquad c_+ = c_- \tag{7.3.11}$$

and so a, c are entire. By a generalization of Liouville theorem (see Exercise 7.3.5 below), since a, c have polynomial growth by (7.3.9) then a, c are polynomials. More precisely, there exists a monic polynomial $p_n(z) = z^n + \cdots$ of degree n and a polynomial $q_n(z)$ of degree $< n$ such that

$$a(z) = p_n(z), \qquad c(n) = q_n(z). \tag{7.3.12}$$

Then

$$b_+(z) - b_-(z) = w(z)p_n(z), \qquad d_+(z) - d_-(z) = w(z)q_n(z), \tag{7.3.13}$$

and, moreover, $b(z), d(z) \to 0$ as $z \to \infty$ because of (7.3.9). By the previous discussion of scalar additive RH problems we can solve the last relations as

$$b(z) = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} p_n(s)w(s)\frac{\mathrm{d}s}{s-z}, \qquad d(z) = \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} q_n(s)w(s)\frac{\mathrm{d}s}{s-z}. \tag{7.3.14}$$

Summarizing: we have proven that there exist a monic degree n polynomial p_n and a degree $< n$ polynomial q_n such that

$$Y(z) = \begin{pmatrix} p_n(z) & \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} p_n(s)w(s)\frac{\mathrm{d}s}{s-z} \\ q_n(z) & \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} q_n(s)w(s)\frac{\mathrm{d}s}{s-z} \end{pmatrix}. \tag{7.3.15}$$

It remains to show that $p_n(z) = P_n(z)$ and $q_n(z) = -\frac{2\pi \mathrm{i}}{h_{n-1}}P_{n-1}(z)$. To this end, it is convenient to stop the proof for a moment and prove a general lemma.

Lemma 7.3.4. *Let p be a polynomial and let $w : \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfying the assumptions above, see equation (7.3.1). Then, for all $k \geq 0$ there exists a constant $C(k) > 0$, independent of z, such that*

$$\left| \int_{\mathbb{R}} p(s)w(s)\frac{\mathrm{d}s}{s-z} + \sum_{i=0}^{k-1} \frac{c_i}{z^{i+1}} \right| \leq \frac{C(k)}{z^{k+1}}, \qquad \textit{for all } z \in \mathbb{C} \setminus \gamma \textit{ such that } |z| \geq 1, \tag{7.3.16}$$

where

$$c_i = \int_{\mathbb{R}} w(s)p(s)s^i \mathrm{d}s. \tag{7.3.17}$$

Proof of lemma. We have the algebraic identity (a geometric sum)

$$\frac{1}{z} + \frac{s}{z^2} + \cdots + \frac{s^{k-1}}{z^k} = \frac{1-(s/z)^k}{z-s}. \tag{7.3.18}$$

Hence

$$\int_{\mathbb{R}} p(s)w(s)\frac{\mathrm{d}s}{s-z} + \sum_{i=0}^{k-1}\frac{1}{z^{i+1}}\underbrace{\int_{\mathbb{R}} s^i p(s)w(s)\mathrm{d}s}_{=:c_i} = -\frac{1}{z^{k+1}}\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s \tag{7.3.19}$$

It remains to show that there exists a constant C_k depending on k only such that

$$\left|\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s\right| \le C_k, \qquad \text{for all } |z|\ge 1. \tag{7.3.20}$$

To this end we need to control the denominator $1-s/z$, and, to this end, let us first assume that z (such that $|z|>1$) also satisfies $\frac{\epsilon}{3} < |\arg z| < \pi - \frac{\epsilon}{3}$, where we recall that we are assuming w analytic in the domain Ω_ϵ in (7.3.1). We have the estimates:

$$\text{if } |s| \le \frac{1}{2}|z| \Rightarrow \left|1-\frac{s}{z}\right| = \frac{|z-s|}{|z|} \ge \frac{\big||z|-|s|\big|}{|z|} \ge \frac{1}{2}, \tag{7.3.21}$$

$$\text{if } |s| \ge \frac{1}{2}|z| \Rightarrow \left|1-\frac{s}{z}\right| \ge |\mathrm{Im}\,\frac{s}{z}| = \frac{|s|\mathrm{Im}\,z}{|z|^2} \ge \frac{1}{2}\sin\frac{\epsilon}{3}. \tag{7.3.22}$$

