Introduction to random matrix theory

Gaëtan Borot[*]

Master course, Universität Bonn

February 15^{th}, 2015

Contents

[*]Max-Planck Institut for Mathematics: gborot@mpim-bonn.mpg.de

One of the goal of random matrix theory (= RMT) was initially to describe the distribution of eigenvalues of large random matrices. They give rise to universal laws quite different from those known for independent random variables (like Gauss law). They appear e.g. in nuclear physics, number theory, statistical physics, ...

Matrix integrals have also many connections to integrable systems, algebraic geometry, combinatorics of surfaces, ... In this master course, we will present basic results and techniques in the study of random matrices, and describe some of their surprising applications.

The prerequisites are a basic command of probability theory, linear algebra, and real and complex analysis.

Among the general references on random matrix theory, I recommend:

- *Random matrices*, M.L. Mehta, 3rd edition, Elsevier (2004). Written by a pioneer of random matrix theory. Accessible, rather focused on calculations and results for exactly solvable models.

- *An introduction to random matrices*, G.W. Anderson, A. Guionnet, O. Zeitouni, Cambridge Studies in Advanced Mathematics **118**, CUP (2010). A level of technicity higher than Mehta. Probability oriented. Self-contained proofs and progression.

- *Topics in random matrix theory*, T. Tao., Graduate Studies in Mathematics **132**, AMS (2012). A series of graduate lectures, yet the exposition makes some parts accessible to master level.

Though I do not follow a book in particular, these monographs were useful in the preparation of this course, and I sometimes borrowed some of their arguments.

Further readings

The content of these references is almost not treated in this course, but they represent a window to more recent uses of random matrix theory:

- *Planar diagrams*, É. Brézin, C. Itzykson, G. Parisi and J.-B. Zuber, Commun. Math. Phys. **59** 35–51 (1978). Physics paper (all statements therein can easily be written to match mathematical standards of rigor), where the combinatorial interpretation of matrix integrals as generating series of maps was first introduced.

- *The Euler characteristics of the moduli space of curves*, J. Harer and D. Zagier, Invent. Math. **85** 457–485 (1986). First appearance of matrix model techniques in algebraic geometry. In the combinatorial part of the paper, Harer-Zagier recursion is derived as an intermediate result – without use of Hermite polynomials – and is solved in several ways.

- *Orthogonal polynomials and random matrices : a Riemann-Hilbert approach*, P. Deift, AMS, Courant Institute of Mathematical Sciences (1998). A course in RMT which gives a good overview of the relations between orthogonal polynomials and unitary invariant ensembles of random matrices. Emphasis on asymptotic analysis.

Figures 1,3 and 8 are reprinted with permission from APS journals: ”Readers may view, browse, and/or download material for temporary copying purposes only, provided these uses are for noncommercial personal purposes. Except as provided by law, this material may not be further reproduced, distributed, transmitted, modified, adapted, performed, displayed, published, or sold in whole or part, without prior written permission from the American Physical Society.”

0 Notations and basic properties

0.1 *Linear algebra*

• $\mathbb{A}$ denotes a field, most frequently $\mathbb{R}$ or $\mathbb{C}$. $\mathscr{M}_N(\mathbb{A})$ is the algebra of $N \times N$ matrices with entries in $\mathbb{A}$, with product $(A \cdot B)_{i,k} = \sum_{j=1}^{N} A_{i,j} B_{j,k}$. The identity and the zero matrix are denoted 1_N and 0_N, or 1 and 0 when the size is obvious. For N even, we sometimes encounter the elementary matrix $J_{N/2}$ of size $N \times N$, made of $N/2$ blocks $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ along the diagonal. $\mathscr{D}_N(\mathbb{A})$ is the set of diagonal matrices, they form a commutative subalgebra of $\mathscr{M}_N(\mathbb{A})$. Unless specified otherwise, $\mathbb{R}^N$ (resp. $\mathbb{C}^N$) is considered as a Euclidean space (resp. a Hilbert space) equipped with the canonical scalar product, denoted $(\cdot|\cdot)$.

• Let $A \in \mathscr{M}_N(\mathbb{C})$. $\lambda \in \mathbb{C}$ is an eigenvalue of A if there exists $v \in \mathbb{C}^N \setminus \{0\}$ such that $Av = \lambda v$. v is then called an eigenvector, and $E_\lambda(A) = \mathrm{Ker}(A - \lambda)$ is the eigenspace of A for the eigenvalue λ. The eigenvalues coincide with the roots of the characteristic polynomial $\det(\lambda - A)$. Therefore, any complex matrix of size N has exactly N eigenvalues counted with multiplicity. The set of eigenvalues of A is its **spectrum**, denoted $\mathrm{Sp}(A)$.

• A^T is the transpose of A, namely $(A^T)_{i,j} = A_{j,i}$. $*$ denotes the complex conjugation. The adjoint is $A^\dagger = (A^T)^*$.

• $A \in \mathscr{M}_N(\mathbb{C})$ is **normal** if it commutes with its adjoint. A is unitary if $AA^\dagger = 1$, is **hermitian** if $A = A^\dagger$, is **antihermitian** if $A = -A^\dagger$, is **symmetric** if $A = A^T$, is **antisymmetric** if $A = -A^T$, is **orthogonal** if $AA^T = 1$, is **symplectic** if N is even and $AJ_{N/2}A^T = J_{N/2}$. We denote $\mathscr{U}_N(\mathbb{C})$ the group of unitary matrices, $\mathscr{H}_N(\mathbb{C})$ the $\mathbb{R}$-vector space of hermitian matrices, $\mathscr{S}_N(\mathbb{A})$ the $\mathbb{A}$-vector space of symmetric matrices with entries in $\mathbb{A}$, $\mathscr{O}_N(\mathbb{A})$ the group of orthogonal matrices with entries in $\mathbb{A}$, $\mathrm{Sp}_N(\mathbb{A})$ the group of symplectic matrices with entries in $\mathbb{A}$ (which only exists when N is even).

• $A \in \mathscr{M}_N(\mathbb{A})$ can be considered as endomorphism of $\mathbb{A}^N$. If we change from the canonical basis to a new basis of $\mathbb{A}^N$ by a matrix P, the new matrix representing the endomorphism is PAP^{-1}. We list various results of reduction of endomorphisms. Normal matrices are characterized by the existence of $U \in \mathscr{U}_N(\mathbb{C})$ and $D \in \mathscr{D}_N(\mathbb{C})$ such that $A = UDU^\dagger$, i.e. they can be diagonalized in an orthonormal basis. Furthermore, A is hermitian iff D has real entries ; A is antihermitian iff D has pure imaginary entries ; A is unitary iff D has entries in $\mathbb{U}$ (the complex numbers of modulus 1) ; A is symmetric real if U is orthogonal real and D is real. If A is an antisymmetric matrix, its rank r is even, and there exists $U \in \mathscr{U}_N(\mathbb{C})$ such that:

$$A = U\,\mathrm{diag}\big(a_1 J_1, \ldots, a_{r/2} J_1, \underbrace{0, \ldots, 0}_{n-r \text{ times}}\big) U^{-1}$$

where the middle matrix is a block diagonal matrix.

• Let A be a normal matrix, $f : D \to \mathbb{C}$ with domain of definition $D \subseteq \mathrm{Sp}\, A$.

If we diagonalize $A = U\mathrm{diag}(\lambda_1, \ldots, \lambda_N)U^\dagger$, we define the matrix

$$f(A) := U\mathrm{diag}\big(f(\lambda_1), \ldots, f(\lambda_N)\big)U^\dagger,$$

and it does not depend on the basis of diagonalization of A.

• If $A \in \mathscr{M}_N(\mathbb{C})$, the eigenvalues of the hermitian matrix $\sqrt{AA^\dagger}$ are the **singular values** of A.

If V is a vector space, the Grassmannian $\mathrm{Gr}_k(\mathbb{V})$ is the set of its k-dimensional subspaces. If $v_1, \ldots, v_k \in V$, $\mathrm{vect}(v_1, \ldots, v_k)$ denotes the subspace of V spanned by the v_i's.

• A N-dimensional vector is identified with a column matrix (size $N \times 1$). The hermitian product of two vectors $v_1, v_2 \in \mathbb{C}^N$ can then be represented:

$$(v_1|v_2) = v_1^\dagger v_2.$$

0.2 *Norms*

• The L^2 norm on $\mathbb{C}^N$ is defined by:

$$|a|_p = \Big(\sum_{i=1}^N |a_i|^p\Big)^{1/p}.$$

In particular, $|a|_\infty = \sup_i |a_i|$ and $|a| := |a|_2$ is the hermitian norm. The **Hölder inequality** states for any $p, q \in [1, \infty]$,

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \Longrightarrow \quad \Big|\sum_{i=1}^N a_i b_i\Big| \leq |a|_p |b|_q.$$

• If $A \in \mathscr{M}_N(\mathbb{C})$, the **spectral radius** is

$$\varrho(A) = \max \mathrm{Sp}\,\sqrt{AA^\dagger} = \sup_{|v|=1} |Av|$$

The L^2 norm is:

$$||A||_2 = \Big(\sum_{i,j=1}^N |A_{i,j}|^2\Big)^{1/2} = \big(\mathrm{Tr}\,AA^\dagger\big)^{1/2}$$

$\mathscr{H}_N \simeq \mathbb{R}^{N^2}$ can also be equipped with its L^2-norm:

$$||A||_{2,\mathbb{R}^{N^2}} = \Big(\sum_{i=1}^N A_{i,i}^2 + \sum_{1 \leq i < j \leq N} |A_{i,j}|^2\Big)^{1/2}$$

We have the obvious comparisons:

$$||A||_2 \leq \sqrt{2}\,||A||_{2,\mathbb{R}^{N^2}}, \qquad ||A||_2 \leq \sqrt{N}\varrho(A), \qquad \varrho(A) \leq ||A||_2$$

• If I be an ideal of $\mathbb{C}[X_1, \ldots, X_n]$, we denote:

$$\mathcal{Z}(I) = \{\mathbf{x} \in \mathbb{C}^n \;:\; \forall Q \in I,\; Q(\mathbf{x}) = 0\}$$

the zero locus of this ideal in $\mathbb{C}^n$. The **Nullstellensatz** is the following statement: if I is an ideal of $\mathbb{C}[X_1, \ldots, X_n]$ and $P \in \mathbb{C}[X_1, \ldots, X_n]$ such that the polynomial function P vanishes on $\mathcal{Z}(I)$, then there exist an integer $r \geq 1$, and two polynomials $Q \in I$ and $R \in \mathbb{C}[X_1, \ldots, X_n]$, such that $P^r = R \cdot Q$.

0.3 *Analysis*

• If A is a subset of a set X, $\mathbf{1}_A$ denotes the indicator function of X, which assumes value 1 on A and 0 on $X \setminus A$.

• $\delta_{j,k}$ is the Kronecker symbol, equal to 1 if $j = k$, to 0 otherwise.

• If $(a_n)_n$ and $(b_n)_n$ are two sequences of complex numbers, we write $a_n \asymp b_n$ if they have the same order, i.e. if there exists $c \neq 0$ such that $a_n \sim cb_n$ when $n \to \infty$. A similar definition can be made for functions.

• **Convexity.** Let V be a finite-dimensional vector space. A subset $X \subseteq V$ is **convex** if for any $t \in [0,1]$ and x, y in X, the combination $(1-t)x + ty$ is also in X. If $x_1, \ldots, x_n \in V$, a linear combination of the form $\sum_{i=1}^N t_i x_i$ with $t_1, \ldots, t_N \in [0,1]$ and $\sum_{i=1}^N t_i = 1$ is called a **convex combination**. A convex combination is trivial if one of the t_i's is equal to 1. A point x in a convex subset $X \subseteq V$ is **extreme** if it cannot be written as a non-trivial convex combination of points in X. Let X be a convex subset of V. A function $f : V \to \mathbb{R}$ is **convex** if for any $t \in [0,1]$ and $x, y \in X$, we have $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$. Convex functions are continuous in the interior of their domain of definition. Obviously, linear combinations of convex functions with positive coefficients are convex functions. The supremum of a family of convex functions also defines a convex function. If X is a compact convex subset of V and $f : X \to \mathbb{R}$ is a convex function, a useful property is that f achieves its maximum in the set of the extreme points of X.

• $L^2(\mathbb{R})$ is the space of complex-valued square-integrable functions on $\mathbb{R}$, equipped with the Lebesgue measure. It is equipped with the hermitian product:

$$(f|g) = \int_{\mathbb{R}} f(x) g^*(x) \mathrm{d}x.$$

If $f \in L^2(\mathbb{R})$, its Fourier transform is $\hat{f} \in L^2(\mathbb{R})$ defined by:

$$\hat{f}(k) = \int_{\mathbb{R}} e^{2\mathrm{i}\pi kx} f(x) \mathrm{d}x.$$

The Fourier transform is an isometry of L^2 spaces. In particular, the scalar product is equivalently expressed:

$$(f|g) = \int_{\mathbb{R}} \hat{f}(k) \hat{g}^*(k) \mathrm{d}k.$$

The inverse Fourier transform is given by:

$$f(x) = \int_{\mathbb{R}} e^{-2i\pi kx} \hat{f}(k) \mathrm{d}k.$$

0.4 *Probability*

• In this section, $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, and X, $(X_n)_{n\geq 0}$ be random variables (= r.v.) with values in a Banach space. This setting will not be made explicit as long as there is no ambiguity from the context. $(X_n)_{n\geq 0}$ is i.i.d if the X_n are independent and identically distributed.
• The repartition function of a real-valued random variable X is $\Phi_X(x) = \mathbb{P}[X \leq x]$, defined for $x \in \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. It is increasing, right-continuous and takes the values $\Phi_X(-\infty) = 0$ and $\Phi_X(\infty) = 1$. Conversely, for any function $\Phi : \overline{\mathbb{R}} \to [0,1]$ that is right-continuous, increasing from $\Phi_X(-\infty) = 0$ to $\Phi(+\infty) = 1$, there exist a probability space and a random variable X on this probability space such that Φ is the repartition function of X. We say that X has a density if Φ_X is differentiable almost everywhere on $\mathbb{R}$. The probability density function (= p.d.f.) $\Phi'_X(x)$ is often denoted $\rho_X(x)$.
• If μ_1, μ_2 are two probability measures on $\mathbb{R}$, the convolution $\mu_1 * \mu_2$ is the unique probability measure such that, if X_1, X_2 are independent r.v. with probability laws μ_1, μ_2, then $X_1 + X_2$ has probability law $\mu_1 * \mu_2$.
• $\operatorname{supp} \mu$ denotes the support of a probability measure μ.
• We will meet several (in general distinct) notions of convergence of r.v. We say that X_n converges to X when $n \to \infty$: **almost surely** (= a.s.) if

$$\mathbb{P}\big(\{\omega \in \Omega, \qquad \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\big) = 1;$$

in L^r norm for some $r \geq 1$ if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^r] = 0$; **in probability** if, for any $\epsilon > 0$, $\lim_{n\to\infty} \mathbb{P}[|X_n - X| \geq \epsilon] = 0$; in law if $\lim_{n\to\infty} \mathbb{P}[X_n \leq x] = \mathbb{P}[X \leq x]$ for all points x of continuity of the right-hand side. Some notions of convergence are stronger than others:

$$\begin{tikzcd}
\boxed{\text{almost surely}} \arrow[r, Rightarrow] & \boxed{\text{in probability}} \arrow[r, Rightarrow] & \boxed{\text{in law}} \\
& \boxed{L^r,\ r \geq 1} \arrow[u, Rightarrow] &
\end{tikzcd}$$

Without further assumptions, all the implications in this diagram are strict. The L^1 convergence is also called **convergence in expectation**.
• A useful, sufficient condition for almost sure convergence is:

$$\forall \epsilon > 0, \qquad \sum_{n\geq 0} \mathbb{P}\big[|X_n - X| \geq \epsilon\big] \text{ converges.}$$

• We say that a sequence of real-valued r.v. $(X_n)_{n\geq 0}$ has Φ for **limit law** if there exist sequences $(a_n)_{n\geq 0}$ and $(b_n)_{n\geq 0}$ such that $a_n X_n + b_n$ converges in

law to a random variable X with repartition function Φ when $n \to \infty$.
• Gauss(μ, σ^2) denotes a Gaussian real-valued r.v. with variance σ^2 and mean μ. Its p.d.f. is

$$\rho(x) = \frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}.$$

Exp(μ) denotes a exponential law with mean μ. Its p.d.f is

$$\rho(x) = \mathbb{1}_{\mathbb{R}_+}(x) e^{-x/\mu}.$$

It is sometimes called a Poisson law. The Cauchy distribution of width $b > 0$ is the law of density:

$$\rho(x) = \frac{1}{\pi(x^2 + b^2)}.$$

• Probabilistic version of the **Jensen inequality**: if X is a real-valued r.v. and $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function, then

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

0.5 *Topology*

• If X, X' are open sets in a normed vector space, a function $f : X \to X'$ is k**-Lipschitz** if

$$\forall x, y \in X, \qquad |f(x) - f(y)| \leq k|x-y|$$

The Lipschitz constant of f is the infimum of constants k for which this inequality holds. We denote Lip_k the set of k-Lipschitz functions, and $\mathrm{Lip}_{k;m}$ its subset of functions bounded by $m > 0$.
• A **Polish space** is a complete, separable (there exists a countable basis of open sets), metric space X. It is also a measurable space equipped with the σ-algebra generated by open sets. $\mathcal{C}_b(X)$ denotes the vector space of real-valued, continuous bounded functions, and $\mathcal{C}_b^c(X)$ its subspace of functions with compact support.
• If X is a Polish space, we denote $\mathcal{M}_1(X)$ the set of probability measures on X. The **weak topology** is defined by the basis of open sets

$$U_{f,\epsilon,x} = \Big\{\mu \in \mathcal{M}_1(X),\ \Big|\int_X f \mathrm{d}\mu - t\Big| < \varepsilon\Big\} \qquad f \in \mathcal{C}_b(X),\ \epsilon > 0,\ t \in \mathbb{R}$$

The **vague topology** on $\mathcal{M}^1(X)$ is generated by the open sets $U_{f,\epsilon,x}$ with $f \in \mathcal{C}_b^c(X)$. In absence of precision, $\mathcal{M}^1(X)$ is equipped with the weak topology.
• $\mathcal{M}_1(X)$ is a Polish space, which is compact iff X is compact. If Y is a countable and dense subset of X, then the set of probability measures on X that are supported on Y is dense in $\mathcal{M}^1(X)$.
• A probability measure $\mu \in \mathcal{M}^1(X)$ is tight if forall $\epsilon > 0$, there exists a compact $K_\epsilon \subseteq X$ such that $\mu(X \setminus K_\epsilon) < \epsilon$. A subset $M \subseteq \mathcal{M}^1(X)$ is tight if every $\mu \in M$ is tight, and the compact K_ε can be chosen independently of μ. Since X is a Polish space, every probability measure is tight. Prokhorov theorem states that a subset M is tight iff $\overline{M}$ is compact.

• The Vasershtein distance is defined by:

$$\forall \mu_1, \mu_2 \in \mathcal{M}^1(X), \qquad \mathfrak{d}(\mu_1, \mu_2) = \sup_{\substack{f \in \mathrm{Lip}_{b;1} \\ 0 \leq b \leq 1}} \left| \int f(\mathrm{d}\mu_1 - \mathrm{d}\mu_2) \right|.$$

It is compatible with the weak topology.
• A **sub-probability measure** on X is a positive measure of total mass ≤ 1. **Helly's selection theorem** states that any sequence of probability measures on a Polish space X admits a subsequence that converges weakly to a sub-probability measure.
• If $f : [a, b] \to \mathbb{C}$, its total variation is defined by:

$$\mathrm{TV}[f] = \lim_{n \to \infty} \sup_{\substack{a \leq a_1 \leq \cdots \leq < a_n \leq b \\ a_0 = a,\ a_{n+1} = b}} \sum_{i=0}^{n} |f(a_{i+1}) - f(a_i)|.$$

• If X is a real-valued r.v., $\mathbb{M}[X]$ is a median of X, i.e. any real number such that:

$$\mathbb{P}[\mathrm{X} < \mathbb{M}[\mathrm{X}]] \leq 1/2, \qquad \mathbb{P}[\mathrm{X} > \mathbb{M}[\mathrm{X}]] \leq 1/2.$$

0.6 *Complex analysis*

• An **entire** function is a holomorphic function on $\mathbb{C}$. **Liouville theorem** states that bounded entire functions are constants, and entire functions bounded uniformly by a polynomial are polynomials.
• Let U be an open subset of $\mathbb{C}$. **Montel theorem** states that, from any bounded sequence $(f_n)_{n \geq 1}$ of holomorphic functions, one can extract a subsequence that converges uniformly on any compact of U to a holomorphic function f.

0.7 *Miscellaneous*

• $\mathfrak{S}_N$ denotes the symmetric group.
• $[\![a, b]\!]$ the set of integers $\{a, a+1, \ldots, b-1, b\}$.
• $\lfloor x \rfloor$ denotes the integer part of x, i.e. the unique integer n such that $n \leq x < n+1$.

0.8 *Special functions and identities*

• The Gamma function can be defined as the unique function $\Gamma : \mathbb{R}_+^* \to \mathbb{R}$ such that $\Gamma(n+1) = n!$ for any integer $n \geq 0$, $\Gamma(s+1) = s\Gamma(s)$ for any $s \in \mathbb{R}$, and $\ln \Gamma$ is convex. For $s > 0$, it admits the integral representation:

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} \mathrm{d}t,$$

and the special values $\Gamma(1/2) = 2\int_0^\infty e^{-u^2} \mathrm{d}u = \sqrt{\pi}$. It can be analytically continued as a holomorphic function in $\mathbb{C} \setminus \mathbb{Z}$, with simple poles at the non-positive integers. It satisfies the reflection formula:

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}$$

For any $m \geq 0$, it admits an asymptotic expansion when $s \to \infty$ away from the negative real axis:

$$\ln \Gamma(s+1) = s \ln s - s + \frac{\ln(2\pi s)}{2} + \sum_{\ell=1}^{m} \frac{B_{2\ell}}{2\ell(2\ell-1)s^{2\ell-1}} + O(s^{-(2m+1)}),$$

where the B_m are the Bernoulli numbers defined by the Taylor expansion at $x = 0$:

$$\frac{x}{e^x - 1} = \sum_{m \geq 0} \frac{B_m}{m!}\, x^m, \qquad B_1 = -\frac{1}{2}, \qquad B_2 = \frac{1}{6}, \qquad B_4 = -\frac{1}{30}, \qquad \ldots$$

The B_{2m+1} vanish for $m \geq 1$, and the sign of B_{2m} is $(-1)^m$.

• The Riemann Zeta function is initially defined for $\mathrm{Re}\, s > 1$ by the convergent series:

$$\zeta(s) = \sum_{n \geq 1} \frac{1}{n^s}.$$

It is of central importance in number theory since it is also given by the infinite product over all primes:

$$\zeta(s) = \prod_{p} \frac{1}{1 - p^{-s}}$$

Using Cauchy residue formula, the series can be represented as:

$$\zeta(s) = e^{-\mathrm{i}\pi s}\Gamma(1-s) \oint \frac{\mathrm{d}t}{2\mathrm{i}\pi}\, \frac{t^{s-1}}{e^t - 1},$$

where the contour surrounds the positive real axis including 0 from $+\infty + \mathrm{i}0^+$ to $+\infty - \mathrm{i}0^+$. By moving the contour and using the analytic properties of the Gamma function, this formula defines ζ as a holomorphic function on $\mathbb{C} \setminus \{1\}$, with a simple pole at $s = 1$. The Zeta function, as a particular case of L-functions in number theory, satisfy a functional relation:

$$\widetilde{\zeta}(s) = \widetilde{\zeta}(1-s), \qquad \widetilde{\zeta}(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s).$$

The special values of the zeta functions for integers are:

$$\forall n \geq 1, \qquad \zeta(-n) = (-1)^n \frac{B_{n+1}}{n+1}, \qquad \zeta(2n) = \frac{(-1)^{n+1} B_{2n}\, \pi^{2n}}{2 \cdot (2n)!}.$$

Since $\Gamma(s/2)$ has poles when s is a negative even integer, while the right-hand side remains finite, $\zeta(s)$ has a simple zero when s is an even negative integer. These are the **trivial** zeroes. As an entire function, the zeta function has at most countably many other zeroes. The **Riemann Hypothesis** (RH) claims that all the non-trivial zeroes are located on the critical line $\mathrm{Re}\, s = 1/2$. By symmetry,

$1/2 + it$ is a zero iff $1/2 - it$ by symmetry. So, one can consider only the zeroes with positive imaginary part, and order them $s_n = 1/2 + it_n$, with strictly increasing t_n. Knowing the location of the zeroes of Riemann Zeta would give fine information on the distribution of prime numbers. Hardy (1914) proved infinitely many zeroes lie on the critical line, and Conrey (1989) shows that more than $4/10$ of the zeroes lie on the critical line.

1 Introduction

1.1 *Independent r.v. versus eigenvalues*

Lecture 1 (1h30)
October 6th, 2014

If A is a $N \times N$ complex random matrix with random entries, its N eigenvalues are random. Even if the entries are chosen independent, the eigenvalues will be strongly correlated. The essential feature of the sequence of eigenvalues $\lambda_1, \ldots, \lambda_N$ of random matrices is that they repel each other (this statement will be made precise throughout the course). Their properties – and the limit laws appearing in the large N limit – are therefore very different from those relevant in the world of independent r.v.

The limit laws one computes in RMT are found in many problems in physics and mathematics, a priori relatively far from the area of random matrices: we will mention in this introduction nuclear physics, statistical physics, quantum chaos, and number theory. A property is said to be **universal**[1] if it holds in a much more general class of models than the one it was originally found. To a large extent, this is the case for local limit laws of random matrices, i.e. those concerning the distribution of λ_i's in small regions where we expect only finitely many eigenvalues when $N \to \infty$. The global limit laws, like the spectral density of the fluctuations of linear statistics, in general are more model-sensitive.

In this introduction, we will present small computations, experimental data and big theorems for illustration.

1.2 *Our three preferred matrix ensembles*

• **(Real) Wigner matrices.** Let $(X_i)_{1 \leq i \leq N}$ and $(Y_{i,j})_{1 \leq i < j \leq N}$ be two i.i.d sequences of r.v. with zero mean, $\mathbb{E}[Y_{1,2}^2] = 1$, and assume all the moments of X_1 and $Y_{1,2}$ are finite. We construct a $\mathscr{S}_N(\mathbb{R})$-valued r.v. by setting:

$$(M_N)_{i,j} = \begin{cases} N^{-1/2}\, Y_{i,j} & i < j \\ N^{-1/2} Y_{j,i} & i > j \\ N^{-1/2} X_i & i = j \end{cases} .$$

M_N is called a real Wigner matrix[2]. This is the topic of Chapter 3.

• **Invariant matrix ensembles.** Let V be a polynomial with real coefficients, such that $x \mapsto e^{-V(x)}$ is integrable on $\mathbb{R}$, called the **potential**. The $\mathbb{R}$-vector spaces $\mathscr{H}_{N,\beta}$ of matrices listed in the table have a canonical basis, and we denote $\mathrm{d}M$ the product of Lebesgue measure of the coefficients of decomposition in this canonical basis. For instance, for symmetric matrices

$$\mathrm{d}M = \prod_{i=1}^{N} \mathrm{d}M_{i,i} \prod_{1 \leq i < j \leq N} \mathrm{d}M_{i,j}.$$

[1] The notion of universality is thus relative to the historical development of knowledge and the points of view on this knowledge.

[2] The rescaling $1/\sqrt{N}$ will ensure that the eigenvalues are typically bounded when $N \to \infty$. For now, this can be seen heuristically: with the given definition, $\mathbb{E}[\mathrm{Tr}\, M_N^2]$ is of order N, but $\mathrm{Tr}[M_N^2]$ is also the sum of the N squared eigenvalues. Unless a small fraction of eigenvalues dominate the others, they all must be typically of order 1.

The invariant matrix ensembles are defined by considering a $\mathscr{H}_{N,\beta}$-valued r.v. drawn from the probability measure:

(1)
$$Z_{N,\beta}^{-1}\, \mathrm{d}M \, \exp\Big(-\frac{N\beta}{2}\,\mathrm{Tr}\, V(M)\Big), \qquad Z_{N,\beta} = \int_{\mathscr{H}_{N,\beta}} \mathrm{d}M \, \exp\Big(-\frac{N\beta}{2}\,\mathrm{Tr}\, V(M)\Big).$$

This measure is invariant under conjugation $M \to UMU^{\dagger}$ by a matrix U in the group $\mathscr{G}_{N,\beta}$ indicated in the table. Invariant ensembles are exactly solvable: many observables can be computed in terms of special functions. Apart from probabilistic techniques, one can use powerful techniques based on orthogonal polynomials (Chapter 6) and integrable PDEs (Chapter 9) for their study.

β	$\mathscr{H}_{N,\beta}$	$\dim_{\mathbb{R}} \mathscr{H}_{N,\beta}$	$\mathscr{G}_{N,\beta}$
1	$N \times N$ real symmetric	$N(N+1)/2$	$\mathscr{O}_N(\mathbb{R})$
2	$N \times N$ hermitian	N^2	$\mathscr{U}_N(\mathbb{C})$
4	$N \times N$ quaternionic self-dual $\simeq 2N \times 2N$ complex M's such that $J_N M = M^T J_N$ and $J_N M = M^* J_N$	$N(2N-1)$	$\mathrm{Sp}(2N, \mathbb{R})$

• **Gaussian β ensembles.** Consider an invariant ensemble with quadratic potential $V(M) = M^2/2$. Then, all the $\mathbb{R}$-linearly independent entries of M are independent Gaussian r.v with zero mean. In the case $\beta = 1$, $\mathrm{Var}(M_{i,i}) = 2/N$ and $\mathrm{Var}(M_{i,j}) = 1/N$ for $i \neq j$, so M is a particular case of real Wigner matrix. In the case $\beta = 2$, $\mathrm{Var}(M_{i,i}) = 1/N$ and $\mathrm{Var}(M_{i,j}) = 1/(2N)$ for $i \neq j$. These ensembles are denoted $\mathrm{G}\beta E$, or more specifically GOE (Gaussian orthogonal ensemble) for $\beta = 1$, GUE (Gaussian unitary ensemble) for $\beta = 2$. Later we will define as well a "GSE" (Gaussian symplectic ensemble) corresponding to $\beta = 4$. The Gaussian ensembles form the intersection between Wigner matrices and invariant ensembles. They are the easiest to compute with: in the framework of exactly solvable models, the special functions involved are related to the Hermite polynomials and their asymptotics (Chapter 6).

1.3 *The framework of quantum mechanics*

Quantum mechanics is a large source of (random) matrices. It may be worth to review its basic language so as to understand the motivations coming from physics.

Summary

• A system in quantum mechanics is described by a **Hilbert space** $\mathscr{V}$, that is a $\mathbb{C}$-vector space equipped with a definite positive hermitian product, such that $\mathscr{V}$ is complete for the topology induced by the hermitian norm $|v| = \sqrt{(v|v)}$. If A is an endomorphism of $\mathscr{V}$, its adjoint $A^{\dagger}$ is the unique endomorphism satisfying $(v|A^{\dagger}w) = (w|Av)$ for all $v, w \in \mathscr{V}$. We say that A is **self-adjoint** if $A = A^{\dagger}$, and that A is unitary if A preserves the hermitian product, i.e.

$AA^\dagger = A^\dagger A = 1$. There are several notions of spectrum if $\dim \mathscr{V}$ is infinite, but we do not need to enter into details here. Let us just say that for compact self-adjoint operators A, the spectral decomposition theorem guarantees that $\mathscr{V}$ is the Hilbert sum of the eigenspaces – defined as in finite dimension – of A.

• Vectors of unit length in $\mathscr{V}$ describe the states of the system, and $\dim \mathscr{V}$ – finite or infinite – is the number of degrees of freedom. The prediction of quantum mechanics of the result of physical measures on the system is in general of probabilistic nature. A physical observable (like the position, the impulsion, the angular momentum, the polarization ...) is represented by a (hopefully compact) self-adjoint operator A, its spectrum – defined as in finite dimension – represents the possible values obtained in a physical measurement. If the system is in a state $v \in \mathscr{V}$, the probability to observe the result $a \in \mathrm{Sp}(A)$ in an experiment measuring A is $|\pi_{E_a(A)}(v)|^2$ (**Born rule** postulate). Here, π_E denotes the orthogonal projection on the subspace $E \subseteq \mathscr{V}$. When v is an eigenvector of A, this probability is 0 or 1 and we say that the physical observables assumes a definite value (the eigenvalue a such that $v \in E_a(A)$) on v. A commutative subalgebra of compact self-adjoint operators on $\mathscr{V}$ can be diagonalized in the same basis. The physical observables forming a commutative algebra can simultaneously assume a definite value, whereas two observables represented by non-commuting operators cannot (Heisenberg uncertainty principle).

• If the system is in a state $v_0 \in V$ at time $t = 0$, its evolution is characterized by prescribing a self-adjoint H operator called the **hamiltonian**, and solving the Schrödinger equation:

$$\mathrm{i}\hbar\partial_t v_t = H v_t.$$

Understanding the dynamics of a quantum system thus amounts to diagonalizing the hamiltonian. The eigenvalues of H are called **energy levels**[3].

Quantum symmetries

• A set of physical observables $\mathscr{A}$ is **complete** if there exists a finite or countable subset of $\{A_i \;\; i \in I\}$ of $\mathscr{A}$ generating a commutative subalgebra, such that for any $(a_i) \in \prod_i \mathrm{Sp}\, A_i$, $\cap_{i\in I} E_{a_i}(A_i)$ is either 0 or 1-dimensional. A quantum system in which there exists a complete set of physical observables is said **completely integrable**. Two systems that cannot be distinguished by physical measurements are considered as equivalent. One can show that, in a completely integrable system, a system in a state v along with physical observables represented by operators A is equivalent to a system in a state v' with physical observables represented by operators A', iff there exists an unitary operator U such that $v' = Uv$ and $A' = UAU^{-1}$. In particular, if one wishes to keep the same representation for physical observables, we are restricted to take $U = u \cdot \mathrm{id}$, with $u \in \mathbb{U}$, i.e. two states are equivalent iff they are colinear. This motivates to consider only completely integrable system in the following.

• If a quantum system has a symmetry group G, it is implemented as an action of the group G on states $v \to v'$ and physical observables $A \to A'$ that gives

[3] To be fair, the eigenspaces should be considered as the "levels" in this picture.

an equivalent system. If one assumes that it is linearly realized, this means the existence of a representation of G by unitary operators K, with $v' = Kv$ and $A' = KAK^{-1}$. This symmetry is preserved by evolution under the Schrödinger equation when all K's commute with H. If a symmetry group G is a Lie group, it induces at the infinitesimal level a representation of its Lie algebra that also commute with K. We can also speak of a (Lie) **algebra** of symmetries.

• For instance, rotational invariance of a quantum system implies the existence of 3 self-adjoint operators $\mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3$ commuting with H, and having commutation relations $[\mathfrak{l}_i, \mathfrak{l}_j] = \mathfrak{l}_k$ for any (i, j, k) which is a cyclic permutation of $(1, 2, 3)$. The **angular momentum** is the operator $\mathfrak{l}^2 = \mathfrak{l}_1^2 + \mathfrak{l}_2^2 + \mathfrak{l}_3^2$, and $\mathscr{L} = \langle \mathfrak{l}^2, \mathfrak{l}_1 \rangle$ form a maximal commutative subalgebra of $\langle \mathfrak{l}_1, \mathfrak{l}_2, \mathfrak{l}_3 \rangle$. It can be used to split $\mathscr{V}$ into eigenspaces $\mathscr{V}_{\ell^2, \ell_1}$ for $\mathscr{L}$: since $[\mathscr{L}, H] = 0$, H leaves those eigenspaces stable, and we are left with the problem of diagonalizing H in the smaller Hilbert spaces $\mathscr{V}_{\ell^2, \ell_1}$.

• The eigenvalues for a maximal commutative subalgebra of the algebra of operators generated by elements of the symmetry group are called the **quantum numbers**. In a quantum system which only has rotational invariance, these are (for example) ℓ^2 and ℓ_1. To summarize, the eigenvalues and eigenvectors of H behave independently for different spaces attached to a different set of quantum numbers.