Therefore, for all $|z|\ge 1$ with $\frac{\epsilon}{3} < |\arg z| < \pi - \frac{\epsilon}{3}$ and all $s\in\mathbb{R}$ we have

$$\left|1-\frac{s}{z}\right| \ge \frac{1}{2}\sin\frac{\epsilon}{3}, \tag{7.3.23}$$

hence, for all $|z|\ge 1$ with $\frac{\epsilon}{3} < |\arg z| < \pi - \frac{\epsilon}{3}$ we have

$$\left|\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s\right| \le \frac{2}{\sin\frac{\epsilon}{3}}\int_{\mathbb{R}}|s|^k|p(s)|w(s)\mathrm{d}s =: C_1. \tag{7.3.24}$$

It remains to extend such estimates in case $0 < |\arg z| < \frac{\epsilon}{3}$ or $0 < |\arg(-z)| < \frac{\epsilon}{3}$. Assume first that $0 < \arg z < \frac{\epsilon}{3}$ or $0 < \arg(-z) < \frac{\epsilon}{3}$; in such case we can use our assumptions on w and obtain

$$\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s = \int_{\mathbb{R}\mathrm{e}^{-\frac{2}{3}\epsilon\mathrm{i}}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s = (\mathrm{e}^{-\frac{2}{3}\epsilon\mathrm{i}})^{k+1}\int_{\mathbb{R}}\frac{(s')^k p(\mathrm{e}^{-\frac{2}{3}\epsilon\mathrm{i}}s')w(\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}}s')}{1-(s'/z')}\mathrm{d}s. \tag{7.3.25}$$

where in the first step we use Cauchy Theorem (we do not cross the singularity $s=z$ with this deformation) and in the second one we change variable $s' = s\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}}$ and set $z' := z\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}}$ which now satisfies $\frac{\epsilon}{3} < |\arg z'| < \pi - \frac{\epsilon}{3}$. Therefore, with completely similar estimates as above, if $|z|\ge 1$ and $0 < \arg z < \frac{\epsilon}{3}$ or $0 < \arg(-z) < \frac{\epsilon}{3}$,

$$\left|\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s\right| \le \frac{2}{\sin\frac{\epsilon}{3}}\int_{\mathbb{R}}|s|^k|p(s\mathrm{e}^{-\frac{2}{3}\epsilon\mathrm{i}})w(s\mathrm{e}^{-\frac{2}{3}\epsilon\mathrm{i}})|\mathrm{d}s =: C_2. \tag{7.3.26}$$

Finally, if $|z|\ge 1$ and $-\frac{\epsilon}{3} < \arg z < 0$ or $-\frac{\epsilon}{3} < \arg(-z) < 0$, by deforming the integration contour to $\mathbb{R}\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}}$ we obtain in a completely similar way

$$\left|\int_{\mathbb{R}}\frac{s^k p(s)w(s)}{1-(s/z)}\mathrm{d}s\right| \le \frac{2}{\sin\frac{\epsilon}{3}}\int_{\mathbb{R}}|s|^k|p(s\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}})w(s\mathrm{e}^{\frac{2}{3}\epsilon\mathrm{i}})|\mathrm{d}s =: C_3. \tag{7.3.27}$$

The proof is complete since C_1, C_2, C_3 depend on w, p, and k only and not on z, so it suffices to take $C(k) := \max\{C_1, C_2, C_3\}$. □

Back to the proof of Theorem 7.3.3. We need to enforce the asymptotics (7.3.9) for $b(z)$ and $d(z)$ given in (7.3.14). Let us start from $z^n b(z) \to 0$. We can write (by the lemma above)

$$b(z) = R(z) + \sum_{i=0}^{n-1} z^{-i-1} \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} s^i p_n(s) w(s) \mathrm{d}s \tag{7.3.28}$$

with $|R(z)| \leq C|z|^{-n-1}$ for all $|z| \geq 1$, i.e.

$$\left| z^n b(z) - \sum_{i=0}^{n-1} z^{n-i-1} \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} s^i p_n(s) w(s) \mathrm{d}s \right| \leq C|z|^{-1} \tag{7.3.29}$$

hence, we need to have $\int_{\mathbb{R}} s^i p_n(s) w(s) \mathrm{d}s = 0$ for all $i = 0, \dots, n-1$, i.e. $p_n = P_n$, in which case

$$|z^n b(z)| \leq C|z|^{-1} \tag{7.3.30}$$

for all $|z| \geq 1$, hence $z^n b(z) \to 0$ as $z \to \infty$.

Similarly, for $d(z)$ we can write

$$d(z) = \widetilde{R}(z) + \sum_{i=0}^{n-1} z^{-i-1} \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} s^i q_n(s) w(s) \mathrm{d}s \tag{7.3.31}$$

with $|\widetilde{R}(z)| \leq \widetilde{C}|z|^{-n-1}$ for all $|z| \geq 1$, i.e.

$$\left| z^n d(z) - \sum_{i=0}^{n-1} z^{n-i-1} \frac{1}{2\pi \mathrm{i}} \int_{\mathbb{R}} s^i q_n(s) w(s) \mathrm{d}s \right| \leq \widetilde{C}|z|^{-1} \tag{7.3.32}$$

hence, this time we need to have $\int_{\mathbb{R}} s^i q_n(s) w(s) \mathrm{d}s = 0$ for all $i = 0, \dots, n-2$ and $\int_{\mathbb{R}} s^{n-1} q_n(s) w(s) \mathrm{d}s = 2\pi \mathrm{i}$, and so we need to take $q_n = \frac{2\pi \mathrm{i}}{h_{n-1}} P_{n-1}$ to have

$$|z^n d(z) - 1| \leq \widetilde{C}|z|^{-1} \tag{7.3.33}$$

for all $|z| \geq 1$, hence $z^n d(z) \to 1$ as $z \to \infty$. The proof is complete. □

Exercise 7.3.5. Let $f(z)$ be an entire function such that the inequality $|f(z)| \leq C|z|^n$ (for some $n \geq 0$ and some $C > 0$) holds true for all $|z| > 1$. Prove that $f(z)$ is a polynomial of z of degree n. (Hints. Method 1: $f(z)$ has a Taylor series convergent everywhere, apply Cauchy inequalities. Method 2: induction on $n \geq 0$, case $n = 0$ is Liouville; for $n \geq 1$, use the identity $f'(z) = \oint_{|s|=2|z|} \frac{f(s)}{(s-z)^2} \frac{\mathrm{d}s}{2\pi \mathrm{i}}$ to show that $|f'(z)| \leq (2^n C)|z|^{n-1}$ for all $|z| > 1$ and use the inductive hypothesis.) //

It is important to observe that the RH problem description of orthogonal polynomials allows one to characterize the orthogonal polynomial P_n, without having to compute all the previous P_ℓ's for $0 \leq \ell < n$. This is especially convenient if we want to analyze the large degree limit of P_n. Thus, the RH method replaces the integral representation of Hermite polynomials which we exploited in the previous chapter, which is not available for other weights (although for a few other special measures there are analogous standard contour integral representations).

It should also be remarked that there is another formula which we encountered for OPRL, which (purely in principle!) would allow us to compute P_n without having to compute the previous P_ℓ's for $0 \leq \ell < n$, for a general weight. This is Heine's formula (cf. second part in Exercise 3.2.8)

$$P_n(z) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 \prod_{i=1}^n \Big((z - x_i) w(x_i) \mathrm{d}x_i \Big) \tag{7.3.34}$$

It should be appreciated that the RH description of OPRL converts such n-fold integral into a characterization where n appear as a parameter.

The following three exercises imply that we can access all relevant quantities for OPRL, and not just the polynomials themselves, like the norming constants h_n, the coefficients b_n, w_n of the three-term recurrence, and the Christoffel–Darboux kernel, via the matrix $Y_n(z)$.

Exercise 7.3.6. Prove that

$$\lim_{z\to\infty} z\left(Y_n(z)z^{-n\sigma_3}\right)_{12} = -\frac{h_n}{2\pi \mathrm{i}}, \tag{7.3.35}$$

$$\lim_{z\to\infty} z\left(Y_n(z)z^{-n\sigma_3}\right)_{21} = -\frac{2\pi \mathrm{i}}{h_{n-1}}. \tag{7.3.36}$$

(Hint: use Lemma 7.3.4.) //

Exercise 7.3.7. Let the three-term recurrence of the OPRL be

$$xP_n(x) = P_{n+1}(x) + b_n P_n(x) + w_n P_{n-1}(x). \tag{7.3.37}$$

Prove that

$$w_n = (Y_1)_{12}(Y_1)_{21}, \qquad b_n = \frac{(Y_2)_{12}}{(Y_1)_{12}} - (Y_1)_{22}, \tag{7.3.38}$$

where

$$Y_1 := \lim_{z\to\infty} z\left(Y_n(z)z^{-n\sigma_3} - \mathbf{1}_2\right), \qquad Y_2 := \lim_{z\to\infty} z^2\left(Y_n(z)z^{-n\sigma_3} - \mathbf{1}_2 - z^{-1}Y_1\right). \tag{7.3.39}$$