• **Spin.** Representations of the Lie algebra are in correspondence – by the exponential map – with representation of the universal cover of the corresponding Lie group. The universal cover of SO(3) is SU(2). There is a degree 2 covering map $\pi : \mathrm{SU}(2) \to \mathrm{SO}(3)$, and $\pi^{-1}(\mathrm{id}) = \{\pm \mathrm{id}\}$. The irreducible unitary representations of SU(2) are finite dimensional (since SU(2) is compact). They are characterized by $l \in \mathbb{N}/2$, and the dimension over $\mathbb{C}$ of the corresponding representation $\mathscr{R}_l$ is $2l + 1$. The representations of SU(2) with half-integer l induce projective representations of SO(3): performing continuously a rotation from angle 0 to angle 2π realizes the transformation $v \to -v$. Since the phase of a state gives an equivalent state, one could think this has no physical effect. However, if two states – described in two different Hilbert spaces – are put in interaction, their relative phase will matter, so this minus sign does affect physics. If the quantum system enjoys the rotational invariance, $\mathscr{V}$ splits into irreducible representations of SU(2), which may include some summands with half-integer[4] spin. Even in systems which do not have a full rotational invariance, the angular momentum $\mathfrak{l}^2$ very often commutes with H, so the notion of spin is defined. Particles described by $\mathscr{R}_l$ are called **fermions** if l is half-integer (electrons, protons and neutrons have spin 1/2), **bosons** if l is integer (photons have spin 1, the Higgs boson has spin 0).

•**Parity.** When the operation of central symmetry (i.e. reverting all coordinates of a particle simultaneously) commutes with the hamiltonian, its eigenvalue ± 1 is a quantum number called the parity.

[4] A half-integer is conventionally defined as 1/2 times an odd integer.

Time reversal

• We consider $\mathscr{V} = \mathbb{C}^N$ in this section. Let C denote the anti-linear operation consisting in taking the complex conjugate of the vectors to its right. Taking the complex conjugate of the Schrödinger equation, we find:

$$\mathrm{i}\hbar\partial_t v_t = H v_t, \qquad -\mathrm{i}\hbar\partial_t C v_t = H^* C v_t.$$

The quantum system is invariant under time reversal $t \to -t$ when there exists a unitary matrix K such that $KH^*K^T = H$. Indeed, if this condition is enforced, the system after time reversal is described by states $v_t' = KCv_t$ and physical observables – including the hamiltonian – $A' = KAK^{-1}$, and it is equivalent to v_t. Since time reversal is involutive, we must have $A'' = A$, and thus by equivalence principles $v'' = \alpha_K v$ for some $\alpha_K \in \mathbb{U}$ and for any $v \in \mathscr{V}$. This imposes $(KC)^2 = \alpha_K$, and since K is unitary and C is the complex conjugation, this implies $K = \alpha_K K^T$. Since the transpose is involutive, we deduce $\alpha_K = \pm 1$. The two cases are possible.

• If $\alpha_K = 1$, we can fix $K = 1$ by an equivalence transformation, i.e. the time reversed state is $v_t' = Cv_t$. The remaining equivalence transformations that preserve this choice are the left multiplication of vectors (and conjugation of physical observables) by real unitary matrices, i.e. orthogonal matrices. In other words, the residual equivalence group is $\mathscr{O}_N(\mathbb{R})$. One can argue this occurs for integer spin or systems with rotational invariance.

• If $\alpha_K = -1$, the dimension N must be even and we can reduce to $K = J_{N/2}$ by an equivalence transformation, i.e. the time reversed state is $v_t' = J_{N/2}Cv_t$. The remaining equivalence transformations that preserve this choice are left-multiplication of vectors (and conjugation of physical observables) by symplectic matrices. The residual equivalence group is $\mathrm{Sp}_N(\mathbb{R})$. In that case, the eigenvalues of H must have even multiplicities. One can argue this occur for half-integer spin without full rotational invariance.

1.4 *From atomic levels to heavy nuclei resonances*

• As an example of quantum system, consider $\mathscr{V} = L^2(\mathbb{R}^3)$ with hamiltonian:

$$H_1 = -\frac{\hbar^2}{2m}\,\Delta_{\vec{x}} - \frac{Z\mathrm{e}^2}{4\pi\epsilon_0\,|\vec{x}|}.$$

where $\vec{x}$ is the coordinate in $\mathbb{R}^3$, $|\vec{x}|$ the distance to the origin, and $\Delta_{\vec{x}}$ the Laplacian in these coordinates. It describes the electrostatic interaction between an electron of mass m and charge $-\mathrm{e}$ and a nuclei of charge $Z\mathrm{e}$ in the referential of center of mass (Bohr model). $\hbar$ is the Planck constant, ϵ_0 the dielectric constant of the void. The hamiltonian can be diagonalized, and all the eigenvalues are simple (therefore the system is completely integrable):

$$\lambda_n = -\frac{Z^2\,R_\infty}{n^2}, \qquad n \in \mathbb{N}^*, \qquad R_\infty = \frac{m\mathrm{e}^4}{32\hbar^2\epsilon_0^2}.$$

R_∞ is the Rydberg constant, its value is approximately $13,6$ eV. For $Z = 1$, this is a good model for the hydrogen atom. $n = 1$ has the lowest energy
• For Z independent electrons attracted by a nuclei of charge Ze, the Hilbert space is $\mathscr{V} = L^2(\mathbb{R}^{3Z})$, the hamiltonian $H_Z = \sum_i H_1^{(i)}$ where $H_1^{(i)}$ is a copy of the hamiltonian H_1 acting on the i-th vector of coordinates in the product $\mathbb{R}^{3Z} = \prod_{i=1}^{Z} \mathbb{R}^3$. Its eigenvalues are

$$\lambda_{n_1,\ldots,n_Z} = -Z^2 R_\infty \Big(\sum_{i=1}^{Z} \frac{1}{n_i^2} \Big), \qquad n_1, \ldots, n_Z \in \mathbb{N}^*.$$

The picture is that one places each of the Z electrons in the energy levels of H_1. Including some **selection rules** constraining the way the levels can be filled, and renormalizing ϵ_0, gives already a good model for the atomic nuclei with $Z \lesssim 20$. For heavier nuclei, the interactions between electrons play an important role, and other terms should be added in the hamiltonian, which cannot be (even approximately) diagonalized easily.
• There exist similar models describing energy levels for protons and neutrons inside nuclei, but they are only predictive for light nuclei[5]. To describe nuclear reactions, one needs to understand how "free" neutrons interact with target heavy nuclei. For certain energies λ_n of the incoming neutron, **resonance** occur: a quasi-bound state neutron-nuclei will form with a lifetime γ_n much larger than the duration of collision. The data of $(E_n + \mathrm{i}\gamma_n)_n$ is important, but since the distribution of energy of the incoming neutrons in a nuclear reaction is wide, their overall statistical properties are more important than their precise individual values. Finding λ_n is similar to diagonalizing the hamiltonian of interactions, which is non self-adjoint: complex eigenvalues $\lambda_n + \mathrm{i}\hbar/\gamma_n$ with $\gamma_n > 0$ being responsible – as seen from the Schrödinger equation – for the decay of the bound state. This picture was well-established in the 30s, but the interactions are so complicated that there is no dream of diagonalizing (even approximately) a realistic hamiltonian. The theoretical physicist Wigner (1951) proposed that, the hamiltonian being so complicated, some statistical properties of its spectrum may well be the same as that of a random one in an ensemble respecting the symmetries of the problem. A hamiltonian is nothing but a large matrix, and it can be argued that rotational invariance and invariance under time reversal of nuclear reactions forces to consider only symmetric matrices. He introduced the very simple – and not at all realistic for nuclear physics – Gaussian ensembles (see § 1.2) of $N \times N$ random matrices in which he could make computations, and he argued that the local properties of the eigenvalues in those ensembles are actually universal, so that they should be compared to statistical properties of the resonance energies.

1.5 *Local spacings*

• **Local spacings for i.i.d.** Let $X_1, \ldots, X_N$ be i.i.d. sequence of real valued r.v., and assume X_1 has a density. We denote $\Omega_{x,s}$ the event "the interval $[x, x+s]$

[5]This important sentence was confusedly omitted during the lecture: nuclear reactions do not involve electrons !

contains none of the X_i's".

$$\mathbb{P}[\Omega_{x,s}] = \sum_{k=0}^{N} \binom{N}{k} \Phi_X(x)(1-\Phi_X(x+s))^{N-k} = \big[1-(\Phi_X(x+s)-\Phi_X(x))\big]^N.$$

When $N \to \infty$, this probability will tend to 0 or 1, unless $(\Phi_X(x+s) - \Phi_X(x))$ is of order $1/N$. So, if we choose $s = \hat{s}/N$, we have $\Phi_X(x+\hat{s}/N) - \Phi_X(x) \sim \hat{s}\rho_X(x)/N$, and:

$$\lim_{N\to\infty} \mathbb{P}[\Omega_{x,\hat{s}/N}] = \exp(-\rho_X(x)\,\hat{s}).$$

So, the local spacings of i.i.d around the point x follow a Poisson law with mean $\rho_X(x)$.

• **Wigner surmise.** Pushing further his idea, Wigner tried to guess the local spacings of heavy nuclei resonances by comparing it to the spacings in the 2×2 Gaussian ensembles ! Accidentally, the $N = 2$ result approximates very well the local spacing distribution of the Gaussian ensembles in the limit $N \to \infty$. The exact result for any N was found by Mehta and can be computed from the results of Chapter 6 and 9. We will do the $N = 2$ computation in the GβE, relying on the result established later (Theorem 4.3) for the p.d.f of the two eigenvalues (λ_1, λ_2):

$$\rho(\lambda_1,\lambda_2) = \frac{1}{Z_\beta(\sigma)}\,|\lambda_1-\lambda_2|^\beta \exp\Big(-\frac{\lambda_1^2+\lambda_2^2}{2\sigma^2}\Big).$$

The normalization constant $Z_\beta(\sigma)$ ensures that $\int_{\mathbb{R}^2} \rho(\lambda_1,\lambda_2)\mathrm{d}\lambda_1\mathrm{d}\lambda_2 = 1$. We compute it by performing the change of variables $(x_1,x_2) = (\lambda_1+\lambda_2, \lambda_1-\lambda_2)$:

$$\begin{aligned} Z_\beta(\sigma) &= \int_{\mathbb{R}^2} |x_2|^\beta \exp\Big(-\frac{x_1^2+x_2^2}{4\sigma^2}\Big)\frac{\mathrm{d}x_1\mathrm{d}x_2}{2} \\ &= (2\sigma)\pi^{1/2}\int_0^\infty x_2^\beta \exp\Big(-\frac{x_2^2}{4\sigma^2}\Big)\mathrm{d}x_2 \\ &= (2\sigma)\pi^{1/2}\cdot\frac{(2\sigma)^{\beta+1}}{2}\int_0^\infty t^{\beta/2-1/2}e^{-t}\mathrm{d}t = \frac{\sqrt{\pi}}{2}\,(2\sigma)^{\beta+2}\,\Gamma(\beta/2+1/2). \end{aligned}$$

The spacing is the r.v. $S = |\lambda_1 - \lambda_2|$, and its p.d.f is:

$$\rho_S(s) = 2\int_0^\infty \rho(\lambda_1+s,\lambda_1)\mathrm{d}\lambda_1 = \frac{(2\sigma)\sqrt{\pi}}{Z_\beta(\sigma)}\,s^\beta \exp\Big(-\frac{s^2}{4\sigma^2}\Big),$$

as follows from the evaluation of the Gaussian integral. The mean spacing is:

$$\mathbb{E}[S] = \int_0^\infty s\rho_S(s)\mathrm{d}s = \int_0^\infty \frac{(2\sigma)^{\beta+1}\,\mathrm{d}t}{2}\,t^{\beta/2}\,e^{-t} = \frac{(2\sigma)\sqrt{\pi}}{Z_\beta(\sigma)}\,\Gamma(\beta/2+1)(2\sigma)^{\beta+2}.$$

Imposing the mean spacing fixes the value of the parameter σ:

$$\mathbb{E}[S] = 1 \quad \Longrightarrow \quad \sigma = \sigma_\beta := \frac{\Gamma(\beta/2+1/2)}{2\,\Gamma(\beta/2+1)}.$$

The corresponding distribution is **Wigner surmise**:

$$\rho_S(s) = C_\beta\, s^\beta\, \exp(-a_\beta s^2),$$

and the values of the constants is listed in the Table below.

symmetry class	β	a_β	C_β
GOE	1	$\pi/4$	$\pi/2$
GUE	2	$4/\pi$	$32/\pi^2$
GSE	4	$64/9\pi$	$262144/729\pi^3$

• **Comparison to nuclear physics.** Nuclear reactions have a time reversal and a rotation invariance. Therefore, the residual equivalence group is $\mathscr{O}(N)$, and the proposition of Wigner was actually to compare the spacings of the resonance energies to those of the eigenvalues of the GOE. This must be done with some care : (1) as we argued in § 1.3, resonance energies associated to different quantum numbers (here: total angular momentum ℓ ; one of the projection of the angular momentum, say ℓ_1 ; and parity p) behave independently ; (2) the average spacing clearly differs between samples, so at least one should first match the mean spacing by rescaling experimental data and fixing $\sigma := \sigma_\beta$. Porter and Rosenzweig (1960) recorded the statistics of consecutive spacings in large sequences of resonance energies with fixed quantum numbers and odd parity. They grouped nuclei in 3 groups (small, medium, and large number of protons), and within each of these groups, built a histogram by averaging the statistics of spacings over quantum numbers and nuclei species (Figure 1). As the size of the nuclei grow, one observes a transition between Poisson statistics – that would occur for i.i.d – and GOE statistics of spacings. The Wigner surmise gives a good prediction for heavy nuclei.
• **Quantum chaos.** Bohigas, Giannoni and Schmit (1980) conjectured that the eigenvalues of the Laplacian in a generic[6] 2d domain have the same local statistics as the eigenvalues of the GOE. Their conjecture was supported by numerical experiments, showing e.g. a good agreement between the Wigner surmise and the distribution of spacings (Figure 2).

1.6 *Zeroes of the Riemann zeta function and GUE*

Lecture 2 (1h30)
October 7th, 2014

Let $1/2 + \mathrm{i}t_n$ the zeroes of Riemann zeta function on the critical line, ordered so that $t_n > 0$ increases with n. Define:

$$F_T(k) = \frac{2\pi}{T \ln T} \sum_{0 \le t_n, t_m \le T} \phi(t_n - t_m)\, \exp(\mathrm{i}k \ln T(t_n - t_m)), \qquad \varphi(x) = \frac{4}{4+x^2}.$$

[6]To my knowledge, the correct assumption for the result to hold is still a matter of discussion among mathematicians.

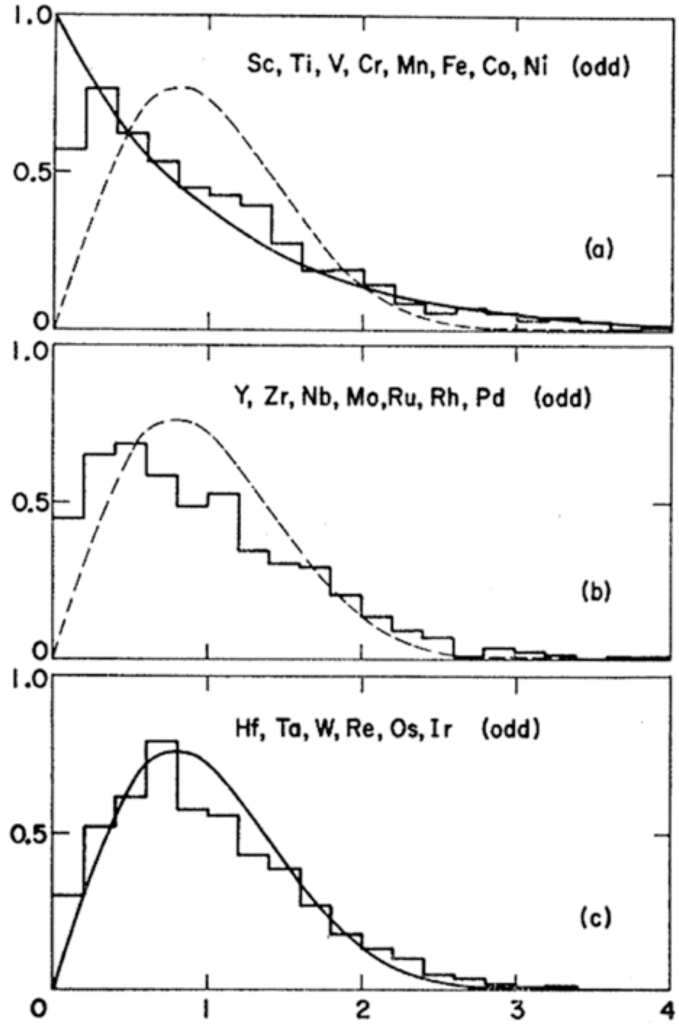

Figure 1: *Reprinted with permission from C.E. Porter and N. Rosenzweig, Phys. Rev.* ***120*** *(1960), 1968-1714 © APS*. The histogram describe averages of spacing statistics – normalized by mean spacing – within 3 groups of nuclei: ^{21}Sc to ^{28}Ni in the first group, ^{39}Y to ^{46}Pd in the second group, ^{72}Hf from ^{77}Ir for the third group. The dashed curve is the Wigner surmise for GOE, the plain line curve on panel (a) is the Poisson law for i.i.d. ^{43}Tc is missing in the second group, because it is an artificial element.

The function $\varphi(x)$ is just here to regularize the sum that we want to study when $T \to \infty$. Assuming Riemann Hypothesis, Montgomery (1970) proved:

$$\forall k \in [-1, 1], \qquad \lim_{T \to +\infty} F_T(k) = |k|$$

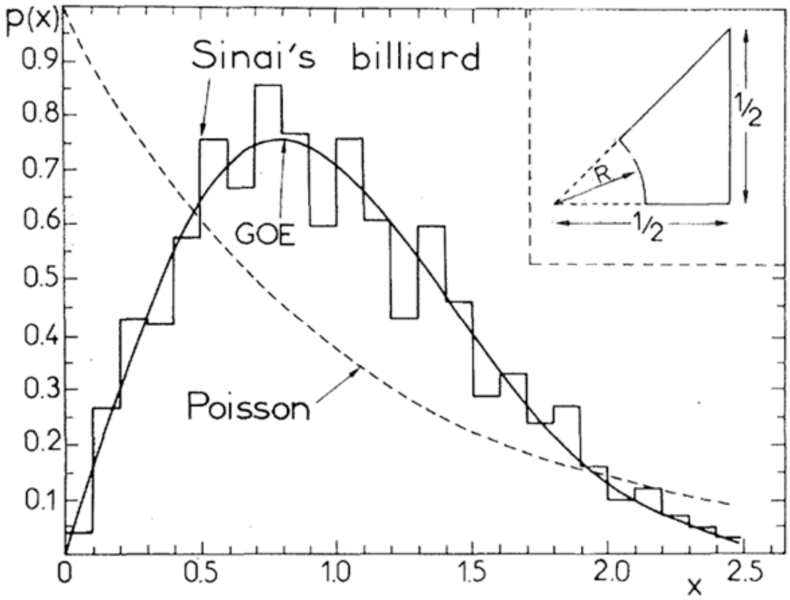

Figure 2: © *Reprinted with permission from O. Bohigas, M.J. Giannoni and C. Schmit, Phys. Rev. Lett.* **52** *(1984) 1-4* ©*APS*. Sinai billiard is the 2d domain on the top right-hand corner. The histogram is the statistics of spacings – normalized by mean spacings – in a large sequence of eigenvalues, and it is compared to Wigner surmise for GOE, and to Poisson law for i.i.d.

uniformly, and conjectured that

$$\forall k \in \mathbb{R}\backslash]-1,1[, \qquad \lim_{T\to+\infty} F_T(k) = 1$$

uniformly on any compact. The conjecture can be reformulated as:

$$\forall k \in \mathbb{R}, \qquad \lim_{T\to+\infty} F_T(k) = F_\infty(k) := 1 + \delta_{k,0} - (1-|k|)\mathbb{1}_{[-1,1]}(k).$$

By convolution of $F_T(k)$ with a continuous function $\hat{f}$ with compact support (or with support in $[-1,1]$), Montgomery's conjecture (or theorem) gives access to statistics of the zeroes of Riemann zeta function probed by the inverse Fourier transform f of $\hat{f}$. Indeed:

$$\int_{\mathbb{R}} F_\infty(k)\hat{f}(-k)\mathrm{d}k = f(0) + \hat{f}(0) - \int_{-1}^{1} (1-|k|)\hat{f}(-k).$$

The last term can be recast in real space by computing the inverse Fourier transform:

$$\begin{aligned}
\int_{-1}^{1}(1-|k|)e^{-2\mathrm{i}\pi kr} &= \int_{0}^{1} 2(1-k)\cos(2\pi kr) = \int_{0}^{1}\frac{2k}{2\pi r}\sin(2\pi kr)\\
&= \Big(\frac{\sin(\pi r)}{\pi r}\Big)^2. \qquad (2)
\end{aligned}$$

Therefore:

$$\int_{\mathbb{R}} F_\infty(k)\hat{f}(-k)\mathrm{d}k = f(0) + \int_{\mathbb{R}} f(r)\Big[1-\Big(\frac{\sin \pi r}{\pi r}\Big)^2\Big]\mathrm{d}r.$$

Besides:

$$\int_{\mathbb{R}} F_T(k)\hat{f}(-k)\mathrm{d}k = \frac{2\pi}{T\ln T}\sum_{0\le t_n,t_m\le T}\varphi(t_n-t_m)f\Big(\frac{\ln T(t_n-t_m)}{2\pi}\Big).$$

We deduce from the uniform convergence in Montgomery's conjecture

$$\begin{aligned}
&\lim_{T\to\infty}\frac{2\pi}{T\ln T}\sum_{0\le t_n,t_m\le T}\varphi(t_n-t_m)f\Big(\frac{\ln T(t_n-t_m)}{2\pi}\Big)\\
=\ & f(0)+\int_{\mathbb{R}} f(r)\Big[1-\Big(\frac{\sin \pi r}{\pi r}\Big)^2\Big]\mathrm{d}r
\end{aligned}$$

Dyson pointed out to Montgomery that the right-hand side is the local correlation function of pairs of eigenvalues in the GUE, also called **sine law** (see Proposition 8.13). This conjecture was later supported by numerical experiments of Odlyzko (e.g. Figure 3), and extended to many-points correlations. The Hilbert-Pólya philosophy – dating back from the 1910s – proposes a reason for the zeroes of the zeta function to be aligned the critical line: it would be so if they were eigenvalues of $1/2+\mathrm{i}Z$ where Z is a self-adjoint operator. Montgomery's result gave a new impetus to this idea, and though such an operator has not been constructed so far, many intuitions coming from random matrix theory results have been imported in number theory to guess properties of L-functions, which sometimes have been later proved independently (see the review of Katz and Sarnak, and the works of Keating).

1.7 *Extreme value statistics for i.i.d.*

For i.i.d.'s, the following result was proved by Gnedenko (1943) and in more restricted generality by Fisher and Tippett (1928):

1.1 THEOREM. *Consider $(X_i)_{1\le i\le N}$ i.i.d sequence of real-valued r.v., and let $M_N=\max_i X_i$ its maximum. If $(M_N)_{N\ge1}$ has a limit law, then its repartition function is – up to affine transformation – of the form:*

- *(Gumbel)* $\Phi_G(t)=\exp(-\exp(-t))$.

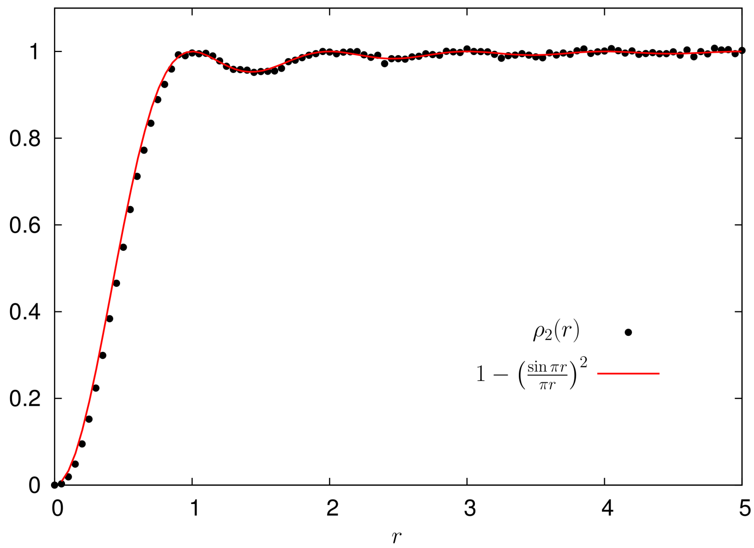

Figure 3: Black dots form a properly renormalized histogram of the pair correlation of a sequence of 10^6 consecutive zeroes of the zeta function, around the 10^{20}-th one (data courtesy of A. Odlyzko, graph courtesy of J.M. Stéphan). The red curve is the sine law found in the GUE.

- *(Fréchet)* $\Phi_{F,\alpha}(t) = \exp(-t^{-\alpha})\mathbb{1}_{\mathbb{R}_+}(t)$ *for some* $\alpha > 0$.
- *(Weibull)* $\Phi_{W,\alpha}(t) = \exp(-(-t)^{\alpha})\mathbf{1}_{\mathbb{R}_-}(t)$ *for some* $\alpha > 0$.

Remark that it is equally relevant for the limit law of the minimum, upon changing X_i to $-X_i$.

We will prove a weaker version of this result, which is enough for practical applications. We assume that X_1 has a continuous density, and write $G(x) = \mathbb{P}[X \geq x]$. We have:

$$\mathbb{P}[M_N \leq x] = \mathbb{P}[X_1, \ldots, X_N \leq x] = \prod_{i=1}^{N} \mathbb{P}[X_i \leq x] = (1 - G(x))^N.$$

This probability tends to 0 or 1 unless we set $x = a_N + b_N t$ with a_N and b_N chosen such that $G(a_N + b_N t)/N \to G^*(t)$ is bounded and non-zero. When it is the case, the reduced r.v. $\hat{M}_N = (M_N - a_N)/b_N$ converges in law to $\hat{M}_\infty$ with repartition function:

$$\mathbb{P}[\hat{M}_\infty \leq t] = \exp(-G^*(t)).$$

We can for instance define a_N uniquely by the condition $G(a_N) = 1/N$. Then, the choice of $b_N = |(\ln G)'(a_N)|^{-1}$ guarantees that $G^*(t)$ will vary at order $O(1)$ when we shift t by order $O(1)$. We say that M_N is of order a_N, with fluctuations of order b_N. We will establish Theorem 1.1 for three types of asymptotic behaviors of the repartition function of X_1:

• **Exponential tails.** $G(x) \asymp x^{\alpha} \exp(-\beta x^{\gamma})$ for some $\alpha \in \mathbb{R}$ and $\beta, \gamma > 0$. This is e.g. the case for waiting times in systems without memory (radioactive emission), for rarely occurring events (Poisson law), high level sport performances, ... We find

$$a_N \asymp (\ln N/\beta)^{1/\gamma}, \qquad b_N \asymp a_N^{1-\gamma} \ll a_N,$$

and we compute:

$$\frac{G(a_N + b_N t)}{G(a_N)} \sim \exp(-\beta\gamma a_N^{\gamma-1} b_N t) \sim \exp(-ct), \qquad N \to \infty.$$

for some constant $c > 0$. Therefore, the fluctuations of the maximum follow – up to an affine transformation – a Gumbel law. It is remarkable that the distribution of the maximum in that case is spiked around $a_N \gg 1$, with a much smaller width $b_N \asymp a_N^{1-\gamma} \ll a_N$. The Gumbel law is e.g. used by insurance companies to estimate their maximal loss if they have to backup customers for rare damages (like serious illnesses or natural catastrophe).

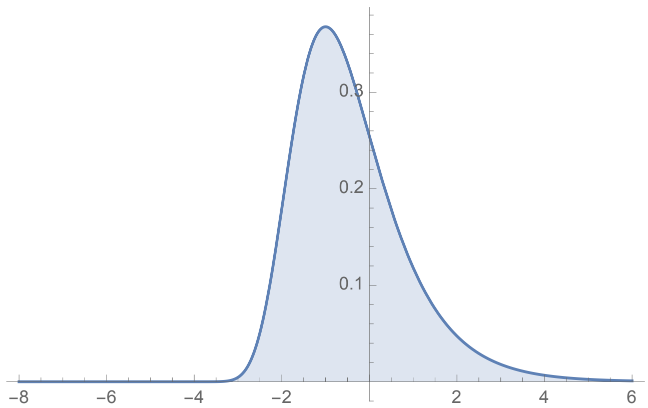

Figure 4: P.d.f. of the Gumbel law: $\Phi'_W(t)$. It has exponential tails at $t \to +\infty$, and doubly-exponential tails on at $t \to -\infty$.

• **Heavy tails.** $G(x)$ decays like a power law: $G(x) \asymp x^{-\alpha}$ for some $\alpha > 0$. This is e.g. the case for the energy released in earthquakes or in other natural catastrophes, or the connectivity of a node in a social network or on Internet. We find:

$$a_N \asymp b_N \asymp N^{\alpha},$$

and we compute:

$$\frac{G(a_N + b_N t)}{G(a_N)} \sim (1 + b_N t/a_N)^{-\alpha} \sim (1 + ct)^{-\alpha} \qquad N \to \infty.$$

for some constant $c > 0$. Therefore, the fluctuations of the maximum follow –

up to an affine transformation – a Fréchet law. Here, it is remarkable that the center of the distribution a_N and the width b_N are of the same order.

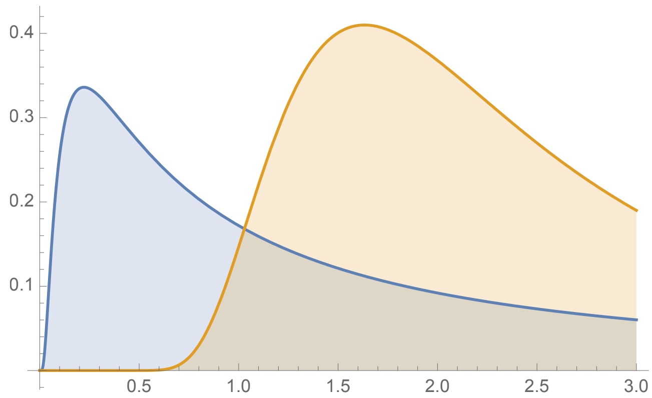

Figure 5: P.d.f. of the Fréchet law: $\Phi'_{W,\alpha}(t)$ for $\alpha = 0.5$ (blue) and 2 (orange). It has exponential tails at $t \to +\infty$, and vanishes non-analyticity at $t \to 0^+$.

• **Bounded r.v.** X_1 is bounded by x_0, and when $x \to x_0^-$, we have $G(x) \asymp (x_0 - x)^\alpha$ for some $\alpha > 0$. This is e.g. the case in trading games, where the losses an individual may take are (usually) bounded by below. We find:

$$x_0 - a_N \asymp b_N \asymp N^{-\alpha},$$

and we compute:

$$\frac{G(a_N + b_N t)}{G(a_N)} \sim \left(1 - \frac{b_N t}{x_0 - a_N}\right)^\alpha \sim (1 - ct)^{-\alpha}, \qquad N \to \infty.$$

for some constant $c > 0$. There, the fluctuations of the maximum follow – up to an affine transformation – a Weibull law. It is not surprising that the maximum is close to (and smaller than) x_0, with fluctuations of order $b_N \ll 1$ when the number N of samples becomes large.

• **Comment.** The energy E released by an earthquake has a heavy tail distribution for an exponent α that can be measured empirically, and thus the maximum energy of a large N of independent earthquakes follow a Fréchet law with center and width $\asymp N^\alpha$. Yet, the human and material damage an earthquake causes is better accounted by its magnitude. The magnitude is proportional to $\ln E$, which has an exponential tail with $\gamma := 1$ and $\beta := \alpha$, hence its maximum is typically of order $\alpha \ln N$, with fluctuations of order $\asymp 1$ described by a Gumbel law.

1.8 *Extreme value statistics for GβE*

The situation is very different for random matrices. The maximum eigenvalue of a matrix of size $N \to \infty$ in the GβE converges to a deterministic constant, and its fluctuations are of order $N^{-2/3}$. The limit law was elucidated by Tracy

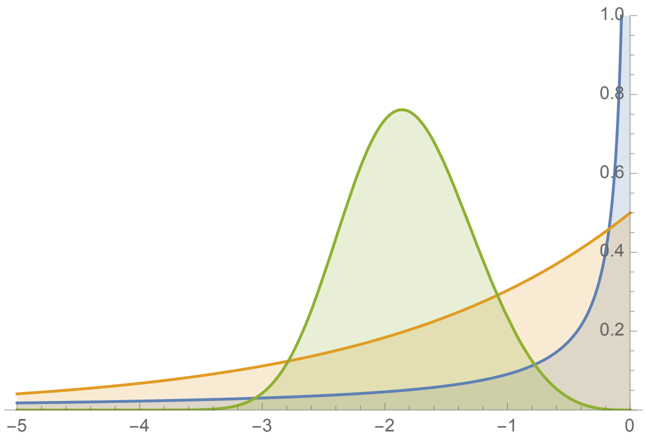

Figure 6: P.d.f. of the Weibull law: $\Phi'_{W,\alpha}(t)$ for $\alpha = 0.25$ (blue), 1 (orange) and 4 (green). It has exponential tails at $t \to -\infty$. At $t \to 0^-$: if $\alpha > 1$, it vanishes diverges a power law ; at $\alpha > 1$, it vanishes like a power law. At $\alpha = 1$, this is the exponential distribution up to reflection $t \to -t$.

and Widom (in 1993 for $\beta = 2$, in 1995 for $\beta = 1,4$), and it involves new special functions. Consider a solution of the Painlevé II (PII) equation

$$q''(s) = 2q^3(s) + sq(s)$$

that matches the growth conditions:

$$q(s) \underset{s\to-\infty}{\sim} \sqrt{-s/2}, \qquad q(s) \underset{s\to+\infty}{\sim} \frac{1}{2\sqrt{\pi}\, s^{1/4}} \exp\Big(-\frac{2s^{3/2}}{3}\Big).$$

PII is an example of non-linear integrable ODE – the meaning of integrability will be explained later in the course (Chapter 9). Flaschka and Newell (1980) have shown existence and uniqueness of $q(s)$, it is called the Hastings-McLeod solution of PII. Then, define:

$$\begin{aligned} E(s) &= \exp\Big(-\frac{1}{2}\int_s^\infty q(t)\mathrm{d}t\Big) \\ H(s) &= \exp\Big(-\frac{1}{2}\int_s^\infty q'(t)^2 - tq^2(t) - q^4(t)\Big) \end{aligned}$$

1.2 theorem. * *In the GβE,*

$$\lim_{N\to\infty} \mathbb{P}_{N,\beta}\Big[\frac{\max_i \lambda_i - 2}{c_\beta N^{2/3}} \geq s\Big] = \mathrm{TW}_\beta(s),$$

where:

$$\begin{aligned} \mathrm{TW}_1(s) &= E(s)H(s), & c_1 &= 1, \\ \mathrm{TW}_2(s) &= H^2(s), & c_2 &= 1, \\ \mathrm{TW}_4(s) &= \frac{E(s)+1/E(s)}{2}\,H(s), & c_4 &= 2^{-2/3}. \end{aligned}$$

TW_β is called the **Tracy-Widom β law**. It is by now well-tabulated[7] and can be used for statistical fits almost as easily as the Gaussian or the Gumbel law.