(Hint: prove the identities $w_n = h_n/h_{n-1}$ and $b_n = \frac{1}{h_n}\int_{\mathbb{R}} x^{n+1}P_n(x)w(x)\mathrm{d}x - \frac{1}{h_{n-1}}\int_{\mathbb{R}} x^n P_{n-1}(x)w(x)\mathrm{d}x$ and use Lemma 7.3.4 and the previous exercise.) //

Exercise 7.3.8. Prove that we can compute the Christoffel–Darboux kernel

$$K_n^{\mathsf{CD}}(x,y) := \sqrt{w(x)}\sqrt{w(y)}\sum_{\ell=0}^{n-1}\frac{P_\ell(x)P_\ell(y)}{h_{\ell-1}} \tag{7.3.40}$$

in terms of $Y_n(z)$ through the formulae

$$K_n^{\mathsf{CD}}(x,y) = -\frac{1}{2\pi \mathrm{i}}\frac{\sqrt{w(x)}\sqrt{w(y)}}{x-y}\left(Y_n(x)^{-1}Y_n(y)\right)_{21}, \tag{7.3.41}$$

$$K_n^{\mathsf{CD}}(x,x) = \frac{w(x)}{2\pi \mathrm{i}}\left(Y_n(x)^{-1}Y_n'(x)\right)_{21}, \tag{7.3.42}$$

where the notation $(A)_{21}$ denotes the entry in the second row and first column of the matrix A. Note that although $Y_n(x)$ is not defined for $x \in \mathbb{R}$ (it has a jump there), these expressions only involve the entries of Y in the first column, i.e. the orthogonal polynomials, and not the entries which do have a jump across $\mathbb{R}$. (Hint: use the Christoffel–Darboux identity and the fact that $\det Y_n(z) = 1$.) //

7.4 Universality in RMT via the Riemann–Hilbert approach

7.4.1 Scaling of Unitary-invariant Ensembles

As we did for the GUE (Section 6.1), to have a meaningful large-size limit, we need to take an appropriate (size-dependent) rescaling of the measure of unitary-invariant ensembles of Hermitian matrices. This is justified by the electrostatic analogy anticipated in Section 6.1 as we now explain in more detail for a general potential V. The latter is assumed to satisfy the usual condition

$$\lim_{x\to\pm\infty}\frac{V(x)}{\log|x|} = +\infty. \tag{7.4.1}$$

Namely, the probability density function for eigenvalues $(x_1,\dots,x_n)$ can be expressed as

$$\frac{1}{\widehat{Z}_n^V}\Delta^2(x_1,\dots,x_n)\exp\Big(-\sum_{i=1}^n V(x_i)\Big) = \frac{1}{\widehat{Z}_n^V}\exp\left(\sum_{1\le i\neq j\le n}\log|x_i-x_j| - \sum_{i=1}^n V(x_i)\right). \tag{7.4.2}$$

Looking at the exponent of the last expression we clearly see two competing effects.

- The probability will be very small if $x_i \to +\infty$, because V is a potential growing at $\pm\infty$.
- The probability will also be very small if the x_i are close to each other.

If we just let $n \to +\infty$ without further rescaling the potential, we can expect one of these effects to dominate on the other; more precisely, there are $\binom{n}{2} = O(n^2)$ contributions from the repulsive part, and only n contributions from the confining part. Thus we are led to rescale the potential function as $V(x) \mapsto nV(x)$. Otherwise, without this rescaling, the eigenvalues x_i would tend as $n \to +\infty$ to spread on the real line due to the repulsive effect and the confining potential would not be strong enough to prevent the eigenvalues to escape to $\pm\infty$.

This argument prompts us to consider the following probability measures on $\mathbf{H}_n$:

$$\frac{1}{\widehat{Z}_n^{nV}} \exp\left(-n \operatorname{tr}\left(V(M)\right)\right) \prod_{i=1}^{n} \mathrm{d}M_{ii} \prod_{1\le i<j\le n} \mathrm{dRe}\, M_{ij} \mathrm{dIm}\, M_{ij}. \tag{7.4.3}$$

(Compare this with (6.1.4).)

7.4.2 Equilibrium measure

It is convenient to introduce the (random) probability measure ("empirical spectral measure")

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \tag{7.4.4}$$

where $x_1, \ldots, x_n$ are the eigenvalues of a random matrix distributed according to (7.4.3). The probability distribution of x_i, and hence of μ_n, is (proportional to)

$$\exp\left(\sum_{1\le i\neq j\le n} \log|x_i - x_j| - n\sum_{i=1}^{n} V(x_i)\right) \\ = \exp\left[-n^2\left(\int_{\mathbb{R}} V(x)\mu_n(\mathrm{d}x) - \int_{\mathbb{R}^2\setminus\Delta} \log|x-y|\,\mu_n(\mathrm{d}x)\mu_n(\mathrm{d}y)\right)\right]. \tag{7.4.5}$$

where $\Delta := \{(x,y) \in \mathbb{R}^2 \,:\, x = y\}$. (Again, the fact that we can factorize n^2 in front is due to the rescaling of the potential.)

This prompts us to introduce the functional

$$I : \mathcal{M}_1 \to \mathbb{R}, \tag{7.4.6}$$

on the set $\mathcal{M}_1$ of Borel (positive) probability measures on $\mathbb{R}$, by

$$I_V[\mu] := \int_{\mathbb{R}} V(x)\mu(\mathrm{d}x) - \int_{\mathbb{R}^2\setminus\Delta} \log|x-y|\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) \tag{7.4.7}$$

Therefore, for each n the probability distribution, in terms of the empirical spectral measure μ_n, is just (up to the normalization constant)

$$\exp(-n^2 I_V[\mu_n]). \tag{7.4.8}$$

It is clear therefore the the most likely configurations, in the large n limit, will tend to the to the measure ν_V realizing the minimum of the functional I_V.

That such a minimum exists is a fundamental result of Potential Theory (essentially due to convexity of the functional I_V) which we state without proof[1].

[1]See Chapter 6 in Deift's book quoted at page 2, or Saff & Totik "Logarithmic Potentials with External Fields", Springer-Verlag, Berlin, 1997.

Theorem 7.4.1. *Suppose V is smooth and $\lim_{x\to\pm\infty}\frac{V(x)}{\log|x|}=+\infty$. There exists a unique minimizing measure $\nu_V\in\mathcal{M}_1$ for which $I_V[\mu]$ achieves the minimum at $\mu=\nu_V$ only.*

Suppose in addition that V is strictly convex, then ν_V is supported in a single interval.

Definition 7.4.2. The minimizing probability measure ν_V is called **equilibrium measure**. //

Summarizing, **we expect the distribution of eigenvalues μ_n (which is random, as the matrix is random) to "converge" to the *deterministic* distribution ν_V.** This is an instance of concentration of measures.

When $V(x)=x^2/2$, the equilibrium measure is the semicircle distribution. The concentration of measures explains the fact that if we plot the eigenvalues of a random GUE matrix we *always* obtain a shape resembling a semicircle (again, in other words, this is true for every realization, not just for an average of many realizations of the random matrix).

7.4.3 g-function and heuristic large degree asymptotics for OPRL with varying weights

We now take $P_{n+j}(z;n)$ to be the $(n+j)$-th monic orthogonal polynomial with respect to the measure $\mathrm{e}^{-nV(x)}\mathrm{d}x$ (note the dependence on n both in the order of the polynomial and on the measure, so that we refer to them by saying that they are OPRL *with varying weigth*).

These (for $j=0,-1$ only) are the polynomials which we are interested in to describe the large n limit of the eigenvalue distribution of the random matrix model (7.4.3), because of the expression of all the relevant statistical quantities in terms of the Christoffel–Darboux kernel and because of the Christoffel–Darboux identity.

By the previous section, we expect the distribution of eigenvalues to converge to to the determininstic probability measure ν_V. Therefore we have the following heuristic approximations for large n, where the symbol $\approx$ is used in a loose sense of approximation,

$$\log P_n(z)=\log\mathbb{E}[\det(z-M)]\approx n\int_{\mathbb{R}}\log(z-x)\nu_V(\mathrm{d}x)=:ng(z) \tag{7.4.9}$$

where the first equality is again Heine formula (cf. part 3 in Exercise 3.2.8). The final relation is the definition of the so-called **g-function**

$$g(z):=\int_{\mathbb{R}}\log(z-x)\nu_V(\mathrm{d}x) \tag{7.4.10}$$

which plays a fundamental role in the asymptotic analysis of OPRL. It is an analytic function of z in $\mathbb{C}\setminus\operatorname{supp}\nu_V$, once we stipulate to take the principal branch of the logarithm in (7.4.10). Among its properties, it is important to note that

$$g'(z)=\int_{\mathbb{R}}\frac{\nu_V(\mathrm{d}x)}{z-x} \tag{7.4.11}$$

is a Cauchy-type integral, hence the equilibrium measure can be obtained from g via

$$\nu_V(\mathrm{d}x)=\psi(x)\mathrm{d}x,\qquad g'_+(x)-g'_-(x)=\begin{cases}2\pi\mathrm{i}\psi(x) & x\in\operatorname{supp}\nu_V,\\ 0, & \text{elsewhere.}\end{cases} \tag{7.4.12}$$

An analysis of the RH problem for OPRL allows us to prove rigorously (7.4.9).