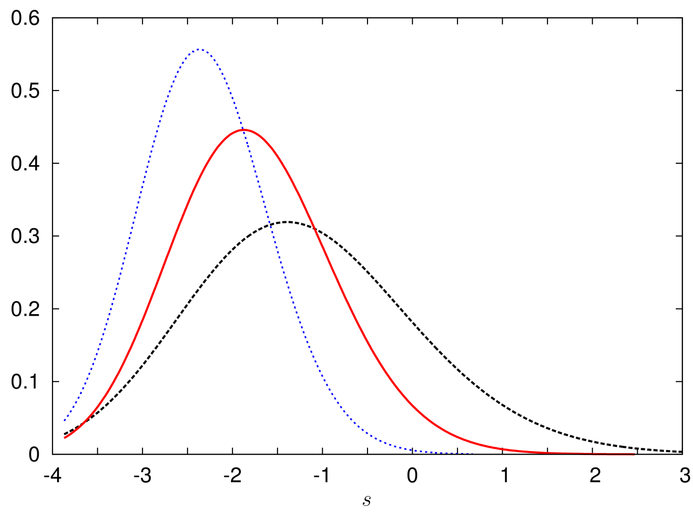

Figure 7: The p.d.f of the Tracy-Widom laws: $-\mathrm{TW}'_\beta(s)$ for $\beta = 1$ (black), $\beta = 2$ (red) and $\beta = 4$ (blue), plot courtesy of J.M. Stéphan. We emphasize that the shape (and thus the mean, skewness, etc.) of the distributions are different for different values of β, thus establishing a clear distinction between the various symmetry classes.

1.9 *Universality of the Tracy-Widom laws*

• The Tracy-Widom β law is expected to be the universal limit law for fluctuations of the maximum of generic ensembles of random matrices in the orthogonal ($\beta = 1$), unitary ($\beta = 2$), or symplectic ($\beta = 4$) symmetry class. But it also found in statistical physics (interface growth, non-intersecting 1d random walks, repulsive particle systems like TASEP,. . .) and in mathematics.
• The physicists Takeuchi and Sano (2010) observed experimentally the Tracy-Widom law in **nematic liquid crystals**. "Nematic" means that the material is made of long molecules whose orientation have long-range correlations, while liquid means that the molecules in the neighborhood of a given one are

[7]However, to evaluate it, one prefers to use another representation of $\mathrm{TW}_\beta(s)$ in terms of Fredholm determinants, obtained in § 8.6.

always changing, i.e. the correlation of positions have short range. In nematic materials, a **topological defect** is a configuration of orientations that winds around a point. In 2d, it occurs for instance when the local orientation rotates like the tangent vector when following a circle throughout the material, in 3d, the Hopf fibration $\phi : \mathbb{S}_3 \to \mathbb{S}_2$ is a configuration of orientations realizing a topological defect. The material studied by Takeuchi and Sano presents two phases: the phase appearing here in gray (resp. black) has a low (resp. high) density of topological defects. If one applies a voltage to the grey phase, one encourages the formation of defects. Once this happens – here at the center of the picture at time $t = 0$ – the black phase takes over the grey phase from this primary cluster of defects. One observes that the interface grows approximately linearly with time t. However, the turbulence driving the system causes some fluctuations from samples to samples. The distribution of this fluctuations of radius from the linear drift fits well with the Tracy-Widom $\beta = 2$ law (GUE), and the matching quality improves for large time (Figure 8). The symmetry class in this case is conditioned by the geometry: a spherical geometry leads to GUE, while a flat interface between two phases would lead to GOE. This result is confirmed in a theoretical model for the interface growth analyzed at $t \to +\infty$ by Sasamoto and Spohn (2010).

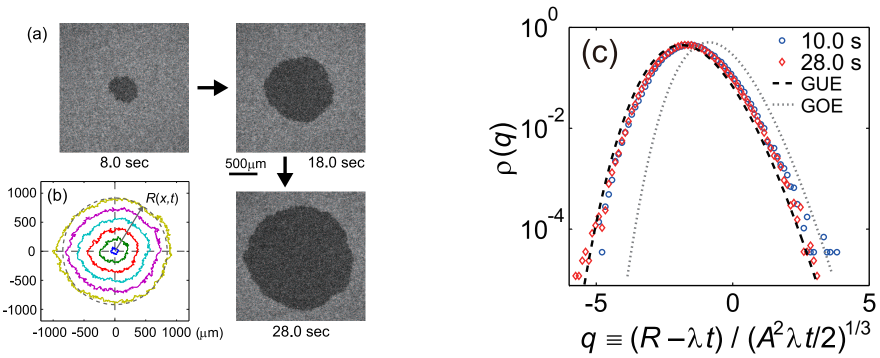

Figure 8: *Reprinted with permission from K. Takeuchi and M. Sano, Phys. Rev. Lett.* ***104*** *230601 (2010)* ©*APS*. Comparison between fluctuations of the radius of a growing interface in nematic liquid crystals and Tracy-Widom laws.

• Let ℓ_N be the length of the longest cycle of a permutation chosen at random, uniformly in $\mathfrak{S}_N$. Vershik and Kerov (1977) proved that $\mathbb{E}[\ell_N] \sim 2\sqrt{N}$ when $N \to \infty$. Baik, Deift and Johansson (2000) went further by exploiting a relation between this problem and an invariant ensemble of random matrices with integer eigenvalues, and showed:

1.3 theorem. *The reduced r.v.* $(\ell_N - 2\sqrt{N})/N^{1/6}$ *converges in law to the Tracy-Widom GUE law.*

1.10 *Conclusion*

Though the hopes concerning universality could look like a wishful thinking in the 50s, they are now supported by many experimental or numerical data, as well as rigorous results. In mathematics, the quest for weaker and weaker assumptions to guarantee that a certain limit law appears led to interesting but difficult problems. Random matrix theory motivated the developments of many techniques of asymptotic analysis since the 80s, useful for other fields (statistical physics, integrable systems, ...). We will illustrate some of them in the course. As a summary, we will encounter:

- for real Wigner matrices: universality of the semi-circle law*, central limit theorem for the fluctuations of linear statistics*, sine kernel distributions in the bulk, Airy kernel distributions at the edge, Tracy-Widom GOE distribution for fluctuations of the maximum.

- for the 1-trace hermitian matrix model : non-universality of the spectral density, central limit theorem only in the 1-cut regime*, sine kernel laws in the bulk**, Airy kernel laws kernel distributions at the generic edge**, Tracy-Widom GUE law for fluctuations of the maximum at a generic edge**.

The statements * will be proved in the course, while ** will be derived in the Gaussian ensembles only (for which explicit computations can be carried out at an elementary level). We will state (part of) the other results, but their proof fall out of the scope of this course.

2 Some matrix inequalities and identities

Lecture 3 (1h30)
October 13th, 2014

2.1 Hadamard inequality

2.1 LEMMA. *Let $v_1, \ldots, v_N \in \mathbb{C}^N$ be N-dimensional vectors.*

$$\det(v_1, \ldots, v_N) \leq \prod_{i=1}^{N} |v_i|.$$

Proof. By homogeneity, we can rescale the vectors and assume $|v_i| = 1$. Let $A = (v_1, \ldots, v_N)$, and $s_1, \ldots, s_N$ its singular values. Since the geometric mean is bounded by the arithmetic mean, we have:

$$|\det A| = \big(\det(AA^{\dagger})\big)^{1/2} = \Big(\prod_{i=1}^{N} s_i\Big)^{1/2} \leq \Big(\frac{1}{N}\sum_{i=1}^{N} s_i\Big)^{N/2}.$$

So:

$$|\det A| \leq \Big(\frac{1}{N}\,\mathrm{Tr}\,AA^{\dagger}\Big)^{N/2} = \Big(\frac{1}{N}\sum_{i=1}^{N} |v_i|^2\Big)^{N/2} = 1,$$

which is the desired inequality. □

Since $|v_i|^2 = \sum_{j=1}^{N} |v_{i,j}|^2 \leq N|v_i|_\infty^2$, we deduce:

2.2 COROLLARY (Hadamard inequality). $\det(v_1, \ldots, v_N) \leq N^{N/2} \prod_{i=1}^{N} |v_i|_\infty$.

It shows that the determinant of a $N \times N$ matrix with bounded coefficients grows much slower than $N!$ when $N \to \infty$.

2.2 Hofmann-Wielandt inequality

Let $A, B \in \mathscr{H}_N(\mathbb{C})$, and $\lambda_1^A \geq \cdots \geq \lambda_N^A$ the eigenvalues of A (idem for B) in decreasing order.

2.3 LEMMA (Hofmann-Wielandt). $\sum_{i=1}^{N} (\lambda_i^A - \lambda_i^B)^2 \leq \mathrm{Tr}\,(A-B)^2$.

The right-hand side of the Hofmann-Wielandt inequality is the matrix L^2 norm $||A-B||_2^2$, and is bounded by $\sqrt{2}||A-B||^2_{2,\mathbb{R}^{N^2}}$.

2.4 COROLLARY. *If we equip $\mathscr{H}_N(\mathbb{C})$ with the L^2-norm, $A \mapsto (\lambda_1^A, \ldots, \lambda_N^A)$ is 1-Lipschitz.*

A fortiori, eigenvalues of a hermitian matrix are continuous functions of its entries[8]. This also justifies, if M is a $\mathscr{H}_N(\mathbb{C})$-valued r.v. in some probability space, that its eigenvalues are r.v. in the same probability space since they are measurable functions of M.

[8] Continuity can be proved alternatively by remarking that the coefficients of the characteristic polynomial are continuous functions of the entries of A, and the roots of a polynomial are continuous functions (use Cauchy residue formula to locate the roots !) of the coefficients.

Proof. It is equivalent to prove $\mathrm{Tr}(AB) \le \sum_{i=1}^{N} \lambda_i^A \lambda_i^B$. Let us diagonalize A and B: we denote U a unitary matrix of change of eigenbasis from A to B, and $\Lambda^A = \mathrm{diag}(\lambda_1^A, \ldots, \lambda_N^A)$ and similarly Λ^B. Then:

$$\mathrm{Tr}(AB) = \mathrm{Tr}(\Lambda^A U^{-1} \Lambda^B U) = \sum_{i,j=1}^{N} \lambda_i^A \lambda_j^B |U_{i,j}|^2.$$

This is a linear (hence convex) function of $S_{i,j} = |U_{i,j}|^2$, and we would like to maximize it over the (convex) set $\mathcal{S}$ of doubly-stochastic matrices, i.e. $S \in \mathscr{M}_N(\mathbb{R})$ satisfying

$$\forall i, j, \qquad S_{i,j} \ge 0, \qquad \sum_{l=1}^{N} S_{i,k} = 1, \qquad \sum_{k=1}^{N} S_{k,j} = 1,$$

By convexity, the maximum is achieved in the set of extreme points of $\mathcal{S}$, i.e. the set of permutation matrices. We thus need to find the maximum over $\sigma \in \mathfrak{S}_N$ of:

$$f(\sigma) = \sum_{i=1}^{N} \lambda_i^A \lambda_{\sigma(i)}^B.$$

The announced result follows if we show that the maximum of f_σ is achieved for $\sigma = \mathrm{id}$. Due to the ordering of eigenvalues, we observe when $i_0 < j_0, k_0$:

$$(3) \qquad 0 \le (\lambda_{i_0}^A - \lambda_{k_0}^A)(\lambda_{i_0}^B - \lambda_{j_0}^B) = \lambda_{i_0}^A \lambda_{i_0}^B + \lambda_{k_0}^A \lambda_{j_0}^B - (\lambda_{k_0}^A \lambda_{i_0}^B + \lambda_{i_0}^A \lambda_{j_0}^B).$$

Now, if $\sigma \neq \mathrm{id}$, let $i_0 := \min \{ i \in [\![1, N]\!], \quad \sigma(i) \neq i \}$. By minimality, $j_0 := \sigma(i_0)$ and $k_0 := \sigma^{-1}(i_0)$ are both greater than i_0. The two last terms of (3) occur in $f(\sigma)$, and thus can be replaced by the two first terms:

$$
\begin{aligned}
(4) \qquad\qquad f_\sigma &= \sum_{i=1}^{i_0 - 1} \lambda_i^A \lambda_i^B + \lambda_{i_0}^A \lambda_{j_0}^A + \cdots + \lambda_{k_0}^A \lambda_{i_0}^B + \cdots \\
(5) \qquad\qquad &\le \sum_{i=1}^{i_0} \lambda_i^A \lambda_i^B + \cdots = f_{\widetilde{\sigma}}.
\end{aligned}
$$

$\widetilde{\sigma}$ is a new permutation that is identity of the larger segment $[\![1, i_0]\!]$, sends k_0 to j_0, and coincides with σ otherwise. By induction, one finally arrives at $f_\sigma \le f_{\mathrm{id}}$. □

2.3 *Convex matrix functions from eigenvalues*

The concentration results we will derive in § 3.3 can be applied only to convex functions of the matrix entries. To some extent, eigenvalues can produce convex functions. If $A \in \mathscr{H}_N(\mathbb{C})$, λ_i^A still denote the eigenvalues in decreasing order.

2.5 lemma (Min-Max characterization of eigenvalues). *For any $i \in [\![1, N]\!]$,*

$$\lambda_i^A = \sup_{V \in \mathrm{Gr}_i(\mathbb{C}^N)} \inf_{\substack{v \in V \\ |v|_2 = 1}} (v|Av) = \inf_{V \in \mathrm{Gr}_{N-i+1}(\mathbb{C}^N)} \sup_{\substack{v \in V \\ |v|_2 = 1}} (v|Av).$$

Proof. Since $\lambda_i^A = -\lambda_{N-1+i}^{-A}$, it is enough to prove the first equality. Let us denote $(e_1, \ldots, e_N)$ an orthonormal basis of eigenvectors of A for its eigenvalues λ_i^A (in decreasing order). By taking $V = \mathrm{vect}(e_1, \ldots, e_i)$, we see that

$$\lambda_i^A \leq \sup_{V \in \mathrm{Gr}_i(\mathbb{C}^N)} \inf (v|Av).$$

So, it remains to show the reverse inequality. For dimensional reasons, any $V \in \mathrm{Gr}_i(\mathbb{C}^N)$ will have a non-zero intersection with the codimension $(i-1)$ subspace $\mathrm{vect}(e_i, \ldots, e_N)$: let $v = \sum_{j=i}^N v_j e_j \in V \setminus \{0\}$. Then:

$$(v|Av) = \sum_{j=i}^{N} \lambda_j^A |v_j|^2 \leq \lambda_i^A.$$

□

In particular, $\lambda_{\max}^A = \lambda_1^A = \sup_{|v|_2=1}(v|Av)$ is the sup of linear (hence convex) functions, so $\lambda_{\max}^A$ is a convex function of the entries of A. The other eigenvalues fail to be convex functions, but something can be said for the partial sums of ordered eigenvalues. First, we define the partial trace of a matrix $A \in \mathscr{M}_N(\mathbb{C})$ over a subspace $V \subseteq \mathbb{C}^N$, denoted $\mathrm{Tr}_V\, A$. Choosing an orthonormal basis $(v_i)_i$ of V, we set:

$$\mathrm{Tr}_V\, A := \sum_i (v_i|Av_i),$$

and this actually does not depend on the choice of the orthonormal basis.

2.6 lemma (Partial sums). *For any $k \in [\![1, N]\!]$:*

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i^A &= \sup_{V \in \mathrm{Gr}_k(\mathbb{C}^N)} \mathrm{Tr}_V\, A, \\ \sum_{i=k+1}^{N} \lambda_i^A &= \inf_{V \in \mathrm{Gr}_k(\mathbb{C}^N)} \mathrm{Tr}_V\, A. \end{aligned}$$

Proof. Again, it is enough to prove the first equality, and we will proceed by proving a double inequality. We keep the notations of the previous proof. Since $\sum_{i=1}^k \lambda_i^A$ is the partial trace over $\mathrm{vect}(\mathrm{e}_1, \ldots, \mathrm{e}_k)$, we have $\sum_{i=1}^k \lambda_i^A \leq \sup_{V \in \mathrm{Gr}_k(\mathbb{C}^N)} \mathrm{Tr}_V\, A$. We now show the converse by induction on the dimension N. For $N = 1$, the result is obvious. Assume it holds in dimension $N-1$. Let $W = \mathrm{vect}(e_2, \ldots, e_N)$. For any $V \in \mathrm{Gr}_k(\mathbb{C}^N)$, $W \cap V$ must contain a subspace V' of dimension $(k-1)$. Let $v_0 \in V$ be a unit vector such that $V \simeq \mathbb{C} \cdot v_0 \oplus V'$

is an orthogonal sum. We have:

$$\mathrm{Tr}_V\, A = \mathrm{Tr}_{V'} A + (v_0|Av_0),$$

where V' is considered as a subspace of $W \simeq \mathbb{C}^{N-1}$, for which we can apply the induction hypothesis (remark that the ordered eigenvalues of A for the eigenvectors in W are $\lambda_2^A, \ldots, \lambda_N^A$). Besides, we have the obvious bound $(v_0|Av_0) \leq \lambda_1^A$. Therefore:

$$\mathrm{Tr}_V\, A \leq \Big(\sum_{i=2}^{k} \lambda_i^A\Big) + \lambda_1^A,$$

which is the desired result in dimension N. We conclude by induction □

Since the partial sums of eigenvalues in decreasing order are suprema of linear (hence convex) functions of A, we deduce:

2.7 corollary. *For any $k \in [\![1, N]\!]$, the function $A \mapsto \sum_{i=1}^{k} \lambda_i^A$ is convex over $\mathscr{H}_N(\mathbb{C})$.*

2.4 *Convex matrix functions from convex real functions*

Another source of convex functions of the matrix entries arise from convex functions on $\mathbb{R}$.

2.8 lemma (Klein lemma). *If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, $A \mapsto \mathrm{Tr}\, f(A)$ is a convex function over $\mathscr{H}_N(\mathbb{C})$.*

Proof. We first assume that f is twice differentiable. In that case, convexity implies:

$$g(x,y) := f(x) - f(y) - (x-y)f'(y) \geq \inf c_f (x-y)^2 \qquad c_f = \inf f''/2 \geq 0.$$

Let us apply this function to two matrices $X, Y \in \mathscr{H}_N(\mathbb{C})$ and compute the trace. To compute the terms involving both X and Y in the trace, we need to introduce as in the proof of Lemma 2.3 a matrix U of change of basis between eigenvectors of X and Y.

$$\mathrm{Tr}\, g(X,Y) = \Big(\sum_{i=1}^{N} f(\lambda_i^X) - f(\lambda_i^Y) + \lambda_i^Y f'(\lambda_i^Y)\Big) - \sum_{i,j=1}^{N} |U_{i,j}|^2 \lambda_i^X f'(\lambda_j^Y).$$

We can for free substitute the identity $\sum_{j=1}^{N} |U_{i,j}|^2 = 1$ to convert the simple sum to a double sum, and we find:

$$\mathrm{Tr}\, g(X,Y) = \sum_{i,j=1}^{N} |U_{i,j}|^2\, g(\lambda_i^X, \lambda_j^Y) \geq 0.$$

Now, to get rid of the term involving f', let us apply this inequality firstly to $(X,Y) \to (A, (A+B)/2)$, secondly to $(X,Y) = (B, (A+B)/2)$, and sum up the two inequalities. The result is:

$$\mathrm{Tr}\, f\Big(\frac{A+B}{2}\Big) \leq \frac{1}{2}\Big(\mathrm{Tr}\, f(A) + \mathrm{Tr}\, f(B)\Big).$$

A function satisfying this property is said **midpoint convex**, and since here $A \mapsto \mathrm{Tr}\, f(A)$ is continuous (because we assume $f\ \mathbb{R} \to \mathbb{R}$ convex hence continuous), this is known to imply convexity. Indeed, using repeatedly this inequality, one can show $\mathrm{Tr}\, f(tX + (1-t)Y) \leq t\,\mathrm{Tr}\, f(X) + (1-t)\mathrm{Tr}\, f(Y)$ for any $t \in [0,1]$ that has a finite dyadic expansion, i.e. $t = \sum_{k=1}^{K} \varepsilon_k\, 2^{-k}$ with some $\varepsilon_k \in \{0,1\}$. Then, one can approximate any $t \in [0,1]$ by truncations of its dyadic expansion, and using the continuity of $\mathrm{Tr}\, f$, conclude to the convexity inequality.

If the assumption of twice differentiability of f is dropped, we can still approximate (for simple convergence) by smooth functions by performing a convolution with a smooth non-negative function h_σ tending to a Dirac when $\sigma \to 0$. For instance:

$$f_\sigma(x) = \int_{\mathbb{R}} f(x+y)\, h_\sigma(-y), \qquad h_\sigma = \frac{\exp(-x^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}}.$$

The positivity of h_σ ensures that f_ε is also convex – as an (infinite) linear combination of convex functions. So, $A \mapsto \mathrm{Tr}\, f_\sigma(A)$ is a convex function for any $\sigma > 0$, and by taking the limit $\sigma \to 0$, one gets the convexity inequality for $A \mapsto \mathrm{Tr}\, f(A)$ in full generality. $\square$

It is also useful to see how the Lipschitz behavior of f carries on to matrix functions:

2.9 lemma. *If $f : \mathbb{R} \to \mathbb{R}$ is k-Lipschitz, then $X \mapsto N^{-1}\,\mathrm{Tr}\, f(X)$ is $\sqrt{2/N}k$-Lipschitz for the L^2-norm on $\mathscr{H}_N(\mathbb{C})$.*

Proof. Let $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ defined by $F(x_1, \ldots, x_N) = N^{-1} \sum_{i=1}^{N} f(x_i)$. We have:

$$\big|\tilde{f}(x) - \tilde{f}(y)\big| \leq N^{-1} \sum_{i=1}^{N} k|x_i - y_i| \leq N^{-1/2} k\, |x-y|^2$$

by Cauchy-Schwarz inequality. Then, if we denote $F : M \mapsto \tilde{f}(\Lambda^{(M)})$, we have:

$$|F(M) - F(M')| \leq N^{-1/2} k \Big(\sum_{i=1}^{N} \big(\Lambda_i^{(M)} - \Lambda_i^{(M')}\big)^2\Big)^{1/2} \leq N^{-1/2} k \big(\mathrm{Tr}(M-M')^2\big)^{1/2}$$

by Hofmann-Wielandt inequality. $\square$

2.5 *Block determinants and resolvent*

Imagine a matrix $M \in \mathscr{M}_N(\mathbb{C})$ is cut into 4 blocks by a decomposition $N = r + (N-r)$:

$$M = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right).$$

Assume A is invertible. A useful trick to compute its determinant in terms of the blocks is to use the factorization into block upper/lower triangular matrices:

$$M = \left(\begin{array}{c|c} A & 0_{r,N-r} \\ \hline C & D - BA^{-1} \end{array}\right)\left(\begin{array}{c|c} 1_r & A^{-1}B \\ \hline 0_{N-r,r} & 1_{N-r} \end{array}\right).$$

Then, by multiplicativity of the determinant:

2.10 LEMMA (Block determinant). $\det M = \det A \cdot \det(D - BA^{-1}C)$.

The result is not 2×2 Cramer's rule $\det(AD - BC)$ – which does not make sense anyway because dimensions of blocks do not match – but close !

One application we will use in the study of Wigner matrices is a nesting property of the resolvent of a matrix.

2.11 DEFINITION. Let $M \in \mathscr{M}_N(\mathbb{C})$. The function $R_M : z \mapsto (z - M)^{-1}$ defined for $z \in \mathbb{C} \setminus \mathrm{Sp}\, M$ is the **resolvent** of M.

$R_M(z)$ is a matrix-valued holomorphic function[9] of z in the complement of the spectrum of M, and it behaves like $R_M(z) \sim N/z$ when $z \to \infty$.

2.12 LEMMA (Resolvent on the diagonal). *Let $M \in \mathscr{H}_N(\mathbb{C})$, and $i \in [\![1, N]\!]$. We denote $M[i] \in \mathscr{M}_{N-1}(\mathbb{C})$ the matrix M with its i-th column and i-th row removed, and $v_i^M \in \mathbb{C}^{N-1}$ the i-th column of M with its N-th entry removed. We have for any $z \notin \mathrm{Sp}\, M$:*

$$R_M(z)_{i,i} = \frac{1}{z - M_{i,i} - \big(v_i^M \big| (z - M[i])^{-1} v_i^M\big)}.$$

This identity can be rewritten:

$$R_M(z)_{i,i} = \frac{1}{z - M_{i,i} - \big(v_i^M \big| R_{M[i]}(z)\, v_i^M\big)}.$$

In the right-hand side appears the resolvent of the matrix $M[i]$ of size $(N-1)$, which was a submatrix of M. This equation thus offers a recursive (on the size N) way to control the resolvent.

Proof. By Cramer's rule for the computation of the inverse:

$$R_M(z)_{i,i} = \frac{\det(z - M[i])}{\det(z - M)}.$$

Besides, after permutations of rows and columns to bring the i-th row and

[9]This means that each entry is a holomorphic function.

i-th column to be the last ones (that does not change the determinant), the denominator reads:

$$\det(z - M) = \det \begin{pmatrix} z - M[i] & -v_i \\ -v_i^\dagger & z - M_{i,i} \end{pmatrix}.$$

To write the bottom-left corner of the matrix, we have used the fact that M is hermitian. Since $z \notin \operatorname{Sp} M$, $(z - M)$ is be invertible, thus $z - M[i]$ must also be invertible. So, we can compute this determinant using its $(N-1)+1$ blocks:

$$\begin{aligned} \det(z - M) &= \det(z - M[i]) \cdot \left(z - M_{i,i} - v_i^\dagger (z - M[i])^{-1} v_i\right) \\ &= \det(z - M[i]) \cdot \left(z - M_{i,i} - (v_i | (z - M[i])^{-1} v_i)\right). \end{aligned}$$

The result follows. □

3 Wigner matrices

Lecture 4 (1h30)
October 14^{th}, 2014

Let M_N be a real Wigner matrix of size N (§ 1.2), and denote its eigenvalues:

$$\Lambda^{(M_N)} = (\lambda_1^{(M_N)}, \ldots, \lambda_N^{M_N}).$$

The **empirical measure** is the random probability measure:

$$L_N^{(M_N)} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i^{(M_N)}}$$

and the **linear statistics** probed by a test function f is the r.v.

$$L_N^{(M_N)}[f] = \int f(x) \mathrm{d}L_N^{(M_N)}(x).$$

This chapter is devoted to the proof of Wigner's theorem:

3.1 THEOREM (Universality of semi-circle law). *$L_N^{(M_N)}$ converges for the weak topology, in probability, to the probability measure μ_{sc} with density*

$$\rho_{\mathrm{sc}}(x) = \mathbf{1}_{[-2,2]}(x)\sqrt{4-x^2}\mathrm{d}x.$$

In other words, for any function f continuous bounded, and any $\epsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\Big[\Big|\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i^{(M_N)}) - \int_{-2}^{2} f(x)\,\rho_{\mathrm{sc}}(x)\mathrm{d}x\Big| > \epsilon\Big] = 0.$$

μ_{sc} is called the **semi-circle law**, referring to the shape of the graph of its density (Figure 9).

3.2 LEMMA. *The odd moments of the semi-circle law vanish, and the even moments are:*

$$\forall k \geq 0, \qquad \mu_{\mathrm{sc}}[X^{2k}] = \mathrm{Cat}(k) = \frac{1}{k+1}\binom{2k}{k}.$$

$\mathrm{Cat}(k)$ are the Catalan numbers: they enumerate planar pairings, trees, and many other combinatorial objects. As we shall see in Chapter **??**, their occurrence in RMT is not accidental.

Proof. The odd moments vanish since ρ_{sc} is even. For the even moments, we use the change of variable $x = 2\cos\theta$:

$$\mu_{\mathrm{sc}}[X^{2k}] = \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}\,\mathrm{d}x = \frac{2^{2k+2}}{2\pi}\int_{0}^{\pi} \cos^{2k}\theta\sin^2\theta = 2^{2k+1}(m_k - m_{k+1}),$$

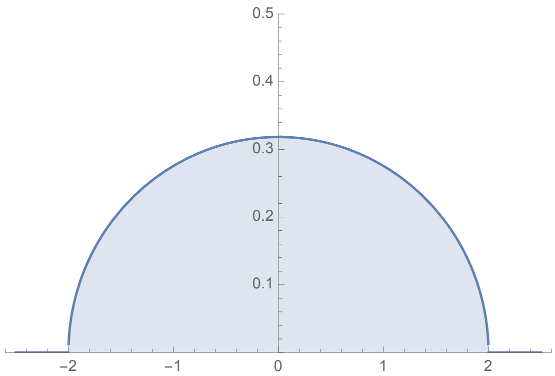

Figure 9: Plot of the density ρ_{sc} of Wigner semi-circle law.

where:

$$m_k := \int_0^\pi \frac{\mathrm{d}\theta}{\pi} \cos^{2k}\theta = 2^{-2k} \sum_{m=0}^{2k} \binom{2k}{m} \int_0^\pi \frac{\mathrm{d}\theta}{\pi} e^{2\mathrm{i}\theta(k-m)} \mathrm{d}\theta = 2^{-2k} \binom{2k}{k}.$$

Hence:

$$\mu_{\mathrm{sc}}[X^{2k}] = 2\Big(\frac{(2k)!}{k!^2} - \frac{1}{4}\frac{(2k+2)!}{(k+1)!^2}\Big) = \frac{2k!}{k!(k+1)!}$$

after reducing at the same denominator. □

Since in Wigner matrices the distribution of the entries has not been specified, Theorem 3.1 can be considered as our first universality result. The proof relies on two important tools, namely the notion of Stieltjes transform, and concentration inequalities. They both have their own interest beyond random matrix theory, so we shall present them independently in § 3.2 and 3.3.

3.1 *Stability under perturbations*

With the Hofmann-Wielandt inequality, one can show that many perturbations or truncations of random matrices do not affect limit laws for eigenvalue statistics. Take $F_N : \mathbb{R}^N \to \mathbb{C}$ a k_N-Lipschitz function, and Δ_N a random matrix. We have:

$$\begin{aligned} \big|F_N(\Lambda^{(M_N+\Delta_N)}) - F_N(\Lambda^{(M_N)})\big| &\leq k_N \Big(\sum_{i=1}^N (\lambda_i^{(M_N+\Delta_N)} - \lambda_i^{(M_N)})^2\Big)^{1/2} \\ &\leq k_N (\mathrm{Tr}\,\Delta_N^2)^{1/2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{P}\Big[\big|F_N(\Lambda^{(M_N+\Delta_N)}) - F_N(\Lambda^{(M_N)})\big| > \epsilon\Big] &\leq \mathbb{P}\Big[\mathrm{Tr}\,(k_N\Delta_N)^2 \geq \epsilon^2\Big] \\ &\leq (k_N/\epsilon)^2\, \mathbb{E}\big[\mathrm{Tr}\,\Delta_N^2\big], \end{aligned} \tag{6}$$

where the last inequality follows from Chebyshev inequality. Subsequently, if $F(\Lambda^{(M_N)})$ converges in probability and $k_N^2 \mathbb{E}\big[\mathrm{Tr}\,\Delta_N^2\big] \to 0$ when $N \to \infty$, then $F(\Lambda^{(M_N+\Delta_N)})$ has the same limit.

This setting can be specialized to linear statistics: if $f : \mathbb{R} \to \mathbb{C}$ is a k-Lipschitz function, we set

$$F_N(x_1, \ldots, x_N) := \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

Then, F is Lipschitz with constant $k_N = kN^{-1/2}$ (by application of Cauchy-Schwarz inequality). So, convergence in probability of $L_N^{(M_N)}[f]$ is unchanged under perturbations Δ_N satisfying:

$$(7) \qquad \mathbb{E}\big[\mathrm{Tr}\,\Delta_N^2\big] \ll N.$$

For instance, this applies when Δ_N is a diagonal matrix of mean 0 whose entries have variance $o(1)$.

3.3 LEMMA. *It is enough to prove Theorem 3.1 for Wigner matrices with vanishing diagonal.*

Another consequence is that, in a random matrix M_N whose entries are of order $N^{-1/2}$, we can actually replace $o(N^2)$ entries by their mean without changing the convergence in probability of linear statistics ! A less shocking restatement is that we can make deterministic a small (= tending to 0 when $N \to \infty$) fraction of the entries to our wish. (7) is also satisfied if Δ_N is a random matrix with bounded rank matrix and spectral radius $\ll N$.

A preliminary result in the same spirit is:

3.4 LEMMA. *It is enough to show Theorem 3.1 assuming the existence of a uniform $C > 0$ such that $|(M_N)_{i,j}| \leq C/\sqrt{N}$*

In particular, in the definition of Wigner matrices, one could waive the conditions that all moments of X_1 and $Y_{1,2}$ are finite.

Proof. Here, M_N denotes a Wigner matrix. Let $C > 0$ and denote $\mathcal{B}(i,j;C)$ the event $\{\sqrt{N}|(M_N)|_{i,j} < C\}$. By Chebyshev inequality and the variance conditions in the definition of a Wigner matrix:

$$\forall i,j \in [\![1,N]\!], \qquad \mathbb{P}[\mathcal{B}(i,j;C)] \leq \frac{2C^2}{N}.$$

We define a new Wigner matrix which has bounded entries:

$$(\widetilde{M}_{N;C})_{i,j} := \frac{(M_N)_{i,j}\mathbf{1}_{\mathcal{B}(i,j;C)} - \mathbb{E}\big[(M_N)_{i,j}\mathbf{1}_{\mathcal{B}(i,j;C)}\big]}{\sigma_{i,j;C}},$$

where $\sigma_{i,j;C}^2 = \mathrm{Var}\big[N^{1/2}(M_N)_{i,j}\mathbf{1}_{\mathcal{B}(i,j,C)}\big]$ was included to match the variance requirements in the definition of a Wigner matrix, and we have $\lim_{C\to\infty} \sigma_{i,j;C} =$

$1 + \delta_{i,j}$. We consider $\Delta_{N;C} = \widetilde{M}_{N;C} - M_N$ and compute:

$$\begin{aligned} \frac{1}{N}\,\mathbb{E}\big[\mathrm{Tr}\,\Delta_N^2\big] & \leq \frac{1}{N}\sum_{i,j=1}^{N} \mathrm{Var}\Big[(M_N)_{i,j}\cdot\Big(\mathbf{1}_{\mathcal{B}(i,j;C)^c} + (\sigma_{i,j;C}^{-1} - 1)\mathbf{1}_{\mathcal{B}(i,j;C)}\Big)\Big] \\ & \leq \frac{1}{N}\sum_{i,j=1}^{N} \mathrm{Var}\big[(M_N)_{i,j}\big] + CN \max_{(i,j)=(1,2),(1,1)} \big|\sigma_{i,j;C}^{-1} - (1+\delta_{i,j})\big|. \end{aligned}$$

Since the off-diagonal entries (resp. the diagonal entries) of M_N are i.i.d. with variance bounded by $2/N$, and there are N^2 terms in the sum, the first sum can be made smaller than any $\epsilon > 0$ by choosing $C_\epsilon > 0$ independent of N but large enough. Besides, since $\sigma_{i,j;C}$ converges to $(1+\delta_{i,j})$, the second term can also be made smaller than ϵ if we take a maybe larger value of C_ϵ. Applying (6) for functions f bounded by 1 and with Lipschitz constant $k \leq 1$, we obtain:

$$\big|L_N^{(\widetilde{M}_{N;C_\epsilon})}[f] - L_N^{(M_N)}[f]\big| \leq 2\epsilon$$

uniformly in N. Therefore, the weak convergence of $L_N^{(\widetilde{M}_{N;C})}$ implies the weak convergence of the original empirical measure $L_N^{(M_N)}$. $\square$

3.2 *Stieltjes transform*

3.5 definition. If $\mu \in \mathcal{M}^1(\mathbb{R})$, we define its **Stieltjes transform** W_μ as a function of $z \in \mathbb{C}\setminus\mathbb{R}$:

$$W_\mu(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu(x)}{z-x}.$$

If $\mathrm{supp}\,\mu \subseteq A$, then W_μ is actually defined for $z \in \mathbb{C}\setminus(\mathrm{supp}\,\mu)$.

As an illustration, let us compute the Stieltjes transform of the semi-circle law:

3.6 lemma. $W_{\mu_{\mathrm{sc}}}(z) = \frac{z-\sqrt{z^2-4}}{2}$.

In this expression, the sign of the squareroot is chosen such that $z \mapsto \sqrt{z}$ is holomorphic in $\mathbb{C}\setminus\mathbb{R}_-$, and it sends $\mathbb{R}_+$ to $\mathbb{R}_+$.