Example 7.4.3. Specializing the general definition (7.4.10), the g-function in the GUE case is given, for $z\in\mathbb{C}\setminus[-2,2]$, as

$$g(z):=\int_{-2}^{2}\log(z-x)\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x. \tag{7.4.13}$$

To compute this integral it is convenient[2] to first compute

$$g'(z)=\int_{-2}^{2}\frac{1}{z-x}\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x=\frac{2}{\pi}\int_0^{\pi}\frac{(\sin\phi)^2}{z+2\cos\phi}\mathrm{d}\phi=\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{(\sin\phi)^2}{z+2\cos\phi}\mathrm{d}\phi \tag{7.4.14}$$

[2]The approach that consists in first computing $g'(z)$ and then integrating is useful to deal with g-functions in general.

where in the first step we use the substitution $x = -2\cos\phi$ and in the second one elementary periodicity properties of sin and cos. Setting $w = \mathrm{e}^{\mathrm{i}\phi}$ we obtain

$$g'(z) = \frac{1}{4\pi \mathrm{i}} \oint_{|w|=1} \frac{(w^2-1)^2}{2w(w^2+2wz-1)} \mathrm{d}w. \tag{7.4.15}$$

and a residue computation (**Exercise**) provides

$$g'(z) = \frac{z-\sqrt{z^2-4}}{2}, \tag{7.4.16}$$

where the branch of the square root is determined by $\operatorname{Im}\sqrt{z^2-4} > 0$ in the upper half-plane and $\operatorname{Im}\sqrt{z^2-4} < 0$ in the lower half-plane (so that $\sqrt{z^2-4}$ is analytic in $\mathbb{C}\setminus[-2,2]$ and $\sim z$ as $z\to\infty$). (See Exercise 7.4.4 for another derivation of the same expression for $g'(z)$.) Integrating this relation we obtain

$$g(z) = \frac{1}{4}\Big(z\big(z-\sqrt{z^2-4}\big) + 4\log\big(\sqrt{z^2-4}+z\big)\Big) + c \tag{7.4.17}$$

where the integration constant c is fixed by the fact that, as $z\to\infty$ we have

$$\begin{aligned}\log(z-x) = \log(z) - \frac{x}{z} + O(z^{-2}) \Rightarrow g(z) &= \log(z)\int_{-2}^{2}\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x - \frac{1}{z}\int_{-2}^{2} x\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x + O(z^{-2}) \\ &= \log(z) + O(z^{-2}).\end{aligned} \tag{7.4.18}$$

and so $c = -\frac{1}{2} - \log 2$. Finally, taking the logarithm in the asymptotics for (scaled) Hermite polynomials in the exterior region $|z| > 2$, given in (6.2.12), we obtain that

$$\log \widetilde{P}_n(z;n) = ng(z) + O(1), \tag{7.4.19}$$

in agreement with the general heuristic relation (7.4.9). As a matter of fact, the asymptotic relation (6.2.12) is true for all $z\in\mathbb{C}\setminus[-2,2]$, not just for real values of z satisfying $|z|>2$; this could be proved by the same standard steepest descent method used in the previous chapter. //

Exercise 7.4.4. The Catalan numbers are defined for any integer $n \geq 0$ as

$$C_n := \frac{1}{n+1}\binom{2n}{n}. \tag{7.4.20}$$

The sequence C_n looks like

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, \dots \tag{7.4.21}$$

The numbers C_n have several important applications in combinatorics.

1. Prove that

$$\int_{-2}^{2} x^{2k}\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x = C_k. \tag{7.4.22}$$

Note that $\int_{-2}^{2} x^{2k+1}\frac{\sqrt{4-x^2}}{2\pi}\mathrm{d}x = 0$ by parity symmetry. (Hint: change of variables $x = -2\cos\phi$.) Deduce that if M is a random GUE matrix, i.e. the probability distribution of its entries is (6.1.4), we have

$$\frac{1}{n}\mathbb{E}[\operatorname{tr}(M^{2k})] \to C_k, \qquad n\to+\infty. \tag{7.4.23}$$

(Hint: use (3.3.27).) It is worth to mention that the convergence of the "moments" $\frac{1}{n}\mathbb{E}[\operatorname{tr}(M^{2k})]$ to the Catalan numbers is essentially equivalent to the converge of $\frac{1}{n}\rho_1(x)$ to the semicircle law. Moreover, the convergence (7.4.23) could be proven directly from a combinatorial arguent using the independence of the entries of M and existence of the first moments; this explains why the semicircle law is characteristic of all Wigner matrix models, cf. Theorem 6.3.4.