Proof. There are basically two methods to compute the Stieltjes transform of a measure with an explicit density. The first one exploits the fact that the coefficients of its asymptotic expansion at $z \to \infty$ away from the real axis is given in terms of the moments:

$$W_{\mu_{\mathrm{sc}}}(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_{\mathrm{sc}}(x)}{z-x} = \sum_{m\geq 0} \frac{1}{z^{m+1}} \int_{\mathbb{R}} x^m \mathrm{d}\mu_{\mathrm{sc}}(x).$$

The series with moments given by Lemma 3.2 gives the result announced. The second method consists in representing the integral over $\mathbb{R}$ as a contour integral around the support, and computing this integral by moving the contours in the complex plane and using Cauchy residue formula. Here, we find:

$$W_{\mu_{\mathrm{sc}}}(z) = \int_{-2}^{2} \frac{\mathrm{d}x}{2\pi} \, \frac{\sqrt{4-x^2}}{z-x} = \oint \frac{\mathrm{d}x}{4\mathrm{i}\pi} \, \frac{\sqrt{x^2-4}}{x-z},$$

and therefore:

$$W_{\mu_{\mathrm{sc}}}(z) = \operatorname*{Res}_{x \to z, \infty} \frac{\sqrt{x^2-4}}{x-z} = \frac{-\sqrt{z^2-4}+z}{2}.$$

□

The Stieltjes transform has the following properties:

(i) $z \mapsto W_\mu(z)$ is a holomorphic function of $z \in \mathbb{C} \setminus (\operatorname{supp}\mu)$.

(ii) $W_\mu(z) \sim 1/z$ when $z \to \infty$ and $\operatorname{Re} z$ remains bounded away from 0.

(iii) $(2\mathrm{i}\pi)^{-1}\big(W_\mu(a-\mathrm{i}b) - W_\mu(a+\mathrm{i}b)\big)$ defines a positive measure on $a \in \operatorname{supp}\mu$ in the limit $b \to 0^+$.

and the limit measure is μ. To show (iii), we first observe a probabilistic interpretation of the Stieltjes transform:

$$a \mapsto \frac{\operatorname{Im} W_\mu(a-\mathrm{i}b)}{\pi} = \frac{W_\mu(a-\mathrm{i}b) - W_\mu(a+\mathrm{i}b)}{2\mathrm{i}\pi} = \int_{\mathbb{R}} \frac{\mathrm{d}\mu(x)}{\pi((x-a)^2+b^2)}$$

is the density of the probability measure $\mu * C_b$, where C_b is a Cauchy law. And, for any continuous bounded function f, $\mu * C_b[f] = \mu[f * C_b]$. Since $f * C_b$ converges pointwise to f and remains uniformly bounded, we have $\lim_{b\to 0} \mu[f * C_b] = \mu[f]$ by dominated convergence.

Conversely, any function W satisfying these three properties is the Stieltjes transform of a probability measure. (iii) allows the reconstruction of the measure μ from its Stieltjes transform, and (ii) is equivalent to the requirement that the total mass of μ is 1.

Like the Fourier transform which converts the question of convergence in law of real-valued r.v. to a question of pointwise convergence of a function, the Stieltjes transform converts the question of convergence of random probability measures in vague topology into the question of pointwise convergence of a random function.

3.7 lemma (Stieltjes continuity, vague[10]). *Let $(\mu_n)_n$ be a sequence of random probability measures on $\mathbb{R}$, and μ a probability measure on $\mathbb{R}$. There is equivalence between:*

(i) *$\mu_n \to \mu$ almost surely (resp. in probability, resp. in expectation[11]) in the vague topology.*

[10]This characterization can be improved, with different arguments, to handle convergence in weak topology. But one can – and we will – avoid using an improved version of Stieltjes continuity to prove Wigner's theorem because the semi-circle law has compact support.

[11]Convergence in law and convergence $L^{r\neq 1}$ of random probability measures do not make sense.

(*ii*) *For any* $z \in \mathbb{C} \setminus \mathbb{R}$, $W_{\mu_n}(z) \to W_\mu(z)$ *almost surely (resp. in probability, resp. in expectation).*

Proof. $(i) \Rightarrow (ii)$. Let $z \in \mathbb{C} \setminus \mathbb{R}$, and $\psi_C : \mathbb{R} \to [0,1]$ be a continuous function with compact support included in $[-2C, 2C]$, and assuming the value 1 on $[-C, C]$. We assume $C > |z|$. We have:

$$\begin{aligned} \big|W_{\mu_n}(z) - W_\mu(z)\big| &\leq \Big|\int_{\mathbb{R}} \frac{1-\psi_C(x)}{z-x}\,\mathrm{d}\mu_n(x)\Big| + \Big|\int_{\mathbb{R}} \frac{1-\psi_C(x)}{z-x}\,\mathrm{d}\mu(x)\Big| + \\ &\quad + \Big|\int_{\mathbb{R}} \frac{\psi_C(x)}{z-x}\mathrm{d}(\mu_n-\mu)(x)\Big| \\ &\leq \frac{2}{C-|z|} + \Big|\int_{\mathbb{R}} \frac{\psi_C(x)}{z-x}\,\mathrm{d}(\mu_n-\mu)(x)\Big|. \end{aligned}$$

For any $\epsilon > 0$, there exists $C_{\epsilon,z} > 0$ such that the first term is smaller than $\epsilon/2$, and by assumption, for $n > n_\epsilon$ independently of C, the second term is smaller than $\epsilon/2$. Hence for $n > n_\epsilon$, $\big|W_{\mu_n}(z) - W_\mu(z)\big| \leq \epsilon$. This is enough to conclude in the various modes of convergence.

$(ii) \Rightarrow (i)$ Let $f \in \mathcal{C}_b^c(\mathbb{R})$. We approximate f by its convolution with a Cauchy law of small width, whose integration against our measures is then controlled by their Stieltjes transforms:

$$\begin{aligned} \big|\mu_n[f] - \mu[f]\big| &\leq \big|\mu_n[f] - \mu_n[f * C_b]\big| + \big|\mu[f] - \mu[f * C_b]\big| \\ &\quad + \Big|\int_{\mathbb{R}} \mathrm{d}a\, f(a)\, \mathrm{Im}\, W_{\mu_n-\mu}(a - \mathrm{i}b)\Big| \end{aligned}$$

For any $\epsilon > 0$, we can find $b_\epsilon > 0$ independent of n such that the two first terms are smaller than $\epsilon/3$ each. By assumption and dominated convergence (applicable because f has compact support), the last term with the choice $b := b_\epsilon$ can then be made smaller than $\epsilon/3$ for n large enough. Hence the conclusion. □

3.3 *Concentration inequalities (basics)*

Lecture 5 (1h30)
October 20th, 2014

Concentration inequalities give estimates for the probability of a r.v. to deviate from a given r.v. or a deterministic variable. They are extremely useful to establish convergence of r.v. We start to prove the most basic concentration inequality as a warmup:

3.8 LEMMA (Höffding lemma). *Let* X *be a centered real-valued r.v., taking values a.s. in a segment of length* $\ell > 0$. *Then* $\mathbb{E}[e^{tX}] \leq e^{\ell^2 t^2/2}$.

3.9 COROLLARY (Höffding inequality). *Let* $X_1, \ldots, X_n$ *be independent real-valued* *r.v., such that X_i values a.s. in a segment of bounded length ℓ_i. Then:*

$$\mathbb{P}\Big[\Big|\sum_{i=1}^{N}(X_i - \mathbb{E}[X_i])\Big| > t\Big] \leq 2e^{-L^{-2}t^2/2}, \qquad L^2 := \sum_{i=1}^{n} \ell_i^2.$$

We can say that $\sum_{i=1}^{N} X_i$ is **concentrated around its mean**, with **subgaussian tails** (to characterize the behavior of the bound when $t \to \infty$).

Proof. (of the Lemma) Denote $I = [a, b]$ the segment in which X takes values a.s., its length is $\ell = b - a > 0$. Since X is centered, we must have $a < 0 < b$. The rescaled r.v. $\tilde{X} = X/(b-a)$ is centered and takes values a.s. in $[-1, 1]$ – we slightly enlarged the segment here. By convexity of the exponential:

$$\forall \tilde{x} \in [-1, 1], \qquad e^{t\tilde{x}} \leq \frac{1-\tilde{x}}{2}\, e^{-t} + \frac{1+\tilde{x}}{2}\, e^{t},$$

and averaging over $\tilde{x}$:

$$\mathbb{E}[e^{t\tilde{X}}] \leq \cosh(t) = \sum_{k \geq 0} \frac{t^{2k}}{(2k)!}.$$

Since $(2k)! \geq 2^k\, k!$, we can actually bound:

$$\mathbb{E}[e^{t\tilde{X}}] = e^{t^2/2},$$

and we obtain the announced result when coming back to X. □

Proof. (of the Corollary) $\overline{X}_i = X_i - \mathbb{E}[X_i]$ is centered and takes values in a segment of length ℓ_i. If we denote $S_n = \sum_{i=1}^{n} \overline{X}_i$, independence and Höffding lemma imply:

$$\mathbb{E}[e^{tS_n}] \leq e^{t^2L^2/2}, \qquad L^2 := \sum_{i=1}^{n} \ell_i$$

Thanks to Markov inequality

$$\forall u, t \in \mathbb{R}, \qquad \mathbb{P}[S_N \geq u] \leq e^{-tu + t^2L^2/2},$$

and optimizing in t, we find:

$$\mathbb{P}[S_N \geq u] \leq e^{-L^{-2}u^2/2}.$$

(The argument of these last four lines is the detail of Chernov inequality). The same argument can be repeated for $-S_N$, hence:

$$\mathbb{P}[|S_N| \geq u] \leq 2\, e^{-L^{-2}u^2/2}.$$

□

There is a more general version of the Höffding inequality for functions of n independent variables that do not vary too much in each variable:

3.10 theorem (McDiarmid inequality). *Let $X = (X_1, \ldots, X_n)$ a sequence of independent real-valued r.v. Let $F : \mathbb{R}^n \to \mathbb{C}$ such that, for any $i \in [\![1, n]\!]$ the total variation of $F(x_1, \ldots, \bullet_i, \ldots, x_n)$ as a function of its i-th variable is bounded by c_i uniformly in the $(n-1)$ other variables. Then:*

$$\mathbb{P}\big[|F(X) - \mathbb{E}[F(X)]| \geq u\big] \leq 2\, e^{-c^{-2}\lambda^2/8}, \qquad c^2 := \sum_{i=1}^n c_i^2.$$

Proof. By decomposing into real and imaginary parts, we can restrict ourselves to real-valued F. We show by induction on the number of variables n that:

$$\mathbb{E}\big[e^{t(F(X_1,\ldots,X_n) - \mathbb{E}[F(X_1,\ldots,X_n)])}\big] \leq e^{2c^2t^2}, \qquad c^2 := \sum_{i=1}^n c_i^2.$$

For $n = 1$, this is Höffding lemma. Assume the result for $n-1$ variables. To show the result for n variables, we condition on the last $(n-1)$-variables and set:

$$Z := \mathbb{E}[F(X)|X_2, \ldots, X_n], \qquad Y := F(X) - Z.$$

We have:

$$\mathbb{E}[e^{tF(X)}|X_2, \ldots, X_n] = \mathbb{E}[e^{tZ}]\, \mathbb{E}[e^{tY}|X_2, \ldots, X_n].$$

The induction hypothesis applies to Z, which is a function of the variables $X_2, \ldots, X_n$, with total variations bounded by $(c_2, \ldots, c_n)$ in the respective variables. So:

$$\mathbb{E}[e^{tZ}] \leq e^{2(\sum_{i=2}^n c_i^2)t^2}.$$

By assumption on the total variation of F with respect to its first variable, Y takes values in $[-c_1, c_1]$. Höffding lemma yields:

$$\mathbb{E}[e^{tY}|X_2, \ldots, X_n] \leq e^{2c_1^2t^2}.$$

Hence the claim for n variables. McDiarmid inequality follows by application of Chernov inequality, as in the previous proof. $\square$

McDiarmid inequality exploited the behavior of a function F of n variables separately in each variable. We can obtain much more powerful results if we take into account the global Lipschitz behavior in n variables. Talagrand inequality is one of those results, and it will be our key to the proof of Wigner theorem.

3.4 *Concentration inequalities (Talagrand)*

The intuition behind concentration is, as quoted from Talagrand: *A random variable that depends (in a "smooth" way) on the influence of many independent variables (but not too much on any of them) is essentially constant*. In other words, independent fluctuations in general do not help each other going far ! A precise statement is Theorem 3.12 below: "smooth dependence" here is Lipschitz behavior, and we need the extra assumption of convexity to make the proof work.

We start with a Lemma whose content is intuitive, but concentrates all the difficulty of the proof.

3.11 lemma. *Let $X_1, \ldots, X_n$ complex-valued r.v., bounded a.s. by K, and $X = (X_1, \ldots, X_n)$. Let $A \subseteq [-K, K]^n$ be a convex set.*

$$\mathbb{E}\big[\exp(d^2(A, X)/16K)\big]\,\mathbb{P}[X \in A] \leq 1.$$

Here, $d(A, \bullet)$ denotes the Euclidean distance in $\mathbb{R}^n$ between A and the point or the set $\bullet$.

3.12 theorem. *Let $X_1, \ldots, X_n$ complex-valued r.v., bounded a.s. by K, and $X = (X_1, \ldots, X_n)$. Let $F : [-K, K]^n \to \mathbb{R}$ be a* ***convex****, k-Lipschitz function (for the euclidean norm at the source). $F(X)$ is concentrated around its median with subgaussian tails:*

$$\mathbb{P}\big[|F(X) - \mathbb{M}[F(X)]| \geq t\big] \leq 4\exp(-t^2/16K^2k_F^2).$$

We remark that the smaller the Lipschitz constant k_F is, the better is the concentration estimate for $F(X)$. It is then a simple step to show that the mean differs from the median only by a finite amount, so we have as well concentration around the mean:

3.13 theorem. *With the assumptions of Theorem 3.12, $F(X)$ is concentrated around its mean with subgaussian tails:*

$$\mathbb{P}\big[|F(X) - \mathbb{E}[F(X)]| \geq t + 8\sqrt{\pi}Kk_F\big] \leq 4\exp(-t^2/16K^2k_F^2).$$

Proof of Lemma 3.11

We follow the original proof of Talagrand. We first need a few definitions. Let Ω be a probability space, and for any $\omega, \omega' \in \Omega$, define:

$$\delta(\omega, \omega') = \begin{cases} 0 \text{ if } \omega = \omega' \\ 1 \text{ otherwise} \end{cases}.$$

We consider Ω^n equipped with a product probability measure – which is the appropriate home for n independent r.v.. We define $D : \Omega^n \times \Omega^n \to \mathbb{R}^n$ by:

$$\forall \xi, \xi' \in \Omega^n, \qquad D(\xi, \xi') = \big(\delta(\xi_1, \xi_1'), \ldots, \delta(\xi_n, \xi_n')\big).$$

It is a comparison vector whose entries are 0 or 1 depending whether ξ_i and ξ_i' coincide or not. Let $\mathcal{A}$ be a subset of Ω^n, and $\xi \in \Omega^n$. We define the set of comparison vectors between configurations in $\mathcal{A}$ and the configuration ξ:

$$U'_{\mathcal{A}}(\xi) = \{s \in \mathbb{R}^n, \qquad \exists \eta \in \Omega^n \quad s = D(\eta, \xi)\}.$$

Eventually, we set:

$$f(\mathcal{A}, \xi) = d(\underline{U'_{\mathcal{A}}(\xi)}, 0),$$

where the underline stands for the operation of taking the convex hull, and d is the euclidean distance between this set and 0 in $\mathbb{R}^n$. Obviously, f takes values in $\{0, 1, \sqrt{2}, \ldots, \sqrt{n}\}$. It is actually more convenient to allow less strict comparison vectors: we define

$$U_{\mathcal{A}}(\xi) := \{s \in \mathbb{R}^n \qquad \exists \eta \in \Omega^b \quad \forall i \in [\![1, n]\!], \quad s_i \geq \delta(\xi_i, \eta_i)\}.$$

If s is an element of this set, we say that an $\eta \in \mathcal{A}$ satisfying the defining property witnesses s. Since in $U_{\mathcal{A}}(\xi)$ we allowed vectors that are more distant from 0 than comparison vectors, we also have:

$$f(\mathcal{A}, \xi) = d(\underline{U_{\mathcal{A}}(\xi)}, 0).$$

We now quantify the relation between the probability to be in $\mathcal{A}$, and large deviations of the "distance" between a random point and $\mathcal{A}$.

3.14 lemma. *Averaging over $\xi \in \Omega^n$, we have $\mathbb{E}[e^{f^2(\mathcal{A},\xi)/4}]\mathbb{P}[\mathcal{A}] \leq 1$.*

Proof. We proceed by induction on the "number of variables" n. If $n = 1$, $U'_{\mathcal{A}}(\xi) \subseteq \{0, 1\}$, and it contains 0 iff $\xi \in \mathcal{A}$. The expectation value in the lemma can be computed by conditioning on ξ belonging to $\mathcal{A}$ or not. If we denote $p = \mathbb{P}[\mathcal{A}]$, we have:

$$\mathbb{E}\big[e^{f^2(\mathcal{A},\xi)/4}\big]\mathbb{P}[\mathcal{A}] = (p + (1-p)e^{1/4})p.$$

Elementary calculus shows that the maximum of the right-hand side over $p \in [0, 1]$ is reached for $p = 1$, hence is equal to 1. Now, assume the claim for n variables, and consider $\mathcal{A} \subseteq \Omega^{n+1}$, $\xi \in \Omega^n$ and $\omega \in \Omega$. By forgetting the last component, we define the projection $\mathcal{B} \subseteq \Omega^n$ of $\mathcal{A}$, and the slice $\mathcal{B}(\omega) \subseteq \Omega^n$ of this projection that are completed to elements of $\mathcal{A}$ by adjunction of ω:

$$\begin{aligned} \mathcal{B} &:= \big\{\eta \in \Omega^n, \qquad \exists \omega' \in \Omega, \qquad (\eta, \omega) \in \mathcal{A}\big\} \\ \mathcal{B}(\omega) &:= \big\{\eta \in \Omega^n, \qquad (\eta, \omega) \in \mathcal{A}\big\} \end{aligned}$$

We collect two elementary observations:

- If $s \in U_{\mathcal{B}(\omega)}(\xi)$, then $(s, 0) \in U_{\mathcal{A}}(\xi, \omega)$. Indeed, there is an $\eta \in \mathcal{B}(\omega)$ witnessing s. By definition of $\mathcal{B}(\omega)$, we have $(\eta, \omega) \in \mathcal{A}$, and this witnesses the vector $(s, 0)$ in $U_{\mathcal{A}}(\xi, \omega)$.

- If $s \in U_{\mathcal{B}}(\xi)$, then $(s, 1) \in U_{\mathcal{A}}(\xi, \omega)$. Indeed, there is an $\eta \in \mathcal{B}$ witnessing s. By definition of $\mathcal{B}$, there is an $\omega' \in \Omega$ such that $(\eta, \omega') \in \mathcal{A}$, and since we do not know if ω coincides with ω' or not, we can at least say that (η, ω') witnesses $(s, 1)$ in $U_{\mathcal{A}}(\xi, \omega)$ (this is where the use of U instead of U' is useful).

Then, for any $s \in U_{\mathcal{B}}(\omega)$, $s' \in U_{\mathcal{B}}$ and $t \in [0, 1]$, we have that $t(s, 0) + (1 - t)(s', 1) \in \underline{U_{\mathcal{A}}(\xi, \omega)}$ remarking that we work with the convex hull. This provides an upper bound:

$$f^2(\mathcal{A},(\xi,\omega)) \leq |t(s,0)+(1-t)(s',1)|^2 = |ts+(1-t)s'|^2+(1-t)^2.$$

Using the convexity of the square of the euclidean norm:

$$f^2\big(\mathcal{A},(\xi,\omega)\big) \leq t|s|^2+(1-t)|s'|^2+(1-t)^2,$$

and then optimizing over the choice of s and s':

$$f^2(\mathcal{A},(\xi,\omega)) \leq tf^2(\mathcal{B}(\omega),\xi)+(1-t)f^2(\mathcal{B},\xi)+(1-t)^2.$$

We can now exponentiate and average over $\xi \in \Omega^n$ (here we use that Ω^{n+1} is equipped with a product measure, so we can keep $\omega \in \Omega$ fixed for the moment):

$$\mathbb{E}_{\Omega^n}\big[e^{f^2(\mathcal{A},(\xi,\omega))/4}\big] \leq e^{(1-t)^2/4}\,\mathbb{E}_{\Omega^n}\big[e^{tf^2(\mathcal{B}(\omega),\xi)/4}\,e^{(1-t)f^2(\mathcal{B},\xi)/4}\big].$$

We then use Hölder inequality to split the expectation values in two, and the induction hypothesis for $\mathcal{B}$ and $\mathcal{B}(\omega)$ which are subsets of Ω^n:

$$\begin{aligned}\mathbb{E}_{\Omega^n}\big[e^{f^2(\mathcal{A},(\xi,\omega))/4}\big] &\leq e^{(1-t)^2/4}\big(\mathbb{E}_{\Omega_n}[e^{f^2(\mathcal{B}(\omega),\xi)/4}]\big)^t\big(\mathbb{E}_{\Omega^n}[e^{f^2(\mathcal{B},\xi)/4}]\big)^{1-t} \\ &\leq e^{(1-t)^2/4}\mathbb{P}_{\Omega^n}[\mathcal{B}(\omega)]^{-t}\mathbb{P}_{\Omega^n}[\mathcal{B}]^{-(1-t)} = \frac{\exp(g_r(t))}{\mathbb{P}_{\Omega^n}[\mathcal{B}]}\end{aligned}$$

with:

$$r := \frac{\mathbb{P}_{\Omega^n}[\mathcal{B}(\omega)]}{\mathbb{P}_{\Omega^n}[\mathcal{B}]},\qquad g_r(t) := \frac{(1-t)^2}{4} - t\ln r.$$

Let us optimize over $t \in [0,1]$. The minimum of g_r over $\mathbb{R}$ is reached at $t_r := 1+2\ln r$. When $r \in [0,e^{-1/2}]$, t_r is non-positive, so the minimum of g_r over $[0,1]$ is reached at $t=0$ and is $1/4$. When $r \geq e^{-1/2}$, we rather have:

$$\min_{t\in[0,1]} g_r(t) = g_r(t_r) = -\ln^2 r + \ln r.$$

We would prefer a simpler bound in terms of r, and an exercise of calculus reveals that:

$$\forall r \in \mathbb{R},\qquad \min_{t\in[0,1]} g_r(t) \leq \ln(2-r).$$

Subsequently:

$$\mathbb{E}_{\Omega^{n+1}}\big[e^{f^2(\mathcal{A},(\xi,\omega))/4}\big] \leq \frac{1}{\mathbb{P}_{\Omega^n}[\mathcal{B}]}\Big(2-\frac{\mathbb{P}_{\Omega^n}[\mathcal{B}(\omega)|\omega]}{\mathbb{P}_{\Omega^n}[\mathcal{B}]}\Big).$$

Finally, we integrate over ω. Since $\bigcup_{\omega\in\Omega} B_\omega = \mathcal{B}$, we obtain:

$$\mathbb{E}_\Omega\big[\mathbb{P}_{\Omega^n}[\mathcal{B}(\omega)]\big] = \mathbb{E}_\Omega\big[\mathbb{P}_{\Omega^{n+1}}\mathbb{P}[\mathcal{A}|\omega]\big] = \mathbb{P}_{\Omega^{n+1}}[\mathcal{A}].$$

Hence:

$$\mathbb{E}_{\Omega^{n+1}}\left[e^{f^2(\mathcal{A},(\xi,\omega))/4}\right] \leq \frac{\phi\big(\mathbb{P}_{\Omega^{n+1}}(\mathcal{A})/\mathbb{P}_{\Omega^n}[\mathcal{B}]\big)}{\mathbb{P}_{\Omega^n}[\mathcal{A}]}, \qquad \phi(x) = x(2-x).$$

Since $\max_{x\in\mathbb{R}} \phi(x) = 1$, we eventually obtain:

$$\mathbb{E}_{\Omega^{n+1}}\left[e^{f^2(\mathcal{A},(\xi,\omega))/4}\right] \leq \frac{1}{\mathbb{P}_{\Omega^{n+1}}[\mathcal{A}]}.$$

□

Consider $X_1, \ldots, X_n$ independent r.v. bounded by $K > 0$, and $A \subseteq [-K, K]^n$. We can always assume X_i's are defined in the same probability space Ω, therefore $X = (X_1, \ldots, X_n)$ is a r.v. defined on Ω^n, and independence means that the latter is equipped with a product probability measure. Lemma 3.14 can be applied to the event $\mathcal{A} = \{\xi \in \Omega^n, \quad X(\xi) \in A\}$, and we now need to compare the funny distance $f(\mathcal{A}, \xi)$ between a set and a point in Ω^n with the usual euclidean distance $d(A, X(\xi))$ between a set and a point in $\mathbb{R}^n$.

3.15 lemma. *Let $\eta \in \Omega^n$. If A convex, then $d(A, X(\eta)) \leq 2Kf(\mathcal{A}, \eta)$.*

Proof. Let $y = X(\eta)$. If $\xi \in \mathcal{A}$, we remark that

$$\forall i \in [\![1, n]\!], \qquad |y_i - X_i(\xi)| = |X_i(\eta_i) - X_i(\xi_i)| \leq 2K\delta(\xi_i, \eta_i).$$

Indeed, if the right-hand side is 0, so must be the left-hand side ; and otherwise, the right-hand side is equal to $2K$, and the claim trivially holds because $|X_i(\xi_i)|$ and $|X_i(\xi_i')|$ are bounded by K. Let us apply this inequality to each of the components of some $\xi^{(1)}, \ldots, \xi^{(k)} \in \mathcal{A}$, and make a linear combination with positive coefficients $t^{(1)}, \ldots, t^{(k)}$ such that $\sum_{j=1}^k t^{(j)} = 1$:

$$\forall i \in [\![1, n]\!], \qquad \Big| y_i - \sum_{j=1}^{k} t^{(j)}\, X(\xi^{(j)}) \Big| \leq 2K \sum_{j=1}^{k} t^{(j)}\, \delta(\xi_i^{(j)}, \eta_i),$$

and in euclidean norm in $\mathbb{R}^n$:

$$\Big| y - \sum_{j=1}^{k} t^{(j)}\, X(\xi^{(j)}) \Big| \leq 2K \left[\sum_{i=1}^{n} \Big(\sum_{j=1}^{k} t^{(j)}\, D(\xi^{(k)}, \eta) \Big)^2 \right]^{1/2}.$$

Since A is assumed convex, $\sum_{j=1}^k t^{(j)}\, X(\xi^{(j)})$ is in A. Therefore, the distance $d(A, y)$ is a lower bound for the left-hand side. On the other hand, the infimum of the bracket in the right-hand side over the $t^{(j)}$'s and $\xi^{(k)}$'s is precisely the distance from 0 to the convex hull of $U'_{\mathcal{A}}(\eta)$. Hence the claim. □

We can now finish the proof of Lemma 3.11. If $A \subseteq [-K, K]^n$ is convex, then:

$$\mathbb{E}[e^{d^2(A,y)/16K^2}] \leq \mathbb{E}[e^{f^2(\mathcal{A},\eta)/4}] \leq \frac{1}{\mathbb{P}[\mathcal{A}]} = \frac{1}{\mathbb{P}[X \in A]}.$$

Proof of Theorem 3.12

Let $F : [-K,K]^n \to \mathbb{R}$ be a convex, k_F-Lipschitz function. Then, the level sets $A_u := \{x \in [-K,K]^n,\quad F(x) \leq u\}$ are convex[12], and we can apply Lemma 3.11. Denoting $\mathcal{A}_u$ the event $\{X \in A_u\}$:

$$\forall u \in \mathbb{R},\qquad \mathbb{E}[e^{d^2(A_u,X)/16K^2}]\mathbb{P}[\mathcal{A}_u] \leq 1.$$

If $u' \geq u$, when the event $\mathcal{A}_{u'}$ is realized the Lipschitz property implies $d(A_u,X) \geq k_F^{-1}|u'-u|$. Therefore:

$$\mathbb{E}\big[e^{d^2(A_u,X)/16K^2}\big|\mathcal{A}_{t'}\big] \geq \exp\big(|u'-u|^2/16K^2k_F^2\big).$$

and:

$$\mathbb{E}\big[e^{d^2(A_u,X)/16K^2}\big] \geq \mathbb{P}[{}^c\mathcal{A}_{u'}]\exp\big(|u'-u|^2/16K^2k_F^2\big).$$

All in all, we obtain:

$$\mathbb{P}[\mathcal{A}_u]\mathbb{P}[{}^c\mathcal{A}_{u'}] \leq \exp\big(-|u-u'|^2/16K^2k_F^2\big).$$

This already looks like a large deviation estimate. If $t \geq 0$, this inequality can be applied $u = \mathbb{M}[F(X)]$ and $u' = \mathbb{M}[F(X)] + t$. In this case $\mathbb{P}[\mathcal{A}_u] \geq 1/2$ (because the inequality is not strict) and $\mathcal{A}_{u'}$ is the event $\{F(X) - \mathbb{M}[F(X)] \geq t\}$, so:

$$\mathbb{P}\big[F(X) - \mathbb{M}[F(X)] \geq t\big] \leq 2\exp\big(-t^2/16K^2k_F^2\big).$$

If we rather choose $u = \mathbb{M}[F(X)] - t$ and $u = \mathbb{M}[F(X)]$, we obtain the same upper bound for the probability of the event $\{F(X) - \mathbb{M}[F(X)] \leq -t\}$. Thus:

$$\mathbb{P}\big[|F(X) - \mathbb{M}[F(X)]| \geq t\big] \leq 4\exp\big(-t^2/16K^2k_F^2\big).$$

Proof of Theorem 3.13

We estimate the difference between the mean and a median using Theorem 3.12:

$$\begin{aligned}
\big|\mathbb{E}\big[F(X) - \mathbb{M}[F(X)]\big]\big| &\leq \mathbb{E}\big[|F(X) - \mathbb{M}[F(X)]|\big] \\
&\leq \int_0^\infty \mathbb{P}[|F(X) - \mathbb{M}[F(X)]| \geq t]\,\mathrm{d}t \\
&\leq \int_0^\infty 4\exp\big(-t^2/16K^2k_F^2\big)\mathrm{d}t
\end{aligned}$$

and the last integral is evaluated to $8\sqrt{\pi}Kk_F$. The conclusion follows by splitting

$$\big|F(X) - \mathbb{E}[F(X)]\big| \leq \big|F(X) - \mathbb{M}[F(X)]\big| + \big|\mathbb{E}\big[F(X) - \mathbb{M}[F(X)]\big]\big|$$

[12] in Theorem 3.12, we could replace the assumption that F is convex by the strictly weaker assumption that its level sets are convex.

and applying Theorem 3.12.

3.5 *Application of concentration in RMT*

Here is a nice application in random matrix theory:

3.16 corollary. *Let M_N be a $N \times N$ random hermitian matrix, whose N^2 $\mathbb{R}$-linearly independent entries are independent r.v. bounded by $C/\sqrt{N}$. Let $f : \mathbb{R} \to \mathbb{R}$, be a **convex**, k_f-Lipschitz function.*

$$\mathbb{P}\Big[\big|L_N[f] - \mathbb{E}[L_N[f]]\big| \leq t + \frac{8Ck_f\sqrt{2\pi}}{\sqrt{N}}\Big] \leq 4\,\exp(-Nt^2/32C^2k_f^2).$$

Proof. If $f : \mathbb{R} \to \mathbb{R}$ is a convex, k_f-Lipschitz function, we have seen that $A \mapsto N^{-1}\,\mathrm{Tr}\, f(A)$ is convex (Klein lemma 2.8) and has a Lipschitz constant bounded by $\sqrt{2/N} \cdot k_f$. The N^2 $\mathbb{R}$-linearly independent entries of M_N to be our random variables, and they are bounded by $C/\sqrt{N}$. The claim follows by application of Theorem 3.12. $\square$

If we are not interested in precise large deviation estimates, we can get rid of the convexity condition by approximation:

3.17 theorem. *Let M_N be a $N \times N$ random hermitian matrix, whose N^2 $\mathbb{R}$-linearly independent entries are independent **centered** r.v. bounded by $C/\sqrt{N}$. Let $f : \mathbb{R} \to \mathbb{C}$ be a continuous bounded function. Then $L_N[f] - \mathbb{E}[L_N[f]]$ converges in probability to 0.*

Proof of Theorem 3.17

If f is only a continuous function without stronger assumption, we can rely on the approximation lemma:

3.18 lemma. *Any continuous $f : [-K, K] \to \mathbb{R}$ can be approximated for the sup norm by the difference of two convex Lipschitz functions.*

Proof. Let $\mathcal{D}$ the set of $f : [-K, K] \to \mathbb{R}$ that are difference of two convex Lipschitz functions. It is clear that constant functions are in $\mathcal{D}$, and that $\mathcal{D}$ separate points (i.e. for any two $x, y \in [-K, K]$, there exist f and g such that $f(x) \neq g(y)$). If we can prove that $\mathcal{D}$ is an algebra, the result would follow from Stone-Weierstraß theorem. Thus, it remains to justify that $\mathcal{D}$ is stable by multiplication, and more specifically, that the property "being a difference of two convex functions" is stable by multiplication[13]. First, we claim that in a decomposition $f = f_1 - f_2$ with f_1, f_2 convex functions, we can always assume $f_i \geq 0$ for $i = 1, 2$. Indeed, since convex functions on a compact set are bounded, there exist constants m_1, m_2 such that $f_i + m_i = \tilde{f}_i$ is non-negative (and obviously convex). So $f = \tilde{f}_1 - \tilde{f}_2 + m_2 - m_1$. If the constant $m = m_2 - m_1$

[13](Actually, since $\mathcal{D}$ contains affine functions, stability under multiplication guarantees that it also contain all polynomials, so we can even approximate any continuous function by the difference of two convex polynomials)

is positive, we add it to $\tilde{f}_1$, it is negative, we subtract it from $\tilde{f}_2$, so as to get a decomposition of f into non-negative convex functions. Now, if $f \in \mathcal{D}$, then $f^2 \in \mathcal{D}$: indeed, if $f = f_1 - f_2$ is a decomposition into non-negative convex functions, then $f^2 = 2(f_1^2 + f_2^2) - (f_1 + f_2)^2$ is also one, thus $f^2 \in \mathcal{D}$. Then, if $f, g \in \mathcal{D}$, we conclude that $fg \in \mathcal{D}$ by writing $fg = [(f+g)^2 - (f-g)^2]/4$.
. ☐

$M \mapsto F(M)$ can also be considered as a function of the vector $\Lambda^{(M)}$ of eigenvalues of M. To use the approximation result, we need to show that the fluctuations of eigenvalues mainly take place in a compact region $[-K, K]$. We can actually show a large deviation estimate:

3.19 lemma. *There exist $c, t_0 > 0$ independent of N, such that:*

$$\forall t \geq t_0, \qquad \mathbb{P}\big[|\lambda_{\max}^{(M_N)}| \geq t\big] \leq e^{-cNt^2}.$$

Proof. Let v be a deterministic, unit vector. We have:

$$(M_N \cdot v)_i = \sum_{j=1}^{N} M_{i,j} v_j,$$

and by our assumptions, $M_{i,j} v_j$ are centered, independent r.v. uniformly bounded by $CN^{-1/2}$. Thanks to Höffding inequality (Corollary 3.9):

$$\mathbb{P}\big[|(M_N \cdot v)_i| \geq t\big] \leq 2\exp(-Nt^2/2C).$$

Hence:

$$\mathbb{P}\big[|M_N \cdot v| \geq t\big] \leq \sum_{i=1}^{N} \mathbb{P}\big[|(M_N \cdot v)_i| \geq t\big] \leq 2N\exp(-Nt^2/2C).$$

The prefactor of N does not hurt: put in the exponential, it becomes $\ln N \ll N$, so up to trading C for some $C' > C$ independent of N in this inequality and assuming $t \geq t_0$ for some $t_0 > 0$ independent of N, we have:

(8) $\quad \mathbb{P}\big[|M_N \cdot v| \geq t\big] \leq \exp(-Nt^2/2C').$

As a matter of fact, we would like to have a similar large deviation result for

$$|\lambda_{\max}^{(M_N)}| = \sup_{|v|=1} |M_N \cdot v|,$$

but of course, the unit vector realizing the sup is random since it depends on M_N, so (8) does not apply directly. However, we can always cover the unit sphere of $\mathbb{R}^N$ by less than c_0^N unit balls of radius $1/2$ for some constant $c_0 > 0$ (this is called a $1/2$-net). If we denote $\{w^{(k)}\}$ the set of their centers, it means that any unit vector $v \in \mathbb{R}^N$ is at a distance less than $1/2$ from at least one

$w^{(k(v))}$, and we can write:

$$|M_N \cdot v| \leq |M_N \cdot w^{(k(v))}| + |M_N \cdot (v - w^{(k(v))})| \leq |M_N \cdot w^{(k(v))}| + \frac{|\lambda_{\max}^{(M_N)}|}{2}.$$

Taking the sup over unit vectors v, we find $\lambda_{\max}^{(M_N)} \leq 2 \sup_k |M_N \cdot w^{(k)}|$, and we can deduce by a naive union bound:

$$\begin{aligned} \mathbb{P}\big[|\lambda_{\max}^{(M_N)}| \geq t\big] & \leq \mathbb{P}\big[\sup_k |M_N \cdot w^{(k)}| \geq t/2\big] \\ & \leq \sum_i \mathbb{P}\big[|M_N \cdot w^{(k)}| \geq t/2\big] \leq c_0^N \exp(-Nt^2/8C'). \end{aligned}$$

By taking t large enough independently of N, there exists a constant $c > 0$ such that $\mathbb{P}\big[|\lambda_{\max}^{(M_N)}| \geq t\big] \leq \exp(-cNt^2)$. $\square$

If $f \neq 0$, we now prove that for $\epsilon, \delta > 0$ small enough independent of N:

$$\text{(9)} \quad \mathbb{P}\big[\overline{L}_N^{(M_N)}[f]\big| \geq \epsilon\big] \leq \delta, \qquad \overline{L}_N^{(M_N)} := L_N^{(M_N)} - \mathbb{E}[L_N^{(M_N)}]$$

for N large enough. We choose $K_\delta > 0$ such that:

$$\mathbb{P}[\mathcal{A}_\delta] \leq \frac{\delta}{4||f||_\infty}, \qquad \mathcal{A}_\delta := \big\{|\lambda_{\max}^{(M_N)}| > K_\delta\big\}.$$

Since $L_N^{(M_N)}$ is a probability measure, we have:

$$\mathbb{P}\big[|\overline{L}_N^{(M_N)}[f]| \geq \epsilon\big] \leq \delta/2 + \mathbb{P}\Big[\big\{|\overline{L}_N^{(M_N)}[f]| \geq \epsilon\big\} \cap \mathcal{A}_\delta^c\Big].$$

Let $\tilde{f} = \tilde{f}_1 - \tilde{f}_2$ be a difference of two Lipschitz convex functions such that $||f - \tilde{f}||_\infty^{[-K_\delta, K_\delta]} \leq \epsilon/4$. On the event $\mathcal{A}_\delta^c$, $L_N^{(M_N)}$ is a probability measure supported on $[-K_\delta, K_\delta]$, so $\big|\overline{L}_N^{(M_N)}[f - \tilde{f}]\big| \leq \epsilon/2$ uniformly in N, and:

$$\mathbb{P}\big[|\overline{L}_N^{(M_N)}[f]| \geq \epsilon\big] \leq \delta/2 + \mathbb{P}\Big[|\overline{L}_N^{(M_N)}[\tilde{f}]| \geq \epsilon/2\big\} \cap \mathcal{A}_\delta^c\Big].$$

Eventually, we can apply the concentration given by Corollary 3.16 to the matrix functions $A \mapsto N^{-1}\,\mathrm{Tr}\,\tilde{f}_i(A)$ for $i = 1, 2$. It shows that $\overline{L}_N^{(M_N)}[\tilde{f}_i]$ for $i = 1, 2$ (hence their difference) converge in probability to 0 when $N \to \infty$. So, for N large enough depending on ϵ and δ:

$$\mathbb{P}\big[|\overline{L}_N^{(M_N)}[\tilde{f}]| \leq \epsilon/2\big] \leq \delta/2.$$

and we arrive to (9).

3.6 *Proof of Wigner's theorem*

Lecture 6 (1h30)
October 21st, 2014

Thanks to Lemma 3.3-3.4, we assume that M_N is a Wigner matrix with zero diagonal and entries $(M_N)_{i,j} \leq C/\sqrt{N}$ for some $C > 0$ independent of i, j, N. In order to prove convergence of the empirical measure $L^{(M_N)}$, we will study the convergence of its Stieltjes transform. We assume throughout this paragraph that $z \in \mathbb{C} \setminus \mathbb{R}$ such that $|\mathrm{Im}\, z|$ remains bounded away from 0, uniformly in N. We set:

$$S_{M_N}(z) := \frac{1}{N} \mathrm{Tr}\, R_{M_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \Big(\frac{1}{z - M_N} \Big)_{i,i} = L^{(M_N)}[\phi_z]$$

where $\phi_z(x) = \frac{1}{z-x}$ and for any matrix M, R_M is the resolvent introduced in Definition § 2.11. If M is hermitian, we have the easy bounds:

$$|S_M(z)| \leq \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{(\mathrm{Re}\, z - \lambda_i^{(M)})^2 + (\mathrm{Im}\, z)^2}} \leq \frac{1}{|\mathrm{Im}\, z|},$$

and:

$$\mathrm{Im}\, (z - S_M(z)) = \mathrm{Im}\, z \left(1 + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(\mathrm{Re}\, z - \lambda_i^{(M)})^2 + (\mathrm{Im}\, z)^2} \right).$$

Besides:

$$|\partial_z S_M(z)| = \left| \frac{1}{N} \sum_{i=1}^{N} \frac{1}{(z - \lambda_i^{(M)})^2} \right| \leq \frac{1}{|\mathrm{Im}\, z|^2}$$

which shows that ϕ_z has a Lipschitz constant bounded by $|\mathrm{Im}\, z|^{-2}$.

Stieltjes transform and almost fixed point equation

3.20 LEMMA. $S_{M_N}(z) = \frac{1}{z - S_{M_N}(z)} + D_N(z)$ *where* $D_N(z) \to 0$ *in probability when* $N \to 0$.

Proof. If M_N is going to have a limit in some sense when $N \to \infty$, its resolvent should "stabilize". Reminding the recursive decomposition of the resolvent on the diagonal (Lemma 2.12) and the simplification $(M_N)_{i,i} = 0$ for all i, we have:

$$S_{M_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z - v_{N,i}^T (z - M_N[i])^{-1} v_{N,i}},$$

where $M_N[i]$ is the matrix of size $(N-1)$ obtained by removing the i-th column and row of M_N, and $v_{N,i}$ is the i-th column vector of M_N in which the i-th component is removed (hence is of dimension $N-1$). "Stabilization" means that we expect:

$$\varepsilon_{N,i}(z) := v_{N,i}^T (z - M_N[i])^{-1} v_{N,i} - S_{M_N}(z)$$

to be small. This suggests to consider the decomposition:

$$S_{M_N}(z) = \frac{1}{z - S_{M_N}(z)} + D_N(z) \tag{10}$$

$$D_N(z) = \frac{1}{N}\sum_{i=1}^{N} \frac{-\varepsilon_{N,i}}{(z - S_N(z))(z - S_N(z) - \varepsilon_{N,i}(z))}, \tag{11}$$

and try to show that $D_N(z) \to 0$ in probability. From our preliminary remarks, we find:

$$\big|D_N(z)\big| \leq \frac{\eta_N(z)}{|\mathrm{Im}\, z|^2} + \frac{\eta_N(z)^2}{|\mathrm{Im}\, z|^3}, \qquad \eta_N(z) := \sup_{1\leq i\leq N} |\varepsilon_{N,i}(z)|,$$

so it is enough to prove that $\eta_N(z) \to 0$ in probability.

Let $\widetilde{M}_N[i]$ be the matrix of size N obtained by filling with 0's the i-th column and row of M_N. It has the same eigenvalues as $M_N[i]$, plus an extra 0 eigenvalue with eigenvector the i-th element of the canonical basis. Therefore:

$$\big|S_{\widetilde{M}_N[i]}(z) - S_{M_N[i]}(z)\big| = \frac{1}{N|\mathrm{Im}\, z|} + \frac{1}{(N-1)|\mathrm{Im}\, z|}.$$

The last term is due to the fact that $M_N[i]$ is a matrix of size $(N-1)$, so we normalized by $1/(N-1)$ instead of $1/N$ the sum defining $S_{M_N[i]}(z)$. Besides, $\widetilde{M}_N[i]$ is a perturbation of M_N by a matrix with $2N-1$ non-zero entries bounded by $C/\sqrt{N}$, hence by Hofmann-Wielandt and the arguments of § 3.1 using the Lipschitz behavior of ϕ_z:

$$\big|S_{M_N}(z) - S_{\widetilde{M}_N[i]}(z)\big| \leq \frac{C\sqrt{2}}{N^{1/2}|\mathrm{Im}\, z|^2}.$$

The new quantity

$$\widetilde{\varepsilon}_{N,i}(z) := v_{N,i}^T R_{M_N[i]} v_{N,i} - \frac{1}{N}\,\mathrm{Tr}\, R_{M_N[i]}(z), \qquad \widetilde{\eta}_N(z) := \sup_{1\leq i\leq N} |\varepsilon_{N,i}(z)|$$

differs only from $\varepsilon_{N,i}(z)$ by the last term, and thanks to the two last estimates, $[\widetilde{\eta}_N(z) - \eta_N(z)] \to 0$ in probability.

So, we are left with the problem of showing that $\widetilde{\eta}_N(z)$ converges to 0 in probability. Let $R_N^{[i]} = R_{M_N[i]}(z)$ and drop the dependence in z to shorten notations. We can decompose $\widetilde{\varepsilon}_{N,i} = \widetilde{\varepsilon}_{N,i}^{\mathrm{diag}} + \widetilde{\varepsilon}_{N,i}^{\mathrm{off}}$ with:

$$\widetilde{\varepsilon}_{N,i}^{\mathrm{diag}} = \frac{1}{N}\sum_{\substack{1\leq k\leq N \\ k\neq i}} \big(|\sqrt{N}(M_N)_{k,i}|^2 - 1\big)\,(R_N^{[i]})_{k,k},$$

$$\widetilde{\varepsilon}_{N,i}^{\mathrm{off}} = \frac{1}{N}\sum_{\substack{1\leq k,l\leq N \\ k,l,i \text{ pairwise distinct}}} \sqrt{N}(M_N)_{k,i}\,\sqrt{N}(M_N)_{l,i}\,(R_N^{[i]})_{k,l}.$$

A first key observation is that the entries of $R_N^{[i]}$ depend on entries of M_N that are not in its i-th column from the entries of the i-th column of M_N, hence they are independent and the $(M_N)_{k,i}$ and $R_{k,l}$ decouple when we take expectation values. A second key observation is that $\mathbb{E}\big[|\sqrt{N}(M_N)_{k,i}|^2\big] = 1$ for $k \neq i$ by the variance condition imposed for off-diagonal elements in the definition of Wigner matrices. Therefore, $\widetilde{\varepsilon}^{\mathrm{diag}}$ and $\widetilde{\varepsilon}^{\mathrm{off}}$ are centered. Our strategy is now to prove that $\mathbb{E}\big[|\widetilde{\varepsilon}^{\bullet}_{N,i}|^k\big] \to 0$ when $N \to \infty$ uniformly in i for some $k \geq 2$, so as to apply Markov inequality:

$$\mathbb{P}\big[|\widetilde{\varepsilon}^{\bullet}_{N,i}|^k \geq t\big] \leq t^{-k}\,\mathbb{E}[|\widetilde{\varepsilon}^{\bullet}_{N,i}|^k],$$

and deduce that $\widetilde{\varepsilon}^{\bullet}_{N,i}$ converges in probability to 0 uniformly in i, and so does $\widetilde{\eta}_{N,i}$, This would end the proof. It turns out that the naive choice $k = 2$ does not work, but $k = 4$ does[14].

We calculate $\mathbb{E}\big[|\widetilde{\varepsilon}_{N,i}|^4\big]$ by developing the sums and using the remark about independence:

$$\mathbb{E}\big[|\widetilde{\varepsilon}_{N,i}|^4\big] \leq \frac{1}{N^4} \sum_{\substack{1 \leq k_1,\ldots,k_4 \leq N \\ k_a \neq i}} \mathbb{E}\Big[\prod_{a=1}^{4}\big(|\sqrt{N}(M_N)_{k,i}|^2 - 1\big)\Big] \cdot \mathbb{E}\Big[\prod_{a=1}^{4}(R_N^{[i]})_{k_a,k_a}\Big].$$

Since the entries of (M_N) are independent, and thanks to the variance condition, the only non-zero terms occur when the indices of summation are identical by pairs, i.e. when $k_a = k_b$ and $k_c = k_d$ for some choice of labeling $\{a,b,c,d\} = \{1,2,3,4\}$. Subsequently:

$$\mathbb{E}\big[|\widetilde{\varepsilon}_{N,i}|^4\big] \leq \frac{1}{N^4} \sum_{\substack{1 \leq m_1,m_2 \leq N \\ m_a \neq i}} \mathbb{E}\Big[\prod_{a=1}^{2}\big(|\sqrt{N}(M_N)_{m_a,i}|^2 - 1\big)^2\Big] \cdot \mathbb{E}\Big[\prod_{a=1}^{2}(R_N^{[i]})^2_{m_a,m_a}\Big].$$

There are only $O(N^2)$ terms in this sum, and by the boundedness condition on the entries of M_N (for the first factor) and the assumption that $\mathrm{Im}\, z$ is bounded away from 0 (for the second factor), each of them uniformly bounded by constant independent of i and N. Thanks to the $1/N^4$ overall factor, the 4-th moment of $\widetilde{\epsilon}^{\mathrm{diag}}_{N,i}$ decays as $O(1/N^2)$, hence $\sup_i \widetilde{\epsilon}^{\mathrm{diag}}_{N,i}$ converges to 0 in probability.

The calculation for $\widetilde{\epsilon}^{\mathrm{off}}_{N,i}$ is similar:

$$\mathbb{E}\big[|\widetilde{\epsilon}^{\mathrm{off}}_{N,i}|\big] \leq N^{-4} \sum_{\substack{1 \leq k_1,l_1,\ldots,k_4,l_4 \leq N \\ k_a \neq l_a \\ k_a,l_a \neq i}} \mathbb{E}\Big[\prod_{a=1}^{4}\sqrt{N}(M_N)_{k_a,i}\cdot\sqrt{N}(M_N)_{l_a,i}\Big] \cdot \mathbb{E}\Big[\prod_{a=1}^{4}(R_N^{[i]})_{k_a,l_a}\Big].$$

Since the entries of (M_N) are centered and independent, the only non-zero terms occur when the indices of summation are identical by pairs. Taking into account the condition $k_a \neq l_a$, this means that for any $a \in \{1,2,3,4\}$, there

[14]Fourth-order moment estimates are actually at the center of versions of Wigner's theorem with weaker assumptions, developed by Tao and Vu.

exists $\{b(a), b'(a)\} \in \{1,2,3,4\}$ distinct from a such that $k_a \in \{k_{b(a)}, l_{b(a)}\}$ and $l_a \in \{k_{b'(a)}, l_{b'(a)}\}$. Hence:

$$\mathbb{E}\big[|\widetilde{\varepsilon}^{\,\mathrm{off}}_{N,i}|^4\big] \quad\leq\quad \frac{1}{N^4} \sum_{\substack{1\leq m_1,\ldots,m_4\leq N \\ m_a\neq i}} \mathbb{E}\Big[\prod_{a=1}^{4} \big|\sqrt{N}(M_N)_{m_a,i}\big|^2\Big]$$
$$\cdot\mathbb{E}\Big[(R_N^{[i]})_{m_1,m_2}(R_N^{[i]})_{m_2,m_1}(R_N^{[i]})_{m_3,m_4}(R_N^{[i]})_{m_4,m_3}\Big].$$

The first expectation value is bounded by C^8, and we find:

$$\big|\mathbb{E}\big[|\widetilde{\varepsilon}^{\,\mathrm{off}}_{N,i}|^4\big]\big| \leq \frac{C^8}{N^4}\mathbb{E}\big[|\mathrm{Tr}\,(R_N^{[i]})^2|^2\big].$$

And:

$$\big|\mathrm{Tr}\,(R_N^{[i]})^2\big| = N\big|S_{\widetilde{M}_N[i]^2}(z)\big| \leq N\,|\mathrm{Im}\,z|^{-1}.$$

This shows that the fourth moment of $\widetilde{\varepsilon}_{N,i}$ is $O(1/N^2)$ uniformly in i. Hence, $\sup_i \widetilde{\varepsilon}^{\,\mathrm{off}}_{N,i}$ converges to 0 in probability. $\square$

Identifying the limit of the Stieltjes transform

3.21 lemma. $\mathbb{E}[S_N(z)]$ *converges pointwise – and* $S_N(z)$ *converges pointwise in probability – to* $W_{\mu_{\mathrm{sc}}}(z) = \frac{z-\sqrt{z^2-4}}{2}$.

Proof. Let $z \in \mathbb{C}\setminus\mathbb{R}$. From the approximate fixed point equation:

$$D_N(z) = S_N(z) - \frac{1}{z - S_N(z)},$$

we deduce that $D_N(z)$ is bounded. Let us recast the equation as

$$S_N(z)(z - S_N(z)) = 1 + D_N(z)S_N(z),$$

take the expectation value and rearrange the result:

$$\mathbb{E}[S_N(z)]\big(z - \mathbb{E}[S_N(z)]\big) = 1 + \mathbb{E}[D_N(z)S_N(z)] + \mathrm{Var}(S_N(z)). \tag{12}$$

Since $D_N(z)$ is bounded and converges in probability to 0, while $S_N(z)$ is bounded, the expectation value in the right-hand side converges to 0 (pointwise in z). Besides, we can apply the concentration result (Corollary 3.17) to M_N and the continuous function $\varphi_z : x \mapsto \frac{1}{z-x}$. It implies that $S_N(z) - \mathbb{E}[S_N(z)]$ converges in probability to 0. Since $S_N(z)$ is also bounded by $|\mathrm{Im}\,z|^{-1}$ uniformly in N, we deduce that $\mathrm{Var}(S_N(z)) \to 0$. Therefore, in the limit $N \to \infty$, the right-hand side of (12) becomes 1, and any limit point of $\mathbb{E}[S_N(z)]$ when $N \to \infty$ satisfy a quadratic equation, whose solutions are $(z \pm \sqrt{z^2-4})/2$.

Now, let $\delta > 0$ be arbitrary, and $U_\delta = \{z \in \mathbb{C},\quad \pm\mathrm{Im}\,z \geq \delta\}$. The function $z \mapsto S_N(z)$ is holomorphic on U_δ and bounded by $|\mathrm{Im}\,z|^{-1}$ uniformly in N. By Montel's theorem, it has limit points for the pointwise convergence, and the limit points are also holomorphic functions on U_δ, bounded by $|\mathrm{Im}\,z|^{-1}$.

This is only compatible with the choice of $-$ sign (for any $z \in U_\delta$) for the squareroot. In particular, there is a unique limit point, so we have:

$$\lim_{N\to\infty} \mathbb{E}[S_N(z)] = \frac{z - \sqrt{z^2-4}}{2}.$$

We recognize the Stieltjes transform of the semi-circle law (Lemma 3.6). $\square$

Weak convergence of the empirical measure

By Stieltjes continuity (Theorem 3.7), the previous result (Lemma 3.21) shows that $L_N = N^{-1}\sum_{i=1}^{N}\delta_{\lambda_i^{(M_N)}}$ converges in probability to the semi-circle law μ_{sc} for the vague topology (test functions = continuous bounded with compact support). For convergence in the weak sense, we want to upgrade it to test functions $f \neq 0$ that do not have compact support. We still know (Theorem 3.17 and its proof) that $L_N[f] - \mathbb{E}\big[L_N[f]\big]$ converges in probability to 0 and that for any $\epsilon > 0$, one can choose $K_\epsilon > 2$ independent of N such that $\mathcal{A}_\epsilon = \big\{|\lambda_{\max}^{(M_N)}| > K_\epsilon\big\}$ has probability less than $\epsilon/6||f||_\infty$. Let $\tilde{f}$ a continuous bounded function with compact support, that coincides with f on $[-K_\delta, K_\delta]$. We write:

$$\big|L_N[f] - \mu_{\mathrm{sc}}[f]\big| \leq (L_N - \mu_{\mathrm{sc}})[\tilde{f}] + (L_N - \mathbb{E}L_N)[f - \tilde{f}] + \mathbb{E}\big[L_N[f - \tilde{f}]\big].$$

The last term is smaller than $\epsilon/3$ by construction, and the second (resp. the first) term converges to 0 in probability by concentration (by convergence in the vague topology). So, $L_N[f] - \mu_{\mathrm{sc}}[f]$ converges as well to 0 in probability.

3.7 *Further results*

Lecture 7 (1h30) October 27^{th}, 2014

Wigner's theorem is a statement about the collective behavior of eigenvalues, but there are many more questions that can be asked. We now describe a few results (without proof) giving finer information on the spectrum of Wigner matrices.

Maximum eigenvalue

From Lemma 3.19, we know that $\mathbb{E}[|\lambda_{\max}^{(M_N)}|]$ is bounded. An easy corollary of Wigner theorem:

3.22 LEMMA. *For any $\delta > 0$, $\mathbb{P}[\lambda_{\max}^{(M_N)} \leq 2-\delta] \to 0$ in the limit $N \to \infty$.*

Proof. Let f_δ a continuous bounded function with compact support included in $]2-\delta, +\infty[$. Multiplying f_δ by a constant, we can assume $\mu_{\mathrm{sc}}[f_\delta] = 1$. Then:

$$\mathbb{P}\big[\lambda_{\max}^{(M_N)} \leq 2-\delta\big] \leq \mathbb{P}\big[L^{(M_N)}[f] = 0\big] \leq \mathbb{P}[(L^{(M_N)} - \mu_{\mathrm{sc}})[f] \geq 1/2],$$

and the latter converges to 0 according to Wigner theorem. $\square$

So, we can conclude that $\liminf \mathbb{E}[\lambda_{\max}^{(M_N)}] \geq 2$, but it is not clear how to derive an upper bound. A priori, a finite number of eigenvalues could be detached from the right end of the support of μ_{sc} without affecting Wigner theorem.

3.23 THEOREM (Bai, Yin, 1988). *$\lambda_{\max}^{(M_N)}$ converges almost surely to a constant c iff $\mathbb{E}[|\sqrt{N}\,(M_N)_{1,2}|^4]$ is uniformly bounded in N. In this case, $c = 2$.*

Since $\lambda_{\max}$ is 1-Lipschitz (as inferred from Hofmann-Wielandt inequality), and a convex function of the matrix entries, we also know by Talagrand concentration that $\lambda_{\max}^{(M_N)}$ is concentrated in a region of width $O(N^{-1/2})$ around its mean. This result is not optimal, since fluctuations of $\lambda_{\max}^{(M_N)}$ are usually of order $N^{-2/3}$. And, they are described by the Tracy-Widom GOE law:

3.24 THEOREM (Soshnikov, 1999). *Assume there exists $c, C > 0$ independent of N such that $\mathbb{E}[e^{c(M_N)_{1,2}^2/N}]$ is uniformly bounded in N, and that all odd moments of $(M_N)_{1,2}$ vanish. Then:*

$$\mathbb{P}[\lambda_{\max}^{(M_N)} \leq 2 + N^{-2/3}s] \to \mathrm{TW}_1(s).$$

This result has a generalization describing the joint distribution of the fluctuation of the k largest eigenvalues for any fixed k when $N \to \infty$.

There exists a more combinatorial proof of Wigner theorem based on computation of moments, and exploiting the independence of the entries of Wigner matrices. To prove Theorem 3.24, Soshnikov performed an analysis of the moments of large degree $\mathrm{Tr}\,(M_N/2)^{\delta N^{2/3}}$ when $N \to \infty$ and δ is fixed. They naturally give access to the distribution of $\lambda_{\max}^{(M_N)}$ at scale $N^{-2/3}$ around the mean value 2, since $\big((\lambda + sN^{-2/3})/2\big)^{\delta N^{2/3}} = 1 + s\delta/2 + o(1)$ for $\lambda = 2$, whereas it decays (resp. grow) exponentially when $\lambda \in [0, 2[$ (resp. $\lambda > 2$).

Local semi-circle law

Wigner theorem allows to probe the spectrum in windows of size $O(1)$ when $N \to \infty$. It does not give access to the number of eigenvalues in a window of size $N^{-\delta}$ for $0 < \delta < 1$ (**mesoscopic scale**), or of size $1/N$ (**microscopic scale**). It is nevertheless true that the semi-circle law can be seen up to microscopic scale:

3.25 THEOREM (Erdös, Schlein, Yau, 2008). *Assume there exists $c > 0$ independent of N such that $\mathbb{E}[e^{c(M_N)_{1,2}^2/N}]$ is uniformly bounded in N. Let $\eta_N \to 0$ while $\eta_N \gg 1/N$, and $x \in]-2, 2[$. Then, the random variable:*

$$\mathcal{N}(x;\eta_N) = \frac{\text{number of eigenvalues in } [x - \eta_N/2, x + \eta_N/2]}{N\eta_N}$$

converges in probability to $\rho_{\mathrm{sc}}(x) = \sqrt{4 - x^2}/2\pi$.

As we have seen, $x \mapsto \operatorname{Im} S_N(x + \mathrm{i}\eta_N)$ is the density of the convolution of the empirical measure $L^{(M_N)}$ with a Cauchy law of width η_N. Allowing $\eta_N \to 0$ amounts to probing the empirical measure at mesoscopic and microsopic scales. In our proof, we very often used bounds proportional to inverse powers of $\operatorname{Im} z$, coming with decaying prefactors of N, so we can only conclude when $\eta_N \gg O(N^{-\delta})$ for $\delta > 0$ small enough. A much finer analysis of the almost fixed point equation is required to arrive to Theorem 3.25.

Central limit theorem

The fluctuation of linear statistics are the random variables of the form $\overline{L}^{(M_N)}[f] = N\big(L^{(M_N)}[f] - \mathbb{E}[L^{(M_N)}[f]]\big)$ for test functions f. As an effect of the strong correlation between eigenvalues[15], this is typically of order 1 – while it would be of order $\sqrt{N}$ if the eigenvalues were independent r.v. More precisely, we have a central limit theorem:

3.26 THEOREM (Lytova, Pastur, 2009). *Let $Y_N = \sqrt{N}(M_N)_{1,2}$. Under the assumptions:*

- $\mathbb{E}[|Y_N|^5]$ *is uniformly bounded in N.*
- *The fourth cumulant $\mathbb{E}[|Y_N|^4] - \big(\mathbb{E}[|Y_N|^2]\big)^2$ vanishes.*
- *The test function $f \in L^2(\mathbb{R})$ is such that $\int_{\mathbb{R}} (1 + |k|^5)\hat{f}(k) < \infty$.*

Then $\overline{L}^{(M_N)}[f]$ converges in law to a centered gaussian variable, with variance:

$$\sigma[f] = \frac{1}{2\pi^2} \int_{[-2,2]^2} \mathrm{d}x_1 \mathrm{d}x_2 \, \frac{4 - x_1 x_2}{\sqrt{(4 - x_1^2)(4 - x_2^2)}} \left(\frac{f(x_1) - f(x_2)}{x_1 - x_2} \right)^2.$$

Generalization to Wishart matrices

3.27 DEFINITION. A Wishart matrix is a $N \times N$ random matrix of the form $M_{N,K} = X_{N,K} X_{N,K}^T$, where X_N is a $N \times K$ matrix with i.i.d. entries of mean 0 and variance $1/N$.

Ensembles of Wishart matrices are useful models for the statistical analysis of large data: they can be compared to covariance matrices. If $K < N$, the rank of M_N is K, therefore M_N has at least $N - K$ deterministic zero eigenvalues. By construction, $M_{N,K}$ are positive symmetric matrices. The methods and results for Wigner matrices can usually be generalized to the case of Wishart matrices. The semi-circle law is replaced by the so-called Marčenko-Pastur law:

3.28 THEOREM (Marčenko, Pastur, 1967). *Assume K depends on N such that $K/N \to \alpha \in \mathbb{R}_+^*$ when $N \to \infty$. The empirical measure $L_{(M_{N,K})}$ converges in probability for*

[15] The configuration of eigenvalues are more ”rigid” than an i.i.d. configuration

the weak topology to the probability measure:

$$\mu_{\mathrm{MP}} = \max(1-\alpha, 0)\delta_0 + \frac{\sqrt{(b_+(\alpha)-x)(x-b_-(\alpha))}}{2\pi x}\,\mathbf{1}_{[b_-(\alpha),b_+(\alpha)]}(x)\,\mathrm{d}x,$$

where δ_0 *is the Dirac mass at* 0, *and* $b_\pm(\alpha) = (1 \pm \sqrt{\alpha})^2$.

4 Invariant ensembles: from matrices to eigenvalues

4.1 *Preliminaries*

We need a few unified notations the values $\beta = 1, 2, 4$. The definitions and basic statements in the unusual case $\beta = 4$ will be reminded in § 4.6 at the end of the chapter. We introduce:

- The fields:

$$\mathbb{K}_\beta = \begin{cases} \beta = 1 & : & \mathbb{R} \\ \beta = 2 & : & \mathbb{C} \\ \beta = 4 & : & \mathbb{H} \end{cases}$$

- The classical vector spaces of matrices:

$$\mathscr{H}_{N,\beta} = N \times N \begin{cases} \beta = 1 & : & \text{symmetric} \\ \beta = 2 & : & \text{hermitian} \\ \beta = 4 & : & \text{quaternionic self-dual} \end{cases}$$

- Their corresponding symmetry groups, i.e. the Lie groups with a left action on $\mathscr{H}_{N,\beta}$ by conjugation:

$$\mathscr{G}_{N,\beta} = N \times N \begin{cases} \beta = 1 & : & \text{orthogonal} \\ \beta = 2 & : & \text{unitary} \\ \beta = 4 & : & \text{quaternionic unitary} \end{cases}$$

  They actually coincide with $\mathscr{U}_N(\mathbb{K}_\beta)$. These are compact Lie groups, since they are defined by closed condition and they are bounded in $\mathscr{M}_N(\mathbb{K}_\beta)$. Their tangent space at identity are the vector space of matrices:

$$\mathscr{T}_{N,\beta} = N \times N \begin{cases} \beta = 1 & : & \text{antisymmetric} \\ \beta = 2 & : & \text{antihermitian} \\ \beta = 4 & : & \text{quaternionic, anti-self-dual} \end{cases}$$

  Their dimension is:

$$\dim \mathscr{G}_{N,\beta} = N + \beta\,\frac{N(N-1)}{2}$$

We have the basic diagonalization result:

4.1 lemma. *For any $A \in \mathscr{H}_{N,\beta}$, there exist $U \in \mathscr{G}_{N,\beta}$ and $D \in \mathscr{D}_N(\mathbb{R}) \subseteq \mathscr{M}_N(\mathbb{K}_\beta)$, such that $A = UDU^{-1}$.*

If $M \in \mathscr{M}_N(\mathbb{K}_\beta)$, we denote L_M (resp. R_M) the left (resp. right) multiplication by M, seen as an endomorphism of $\mathscr{M}_N(\mathbb{K}_\beta)$.

4.2 lemma. *If $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of M, then in the real vector space $\mathscr{M}_N(\mathbb{K}_\beta)$:*

- *the eigenvalues of L_M (resp. $L_M - R_M$) are the λ_i' for $i \in [\![1, N]\!]$ with multiplicity $N\beta$.*

- *the eigenvalues of $(L_M - R_M)$ are the $\lambda_i - \lambda_j$ for $i, j \in [\![1, N]\!]$ with multiplicity β.*

Proof. Any $M \in \mathscr{M}_N(\mathbb{K}_\beta)$ can be written in the form $M = PTP^{-1}$ with $P \in \mathscr{M}_N(\mathbb{K}_\beta)$ invertible and $T \in \mathscr{M}_N(\mathbb{K}_\beta)$ upper triangular, with real coefficients $\lambda_1, \ldots, \lambda_N$ on the diagonal. Since $L_M = L_P L_T L_P^{-1}$, L_M and L_T have the same eigenvalues. We denote $\mathrm{e}^{(\alpha)}$ with $\alpha \in [\![1, \beta]\!]$ the canonical basis of $\mathbb{K}_\beta$ considered as a real vector space, and $\ell^{(\alpha)}$ the dual basis. We also denote $E_{i,j}$ the canonical matrices filled with 0's, except for a 1 at the entry (i, j). The family of matrices

$$E_{i,j}^{(\alpha)} := \mathrm{e}^{(\alpha)} \cdot E_{i,j}, \qquad \alpha \in [\![1, \beta]\!], \quad i, j \in [\![1, N]\!]$$

defines a basis of the real vector space $\mathscr{M}_N(\mathbb{K}_\beta)$. We put the lexicographic order on the triples (i, j, α), and we can read in this basis the matrix of the endomorphism L_T:

$$\ell^{(\gamma)}\big[L_T(E_{i,j}^{(\alpha)})_{k,l}\big] = \delta_{j,l}\, \ell_\beta^{(\gamma)}\big[T_{k,i}\mathrm{e}^{(\alpha)}\big].$$

It has zero entries when $i > k$, therefore it is block-upper triangular. The eigenvalues can be read as eigenvalues of the blocks on the diagonal, i.e. the blocks with $i = k$. Since $T_{i,i} = \lambda_i$ is real and the entries vanish if $j \neq l$, these blocks are actually diagonal, and the diagonal elements are:

$$\ell^{(\alpha)}\big[L_T(E_{i,j}^{(\alpha)})_{i,j}\big] = \lambda_i.$$

Since the rows/columns of the blocks are indexed by $j \in [\![1, N]\!]$ and $\alpha \in [\![1, \beta]\!]$, the eigenvalue λ_i appears $N\beta$ times.

If we want to find the spectrum of $L_M - R_M$, we observe that left and right multiplication always commute, and we write $L_M - R_M = L_P R_P^{-1}(L_T - R_T) R_P L_P^{-1}$. Therefore, we only need to find the spectrum of $L_T - R_T$. The matrix elements of R_T are:

$$\ell^{(\gamma)}\big[R_T(E_{i,j}^{(\alpha)})_{k,l}\big] = \delta_{i,k}\ell^{(\gamma)}\big[\mathrm{e}^{(\alpha)}T_{j,l}\big],$$

which is zero when $j > l$. Arguing as for L_T, we find that this matrix is upper-triangular, with diagonal elements:

$$\ell^{(\alpha)}\big[L_T(E_{i,j}^{(\alpha)})_{i,j}\big] = \lambda_j.$$

Observe that we used the same basis to bring L_T and R_T to upper-triangular form, so we can conclude that $(\lambda_i - \lambda_j)$ for $i, j \in [\![1, N]\!]$ repeated β times (one for each index α) are the eigenvalues of $L_T - R_T$. $\square$

4.2 *Diagonalization*

Lecture 8 (1h30)
October 28th, 2014

If $F : \mathscr{H}_{N,\beta} \to \mathbb{C}$ is invariant under conjugation by $\mathscr{G}_{N,\beta}$, it means that $f(M)$ only depends on the eigenvalues of M, and we identify it with a function $f : \mathbb{R}^N \to \mathbb{C}$, with same regularity as F. If we want to integrate F over a space of matrices, we should then be able to integrate out the eigenvectors. This leaves us with a remarkable form for the integration over eigenvalues:

4.3 THEOREM. *There exists a universal constant $v_{N,\beta} > 0$, such that, for any smooth function f on $\mathscr{H}_{N,\beta}$ with exponential decay at infinity,*

$$\int_{\mathscr{H}_{N,\beta}} \mathrm{d}M\, F(M) = \frac{v_{N,\beta}}{N!} \int_{\mathbb{R}^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^{\beta} f(\lambda_1, \ldots, \lambda_N).$$

Although writing a neat proof requires setting up a bit of differential geometry, it should not hide the key point, which is the computation of a Jacobian of the change of variable from the entries of a matrix to its eigenvalues and eigenvectors. Differentiating naively $M = UDU^{-1}$, we get $\partial M = [(\partial U)U^{-1}, M] + U \cdot \mathrm{d}D \cdot U^{-1}$, and we need to compute the determinant of the operator

$$(\mathfrak{u}, \mathfrak{d}) \mapsto U\big([\mathfrak{u}, D] + \mathfrak{d}\big)U^{-1}$$

in suitable vector spaces and basis. If the reader wants to avoid the technical aspects of the proof, I advise to jump directly to (14). As a matter of fact, the constant can be identified:

$$(13) \quad v_{N,\beta} = \frac{\mathrm{Vol}(\mathscr{G}_{N,\beta})}{[\mathrm{Vol}(\mathscr{G}_{1,\beta})]^N}.$$

once the notion of Riemannian volume for submanifolds of $\mathscr{M}_N(\mathbb{K}_\beta)$ is properly defined. This is done at the end of proof of Theorem 4.3, and requires elementary knowledge of Riemannian geometry.