2. Directly from the definition (7.4.20) show that

$$C_n = 2(-4)^n \binom{1/2}{n+1} \tag{7.4.24}$$

and deduce, for all $y \in \mathbb{C}$ satisfying $|y| < 1/4$, the formula

$$\sum_{n\geq 0} C_n y^n = \frac{1-\sqrt{1-4y}}{2y}. \tag{7.4.25}$$

(Hint: recall that the general binomial coefficient is defined as $\binom{\alpha}{k} := \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$ and we have the binomial theorem (rather, a Taylor expansion) $(1+y)^\alpha = \sum_{k\geq 0}\binom{\alpha}{k} y^k$, valid analytically for $|x| < 1$.)

3. Prove that, for $|z| > 2$.

$$\int_{-2}^{2} \frac{\sqrt{4-x^2}}{2\pi}\frac{\mathrm{d}x}{z-x} = \frac{z-\sqrt{z^2-4}}{2}, \tag{7.4.26}$$

with the branch of the square roots chosen as explained after equation (7.4.16). By an analytic continuation argument prove that the formula is true for all $z \in \mathbb{C} \setminus [-2,2]$. (Hint: develop $\frac{1}{/}(z-x)$ as $z \to \infty$ with a geometric series and use the previous points.)

//

7.4.4 Formal statements

Under the suitable assumptions for V (also depending on the approach that one uses) we have the following results.

Theorem 7.4.5. *Assume V is strictly convex, so that the equilibrium measure $\nu_V(\mathrm{d}x) = \psi(x)\mathrm{d}x$ exists and it is supported on a single (bounded) interval $[a,b]$. Set*

$$\zeta := \frac{z-a}{z-b}. \tag{7.4.27}$$

As $n \to +\infty$ and $z \in \mathbb{C} \setminus [a,b]$ we have

$$P_n(z;n) = \frac{1}{2}\left(\zeta^{1/4} + \zeta^{-1/4} + O(n^{-1})\right)\exp\left(n\int_a^b \log(z-x)\psi(x)\mathrm{d}x\right). \tag{7.4.28}$$

As $n \to +\infty$ and $a < z < b$ we have

$$\begin{aligned} P_n(z;n) = \Bigg[&|\zeta|^{1/4}\cos\left(n\pi\int_z^b \psi(x)\mathrm{d}x - \frac{\pi}{4}\right) \\ &+|\zeta|^{-1/4}\cos\left(n\pi\int_z^b \psi(x)\mathrm{d}x + \frac{\pi}{4}\right) + O(n^{-1})\Bigg]\exp\left(n\int_a^b \log|z-x|\psi(x)\mathrm{d}x\right). \end{aligned} \tag{7.4.29}$$

As in the GUE case, we also have estimates when z approaches the endpoints of the equilibrium measure, which (in the generic case) involve the Airy function as well.

At the level of random matrix models, for the unitary-invariant ensemble (7.4.3) we have.

Theorem 7.4.6. *1. We have*

$$\frac{1}{n}\rho_1(x) \to \psi(x), \qquad n \to +\infty \tag{7.4.30}$$

where $\psi(x)\mathrm{d}x = \nu_V(\mathrm{d}x)$ is the equilibrium measure.

2. For all x_0 in the interior of the support of ν_V we have

$$\frac{1}{n\psi(x_0)}K_n\left(x_0 + \frac{\xi}{n\psi(x_0)}, x_0 + \frac{\eta}{n\psi(x_0)}\right) \to K^{\mathsf{sine}}(\xi,\eta) \qquad \text{"Universality in the bulk"} \tag{7.4.31}$$

3. Let b be a right-endpoint of the support of ν_V which is generic *in the sense that for some $c > 0$ we have*

$$\psi(x) \sim \frac{c}{\pi}\sqrt{b-x}, \qquad \text{as } x \nearrow b. \tag{7.4.32}$$