4.4 COROLLARY. *If M is a random matrix whose entries joint p.d.f. is $Z_{N,\beta}^{-1}\,\mathrm{d}M\,F(M)$ where F is a non-negative function, then the joint p.d.f. for the ordered eigenvalues $\lambda_1 \geq \cdots \lambda_N$ is*

$$\frac{v_{N,\beta}}{N!\, Z_{N,\beta}} \prod_{i=1}^{N} \mathrm{d}\lambda_i \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^{\beta} f(\lambda_1, \ldots, \lambda_N).$$

Proof of Theorem 4.3

In a vector space V with a canonical, ordered, basis (like $\mathscr{H}_{N,\beta}$ or $\mathbb{R}^N$), the Lebesgue volume form $\omega_{\mathrm{Leb}}^{V}$ is the ordered exterior product of dual basis of linear forms. We will drop V and just write ω_{Leb} since the relevant vector space can be read obviously from the context. We wish to compute $\mathcal{I}_{N,\beta}[f] = \int_{\mathscr{H}_{N,\beta}} \omega^{\mathrm{Leb}} f$.

Matrices $M \in \mathscr{H}_{N,\beta}$ can be diagonalized in the form $M = UDU^{-1}$, where

$U \in \mathscr{G}_{N,\beta}$ is a matrix of eigenvectors, and $D \in \mathscr{D}_N(\mathbb{R})$ is a diagonal matrix whose entries are the eigenvalues $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$. $M = VD'V^{-1}$ is another decomposition iff there exists a permutation matrix $P \in \mathfrak{S}_N$ such that $D' = PDP^{-1}$, and the matrix $U^{-1}VP \in \mathscr{G}_{N,\beta}$ commute with D. If the eigenvalues are simple[16], it means that $U^{-1}VP$ is a diagonal matrix, i.e. an element of $\mathscr{G}_{1,\beta}^N$. To compute $\mathcal{I}_{N,\beta}[f]$, it suffices to integrate $f\omega^{\mathrm{Leb}}$ over the subset $(\mathscr{H}_{N,\beta})_\Delta$ of matrices with simple eigenvalues, since it is open and dense in $\mathscr{H}_{N,\beta}$. Let us denote $(\mathbb{R}^N)_\Delta$ the set of vectors in $\mathbb{R}^N$ with pairwise distinct components in increasing order, that we identify with diagonal matrices with simple and ordered eigenvalues. The previous argument shows that the smooth map:

$$\begin{array}{rcl} \Phi\,:\,(\mathbb{R}^N)_\Delta \times \mathscr{G}_{N.\beta} & \longrightarrow & (\mathcal{H}_{N,\beta})_\Delta \\ (D,U) & \longmapsto & UDU^{-1} \end{array}$$

is a fibration whose fiber is the Lie group $\mathscr{G}_{1,\beta}^N$. This group is naturally embedded as a closed subgroup of $\mathscr{G}_{N,\beta}$, and acts by right multiplication on $\mathscr{G}_{N,\beta}$, so the quotient $\widetilde{\mathscr{G}}_{N,\beta} = \mathscr{G}_{N,\beta}/\mathscr{G}_{1,\beta}^N$ is smooth. If we denote π the map projecting to the quotient, we have a commutative diagram of smooth maps:

$$\begin{tikzcd} (\mathbb{R}^N)_\Delta \times \mathscr{G}_{N,\beta} \arrow[r, "\Phi"] \arrow[d, "\pi"] & (\mathscr{H}_{N,\beta})_\Delta \\ \mathscr{B}_{N,\beta} = (\mathbb{R}^N)_\Delta \times \widetilde{\mathscr{G}}_{N,\beta} \arrow[ur, "\widetilde{\Phi}"'] & \end{tikzcd}$$

and the induced map $\widetilde{\Phi}$ is a smooth diffeomorphism between $(\mathbb{R}^N)_\Delta \times \widetilde{\mathscr{G}}_{N,\beta}$ and $(\mathscr{H}_{N,\beta})_\Delta$. The tangent space of $(\mathbb{R}^N)_\Delta \times \mathscr{G}_{N,\beta}$ at the point (D,U) is identified with $\mathbb{R}^N \oplus T_{\mathrm{id}}L_U(\mathscr{T}_{N,\beta})$, and the differential of π sends it onto $\mathbb{R}^N \oplus L_U(\mathscr{T}_{N,\beta}^{(0)})$, where the latter is the set of matrices in $\mathscr{T}_{N,\beta}$ with vanishing diagonal. We then compute the differential:

$$\begin{array}{rcl} T_{(D,U)}\widetilde{\Phi}\,:\,\mathbb{R}^N \oplus T_{\mathrm{id}}L_U(\mathscr{T}_{N,\beta}^{(0)}) & \longrightarrow & \mathscr{H}_{N,\beta} \\ (\mathfrak{d}\ ,\ \mathfrak{u}) & \longmapsto & U\big((R_D - L_D)[\mathfrak{u}] + \mathfrak{d}\big)U^{-1} \end{array} \tag{14}$$

and the Jacobian we need to compute is the determinant of this operator in the canonical basis for the source and the target vector spaces. It has the same determinant as the operator $(\mathfrak{u}, \mathfrak{d}) \mapsto (R_D - L_D)[\mathfrak{u}] + \mathfrak{d}$ in canonical basis. The target space (and its canonical basis) splits into purely diagonal and purely off-diagonal elements:

$$\mathscr{H}_{N,\beta} = \mathscr{H}_{N,\beta}^{\mathrm{diag}} \oplus \mathscr{H}_{N,\beta}^{\mathrm{off}},$$

[16]The commutant of D in general is the subgroup of $\mathscr{G}_{N,\beta}$ leaving stable all the eigenspaces of D: if the eigenvalues are not simple, some eigenspaces have dimension larger than 1 and the commutant contains more than just diagonal matrices.

and we observe – directly or reminding the proof of Lemma 4.2 – that

$$(R_D - L_D)(\mathscr{H}^{\mathrm{diag}}_{N,\beta}) = 0,$$

while $R_D - L_D$ restricted to $\mathscr{H}^{\mathrm{off}}_{N,\beta}$ has the eigenvalues $(\lambda_j - \lambda_i)$ for $1 \leq i < j \leq N$ repeated β times. Therefore, the operator takes an upper-triangular form, and we have:

$$\det[T_{(D,U)}\widetilde{\Phi}] = \prod_{1 \leq i < j \leq N} (\lambda_j - \lambda_i)^{\beta}. \tag{15}$$

In particular, the absolute value of (15) is the Jacobian, and it does not depend on U. Since the eigenvalues are labeled in increasing order in $(\mathbb{R}^N)_\Delta$, the sign of this expression is positive, meaning that the orientations of $\widetilde{\Phi}\big[(\mathbb{R}^N)_\Delta \times \widetilde{\mathcal{G}}_{N,\beta}\big]$ and $(\mathscr{H}_{N,\beta})_\Delta$ coincide. Therefore, we have:

$$\mathcal{I}_{N,\beta}[F] = \int_{(\mathbb{R}^N)_\Delta \times \widetilde{\mathscr{G}}_{N,\beta}} (F \circ \widetilde{\Phi})\ (\widetilde{\Phi}^* \omega_{\mathrm{Leb}}) = \Big(\int_{(\mathbb{R}^N)_\Delta} f \omega_{\mathrm{Leb}}\Big)\Big(\int_{\widetilde{\mathscr{G}}_{N,\beta}} \widetilde{\Omega}_{N,\beta}\Big). \tag{16}$$

We have used Fubini theorem in the second expression. Since our assumption on F implies that f does not depend on the order of eigenvalues, we can also symmetrize this formula, and add to the integral the zero contribution coming from maybe non-distinct eigenvalues:

$$\mathcal{I}_{N,\beta}[F] = \frac{v_{N,\beta}}{N!} \int_{\mathbb{R}^N} f \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^{\beta}\, \omega_{\mathrm{Leb}}, \qquad v_{N,\beta} := \int_{\widetilde{\mathscr{G}}_{N,\beta}} \widetilde{\Omega}_{N,\beta}.$$

Identification of the constant (13)

The proof can be continued to identify the geometric meaning of the constant $v_{N,\beta}$. The volume form $\widetilde{\Omega}_{N,\beta}$ (at the point U) that appear in the derivation of (16) is the pull-back by $T_{\mathrm{id}} L_U$ of the Lebesgue volume form at 0 on $\mathscr{T}^{(0)}_{N,\beta}$. We remark that the Lebesgue volume form on a vector space (considered as a manifold) equipped with a canonical basis is the Riemannian volume form for the canonical flat metric on that manifold. This applies to $\mathscr{M}_N(\mathbb{K}_\beta)$, $\mathscr{T}_{N,\beta}$ and $\mathscr{T}^{(0)}_{N,\beta}$. One can also check that if $U \in \mathscr{G}_{N,\beta}$, the diffeomorphism L_U acts on those spaces as an isometry for the canonical metric. Therefore, if we complete the canonical basis of $\mathscr{T}^{(0)}_{N,\beta}$ by adding elements $\mathfrak{w}_1, \ldots, \mathfrak{w}_r$ to obtain the canonical basis of $\mathscr{M}_N(\mathbb{K}_\beta)$, $L_U \mathfrak{w}_1, \ldots, L_U \mathfrak{w}_r$ is an orthonormal frame for $\widetilde{\mathscr{G}}_{N,\beta}$ considered as a submanifold of $\mathscr{M}_N(\mathbb{K}_\beta)$. Combining all those observations:

$$[\Omega_{N,\beta}]_U(\mathfrak{u}_1, \ldots, \mathfrak{u}_t) = \omega^{\mathscr{M}_N(\mathbb{K}_\beta)}_{\mathrm{Leb}}(\mathfrak{u}_1, \ldots, \mathfrak{u}_t, \mathfrak{w}_1, \ldots, \mathfrak{w}_r)$$

is the Riemannian volume form for the metric on $\widetilde{\mathscr{G}}_{N,\beta} \subseteq \mathscr{M}_N(\mathbb{K}_\beta)$ induced by the canonical metric on $\mathscr{M}_N(\mathbb{K}_\beta)$. Consequently, $v_{N,\beta}$ is indeed the Riemannian volume of $\widetilde{\mathscr{G}}_{N,\beta}$ considered as a submanifold of the Riemannian manifold $\mathscr{M}_N(\mathbb{K}_\beta)$. Similarly:

- the volume form $\Omega_{N,\beta}$ (at the point U) defined as the push-forward by L_U of the Lebesgue volume form at 0 of $\mathscr{T}_{N,\beta}$, must coincide with the Riemannian volume form of the submanifold $\mathscr{G}_{N,\beta} \subseteq \mathscr{M}_N(\mathbb{K}_\beta)$.
- $\Omega_{N,\beta}$ can be written – as one checks on the tangent space at identity and push-forward with L_U:

$$[\widetilde{\Omega}_{N,\beta}]_U(\mathfrak{u}_1,\ldots,\mathfrak{u}_r) = [\Omega_{N,\beta}]_U(\mathfrak{u}_1,\ldots,\mathfrak{u}_r,\mathfrak{e}_1,\ldots,\mathfrak{e}_N),$$

  where $\mathfrak{e}_1,\ldots,\mathfrak{e}_N$ is an orthonormal basis consisting of diagonal matrices that complete the canonical basis of $\mathscr{T}^{(0)}_{N,\beta}$ to the canonical basis of $\mathscr{T}_{N,\beta}$.
- One can check that the diffeomorphism R_U is an isometry of $\mathscr{M}_N(\mathbb{K}_\beta)$ for any $U \in \mathscr{G}_{N,\beta}$. A fortiori, for any $U \in \mathscr{G}^N_{1,\beta} \subseteq \mathscr{G}_{N,\beta}$, R_U acts by isometries on the submanifold $\mathscr{G}_{N,\beta} \subseteq \mathscr{M}_N(\mathbb{K}_\beta)$ equipped with the induced metric.

All in all, the quotient map $\pi : \mathscr{G}_{N,\beta} \to \widetilde{\mathscr{G}}_{N,\beta}$ is a smooth fibration whose fibers are isometric $\mathscr{G}^N_{1,\beta}$'s. Therefore, we have a further decomposition, using Fubini and naturality of the embedding $\mathscr{G}^N_{1,\beta} \hookrightarrow \mathscr{G}_{N,\beta}$:

$$\int_{\mathscr{G}_{N,\beta}} \Omega_{N,\beta} = \Big(\int_{\widetilde{\mathscr{G}}_{N,\beta}} \widetilde{\Omega}_{N,\beta}\Big)\Big(\int_{\mathscr{G}^N_{1,\beta}} \Omega^{\wedge N}_{1,\beta}\Big),$$

which turns into the claimed (13). **Remark:** Since both L_U and R_U act by isometries on $\mathscr{M}_N(\mathbb{K}_\beta)$, this is also true for the Riemannian submanifold $\mathscr{G}_{N,\beta}$, which is a compact Lie group. It is a classical (and usually non constructive) theorem that there exists a unique left and right invariant Riemannian metric on a compact Lie group for which the total volume of the group is 1: it is called the **Haar metric**. Therefore $\Omega_{N,\beta}$ is proportional to the Haar measure – and provides the quickest way to compute it.

Eigenvalues joint p.d.f.

Let $F : \mathscr{H}_{N,\beta} \to \mathbb{R}$ be an integrable smooth function on $\mathscr{H}_{N,\beta}$ invariant under conjugation by $\mathscr{G}_{N,\beta}$. So, $F(M)$ depends only on the eigenvalues of M, and we denote $f : \mathbb{R}^N \to \mathbb{R}$ the corresponding symmetric function of N variables. We have shown:

$$Z_{N,\beta}[F] = \int_{\mathscr{H}_{N,\beta}} F\,\omega_{\text{Leb}} = v_{N,\beta} \int_{(\mathbb{R}^N)_\Delta} f\,\omega_{\text{Leb}}$$

The joint p.d.f of the eigenvalues ρ – if it exists – is characterized as the smooth function defined on:

$$\overline{(\mathbb{R}^N)_\Delta} = \{\lambda \in \mathbb{R}^N, \qquad \lambda_1 \leq \cdots \leq \lambda_N\},$$

such that, for any continuous bounded function g with compact support:

$$\int_{(\mathbb{R}^N)_\Delta} g(\lambda)\,\rho(\lambda) \prod_{i=1}^{N} \mathrm{d}\lambda_i = \frac{Z_{N,\beta}[FG]}{Z_{N,\beta}[F]}.$$

We identify:

$$\rho(\lambda_1,\ldots,\lambda_N) = \prod_{i=1}^{N} f(\lambda_1,\ldots,\lambda_N) \prod_{1\leq i<j\leq N} |\lambda_i - \lambda_j|^\beta.$$

4.3 *The Selberg integral*

Lecture 9 (1h30)
November 3rd, 2014

We will see in Chapter 5 that many exact computations can be done in the invariant ensembles, i.e. $\beta \in \{1,2,4\}$. The Selberg integral is one of the few explicit computations that can be done for arbitrary values of β

$$\mathcal{I}_N(a,b,\gamma) = \int_{[0,1]^N} \big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^{2\gamma} \prod_{i=1}^{N} \lambda_i^{a-1}(1-\lambda_i)^{b-1}\mathrm{d}\lambda_i, \qquad \gamma = \beta/2.$$

4.5 THEOREM. *Let* $N \geq 1$ *and* $a,b,\gamma \in \mathbb{C}$ *such that* $\mathrm{Re}\,a > 0$, $\mathrm{Re}\,b > 0$ *and* $\mathrm{Re}\,\gamma > \max[-1/N, -\mathrm{Re}\,a/(N-1), -\mathrm{Re}\,b/(N-1)]$. *The integral converges and is equal to:*

$$\mathcal{I}_N(a,b,\gamma) = \prod_{m=1}^{N} \frac{\Gamma(1+m\gamma)}{\Gamma(1+\gamma)}\,\frac{\Gamma\big(a+\gamma(m-1)\big)\Gamma\big(b+\gamma(m-1)\big)}{\Gamma\big(a+b+\gamma(N+m-2)\big)}.$$

Various formulas can be derived by taking limits in a and b. The most useful one is:

4.6 COROLLARY (Selberg Gaussian integral).

$$\int_{\mathbb{R}^N} \big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^\beta \prod_{i=1}^{N} e^{-c\lambda_i^2/2}\mathrm{d}\lambda_i = (2\pi)^{N/2}\, c^{-(N/4)(\beta(N-1)+2)} \prod_{m=1}^{N} \frac{\Gamma(1+\gamma m)}{\Gamma(1+\gamma)}.$$

The dependence in c in this formula is obvious by rescaling the integration variables.

Proof of Theorem 4.5

We follow Selberg's original proof. The strategy is to prove the result for integer γ, and then extending it by results of analytic continuation. For $N = 1$, this is the well-known Euler Beta integral:

$$I_1(a,b,\gamma) = \int_0^1 \mathrm{d}\lambda\, \lambda^{a-1}(1-\lambda)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Let us start with the intermediate formula:

4.7 lemma. *For any $\gamma \in \mathbb{N}$, there exists a constant $c_N(\gamma)$ independent of a and b such that:*

$$(17)\quad \mathcal{I}_N(\gamma) = c_N(\gamma) \prod_{m=1}^{N} \frac{\Gamma\big(a+\gamma(m-1)\big)\Gamma\big(b+\gamma(m-1)\big)}{\Gamma\big(a+b+\gamma(N+m-2)\big)}.$$

Proof. We will exploit the symmetries of the problem. Since γ is an integer, the factor $|\Delta(x_1,\ldots,x_N)|^{2\gamma}$ is a polynomial, which is symmetric in its N variables, and homogeneous of degree $\gamma N(N-1)$. Firstly, let us decompose it:

$$(18)\quad \big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^{\gamma} = \sum_{\substack{j_1+\cdots j_N \\ =\gamma N(N-1) \\ j_1\leq\cdots\leq j_N}} c^{(N)}_{j_1,\ldots,j_N} \lambda_1^{j_1}\cdots\lambda_N^{j_N} + \cdots$$

where the $\cdots$ means that we add the terms obtained by symmetrizing in the λ's. We make the simple observation that $j_N \leq \gamma(N-1)$, since the sum of j's is equal to $\gamma N(N-1)$ and j_N is their maximum. Secondly, if $m \in [\![1,N]\!]$, we can also factor a Vandermonde determinant of the m first variables, so there exists a polynomial $P_{m,N}(\lambda_1,\ldots,\lambda_N)$ such that:

$$\begin{aligned}
\big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^{2\gamma} & = \big|\Delta(\lambda_1,\ldots,\lambda_m)\big|^{2\gamma} P_{N,m}(\lambda_1,\ldots,\lambda_N) \\
& = \Big(\sum_{\substack{k_1+\cdots+k_m \\ =\gamma m(m-1) \\ k_1\leq\cdots\leq k_m}} c^{(m)}_{k_1,\ldots,k_m} \prod_{i=1}^{m} \lambda_i^{k_i} + \cdots\Big) P_{N,m}(\lambda_1,\ldots,\lambda_N).
\end{aligned}$$

We deduce that the $c^{(m)}_{k_1,\ldots,k_m}$ contribute to $c^{(N)}_{j_1,\ldots,j_N}$ for $k_i \leq j_i$. In particular, we deduce from our simple remark that the non-zero terms appearing in (19) must have $\gamma(m-1) \leq j_m$, for any m. Thirdly, we notice the palindromic symmetry:

$$\begin{aligned}
\big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^{2\gamma} & = \Big[\prod_{i=1}^{N} \lambda_i^{2\gamma(N-1)}\Big]\big|\Delta(1/\lambda_1,\ldots,1/\lambda_N)\big|^{2\gamma} \\
& = \sum_{\substack{\ell_1+\cdots+\ell_N \\ =\gamma N(N-1) \\ \ell_1\leq\cdots\ell_N}} c^{(N)}_{\ell_1,\ldots,\ell_N} \prod_{i=1}^{N} \lambda_i^{2\gamma(N-1)-\ell_i} + \cdots
\end{aligned}$$

and this shows that $\ell_i = 2\gamma(N-1) - j_i$ for some collection of j's appearing in (18). Using the same argument as before, we conclude to the bounds:

$$(19)\quad \forall m \in [\![1,N]\!], \qquad \gamma(m-1) \leq j_m \leq \gamma(N+m-2)$$

for the indices appearing in (18). If we plug this expansion into the Selberg integral, we obtain a linear combination of a product of N Euler Beta integrals:

$$\mathcal{I}_N(a,b,\gamma) = \sum_{\mathbf{j}} c_{\mathbf{j}}^{(N)} \prod_{m=1}^{N} \frac{\Gamma(a+\gamma j_m)\Gamma(b)}{\Gamma(a+b+j_m)}.$$

The strategy is then to factor out as much as possible of the Gamma functions – using the bound (19) – , so that the remaining factor is a polynomial in a. This procedure breaks the symmetry between a and b that is seen by the change of variable λ's $\mapsto (1-\lambda)$'s:

$$\mathcal{I}_N(a,b,\gamma) = \mathcal{I}_N(b,a,\gamma). \tag{20}$$

So, we also factor out the corresponding terms in b to respect the symmetry:

$$\mathcal{I}_N(a,b,\gamma) = \Big[\prod_{m=1}^{N} \frac{\Gamma\big(a+\gamma(m-1)\big)\Gamma\big(b+\gamma(m-1)\big)}{\Gamma\big(a+b+(N+m-2)\big)}\Big] \frac{Q_N(a,b,\gamma)}{R_N(b,\gamma)},$$

with:

$$\begin{aligned} Q_N(a,b,\gamma) &= \sum_{j_1,\ldots,j_N} c_{j_1,\ldots,j_N}^{(N)} \prod_{m=1}^{N} \frac{\Gamma(a+j_m)}{\Gamma\big(a+\gamma(m-1)\big)} \frac{\Gamma\big(a+b+\gamma(N+m-2)\big)}{\Gamma(a+b+\gamma j_m)}, \\ R_N(b,\gamma) &= \prod_{m=1}^{N} \frac{\Gamma\big(b+\gamma(m-1)\big)}{\Gamma(b)}. \end{aligned}$$

As we desired, the outcome of this decomposition is that Q_N and R_N are polynomials in their variables a and/or b. Since $\sum_{m=1}^{N} j_m = \gamma N(N-1)$, we count:

$$\begin{aligned} \deg_b Q_N(a,b,\gamma) &\leq \sum_{m=1}^{N} \big(\gamma(N+m-2) - j_m\big) = \gamma\frac{N(N-1)}{2}, \\ \deg_b R_N(b,\gamma) &= \sum_{m=1}^{N} \gamma(m-1) = \gamma\,\frac{N(N-1)}{2}. \end{aligned}$$

The first line is an inequality and maybe not an equality because cancellations of terms could occur from the sum over j_i's. Since a and b play a symmetric role in $\mathcal{I}_N$, we must have:

$$c_N(a,b,\gamma) := \frac{Q_N(a,b,\gamma)}{R_N(b,\gamma)} = \frac{Q_N(b,a,\gamma)}{R_N(a,\gamma)}.$$

The previous identity can be rewritten:

$$Q_N(a,b,\gamma)R_N(a,\gamma) = Q_N(b,a,\gamma)R_N(b,\gamma),$$

and we observe that the left-hand side has degree smaller or equal to $\gamma N(N-1)/2$ in b. Since $R_N(b,\gamma)$ on the right-hand side already has degree $\gamma N(N-1)/2$, it shows that $Q_N(b,a,\gamma)$ actually does not depend on b. Hence $c_N(a,b,\gamma)$ does not depend on b, and the same reasoning if we exchange a and b shows that $c_N(a,b,\gamma) := c_N(\gamma)$ depends neither on a nor on b. $\square$

To compute $c_N(\gamma)$, we will prove a recursion in N for the Selberg integral with special values of a and b.

4.8 LEMMA. *For $N \geq 2$, we have:*

$$\mathcal{I}_N(1,1,\gamma) = \frac{\mathcal{I}_{N-1}(1,1+2\gamma,\gamma)}{1+\gamma(N-1)}.$$

Proof. Since the integrand is symmetric in its N variables, the integral over $[0,1]^N$ is N times the integral over the N-uples $(\lambda_1,\ldots,\lambda_N)$ such that $\lambda_i \leq \lambda_1$ for $i \in [\![2,N]\!]$:

$$\mathcal{I}_N(1,1,\gamma) = N\int_0^1 \mathrm{d}\lambda_1 \int_{[0,\lambda_1]^{N-1}} \prod_{i=1}^{N-1} \mathrm{d}\lambda_i\, \big|\Delta(x_2,\ldots,x_N)\big|^{2\gamma}.$$

We exploit the homogeneity of the Vandermonde and rescale $y_i = \lambda_i/\lambda_1$ for $i \in [\![2,N]\!]$ to find:

$$\mathcal{I}_N(1,1,\gamma) = N\int_0^1 \mathrm{d}\lambda_1 \lambda_1^{N-1+\gamma N(N-1)} \int_{[0,1]^{N-1}} \prod_{i=2}^{N} \mathrm{d}y_i\,(1-y_i)^{2\gamma}\,\big|\Delta(y_2,\ldots,y_N)\big|^{2\gamma}$$

and the first integral is factored out and easy to compute. $\square$

If we insert Lemma 4.7 in both sides of the equality in Lemma 4.8, we obtain after simplification:

$$c_N(\gamma) = \frac{\Gamma(1+\gamma N)}{\Gamma(1+\gamma)}\, c_{N-1}(\gamma).$$

With the initial condition $c_1(\gamma) = 1$ that can be identified by comparing Lemma 4.7 with the Euler Beta integral, we deduce:

$$c_N(\gamma) = \prod_{m=1}^{N} \frac{\Gamma(1+\gamma m)}{\Gamma(1+\gamma)},$$

and the value of Selberg integral (17) for all $\gamma \in \mathbb{N}$.

We would like to extend this formula to "all possible values" of γ. Let us denote $f(\gamma) = \mathcal{I}_N(a,b,\gamma)$, and $g(\gamma)$ the function of γ in the right-hand side of (17). f and g are obviously holomorphic functions on:

$$\mathbb{C}_+ := \{\gamma \in \mathbb{C}, \quad \mathrm{Re}\,\gamma > 0\},$$

and continuous on $\overline{\mathbb{C}_+}$. f is uniformly bounded on $\mathbb{C}_+$, since:

$$\big|f(\gamma)\big| \leq \int_{[0,1]^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i\, \lambda_i^{\mathrm{Re}\,a-1}(1-\lambda_i)^{\mathrm{Re}\,b-1} = \Big[\frac{\Gamma(\mathrm{Re}\,a)\Gamma(\mathrm{Re}\,b)}{\Gamma\big(\mathrm{Re}(a+b)\big)}\Big]^N$$

It is also true that g is uniformly bounded on $\mathbb{C}_+$, although that requires some more calculus. This is more than enough to apply Carlson's theorem (Theorem 4.12 in § 4.5), and conclude that $f(\gamma) = g(\gamma)$ over the domain of definition and analyticity of these two functions – which contains $\mathbb{C}_+$. Let us now justify:

4.9 lemma. *$g(\gamma) \in O(\gamma^{-N/2})$ when $\gamma \to \infty$ in $\overline{\mathbb{C}_+}$.*

Proof. We write in an obvious decomposition $g(\gamma) = \prod_{m=1}^{N} g_m(\gamma)$, and compute the asymptotics of the factors by Stirling formula. The result is:

$$g_m(\gamma) \underset{\gamma \to \infty}{\sim} C_m \, \gamma^{-1/2} \Big[\frac{m^m (m-1)^{2(m-1)}}{(N+m-2)^{N+m-2}}\Big]^{\gamma} e^{(\gamma \ln \gamma - \gamma)(2m-1-N)}.$$

for some irrelevant constant $C_m > 0$, and we remind the convention $0^0 = 1$. In the product over $m \in [\![1, N]\!]$, this last factor disappears since $\sum_{m=1}^{N}(2m-1-N) = 0$, and we find:

$$g(\gamma) \underset{\gamma \to \infty}{\sim} C \, \gamma^{-N/2} \Big[\frac{N}{\big(4(1-1/N)\big)^{N-1}} \prod_{m=2}^{N-1} \frac{m^m (m-1)^{m-1} (N-m)^{N-m}}{(N+m-2)^{N+m-2}}\Big]^{\gamma}.$$

To arrive to this form, we have performed the change of index $m \to N-m$ in one of the product $\prod_{m=1}^{N}(m-1)^{m-1}$, and put apart contributions from $m = 1$ and $m = N$. We need to check that the quantity in the brackets is smaller than 1 in order to prove that g is bounded. Since $N \geq 2$, it is obvious that:

$$\frac{N}{\big(4(1-1/N)\big)^{N-1}} \leq N \, 2^{-(N-1)} \leq 1$$

Let us set:

$$\phi(x) := x \ln x + (x-1) \ln(x-1) + (N-x) \ln(N-x) - (N+x-2) \ln(N+x-2).$$

It is less obvious but we now justify that $\phi(x) \leq 0$ for $x \in [2, N-1]$ and $N \geq 2$. Indeed, we calculate the second derivative:

$$\phi''(x) = \frac{x^2 + N(N-2)(2x-1)}{x(x-1)(N-x)(N+x-2)} > 0.$$

We see that $\phi(x)$ is convex, so it must achieves its maximum on the boundary of its domain, i.e. $x = 2$ or $x = N-1$. And we have:

$$\begin{aligned} \phi(2) &= 2 \ln 2 + (N-2) \ln(N-2) - N \ln N \\ &= N\Big[\frac{2}{N} \ln\Big(\frac{2}{N}\Big) + \frac{N-2}{N} \ln\Big(\frac{N-2}{N}\Big)\Big] < 0 \end{aligned}$$

and similarly:

$$\phi(N-1) = (N-1) \ln(N-1) + (N-2) \ln(N-2) - (2N-3) \ln(2N-3) < 0.$$

$\square$

Domain of analyticity of the Selberg integral

We would like to determine the maximal domain on which $\mathcal{I}_N(a,b,\gamma)$ is an absolutely convergent integral for complex-valued a, b and γ. The integrability at $x_i \to 0$ or 1 imposes $\mathrm{Re}\, a > 0$ and $\mathrm{Re}\, b > 0$, and let us assume these two conditions are fulfilled. When $\mathrm{Re}\,\gamma < 0$, we have to estimate the contribution to the integral of a vicinity of the $\lambda_1 = \lambda_2 = \ldots = \lambda_N$, at (still !) at the vicinity of the boundary of $[0,1]^N$. We shall focus on the diagonal, and leave the study at the boundaries to the reader. Let us fix $\eta > 0$, and:

$$D(\epsilon) := \big\{\lambda \in [\eta, 1-\eta]^N, \quad \forall i,j \in [\![1,N]\!],\ i \neq j \Rightarrow |\lambda_i - \lambda_j| > \epsilon\big\}.$$

$\mathcal{I}_N(a,b,\gamma)$ is absolutely convergent in $[\eta, 1-\eta]^N$ on the diagonal when:

$$\mathcal{J}(\epsilon) = \int_{D(\epsilon)} \prod_{i=1}^{N} \mathrm{d}\lambda_i\, \lambda_i^{\mathrm{Re}\,a-1}(1-\lambda_i)^{\mathrm{Re}\,b-1}\big|\Delta(\lambda_1,\ldots,\lambda_N)\big|^{2\mathrm{Re}\,\gamma}$$

remains bounded when $\epsilon \to 0$. For any $m \in [\![1, N(N-1)/2]\!]$, we introduce $F_m(\epsilon) \subseteq [\eta, 1-\eta]^N$ the set of λ's such that there exists a subset $J \subseteq [\![1,N]\!]$ with m elements, for which:

- for any $i \in J$, there exists $j \in J$ such that $|\lambda_i - \lambda_j| \leq \epsilon$.
- for any $i \notin J$ and any $j \in [\![1,N]\!]$, $|\lambda_i - \lambda_j| > \epsilon$.

$F_m(\epsilon)$ is a measurable set with Lebesgue volume satisfying:

$$c\,\epsilon^{m-1} \leq \mathrm{Vol}(F_m(\epsilon)) \leq c'\epsilon^{m-1}$$

for some constants $c, c' > 0$ independent of ϵ. We observe:

$$D(\epsilon/2) \setminus D(\epsilon) = D(\epsilon/2) \cap \Big(\bigcup_{m=0}^{N} F_m(\epsilon)\Big).$$

And for $\lambda \in D(\epsilon/2) \cap F_m(\epsilon)$, we can estimate:

$$\epsilon^{N(N-1)\mathrm{Re}\,\gamma} \leq \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^{2\mathrm{Re}\,\gamma} \leq (\epsilon/2)^{2m\mathrm{Re}\,\gamma}\,\epsilon^{(N(N-1)/2-m)2\mathrm{Re}\,\gamma}.$$

Therefore:

$$C\,\epsilon^{N-1+N(N-1)\mathrm{Re}\,\gamma} \leq \mathcal{J}(\epsilon/2) - \mathcal{J}(\epsilon) \leq C'\,\epsilon^{N-1+N(N-1)\mathrm{Re}\,\gamma}$$

for some constants $C, C' > 0$ independent of ϵ. Applying this double inequality to $\epsilon = 2^{-\ell}$ and summing over ℓ, we conclude that $\mathcal{J}(\epsilon)$ remains bounded iff $1 + N\,\mathrm{Re}\,\gamma < 0$, which is one of the three integrability conditions announced.

Proof of Corollary 4.6

Consider $a = b = 1 + cL^2/8$ and perform the change of variables $x_i = 1/2 + \lambda_i/L$, in the limit $L \to \infty$. We have:

$$f_L(\lambda) = \big[(1/2 + \lambda/L)(1/2 - \lambda/L)\big]^{cL^2/8} \underset{L\to\infty}{\sim} 4^{-cL^2/8}\, e^{-c\lambda^2/2},$$

and therefore:

$$\begin{aligned}
& \mathcal{I}_N(1 + cL^2/8, 1 + cL^2/8, \gamma) \\
= \;& L^{-N(1+\gamma(N-1))} \int_{[-L/2,L/2]^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i\, f_L(\lambda_i)\, \big|\Delta(\lambda_1, \ldots, \lambda_N)\big|^{2\gamma} \\
\sim \;& L^{-N(1+\gamma(N-1))}\, 4^{-NcL^2/8} \int_{\mathbb{R}^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i\, e^{-c\lambda_i^2/2}\, \big|\Delta(\lambda_1, \ldots, \lambda_N)\big|^{2\gamma}.
\end{aligned}$$

The constant prefactor in the right-hand side is the Gaussian Selberg integral we want to calculate. We need to compute the asymptotics of the Selberg integral in the first line using its expression in terms of Gamma functions, and the Stirling formula. This leads after simplification to the announced result.