Then we have

$$\frac{1}{(cn)^{2/3}}K_n\left(b + \frac{\xi}{(cn)^{2/3}}, b + \frac{\eta}{(cn)^{2/3}}\right) \to K^{\mathsf{Airy}}(\xi,\eta) \qquad \text{"Universality at the edge"} \tag{7.4.33}$$

where K^{sine} and K^{Airy} are the Sine and Airy kernels encountered above;

$$K^{\mathsf{sine}}(\xi,\eta) := \frac{\sin(\pi(\xi-\eta))}{\pi(\xi-\eta)}, \qquad K^{\mathsf{Airy}}(\xi,\eta) := \frac{\mathrm{Ai}(\xi)\mathrm{Ai}'(\eta) - \mathrm{Ai}'(\xi)\mathrm{Ai}(\eta)}{\xi-\eta}. \tag{7.4.34}$$

The limits in (7.4.31) and (7.4.33) are for $n \to +\infty$, uniformly for ξ,η in compact sets of the real line.

This has to be compared with Theorems 6.3.1, 6.3.5, and 6.3.6 for the GUE case.

7.4.5 Sketch of the proofs via the Riemann–Hilbert method

The starting point is the following RH problem, for an integer $n \geq 1$.

RH1 $Y_n(z)$ is analytic in $\mathbb{C} \setminus \mathbb{R}$

RH2 $Y_{n,\pm}(z)$ exist for all $z \in \mathbb{R}$ and are related as

$$Y_{n,+}(z) = Y_{n,-}(z)\begin{pmatrix} 1 & \exp(-nV(z)) \\ 0 & 1 \end{pmatrix}. \tag{7.4.35}$$

RH3 We have $Y_n(z)z^{-n\sigma_3} \to I$ as $z \to \infty$.

We know from Theorem 7.3.3 that $P_n(z;n) = (Y(z))_{11}$ and from Exercise 7.3.8 that $K_n(x,y) = \frac{\mathrm{e}^{-\frac{1}{2}(V(x)+V(y))}}{2\pi\mathrm{i}(x-y)}(Y_n(x)^{-1}Y_n(y))_{21}$. So it is enough for both theorems to establish large n asymptotics for the RH solution $Y_n(z)$.

The idea (the so called "nonlinear steepest descent" scheme developed by Deift and Zhou in the 1990s, with earlier contributions by A. Its in the 1980s) is to reduce, by *explicit* transformations of $Y_n(z)$, the RH conditions into RH conditions of a small-norm RH problem in the sense of the small-norm Theorem (Theorem 7.2.7), where n plays the role of the parameter t. Then, the small-norm Theorem allows to control the non-explicit part of $Y_n(z)$, which will tend to $\mathbf{1}_2$ as $n \to +\infty$, providing the error terms in the above theorems, and the explicit transformations will give the leading contributions in the asymptotics.

Let us just mention the main steps:

1. The first obvious difficulty is that in the RH problem for Y_n the normalization condition RH3 is not of the form $Y_n(z) \to \mathbf{1}$. Note that a simple transformation $Y_n(z) \mapsto \widetilde{Y}_n(z) := Y_n(z) z^{-n\sigma_3}$ does not solve the problem but rather transfers it from $z = \infty$ to a singularity at $z = 0$. The heuristic approximation (7.4.9) and the g-function help us here: since $g(z) \sim \log z$ as $z \to \infty$, the correct transformation is

$$Y_n(z) \mapsto S_n(z) := Y_n(z) \mathrm{e}^{-ng(z)\sigma_3}. \tag{7.4.36}$$

2. By a careful analysis of the jumps of $S_n(z)$ (note that boh Y, g have jumps across the real axis) and exploiting the analytic properties of g (stemming from its origin in a minimization problem), one deforms the contour ("opening of the lenses") in such a way that the jumps are small as $n \to +\infty$ (in the appropriate norms as the small-norm theorem prescripts) everywhere except at the support of the equilibrium measure and in small disks near the endpoints.

3. Neglecting the small jumps, we can explicitly solve the RH problems, obtaining the so-called parametrices: this will involve the conformal transformation $(\frac{z-a}{z-b})^{\pm 1/4}$ explaining its appearance in the final asymptotics for the OPRL, and Airy functions in the small disks near the endpoints of the equilibrium measure.

4. Finally, one proves that the final RH problem is small-norm, and tracking back all the transformations to $Y_n(z)$ obtains large n asymptotics for it. By algebraic manipulations one obtains large n asymptotics for the OPRL and the CD kernel.

We content ourselves with this informal description, and refer the reader to the literature for more details (e.g., with reference to the bibliography at page 2, good and very complete sources are Deift's book and Its' article).