4.4 *Consequence: volumes of symmetry groups*

Lecture 10 (1h30)
November 10th, 2014

On the other hand, for $\beta \in \{1, 2, 4\}$, this integral is – up to the volume factor $v_{N,\beta}/N!$ – the partition function of the Gaussian invariant ensembles, which can be computed independently since the elements of the matrices are decoupled. As a by-product, we obtain an expression for the volume of the symmetry groups:

4.10 THEOREM.

$$\mathrm{Vol}[\mathscr{U}_N(\mathbb{R})] = \frac{\pi^{N(N-3)/4}\, N!}{\prod_{m=1}^{N} \Gamma(1 + m/2)}, \tag{21}$$

$$\mathrm{Vol}[\mathscr{U}_N(\mathbb{C})] = \frac{\pi^{N(N+1)/2}\, N!}{\prod_{m=1}^{N} m!}, \tag{22}$$

$$\mathrm{Vol}[\mathscr{U}_N(\mathbb{H})] = \frac{\pi^{N(N+3/2)}\, N!}{\prod_{m=1}^{N} (2m)!}. \tag{23}$$

Proof. According to diagonalization (Theorem 4.3), the Gaussian Selberg integrals for $\beta \in \{1, 2, 4\}$ coincide with:

$$\begin{aligned}
Z_{N,\beta}^{\mathrm{Gauss}} &:= \int_{\mathscr{H}_{N,\beta}} \mathrm{d}M\, e^{-\mathrm{Tr}\, M^2/2} \\
&= \frac{\mathrm{Vol}[\mathscr{U}_N(\mathbb{K}_\beta)]}{N!\, \mathrm{Vol}^N[\mathscr{U}_1(\mathbb{K}_\beta)]} \int_{\mathbb{R}^N} |\Delta(\lambda_1, \ldots, \lambda_N)|^\beta \prod_{i=1}^{N} e^{-c_\beta \lambda_i^2/2} \mathrm{d}\lambda_i,
\end{aligned}$$

where $c_1 = c_2 = 1$, and $c_4 = 2$ since quaternionic self-dual matrices considered as $2N \times 2N$ matrices have eigenvalues $\lambda_1, \ldots, \lambda_N$ of even multiplicity. $\mathscr{U}_1(\mathbb{K}_\beta)$ is just the unit spheres of $\mathbb{K}_\beta$: for $\beta = 1$, this is $\{\pm 1\}$ and has volume 2 ; for $\beta = 2$, this is the unit circle in $\mathbb{C}$, which has volume 2π. For $\beta = 4$, this is $\mathbb{S}_3$ and has volume $2\pi^2$. Eventually, the right-hand side is a product of one-dimensional integrals, namely:

$$\begin{aligned} Z_{N,\beta}^{\text{Gauss}} &= \prod_{i=1}^{N} \Big(\int_{\mathbb{R}} \mathrm{d}M_{i,i}\, e^{-M_{i,i}^2/2} \Big) \prod_{1 \leq i < j \leq N} \prod_{\alpha=1}^{\beta} \Big(\int_{\mathbb{R}} \mathrm{d}M_{i,j}^{(\alpha)} e^{-(M_{i,j}^{(\alpha)})^2} \Big) \\ &= \pi^{N/2}\, (2\pi)^{\beta N(N-1)/2}. \end{aligned} \tag{24}$$

The last factor comes from the fact that off-diagonals terms of a matrix in $\mathscr{H}_{N,\beta}$ are elements of $\mathbb{K}_\beta$, hence consist of β independent real coefficients. Comparison with Selberg formula and some easy algebra put the result in the announced form. □

We will see in Chapter 5 that there is actually a simpler – but more miraculous – way to compute these integrals for $\beta \in \{1, 2, 4\}$ than Selberg's. It gives another proof of Theorem 4.10.

4.5 *Appendix: Phragmén-Lindelöf principle and applications*

The Phragmén-Lindelöf principe is a classical but very useful result in complex analysis, showing that in the world of analytic functions, strong upper bounds on the boundary of a domain, and loose upper bounds inside, can be improved to strong upper bounds everywhere. If U is an open set of $\mathbb{C}$, we denote $\mathscr{C}(U)$ the set of holomorphic functions on U that are continuous on $\overline{U}$. For $a < b$, we denote $S_{a,b} \subseteq \mathbb{C}$ be the sector between the angles a and b.

4.11 theorem (Phragmén-Lindelöf principle). *Let $f \in \mathscr{C}(S_{a,b})$. Assume there exist $A, B > 0$ satisfying:*

- *for any $z \in \partial S_{a,b}$, we have $|f(z)| \leq 1$;*
- *there exist $A, B > 0$ and $d \in]0, \pi/(b-a)[$ such that, for any $z \in S_{a,b}$, we have $|f(z)| \leq A e^{B|z|^d}$;*

Then $|f(z)| \leq 1$ for any $z \in \overline{S_{a,b}}$.

Proof. After a rotation, we can always assume $b = -a = \theta/2$. Let us fix $\epsilon > 0$ and $d < c < \pi/\theta$, and consider $z = re^{\mathrm{i}\varphi}$ with $\varphi \in [a, b]$. For $z \in \partial S_{a,b}$, we have the strong bound $|\tilde{f}(z)| \leq e^{-\epsilon r^c \cos(c\theta/2)} \leq 1$, for the condition $c < \pi/\theta$ implies $\cos(c\theta/2) > 0$. For $z \in S_{a,b}$, we rather have $|\tilde{f}(z)| \leq A e^{-\epsilon r^c \cos(c\varphi) + Br^d}$, where $\cos(c\varphi) > 0$ and $d < c$ according to our choice of c. Therefore, $\tilde{f}(z)$ decays when $z \to \infty$ in $S_{a,b}$. According to the maximum modulus principle, $|\tilde{f}(z)|$ must reached its maximum for $z \in \partial S_{a,b}$, which means that $|\tilde{f}(z)| \leq 1$ for any $z \in \overline{S}_{a,b}$. Taking the limit $\epsilon \to 0$, we deduce $|f(z)| \leq 1$ for any $z \in \overline{S_{a,b}}$. □

Let $\mathbb{C}_+ = S_{-\pi/2,\pi/2}$ be the right-half plane. If two functions $f, g \in \mathscr{C}(\mathbb{C}_+)$ coincide on all integers, it is not necessarily true that $f = g$, as we could have e.g. $f(z) = g(z) + \sin(\pi z)$. But we add the assumption that $f - g$ grows slightly less than $\sin(\pi z)$, it becomes true. For instance, this type of result allows – sometimes – to justify or explain the failure of the replica method in statistical physics. Here, we applied it to show that the expression found the Selberg at all integers is valid in all $\mathbb{C}_+$.

4.12 THEOREM (Carlson's theorem). *Let $f \in \mathscr{C}(\mathscr{S}_{-\pi/2,\pi/2})$. Assume that f vanishes on $\mathbb{N}$, and the existence of $c < \pi$ such that, for any $z \in S_{a,b}$, we have $|f(z)| \leq Ae^{c|z|}$. Then $f = 0$.*

Proof. Since f has at least a simple zero at all integers,

$$f_1(z) := \frac{f(z)}{\sin(\pi z)},$$

still defines an element of $\mathscr{C}(\mathbb{C}_+)$. Let us examine upper bounds for $|f_1|$. Though f_1 has a finite value at all integers, we need to show that those values do not grow too quickly. For this purpose, we observe that the function $s(z) = 1/\sin(\pi z)$ decays like $O(e^{-\pi|\mathrm{Im}\, z|})$ when $z \in \overline{\mathbb{C}_+}$ and $\mathrm{Im}\, z \to \infty$. If $k \in \mathbb{N}$, we denote Γ_k the circle of radius $k + 1/2$, and $U \subseteq \mathbb{C}_+$ the subset of points at distance less than $1/4$ from $\bigcup_{k\geq 0} \Gamma_k$. Since s remains bounded on U, we have $|f_1(z)| \in O(e^{c|z|})$ uniformly for $z \in U$. But since f_1 decays exponentially at infinity on Γ, we can use Cauchy residue formula for any $z \in \mathbb{C}_+$ such that $k - 1/2 < |z| < k + 1/2$:

$$f_1(z) = \int_{\tilde{\Gamma}_k} \frac{\mathrm{d}\zeta}{2\mathrm{i}\pi} \frac{f_1(\zeta)}{\zeta - z},$$

where the contour is:

$$\tilde{\Gamma}_k = \Gamma_k \cup \big[\mathrm{i}(k+1/2), \mathrm{i}(k-1/2)\big] \cup (-\Gamma_{k-1}) \cup \big[-\mathrm{i}(k-1/2), -\mathrm{i}(k+1/2)\big].$$

This shows that $|f_1(z)| \in O(e^{c(k+1/2)})$ for $k - 1/2 < |z| < k + 1/2$ uniformly in k, and therefore $|f_1(z)| \in O(e^{c|z|})$ in the whole $\mathbb{C}_+$. To summarize, $|f_1(z)|$ decays as $O(e^{-\pi|\mathrm{Im}\, z|})$ along the imaginary axis, while it is bounded by $O(e^{c|z|})$ of the right-half plane. We will show that f_1 must be zero.

To come closer to the framework of Phragmén-Lindelöf principle (=PL), we define:

$$f_2(z) := e^{-[c - \mathrm{i}(\pi - c)]z} f_1(z).$$

This new function is tailored to be uniformly bounded on the imaginary axis and on the positive real axis, while it is $O(e^{c'|z|})$ for some constant $c' > 0$. Since the power of $|z|$ in the exponential is $1 < \pi/(2/\pi)$, we can apply PL to f_2 in each quarter plane $S_{0,\pi/2}$ and $S_{-\pi/2,0}$ and find that:

$$\forall z \in \overline{\mathbb{C}_+}, \qquad |f_2(z)| \leq A$$

for some uniform constant A. Remarkably, this reasoning is valid whatever the value of c' is. We exploit this fact by setting, for $M > 0$:

$$f_3(z) := e^{Mz} f_2(z). \tag{25}$$

We have, for $z = re^{\mathrm{i}\varphi}$ in the right half-plane:

$$|f_3(z)| \leq A' e^{\left[(M+c)\cos\varphi - (\pi - c)|\sin\varphi|\right] r} = e^{-B_M \sin(|\varphi| - \varphi_M) r}$$

for an irrelevant $B_M > 0$, and with $\varphi_M = \arctan[(M+c)/(\pi - c)]$. It shows that:

$$\forall z \in \overline{S_{\varphi_M,\pi/2}} \cup \overline{S_{-\pi/2,-\varphi_M}}, \qquad |f_3(z)| \leq A,$$

and we stress that A is independent of M. In particular,$|f_3(z)|$ is bounded by A in the directions $\varphi = \pm\varphi_M$. Besides, we still have $|f_3(z)| \leq Ae^{B|z|}$ for $z \in S_{-\varphi_M,\varphi_M}$. For M large enough, the width of this angular sector is strictly larger than $\pi/2$, so we can apply PL to conclude that $|f_3(z)| \leq A$ in this sector, and thus:

$$\forall z \in \overline{\mathbb{C}_+}, \qquad |f_3(z)| \leq A.$$

Reminding (25), we obtain that $f_2 \equiv 0$ taking the limit $M \to \infty$. □

4.6 *Appendix: quaternions and quaternionic matrices*

The algebra of quaternions

The algebra of quaternion is denoted $\mathbb{H}$. It is defined as an algebra over $\mathbb{R}$ by the following properties:

- As a real vector space, it is spanned by 4 independent vectors, denoted $\mathbf{1}$ (the unit of the algebra), $\mathbf{I}$, $\mathbf{J}$ and $\mathbf{K}$.
- The generators satisfy the relations $\mathbf{IJ} = -\mathbf{JI} = \mathbf{K}$ and the ones obtained by cyclic permutations of $(\mathbf{I}, \mathbf{J}, \mathbf{K})$.

One readily checks this algebra is associative. Any $q \in \mathbb{H}$ can be decomposed:

$$q = q^{(1)} \cdot \mathbf{1} + q^{(I)} \cdot \mathbf{I} + q^{(J)} \cdot \mathbf{J} + q^{(K)} \cdot \mathbf{K}, \qquad q^{(1)}, q^{(I)}, q^{(J)}, q^{(K)} \in \mathbb{R}.$$

It is equipped with an $\mathbb{R}$-linear involution $q \mapsto \overline{q}$:

$$\overline{q} = q^{(1)} \cdot \mathbf{1} - q^{(I)} \cdot \mathbf{I} - q^{(J)} \cdot \mathbf{J} - q^{(K)} \cdot \mathbf{K},$$

$\overline{q}$ is called the **dual** quaternion of q, and one can check the property $\overline{q_1 q_2} = \overline{q_2}\,\overline{q_1}$. We say that q is **scalar** (resp. **purely quaternionic**) if $q = \overline{q}$ (resp. $q = -\overline{q}$). For instance, $\mathbf{I}, \mathbf{J}, \mathbf{K}$ are purely quaternionic. The euclidean norm on $\mathbb{H}$ considered as a vector space has the remarkable property to be multiplicative:

$$|q_1 q_2| = |q_1| \cdot |q_2|.$$

So, the unit sphere of quaternions for this norm – denoted $\mathscr{U}_1(\mathbb{H})$ – forms a group.

We can also consider the algebra of complex quaternions[17] $\mathbb{H} \otimes \mathbb{C}$, i.e. elements of the form (4.6) with $q^{\bullet} \in \mathbb{C}$. The definition of the dual is extended by linearity, the euclidian norm is replaced by the hermitian norm, and all the properties mentioned above continue to hold.

Representation by 2×2 matrices

It is possible to realize $\mathbb{H}$ as a subalgebra of $\mathscr{M}_2(\mathbb{C})$. We consider the endomorphism of vector spaces $\Theta_1 : \mathbb{H} \to \mathscr{M}_2(\mathbb{C})$ defined by $\Theta_1(\mathbf{1}) = 1$ and:

$$\Theta_1(\mathbf{I}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \Theta_1(\mathbf{J}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = J_1, \qquad \Theta_1(\mathbf{K}) = \begin{pmatrix} 0 & \mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}.$$

In other words:

$$\Theta_1(q) = \begin{pmatrix} q^{(1)} + \mathrm{i}q^{(I)} & q^{(J)} + \mathrm{i}q^{(K)} \\ -q^{(J)} + \mathrm{i}q^{(K)} & q^{(1)} - \mathrm{i}q^{(I)} \end{pmatrix}.$$

One can check that the relations between generators are preserved, so Θ_1 is an morphism of algebras. It is clearly injective, but is not surjective, since we have:

$$\operatorname{Im}\Theta_1 = \Big\{\left(\begin{smallmatrix} w & z \\ -z^* & w^* \end{smallmatrix}\right), \quad (z,w) \in \mathbb{C}^2\Big\} = \big\{A \in \mathscr{M}_2(\mathbb{C}), \quad AJ_1 = J_1A^*\big\}.$$

We therefore have an isomorphism of algebras $\mathbb{H} \simeq \operatorname{Im}\Theta_1$. All the definitions in $\mathbb{H}$ can be rephrased in terms of 2×2 matrices. The dual of $A = \Theta_1(q)$ is by definition $\overline{A} := \Theta_1(\overline{q})$. It can be rewritten in terms of matrix operations:

(26) $\quad \overline{A} = J_1 A^T J_1^{-1}, \quad$ or equivalently $\overline{A} = A^{\dagger}$.

The scalar part q is extracted as $q^{(1)} = \frac{1}{2}\mathrm{tr}\,\Theta_1(q)$, while the norm is:

$$|q| = \det\Theta_1(q)$$

which explains its multiplicativity. We remark that the columns of a matrix in $\operatorname{Im}\Theta_1$ are orthogonal for the hermitian product on $\mathbb{C}^2$. Therefore, $\Theta_1(\mathscr{U}_1(\mathbb{H})) = \mathrm{SU}(2)$.

If we consider complex quaternions, we rather have $\mathbb{H} \otimes \mathbb{C} \simeq \mathscr{M}_2(\mathbb{C})$. All the relations above continue to hold – because they are $\mathbb{C}$-linear –, except that the dual of a complex quaternion is not anymore represented by the adjoint of the corresponding 2×2 matrix – because complex conjugation is $\mathbb{R}$-linear but not $\mathbb{C}$ linear.

Quaternionic matrices

If we tensor (over $\mathbb{R}$) the previous construction by $\mathscr{M}_N(\mathbb{R})$, we obtain a morphism of algebras $\Theta_N : \mathbb{H} \otimes \mathscr{M}_N(\mathbb{R}) \to \mathscr{M}_N(\mathbb{C}) \otimes \mathscr{M}_N(\mathbb{R})$, and with natural

[17]Beware, our terminology differs from Mehta's book, where elements of $\mathbb{H}$ are called "real quaternions", and elements of $\mathbb{H} \otimes \mathbb{C}$ are called "quaternions".

identifications of the source and the target:

$$\Theta_N \,:\, \mathscr{M}_N(\mathbb{H}) \longrightarrow \mathscr{M}_{2N}(\mathbb{C}).$$

$\mathscr{M}_N(\mathbb{H})$ is the algebra of $N \times N$ **quaternionic matrices**. If Q is a quaternionic matrix, we can also build the four $N \times N$ matrices of real coefficients in the decomposition (4.6), denoted $Q^{(1)}$, $Q^{(I)}$, $Q^{(J)}$ and $Q^{(K)}$. $\mathbb{H}$ itself is embedded in $\mathscr{M}_N(\mathbb{H})$ by tensoring generators with the identity matrix of $\mathscr{M}_N(\mathbb{R})$, i.e. putting the same quaternion on all diagonal entries of an $N \times N$ matrix. For instance:

$$\Theta_N(\mathbf{J} \otimes 1) = J_N = \begin{pmatrix} \begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix} & & \\ & \ddots & \\ & & \begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix} \end{pmatrix}.$$

Since all linear and multiplicative operations are compatible with extension by the algebra $\mathscr{M}_N(\mathbb{R})$, we immediately find:

$$\mathscr{M}_N(\mathbb{H}) \simeq \mathrm{Im}\,\Theta_N = \big\{A \in \mathscr{M}_{2N}(\mathbb{C}), \quad AJ_N = J_N A^*\big\}.$$

If $Q \in \mathscr{M}_N(\mathbb{H})$, we define its dual $\overline{Q}$ by:

$$\forall i,j \in [\![1,N]\!], \qquad (\overline{Q})_{i,j} = \overline{Q_{i,j}}.$$

We see that this is the analog of the adjoint in the world of quaternions – the involution "complex conjugation" is replaced by "quaternion dual". The effect on the corresponding $2N \times 2N$ matrices is immediately deduced from (26) by extension. If we define the dual of $A = \Theta_N(Q)$ to be $\Theta_N(\overline{Q})$, we then have:

(27) $\overline{A} = J_N A^T J_N^{-1}.$

4.13 definition. The set of **quaternionic self-dual matrices** is:

$$\mathscr{H}_N(\mathbb{H}) = \mathscr{H}_{N,4} := \big\{Q \in \mathscr{M}_N(\mathbb{H}), \quad \overline{Q} = Q\big\}.$$

This condition is equivalent to require that $Q^{(1)}$ is real symmetric, and the three matrices $Q^{(I)}, Q^{(J)}, Q^{(K)}$ are real antisymmetric. The set of **quaternionic unitary matrices** is:

$$\mathscr{U}_N(\mathbb{H}) = \mathscr{G}_{N,4} := \big\{U \in \mathscr{M}_N(\mathbb{H}), \quad U\overline{U} = 1\big\}.$$

Quaternionic self-dual matrices form a vector space, while quaternionic unitary matrices form a group. Using (27), they can be easily be characterized in terms of $2N \times 2N$ matrices:

- Q is quaternionic self-dual iff $J_N\Theta_N(Q)$ is complex antisymmetric.
- U is quaternionic unitary iff $\Theta_N(Q)$ is a (complex) unitary and symplectic matrix.

The main interest about quaternionic self-dual matrices comes from their diagonalization property:

4.14 lemma. *If $Q \in \mathscr{H}_N(\mathbb{H})$, there exist $U \in \mathscr{U}_N(\mathbb{H})$ and scalar quaternions $q_1, \ldots, q_N$ such that $Q = U\mathrm{diag}(q_1, \ldots, q_N)U^{-1}$. In particular, the eigenvalues of $\Theta_N(Q)$ are the $q_i^{(1)}$ with multiplicity 2 for $i \in [\![1, N]\!]$.*

Proof. The proof is very similar to the diagonalization of self-adjoint matrices. If Q is a self-dual matrix, we will actually diagonalize $A = \Theta_N(Q)$. Denote s the canonical symplectic bilinear form, and b the canonical hermitian bilinear form on $\mathbb{C}^{2N}$:

$$s(X, Y) = X^T J_N Y, \qquad b(X, Y) = X^\dagger Y.$$

They are related by $s(X, Y) = -b(J_N X^*, Y)$. Self-duality means $A^T J_N = J_N A$, and this implies that A preserves s:

$$s(AX, Y) = X^T A^T J_N Y = X^T J_N AY = s(X, AY).$$

But since $A \in \mathrm{Im}\, \Theta_N$, we know $A^* J_N = J_N A$ and deduce that A also preserves b:

$$b(X, AY) = -s(J_N X^*, AY) = -s(AJ_N X^*, Y) = -s(J_N A^* X^*, Y) = b(AX, Y).$$

in particular A is hermitian, so has real eigenvalues. We can already diagonalize it with a unitary matrix, but it is useful to remind the source of this result. If X and Y are two eigenvectors with respective eigenvalues λ and μ, we can compute in two different ways $s(X, AY) = \mu s(X, Y)$ and $s(X, AY) = s(AX, Y) = \lambda s(X, Y)$. Similarly, $b(X, AY) = \mu b(X, Y)$ but also $b(X, AY) = b(AX, Y) = \lambda Y$. Therefore, if X and Y are in different eigenspaces, we must have $s(X, Y) = b(X, Y) = 0$. Let us pick up an eigenvector X_1 for an eigenvalue λ_1 of A. Since s is non-degenerate, there exists a non-zero vector X_1' such that $s(X_1, X_1') \neq 0$. Denoting X_1'' its projection to the eigenspace $E_\lambda(A)$, we have $s(X_1, X_1'') = s(X_1, X_1') \neq 0$, so we obtained another non-zero vector in $E_\lambda(A)$ that is not collinear to X. Upon a change of basis $(X_1, X_1') \to (aX_1 + bX_1'', cX_1 + dX_1'')$, we can always enforce that X_1 and X_1'' are orthonormal, and $s(X_1, X_1'') = 1$. If we complete (X_1, X_1'') by vectors $(Y_1, \ldots, Y_{2N-2})$ to form a basis of $\mathbb{C}^{2N}$ respecting the decomposition in eigenspaces, we see that A leaves stables $\mathrm{vect}(Y_1, \ldots, Y_{2N-2}) \simeq \mathbb{C}^{2N-2}$. By induction, we construct in this way a unitary and symplectic basis $(X_1, X_1'', \ldots, X_N, X_N'')$ of eigenvectors of A, i.e. such that:

$$\forall i, j \in [\![1, N]\!], \qquad \begin{array}{ll} s(X_i, X_j'') = \delta_{i,j} & s(X_i, X_j) = s(X_i'', X_j'') = 0 \\ b(X_i, X_j) = b(X_i'', X_j'') = \delta_{i,j} & b(X_i, X_j'') = 0. \end{array}$$

The proof also shows that the eigenvalues come by pairs. In other words, we have a matrix $U \in \mathscr{M}_{2N}(\mathbb{C})$ of eigenvectors satisfying $U^\dagger U = 1$ and $U^T J_N U = J_N$, and such that :

$$A = U\mathrm{diag}(\lambda_1, \lambda_1, \ldots, \lambda_N, \lambda_N)U^{-1}.$$

□

5 Toolbox for exact computations in invariant ensembles

In this Chapter, μ is a positive measure on $\mathbb{R}$ such that all integrals considered are absolutely convergent, and A is a measurable subset of $\mathbb{R}$. Unless specified otherwise, all expectation values are considered with respect to the probability measure on A^n:

$$\frac{1}{Z_{n,\beta}[\mu]} \prod_{i=1}^{n} \mathrm{d}\mu(x_i)\, \big|\Delta(x_1,\ldots,x_n)\big|^{\beta}, \qquad \beta = 1,2,4.$$

and the partition function is given by:

$$Z_{n,\beta}[\mu] = \int_{A^n} \prod_{i=1}^{n} \mathrm{d}\mu(x_i)\, \big|\Delta(x_1,\ldots,x_n)\big|^{\beta}.$$

This is the probability measure on the eigenvalues on a random matrix drawn from the invariant ensembles. In this case, assuming that all the moments of μ are finite guarantees that all integrals are absolutely convergent. As we shall see, $\beta = 2$ is always the simplest case and feature some determinantal structures ; $\beta = 4$ is the next simplest case and features some pfaffian structure ; $\beta = 1$ also has a pfaffian structure, but it is often more cumbersome because the answer depends on the parity of n.

5.1 *Multilinear algebra*

Vandermonde determinant

The partition function of the invariant ensembles is a n-dimensional integral involves the power $\beta \in \{1,2,4\}$ of the factor:

$$\Delta(x_1,\ldots,x_n) = \prod_{1\leq i<j\leq n} (x_j - x_i).$$

This is the expression for the **Vandermonde determinant**:

5.1 lemma.

$$\Delta(x_1,\ldots,x_n) = \det_{1\leq i,j\leq n} x_i^{j-1}$$

.

Proof. The right-hand side is an antisymmetric polynomial in the variables $x_1,\ldots,x_n$, of degree $(n-1)$ in x_i. And, for any pair $\{i,j\}$ it has a simple zeroes when $x_i = x_j$ since two columns of the determinant are equal. So, we can factor out:

$$\det_{1\leq i,j\leq n} x_i^{j-1} = \prod_{1\leq i<j\leq n} (x_j - x_i)\cdot P_n(x_1,\ldots,x_n),$$

but a degree inspection shows that $P_n(x_1, \ldots, x_n)$ does not depend on the x's. Comparing the coefficient of the leading monomial $x_1^{n-1} x_2^{n-2} \cdots x_n^0$, we find $P \equiv 1$. □

We say that a family of non-zero polynomials $(Q_j)_{j \geq 0}$ is **monic** if the coefficient of the leading term is 1, and **staggered** if $\deg Q_j = j$. A staggered family provides, for any $n \geq 0$, a basis $(Q_j)_{0 \leq j \leq n}$ for the vector space of polynomials of degree less or equal than n. By operations on columns of the Vandermonde determinant, we can also write:

$$\Delta(x_1, \ldots, x_n) = \det_{1 \leq i,j \leq n} Q_{j-1}(x_i).$$

Pfaffians

If A is an antisymmetric matrix of size n, let $m := \lfloor n/2 \rfloor$. The **Pfaffian** of A is defined as:

$$\mathrm{pf}(A) = \frac{1}{2^m m!} \sum_{\sigma \in \mathfrak{S}_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{m} A_{\sigma(2i-1),\sigma(2i)}.$$

Notice that if n is odd, the index $\sigma(n)$ is not involved in the sum. We can always convert a pfaffian of size $2m+1$ into a pfaffian of even size $2m+2$:

$$\mathrm{pf}(A) = \mathrm{pf}\left(\begin{array}{c|c} A & \begin{matrix} 1 \\ \vdots \\ 1 \end{matrix} \\ \hline -1 \ \ldots \ -1 & 0 \end{array}\right).$$

The basic properties of the pfaffian:

- $(\mathrm{pf}\, A)^2 = \det A$.
- $\mathrm{pf}(PAP^T) = \det(P)\, \mathrm{pf}(A)$.
- $\mathrm{pf}(A^T) = (-1)^m\, \mathrm{pf}(A)$
- Column expansion assuming n even. Let us denote $A^{[i,j]}$ the matrix of size $(n-2)$ obtained by removing the i-th and j-th lines and columns. Then $\mathrm{pf}(A) = \sum_{i=1}^{n} (-1)^{i+1} A_{i,2m}\, \mathrm{pf}\, A^{[i,2m]}$.

It is also useful to remark:

5.2 lemma. *Assume n is even. For any $\alpha \in \mathscr{M}_n(\mathbb{C})$, and any complex antisymmetric $n \times n$ matrices A and B:*

$$\mathrm{pf}_{2n} \begin{pmatrix} A & A\alpha^T \\ \alpha A & \alpha A \alpha^T + B \end{pmatrix} = \mathrm{pf}\, A \cdot \mathrm{pf}\, B.$$

Proof. By column operations, we have:

$$\det_{2n} \begin{pmatrix} A & A\alpha^T \\ \alpha A & \alpha A \alpha^T + B \end{pmatrix} = \det_{2n} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \det A \cdot \det B.$$

Taking the squareroot of this identity:

$$\operatorname*{pf}_{2n}\begin{pmatrix} A & A\alpha^T \\ \alpha A & \alpha A\alpha^T + B \end{pmatrix} = \varepsilon(\alpha)\,\mathrm{pf}\,A\cdot\mathrm{pf}\,B.$$

for some ε taking values in $\{\pm 1\}$. The sign must be independent of α for the left-hand side is a continuous function of α. In particular, for $\alpha = 0$, the left-hand side is computed as a block pfaffian and gives $\varepsilon = 1$. □

Quaternionic determinant

There are several, inequivalent, notions of determinants for matrices with entries in a non-commutative ring. The quaternionic determinant we will define is one of them, and turns out to be a convenient way to deal with pfaffians. If A is quaternionic matrix and $\gamma = (i(1) \to \cdots i(r) \to i(1))$ is a cyclic permutation of r elements in $[\![1, n]\!]$, we define:

(28) $$\Pi_\gamma(A) = \big[A_{i(1)i(2)}\cdots A_{i(r)i(1)}\big]^{(1)} \in \mathbb{C}.$$

We remind that $q^{(1)} = \mathrm{Tr}\,\Theta_1(q)/2$ is the scalar part of the quaternion q. Since the trace is cyclic, $\Pi_\gamma(A)$ does not depend on the origin $i(1)$ chosen for the cycle γ. If A is a quaternionic self-dual matrix, the product $A_{i(1)i(2)}\cdots A_{i(r)i(1)}$ is actually a scalar, so we can omit $^{(1)}$.

5.3 definition. If $A \in \mathscr{M}_n(\mathbb{H})$, its quaternionic determinant is defined by:

$$\det_{\mathbb{H}} A = \sum_{\sigma\in\mathfrak{S}_n}\varepsilon(\sigma)\prod_{\gamma}P_\gamma(A).$$

The product is indexed by the disjoint cycles γ appearing in the factorization of σ.

If we consider $A \in \mathscr{M}_n(\mathbb{C})$ as a quaternionic matrix, we have $\det_{\mathbb{H}} A = \det A$. Obviously, if $A \in \mathscr{M}_n(\mathbb{H})$, we have $\det_{\mathbb{H}} A = \det_{\mathbb{H}}\overline{A}$. However, a main difference with the commutative case is that in general:

$$\det_{\mathbb{H}}(A\cdot B) \neq \det_{\mathbb{H}} A\cdot\det_{\mathbb{H}} B.$$

There exists a Cramer-type formula to compute the inverse of a self-dual quaternionic matrix:

5.4 lemma. *If $A \in \mathscr{H}_n(\mathbb{H})$, define the matrix:*

$$B_{ij} = \sum_{\substack{\sigma\in\mathfrak{S}_n \\ \sigma(j)=i}}\varepsilon(\sigma)\Big[\prod_{\substack{\gamma \\ i\notin\gamma}}P_\gamma(A)\Big]\cdot A_{ia(1)}A_{a(1)a(2)}\cdots A_{a(r)j}\,.$$

where $(i \to a(1) \to a(r) \to j \to i)$ is the cycle of σ containing i. B is self-dual, and:

$$A\cdot B = B\cdot A = (\det_{\mathbb{H}} A)\,\mathbf{1}_n\,.$$

Proof. Since $\overline{A_{mn}} = A_{nm}$, taking the quaternionic dual of the right-hand side defining B_{ij} amounts to reversing the cycle containing i. So, $\overline{B_{ij}} = B_{ji}$, i.e. B is self-dual. We will prove $(A \cdot B)_{ij} = (\det_{\mathbb{H}} A) \cdot \delta_{i,j}$ for any $i, j \in [\![1, n]\!]$. The proof of $(B \cdot A) = (\det_{\mathbb{H}} A) \cdot \delta_{i,j}$ is similar and is left as exercise. For $i = j$, we have:

$$(A \cdot B)_{ii} = \sum_{k=1}^{n} \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(k)=i}} \Big[\prod_{\substack{\gamma \\ i \notin \gamma}} P_\gamma(A)\Big] \cdot P_{\gamma_i(A)}\,.$$

where γ_i is the cycle containing i. Since k is arbitrary here, we are actually summing over all permutations $\sigma \in \mathfrak{S}_n$, and γ_i plays the same role as the other cycles: we recognize $\det_{\mathbb{H}} A$. If $i \neq j$, we have:

$$(A \cdot B)_{ij} = \sum_{k=1}^{n} \sum_{\substack{\sigma \in \mathfrak{S}_n \\ \sigma(j)=k}} \Big[\prod_{\substack{\gamma \\ j \notin \gamma}} P_\gamma(A)\Big] \cdot A_{ik}A_{ka(1)} \cdots A_{a(r)j}\,.$$

where $\gamma_0 = (k \to a(1) \to \cdots \to a(r) \to j \to k)$. When in the permutation σ, $i = a(s)$ belongs to the cycle containing j, the last product is:

$$A_{ik}A_{ka(1)} \cdots A_{a(r)j} = \big[A_{ik}A_{ka(1)} \cdots A_{a(s-1)i}\big]\, A_{ia(s+1)} \cdots A_{a(r)j}$$

But this term also appear when we consider the permutation $\tilde{\sigma}$, obtained from σ by breaking the cycle:

$$\big(k \to a(1) \cdots a(s-1) \to i \to a(s+1) \to \cdots a(r) \to j \to k\big)$$

into the two disjoint cycles:

$$\big(j \to a(s+1) \to \cdots \to a(r) \to j\big) \text{ and } \big(i \to k \to a(1) \to \cdots \to a(s-1) \to i\big).$$

Since $\varepsilon(\sigma) = -\varepsilon(\tilde{\sigma})$, these two contributions must cancel, hence the claim. $\square$

This result is particularly useful in the proof of:

5.5 proposition. *If $A \in \mathscr{M}_n(\mathbb{H})$:*

$$\det_{\mathbb{H}} A\overline{A} = \det \Theta_n(A).$$

If furthermore A is self-dual:

$$\det_{\mathbb{H}} A = \mathrm{pf}\, J_n \Theta_n(A), \qquad (\det_{\mathbb{H}} A)^2 = \det \Theta_n(A).$$

Proof. When A is self-dual, $J_n\Theta_n(A)$ is antisymmetric, so the right-hand side of the second equality makes sense. We claim it is enough to prove the third equality. Indeed, if we have it for any A, the identity:

$$\big(\mathrm{pf}\, J_n\Theta_n(A)\big)^2 = \det J_n\Theta_n(A) = \det J_n \cdot \det \Theta_n(A) = \det \Theta_n(A)$$

implies that there exists $\varepsilon : \mathscr{H}_n(\mathbb{H}) \to \{\pm 1\}$ such that:

$$\det_{\mathbb{H}} A = \varepsilon(A) \cdot \mathrm{pf}\ J_n \Theta_n(A).$$

Since the determinant and the pfaffian are continuous functions of the matrix entries, $\varepsilon(A)$ does not depend on A. And for the identity matrix, it is equal to 1, so we get the second equality. Besides, the first equality follows from the second equality, since for any quaternionic matrix A, $A\overline{A}$ is self-dual, and we can compute:

$$\begin{aligned}
\det_{\mathbb{H}}(A\overline{A}) &= \mathrm{pf}\, J_n \Theta_n(A\overline{A}) = \mathrm{pf}\, J_n \Theta_n(A)\Theta_n(\overline{A}) \\
&= \mathrm{pf}\, J_n \Theta_n(A) J_n \big(\Theta_n(A)\big)^T J_n^{-1} = \mathrm{pf}\, \Theta_n(A) J_n \big(\Theta_n(A)\big)^T \\
&= \det \Theta_n(A) \cdot \mathrm{pf}\, J_n = \det \Theta_n(A).
\end{aligned}$$

Let us proceed with the proof of the third equality. Taking the $2n \times 2n$ matrix representation of the Cramer formula in Lemma 5.4, we find:

$$\Theta_n(A)\Theta_n(B) = \Theta_n(B)\Theta_n(A) = \det_{\mathbb{H}} A \cdot \mathbf{1}_{2n}.$$

Therefore, if $\det_{\mathbb{H}} A \neq 0$, then $\Theta_n(A)$ is invertible, i.e. $\det \Theta_n(A) \neq 0$. The previous statement can be reformulated by saying that $\det_{\mathbb{H}} A$ – seen as a polynomial in the (linearly independent) complex entries of A – vanishes on the zero locus of the polynomial ideal generated by $\det \Theta_n(A)$. According to the Nullstellensatz (see § 0.1), there exist an integer $r \geq 1$ and a polynomial $R(A)$ such that:

$$\big(\det_{\mathbb{H}} A\big)^r = R(A) \det \Theta_n(A).$$

Degree comparison imposes $r \geq 2$. By dividing both sides by a suitable power of $\det_{\mathbb{H}} A$, one can always assume that R has no common factor with $\det_{\mathbb{H}} A$. Since $\Theta_n(A)$ is homogeneous of degree $2n$ and while $\det_{\mathbb{H}} A$ is homogeneous of degree n, we must have $r = 2$. Degree considerations also show that $R(A)$ must be a constant, which is evaluated to 1 for the identity matrix. $\square$

5.2 *Operators with kernels*

We will consider μ a positive measure on A with finite mass.

Setting

The data of a function $K \in L^2(A^2, \mu^{\otimes 2}) \otimes \mathbb{K}_\beta$ defines an endomorphism of $L^2(A, \mu) \otimes \mathbb{K}_\beta$ also denoted K, by the formula:

$$(Kf)(x) = \int_{\mathbb{A}} K(x,y)\, f(y) \mathrm{d}\mu(y).$$

The function $(x,y) \mapsto K(x,y)$ is called the *kernel* of the operator K. The square-integrability of the kernel implies that K is a continuous operator in L^2 norm:

$$||Kf||_{L^2(A,\mu)} \leq ||K||_{L^2(A^2,\mu^{\otimes 2})} \cdot ||f||_{L^2(A,\mu)}.$$

Properties of the kernel or of the operator will be attributed indifferently to both of them. We say that K is self-dual if $\overline{K(x,y)} = K(y,x)$, and that it is **trace-class** if $x \mapsto (K(x,x))^{(1)}$ exists in $L^1(A,\mu)$. In this case, we set:

$$\mathrm{Tr}\, K = \int_A (K(x,x))^{(1)}\, \mathrm{d}\mu(x).$$

If $K \in L^\infty(A^2, \mu^{\otimes 2})$, then K is trace-class.

Fredholm determinant

Let K be a trace-class operator. For any $\ell \geq 1$, K^ℓ is also trace-class, and:

$$\begin{aligned} \frac{\mathrm{Tr}\ K^\ell}{\ell} &= \frac{1}{\ell}\int_{A^\ell} \big[K(x_1,x_2)\cdots K(x_\ell,x_1)\big]^{(1)} \prod_{i=1}^{\ell} \mathrm{d}\mu(x_i) \\ &= \frac{1}{\ell!}\sum_{\gamma=\ell-\mathrm{cycle}} P_\gamma\big[(K(x_i,x_j))_{1\leq i,j\leq \ell}\big] \prod_{i=1}^{\ell}\mathrm{d}\mu(x_i). \end{aligned}$$

In the second line, we have exploited the freedom to relabel the integration variables. If γ is a cycle, we denote $\ell(\gamma)$ its length, and thus $\varepsilon(\gamma) = (-1)^{\ell(\gamma)+1}$. Consider the formal series in t:

$$\begin{aligned} F(K;t) &:= \exp\Big(-\sum_{\ell\geq 1}\frac{t^\ell}{\ell}\ \mathrm{Tr}\ K^\ell\Big) \\ &= \sum_{\substack{k\geq 0 \\ n\geq 0}} \frac{1}{k!} \sum_{\substack{\gamma_1,\ldots,\gamma_k \\ \text{ord. disjoint cycles} \\ \sum_i \ell(\gamma_i)=n}} \prod_{i=1}^{k} \frac{(-t)^{\ell(\gamma_i)}}{\ell(\gamma_i)!} \int_{A^n} P_\gamma\big[(K(x_i,x_j))_{ij}\big] \prod_{i=1}^{n}\mathrm{d}\mu(x_i). \end{aligned}$$

By ordered disjoint cycles, we mean that γ_1 is a cyclic permutation of $[\![1,\ell(\gamma_1)]\!]$, γ_2 is a cyclic permutation of $[\![\ell(\gamma_1)+1, \ell(\gamma_1)+\ell(\gamma_2)]\!]$, etc. We can again decide to relabel the variables of integrations to un-order the cycles, at the price of dividing by the number of ways to do so, namely:

$$\frac{n!}{\ell(\gamma_1)!\cdots\ell(\gamma_n)!}\,.$$

The sum over unordered disjoint cycles of total length n reconstitutes the sum over all permutations $\sigma \in \mathfrak{S}_n$, and the sign reconstitutes $(-1)^n\varepsilon(\sigma)$. Therefore:

$$F(K;t) = \sum_{n\geq 0}\frac{(-t)^n}{n!}\int_{A^n}\big(\operatorname*{det_{\mathbb{H}}}_{1\leq i,j\leq n} K(x_i,x_j)\big)\prod_{i=1}^{n}\mathrm{d}\mu(x_i)$$

By the Hadamard inequality, the coefficient of t^n is bounded in absolute value by:

$$\frac{n^{n/2}}{n!}\big[\mu(A)||K||_\infty\big]^n$$

Hence, the series converges absolutely for any $t \in \mathbb{C}$, and defines an entire function of t. This leads to the definition:

5.6 definition. Let K be a L^∞ kernel. The Fredholm determinant of K is by definition:

$$\mathrm{Det}_{\mathbb{H}}(I-K) := F(K;1) = \sum_{n\geq 1} \frac{(-1)^n}{n!} \int_{A^n} \big(\det_{\mathbb{H}}{}_{1\leq i,j\leq n} K(x_i,x_j)\big) \prod_{i=1}^n \mathrm{d}\mu(x_i).$$

If the base field is $\mathbb{K}_\beta = \mathbb{R}$ or $\mathbb{C}$, the subscript $\mathbb{H}$ is unnecessary on both sides of the formula. This notion of determinant for operators acting on L^2 spaces generalizes the notion of determinant of matrices of finite size. Indeed, assume there exists an integer k such that:

$$K(x,y) = \sum_{m=1}^k a_m\, B_m(x)\, C_m(y),$$

where a's are complex numbers, and B's and C's satisfy the orthogonality relations:

$$\int_A \mathrm{d}\mu(x)\, C_m(x) B_n(x) = \delta_{m,n} \in \mathbb{H}.$$

Then, the operator K has rank k, and sends B_m to $a_m B_m$ for $m \in [\![1,k]\!]$. Let $V \subseteq L^2(A,\mu)$ be the subspace of dimension k generated by the B's, and introduce $K_{|V} : V \to V$ the restriction of K, and $\tilde{K}_{|V}$ its matrix in the basis of the B's. The trace of K – as an endomorphism of $L^2(A,\mu)$ – coincides with the trace of $\tilde{K}_{|V}$ – as a finite size matrix – and is equal to $\sum_{m=1}^k a_m^n$. Using the initial formula for $F(K;1)$:

$$\mathrm{Det}_{\mathbb{H}}(I-K) = \exp\Big(- \sum_{m=1}^k \sum_{n\geq 1} a_m^n \Big) = \prod_{m=1}^k (1-a_m) = \det(\mathrm{id} - \tilde{K}_{|V}).$$

Hadamard inequality allows the proof of:

5.7 lemma. *For a complex valued kernel, $K \mapsto \mathrm{Det}(I-K)$ is continuous for the sup norm.*

Proof. Let A and B be complex matrices of size n, and v_i^A denote the i-th column of the matrix A. We define the matrices:

$$B^{[0]} = A, \qquad B^{[i]} = \big(v_1^B, \ldots, v_i^B, v_{i+1}^A, \ldots, v_n^A\big) \quad i \in [\![1,n]\!],$$

and:

$$C^{[i]} = \big(v_1^B, \ldots, v_i^B, v_{i+1}^A - v_{i+1}^B, v_{i+2}^B, \ldots, v_n^B\big), \quad i \in [\![0,n-1]\!].$$

Since the determinant is linear with respect to the columns, we have:

$$\det A - \det B = \sum_{i=0}^{n-1} \det B^{[i]} - \det B^{[i+1]} = \sum_{i=0}^{n-1} \det C^{[i]}.$$

Therefore, by Hadamard inequality:

$$\big|\det A - \det B\big| \leq n^{n/2+1}||A-B||_\infty \max\big(||A||_\infty, ||B||_\infty\big)^{n-1}.$$

Now, if K and L are two complex valued L^∞ kernels, we can deduce from the definition of the Fredholm determinant:

$$\big|\operatorname{Det}(I-K) - \operatorname{Det}(I-L)\big| \leq ||K-L||_\infty \left(\sum_{n\geq 0} \frac{n^{n/2+1}}{n!}\big(||K||_\infty, ||L||_\infty\big)^{n-1}\mu(A)^n\right).$$

This implies that $K \mapsto \operatorname{Det}(I-K)$ is Lipschitz on bounded balls in $L^\infty(A^2, \mu)$. A fortiori, it is continuous. □

We shall admit that the quaternionic Fredholm determinant is also continuous for the sup norm on self-dual kernels. However, it cannot be proved by this method. Indeed, the proof of Hadamard inequality uses the multiplicativity of the determinant, thus does not extend to quaternionic determinants.

(Quasi) projectors

5.8 definition. A self-dual kernel K is a quasi-projector if there exists a constant quaternion λ such that:

$$K \circ K = K + [\lambda, K].$$

If $\lambda = 0$, K is a projector.

Our interest in (quasi) projectors come from the integration lemma:

5.9 proposition. *Let K be a self-dual, trace-class kernel which is a quasi projector. Let $c = \operatorname{Tr} K$. We have for any $k \geq 1$:*

$$\int_{A^{n-k}} \Big(\operatorname*{det_{\mathbb{H}}}_{1\leq i,j\leq n} K(x_i,x_j)\Big)\prod_{i=k+1}^{n}\mathrm{d}\mu(x_i) = \frac{\Gamma(c-k+1)}{\Gamma(c-n+1)}\cdot \operatorname*{det_{\mathbb{H}}}_{1\leq i,j\leq k} K(x_i,x_j).$$

Proof. We first prove a one-step integration. We have:

$$\begin{aligned} &\int_A \operatorname*{det_{\mathbb{H}}}_{1\leq i,j\leq n} K(x_i,x_j)\mathrm{d}\mu(x_n) \\ &= \sum_{\sigma\in\mathfrak{S}_n}\varepsilon(\sigma)\Big[\prod_{\substack{\gamma\\ n\notin\gamma}} P_\gamma\big[K(x_i,x_j)_{ij}\big]\Big]\cdot P_{\gamma_n}\big[K(x_i,x_j)_{ij}\big]\mathrm{d}\mu(x_n), \end{aligned} \tag{29}$$

where γ_n is the cycle of σ containing n. If n is a fixed point, the last factor is:

$$\int_A K(x,x)\mathrm{d}\mu(x) = c.$$

Deleting the point n from σ, we obtain a contribution:

$$c \sum_{\tilde{\sigma} \in \mathfrak{S}_{n-1}} \varepsilon(\tilde{\sigma}) \prod_{\gamma} P_{\gamma}\big[K(x_i, x_j)_{1 \leq i,j \leq n-1}\big].$$

If n is not a fixed point, γ_n is of the form $(n \to i(1) \to \cdots \to i(r) \to n)$, and we rather have the factor:

$$\bigg[\Big(\int_A K(x_{i(r)}, x_n) K(x_n, x_{i(1)}) \mathrm{d}\mu(x_n)\Big) K(x_{i(1)}, x_{i(2)}) \cdots K(x_{i(r-1)}, x_{i(r)})\bigg]^{(1)}.$$

The integral precisely computes the kernel $(K \circ K)(x_{i(r)}, x_{i(1)})$. Since K is a quasi projector, we can replace it with $K + \lambda K - K\lambda$. Since K is self-dual, the term associated to σ and containing λK actually coincides with the term associated to σ^{-1}, but the latter comes with a minus sign. So, the terms involving λ cancel out, and we are left with a contribution:

$$\sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) \Big[\prod_{\substack{\gamma \\ n \notin \gamma}} P_{\gamma}\big[K(x_i, x_j)_{ij}\big]\Big] \cdot P_{\tilde{\gamma}_n}\big[K(x_i, x_j)_{ij}\big].$$

Here, $\tilde{\gamma}_n = (i(1) \to \cdots i(r) \to i(1))$ is the cycle γ_n in which n has been jumped. Replacing γ_n by $\tilde{\gamma}$, we obtain from σ a permutation $\tilde{\sigma}$ of $[\![1, n-1]\!]$. Besides, we have $\varepsilon(\sigma) = -\varepsilon(\tilde{\sigma})$ since we decreased by 1 the length of a cycle, and there are exactly $n-1$ permutations σ leading to the same $\tilde{\sigma}$ – corresponding to the number of ways to insert the element n back into $\tilde{\sigma}$. We thus find the contribution:

$$-(n-1) \sum_{\tilde{\sigma} \in \mathfrak{S}_{n-1}} \varepsilon(\tilde{\sigma}) \prod_{\gamma} P_{\gamma}\big[K(x_i, x_j)_{1 \leq i,j \leq n-1}\big].$$

Putting the two terms together, we arrive to:

$$\int_A \big(\operatorname*{det_{\mathbb{H}}}_{1 \leq i,j \leq n} K(x_i, x_j)\big) \mathrm{d}\mu(x_n) = (c - n + 1) \operatorname*{det_{\mathbb{H}}}_{1 \leq i,j \leq n-1} K(x_i, x_j).$$

We conclude by recursion on the number of integrations. □

5.10 corollary. *If K is a self-dual, trace-class, quasi projector with* $\mathrm{Tr}\, K = n$, *we have for any* $k \in [\![1, n]\!]$:

$$\frac{1}{(n-k)!} \int_{A^{n-k}} \big(\operatorname*{det_{\mathbb{H}}}_{1 \leq i,j \leq n} K(x_i, x_j)\big) \prod_{i=k+1}^{n} \mathrm{d}\mu(x_i) = \operatorname*{det_{\mathbb{H}}}_{1 \leq i,j \leq k} K(x_i, x_j)$$

6 Invariant ensembles: partition function and (skew)-orthogonal polynomials

6.1 *Unitary invariant ensembles*

6.1 lemma (Master integration formula). *Let $(\phi_i)_{i\geq 1}$ and $(\psi_j)_{j\geq 1}$ be two families of continuous functions. We have:*

$$\int_{A^n} \prod_{i=1}^{n} \mathrm{d}\mu(\lambda_i) \det_{1\leq i,j\leq n} \phi_i(\lambda_j) \cdot \det_{1\leq i,j\leq n} \psi_i(\lambda_j) = n! \det_{1\leq i,j\leq n} \Big[\int_A \mathrm{d}\mu(\lambda)\, \phi_i(\lambda)\, \psi_j(\lambda)\Big].$$

Proof. We denote $\mathcal{Z}_{n,2}[\mu]$ the integral in the left-hand side. By definition of the determinant and Fubini theorem:

$$\mathcal{Z}_{n,2}[\mu] = \sum_{\sigma,\tau\in\mathfrak{S}_n} \varepsilon(\sigma)\varepsilon(\tau) \prod_{i=1}^{n} \int_{\mathbb{A}} \mathrm{d}\mu(\lambda_i)\, \phi_{\sigma(i)}(\lambda_i)\, \psi_{\tau(i)}(\lambda_i).$$

Since we taking the product where all i's are playing the same role, the right-hand side only depends on the permutation $\tilde{\tau} = \tau\circ\sigma^{-1}$. We also remark $\varepsilon(\tilde{\tau}) = \varepsilon(\sigma)\varepsilon(\tau)$. Changing the summation variables (σ,τ) to $(\sigma,\tilde{\sigma})$, we get $n!$ times the same term by summing over σ, and:

$$\mathcal{Z}_{n,2}[\mu] = n! \sum_{\tilde{\tau}\in\mathfrak{S}_n} \varepsilon(\tilde{\tau}) \prod_{i=1}^{n} \int_{A^n} \mathrm{d}\mu(\lambda)\, \phi_i(\lambda)\, \psi_{\tilde{\tau}(i)}(\lambda),$$

which is the result announced. □

The integrand for invariant ensembles with $\beta = 2$ has this structure, since it is a product of two Vandermonde determinants:

$$\Delta(\lambda_1,\ldots,\lambda_n) = \det_{1\leq i,j\leq n} \phi_i(\lambda_j) \cdot \det_{1\leq i,j\leq n} \psi_i(\lambda_j), \qquad \phi_i(\lambda) = \psi_i(\lambda) = \lambda^{i-1}.$$

We could also choose $\phi_i = \psi_i$ any family of monic staggered polynomials $(q_i)_{i\geq 0}$.

6.2 proposition. *The partition function of unitary invariant ensembles takes a determinantal form:*

$$\begin{aligned} Z_{n,2}[\mu] &= n! \det_{1\leq i,j\leq n} M_{i+j}, \qquad M_k = \int_A \mathrm{d}\mu(\lambda)\, \lambda^k \\ &= n! \det_{1\leq i,j\leq n} \Big[\int_A \mathrm{d}\mu(x)\, q_{i-1}(\lambda)\, q_{j-1}(\lambda)\Big]. \end{aligned}$$

A determinant of the form $\det_{i,j}(M_{i+j})$ is called a **Hankel determinant**, and M_k here is the k-th moment of μ.

A cunning choice of q_i puts the matrix inside the determinant in diagonal form.

6.3 definition. We define a scalar product on the space of polynomials:

$$\langle f, g \rangle_2 = \int_A \mathrm{d}\mu(\lambda)\, f(\lambda)\, g(\lambda).$$

We say that $(p_i)_{i\geq 0}$ are orthogonal polynomials (for the measure μ) if it is a monic staggered family of polynomials and:

$$\forall i,j \geq 0, \qquad \langle p_i, p_j \rangle_2 = h_i \delta_{i,j}.$$

The fact that $\langle \cdot, \cdot \rangle_2$ is positive definite follows from positivity of μ. It also implies that orthogonal polynomials are unique and $h_i > 0$ for all i since they are the squared-norms of p_i. Gram-Schmidt orthogonalization shows that orthogonal polynomial exist. More precisely, for any $n \geq 0$, we can obtain the orthogonal polynomials $p_0, \ldots, p_{n-1}$ by orthogonalization in $\mathbb{R}_{n-1}[X]$ – equipped with this scalar product – of the canonical basis $1, \ldots, X^{n-1}$.

6.4 proposition.

$$\mathcal{Z}_{n,2}[\mu] = n! \prod_{i=0}^{n-1} h_i.$$

6.2 *Quaternionic unitary invariant ensembles*

6.5 lemma (Master integration formula). *Let $(\phi_i)_{i\geq 1}$ and $(\psi_j)_{j\geq 1}$ be two families of continuous functions.*

$$\int_{A^n} \prod_{i=1}^{n} \mathrm{d}\mu(\lambda_i) \det_{\substack{1\leq i\leq 2n \\ 1\leq j\leq n}} \big[\phi_i(\lambda_j)\ \ \psi_i(\lambda_j)\big] = n! \operatorname*{pf}_{1\leq i,j\leq 2n} \Big[\int_A \mathrm{d}\mu(\lambda)\big(\phi_i(\lambda)\psi_j(\lambda) - \phi_j(\lambda)\psi_i(\lambda)\big)\Big].$$

Proof. We denote $\mathcal{Z}_{n,4}[\mu]$ the left-hand side. By definition of the determinant and Fubini theorem:

$$\mathcal{Z}_{n,4}[\mu] = \sum_{\sigma\in\mathfrak{S}_{2n}} \varepsilon(\sigma) \prod_{i=1}^{n} \int_A \mathrm{d}\mu(\lambda_i)\, \phi_{\sigma(2i-1)}(\lambda_i)\, \phi_{\sigma(2i)}(\lambda_i)$$

since the variable λ_i appears in $(2i-1)$-th and $2i$-th position for $i \in [\![1, n]\!]$. We almost recognize the structure of the pfaffian, for a non anti-symmetric matrix. Since the signature of σ and the signature of $\sigma \circ (2i-1\ 2i)$ are opposite, we can also anti-symmetrize this formula:

$$\mathcal{Z}_{n,4} = \frac{1}{2^n} \sum_{\sigma\in\mathfrak{S}_{2n}} \varepsilon(\sigma) \prod_{i=1}^{n} \int_A \mathrm{d}\mu(\lambda_i)\big(\phi_{\sigma(2i)}(\lambda_i)\psi_{\sigma(2i-1)}(\lambda_i) - \phi_{\sigma(2i-1)}(\lambda_i)\psi_{\sigma(2i)}(\lambda_i)\big),$$

which entails the claim. □

This can be applied to the $\beta = 4$ invariant ensembles using:

6.6 lemma (Confluent Vandermonde). *For any monic staggered polynomials $(q_i)_{i\geq 0}$:*

$$\big(\Delta(\lambda_1,\ldots,\lambda_n)\big)^4 = \det_{\substack{1\leq i\leq 2n\\ 1\leq j\leq n}}\big[q_{i-1}(\lambda_j)\ \ q'_{i-1}(\lambda_j)\big].$$

Proof. To justify this, we consider first:

$$\begin{aligned}\Delta(\lambda_1,x_1,\ldots,\lambda_n,x_n) &= \Delta(\lambda_1,\ldots,\lambda_n)\cdot\Delta(x_1,\ldots,x_n)\cdot\prod_{1\leq i<j\leq n}(\lambda_j-x_i)(x_j-\lambda_i)\prod_{j=1}^{n}(x_j-\lambda_j)\\ &= \det_{\substack{1\leq i\leq 2n\\ 1\leq j\leq n}}\big[q_{i-1}(\lambda_j)\ \ q_{i-1}(x_j)\big].\end{aligned}$$

These two equalities allow two independent computations of the "confluent Vandermonde":

$$\tilde{\Delta}(\lambda_1,\ldots,\lambda_n) := \lim_{\substack{x_j\to\lambda_j\\ 1\leq j\leq n}}\frac{\Delta(\lambda_1,x_1,\ldots,\lambda_n,x_n)}{\prod_{j=1}^{n}(x_j-\lambda_j)}.$$

From the first equality, we immediately get:

$$\tilde{\Delta}(\lambda_1,\ldots,\lambda_n) = \big(\Delta(\lambda_1,\ldots,\lambda_n)\big)^4.$$

Now consider the second inequality, where a determinant appears. The confluence operation amounts to applying $\prod_{j=1}^{n}\partial_{x_j}$ and evaluate at $x_j=\lambda_j$. Since x_j only appears in the $2j$-th column, and the determinant is a linear function of each of its column, the differentiation ∂_{x_j} can be applied to the $2j$-th column for $j\in[\![1,n]\!]$, thus giving:

$$\tilde{\Delta}(\lambda_1,\ldots,\lambda_n) = \det_{\substack{1\leq i\leq 2n\\ 1\leq j\leq n}}\big[q_{i-1}(\lambda_j)\ \ q'_{i-1}(\lambda_j)\big].$$

□

6.7 proposition. *The partition function of the quaternionic unitary invariant ensembles takes a pfaffian form:*

$$\begin{aligned} Z_{n,4}[\mu] &= n!\ \operatorname*{pf}_{1\leq i,j\leq 2n}\Big[\int_A \mathrm{d}\mu(\lambda)\,(j-i)\,\lambda^{i+j-3}\Big]\\ &= n!\ \operatorname*{pf}_{1\leq i,j\leq 2n}\Big[\int_A \mathrm{d}\mu(\lambda)\big(q_{i-1}(\lambda)q'_{j-1}(\lambda)-q'_{i-1}(\lambda)q_{j-1}(\lambda)\big)\Big].\end{aligned}$$

Now, the simplest form we can hope to reach by a clever choice of q_i's is the pfaffian of a matrix formed by 2×2 blocks on the diagonal.

6.8 definition. We define a skew-symmetric bilinear form on the space of polynomials:

$$\langle f, g\rangle_4 = \int_A \mathrm{d}\mu(\lambda)\big(f(\lambda)\,g'(\lambda) - f'(\lambda)g(\lambda)\big).$$

We say that $(p_i)_{i\geq 0}$ are $\beta = 4$ skew-orthogonal polynomials (for the measure μ) if it is a monic staggered family of polynomials and:

$$\forall i,j \geq 0, \qquad \left\{ \begin{array}{l} \langle p_{2i}, p_{2j+1}\rangle_4 = -\langle p_{2i+1}, p_{2j}\rangle_4 = h_i\delta_{i,j} \\ \langle p_{2i}, p_{2j}\rangle_4 = \langle p_{2i+1}, p_{2j+1}\rangle_4 = 0 \end{array} \right.$$

The existence of $\beta = 4$ skew-orthogonal polynomials when μ is a positive measure will be justified – constructively – in Proposition 6.16. We can already remark that they are not unique, since we can add to p_{2i+1} any multiple of p_{2i} without changing the skew-orthogonality relations. However, h_i is uniquely defined, and called the pseudo-squared norm of p_i. We deduce:

6.9 proposition.

$$Z_{n,4}[\mu] = n!\,\prod_{i=0}^{n-1} h_i.$$

Since $Z_{n,4}[\mu]$ was the integral of a positive measure on $\mathbb{R}^n$, we deduce that:

$$h_n = (n+1)\,\frac{Z_{n+1,4}[\mu]}{Z_{n,4}[\mu]} > 0.$$

and in particular, the $\beta = 4$ skew-symmetric bilinear form is non-degenerate on $\mathbb{R}[X]$. These two last facts are not at all obvious from the definition of the bilinear form.

6.3 *Orthogonal ensembles*

6.10 lemma (Master integration formula). *Let $s \in L^\infty(A^2)$ be real-valued, such that $s(\lambda_1,\lambda_2) = -s(\lambda_2,\lambda_1)$, and $(\phi_i)_{i\geq 1}$ be a family of continuous functions. Let us denote:*

$$\mathcal{Z}_{n,1}[\mu] = \int_{A^n} \prod_{i=1}^n \mathrm{d}\mu(\lambda_i) \operatorname*{pf}_{1\leq i,j\leq n} s(\lambda_j,\lambda_i)\cdot \det_{1\leq i,j\leq n} \phi_i(\lambda_j).$$

If n is even, we have a pfaffian of size n:

$$\mathcal{Z}_{n,1}[\mu] = n!\,\operatorname*{pf}_{1\leq i,j\leq n} S_{i,j}, \qquad S_{i,j} = \int_{A^2} \mathrm{d}\mu(x)\mathrm{d}\mu(y)\,s(y,x)\,\phi_i(x)\phi_j(y)\,.$$

If n is odd, we have a pfaffian of size $(n+1)$:

$$\mathcal{Z}_{n,1}[\mu] = n!\,\operatorname*{pf}_{1\leq i,j\leq n}\left[\begin{array}{c|c} S_{i,j} & v_i \\ \hline -v_j & 0 \end{array}\right], \qquad v_i = \int \mathrm{d}\mu(x)\,\phi_i(x).$$

Proof. Let $m = \lfloor n/2 \rfloor$. We replace the pfaffian and the determinant by their definition:

$$\mathcal{Z}_{n,1}[\mu] = \frac{1}{2^m\, m!} \sum_{\sigma,\tau \in \mathfrak{S}_n} \varepsilon(\sigma)\varepsilon(\tau) \int_{A^n} \prod_{i=1}^{m} s(\lambda_{\sigma(2i)}, \lambda_{\sigma(2i-1)}) \prod_{i=1}^{n} \phi_{\tau(i)}(x_i)\, \mathrm{d}\mu(x_i).$$

Since the integration variables all play the same role, we can relabel them $\sigma(i) \to i$. Then, the variables $(\lambda_{2i-1}, \lambda_{2i})$ are coupled via s for each $i \in [\![1, m]\!]$, and we can use Fubini theorem to write the result as a sum of products of m double integrals. If n is odd, this must be amended since the variable λ_n – formerly labeled $\lambda_{\sigma(n)}$ – stays alone and gives an extra 1-dimensional integral. All in all, denoting $\tilde{\tau} = \sigma \circ \tau^{-1}$:

$$\begin{aligned} \mathcal{Z}_{n,1}[\mu] &= \frac{1}{2^m\, m!} \sum_{\sigma,\tilde{\tau} \in \mathfrak{S}_n} \varepsilon(\tilde{\tau}) \prod_{i=1}^{m} s(\lambda_{2i}, \lambda_{2i-1})\, \phi_{\tilde{\tau}(2i)}(\lambda_{2i}) \phi_{\tilde{\tau}(2i-1)}(\lambda_{2i-1}) \mathrm{d}\mu(\lambda_{2i-1}) \mathrm{d}\mu(\lambda_{2i}) \\ & \quad \times \int_A \phi_{\tilde{\tau}(n)}(\lambda_n) \mathrm{d}\mu(\lambda_n) \quad \text{if } n \text{ odd}. \end{aligned}$$

The terms do not depend on σ so we get a factor of $n!$. For n even, the sum over $\tilde{\tau}$ is precisely the pfaffian of the matrix S. For n odd, let us extend $\tilde{\tau}$ to a permutation of $(n+1)$ indices by setting $\tilde{\tau}(n+1) = n+1$, and define the $(n+1) \times (n+1)$ matrix:

$$\tilde{S} = \left(\begin{array}{c|c} S_{i,j} & v_i \\ \hline -v_j & 0 \end{array} \right), \qquad v_i = \int_A \mathrm{d}\mu(x)\, \phi_i(x).$$

We recognize:

$$\mathcal{Z}_{n,1}[\mu] = \frac{n!}{2^m\, m!} \sum_{\substack{\tilde{\tau} \in \mathfrak{S}_{n+1} \\ \tilde{\tau}(n+1) = n+1}} \varepsilon(\tilde{\tau}) \prod_{i=1}^{m+1} \tilde{S}_{\tilde{\tau}(2i-1), \tilde{\tau}(2i)}.$$

The terms in this sum only depends on the set of pairs $\{\{\tilde{\tau}(2i-1), \tilde{\tau}(2i)\},\ 1 \leq i \leq n+1\}$, and there are precisely $2^{m+1}(m+1)!$ permutations giving the same set of pairs (corresponding to changing the order in which the pair appears, and labeling the two elements in a given pair). In particular, there is one pair that contains $(n+1)$, and allowing for relabeling of this element, i.e. waiving the condition $\tilde{\tau}(n+1) = n+1$, gives $(n+1) = 2(m+1)$ copies of the same term. Thus:

$$\mathcal{Z}_{n,1}[\mu] = \frac{n!}{2^{m+1}\,(m+1)!} \sum_{\tilde{\tau} \in \mathfrak{S}_n} \varepsilon(\sigma) \prod_{i=1}^{m+1} \tilde{S}_{\tilde{\tau}(2i-1), \tilde{\tau}(2i)} = n! \underset{1 \leq i,j \leq n+1}{\mathrm{pf}} \tilde{S}_{i,j}.$$

□

The annoying feature of the $\beta = 1$ ensembles is the absolute value around the Vandermonde:

$$\begin{aligned}\left|\Delta(\lambda_1,\ldots,\lambda_n)\right| &= \prod_{1\leq i<j\leq n}\mathrm{sgn}(\lambda_j-\lambda_i)\cdot\Delta(x_1,\ldots,x_n)\\ &= \prod_{1\leq i<j\leq n}\mathrm{sgn}(\lambda_j-\lambda_i)\cdot\det_{1\leq i,j\leq n} q_{i-1}(\lambda_j).\end{aligned}\tag{30}$$

Remarkably, the product of signs can be written as a pfaffian, so we recognize the structure of Lemma 6.10.

6.11 lemma (de Bruijn, 1955). *Let X be a completely ordered set, and $x_1,\ldots,x_n$ be pairwise disjoint elements of X. We define* $\mathrm{sgn}(x,y)=1$ *if $x>y$, and -1 if $y<x$. We have:*

$$\prod_{1\leq i<j\leq n}\mathrm{sgn}(x_j,x_i)=\mathop{\mathrm{pf}}_{1\leq i,j\leq n}\mathrm{sgn}(x_j,x_i).$$

Proof. The two sides are antisymmetric functions of the x_i's, so it is enough to prove the lemma for $x_1<\ldots<x_n$. In this case, the left-hand side is 1. The identity for n even implies the identity for n odd. Indeed, if n is odd, we can add to X a variable ∞ which is larger than all elements of X, and set $x_{n+1}=\infty$, and both sides of the identity remain the same under this operation (since all the new signs are $+1$). Now, let us assume $n=2m$ even. We denote S the $n\times n$ matrix with entries $S_{i,j}$, of which we want to compute the pfaffian. We introduce the $n\times n$ matrix P:

$$\forall i,j\in[\![1,n]\!],\qquad P_{i,j}=\begin{cases}1 & \text{if } i>j\\ -1 & \text{if } i=2k-1,\ j=2k,\ k\in[\![1,m]\!]\\ 0 & \text{otherwise}\end{cases}.$$

If we can show $S=PJ_mP^T$, the properties of the pfaffian imply

$$\mathrm{pf}\,S=\det P\cdot\mathrm{pf}\,J_m=\det P=1.$$

Since S and PJ_mP^T are antisymmetric matrices, it is enough to check the identity for the lower off-diagonal terms. For $i>j$, we compute:

$$(PJ_mP^T)_{i,j}=\sum_{k=1}^{m}P_{i,2k-1}P_{j,2k}-P_{i,2k}P_{j,2k-1}=\sum_{k=1}^{\lfloor i/2\rfloor}P_{j,2k}-\sum_{k=1}^{\lfloor (i-1)/2\rfloor}P_{j,2k-1}.$$

The -1 entry in $P_{i,\bullet}$ can contribute only when i is odd, and to the k-th terms such that $i=2k-1$. However, since we assumed $j<i$, we have $P_{j,i=2k-1}=0$ in this case, so we never find an extra term in (6.3). The remaining sums contain only 1's, except for the contribution of the -1 in the first sum when j is odd:

$$(PJ_mP^T)_{i,j}=\left\lfloor\frac{j-1}{2}\right\rfloor-\left\lfloor\frac{j}{2}\right\rfloor+\begin{cases}-1 & \text{if } j \text{ odd}\\ 0 & \text{if } j \text{ even}\end{cases}$$

This is in any case equal to -1, which coincides with $\mathrm{sgn}(x_j-x_i)$. ☐

6.12 proposition. *The partition function of the orthogonal invariant ensembles takes a pfaffian form, whose size depends on the parity of n.*

$$n \text{ even} \qquad Z_{n,1}[\mu] = n! \operatorname*{pf}_{1 \leq i,j \leq n} S_{i,j}\,,$$
$$n \text{ odd} \qquad Z_{n,1}[\mu] = n! \operatorname*{pf}_{1 \leq i,j \leq n+1} \tilde{S}_{i,j}\,,$$

where:

$$S = \Big(\int_{A^2} \mathrm{d}\mu(\lambda_1)\mathrm{d}\mu(\lambda_2)\,\mathrm{sgn}(\lambda_2 - \lambda_1) q_{i-1}(\lambda_1) q_{j-1}(\lambda_2) \Big)_{1 \leq i,j \leq n},$$

and:

$$\tilde{S} = \left(\begin{array}{c|c} S & \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} \\ \hline -v_1 \ \ldots \ -v_n & 0 \end{array} \right), \qquad v_i = \int_A \mathrm{d}\mu(\lambda)\, q_{i-1}(\lambda).$$

As for $\beta = 4$, the result is simplified by the choice of suitable skew-orthogonal polynomials.

6.13 definition. We define the skew-symmetric bilinear form on $\mathbb{R}[X]$:

$$\langle f, g \rangle_1 = \int_{A^2} \mathrm{d}\mu(\lambda_1)\mathrm{d}\mu(\lambda_2)\mathrm{sgn}(\lambda_2 - \lambda_1)\, f(\lambda_1) f(\lambda_2).$$

We say that $(p_i)_{i \geq 0}$ are $\beta = 1$ skew-orthogonal polynomials (for the measure μ) if it is a monic staggered family of polynomials and:

$$\forall i, j \geq 0, \qquad \begin{cases} \langle p_{2i,2j+1} \rangle_1 = -\langle p_{2i+1,2j} \rangle_1 = h_i \delta_{i,j} \\ \langle p_{2i}, p_{2j} \rangle_1 = \langle p_{2i+1}, p_{2j+1} \rangle_1 = 0 \end{cases}$$

The existence of $\beta = 1$ skew orthogonal polynomials will be justified – constructively – in Proposition § 6.18. As for $\beta = 4$, they are not unique since we can shift $p_{2i+1} \to p_{2i+1} + c_i\, p_{2i}$ for any constant c_i.

6.14 proposition.

$$n \text{ even} \qquad Z_{n,1}[\mu] = n! \prod_{i=0}^{n/2-1} h_i\,,$$

$$n \text{ odd} \qquad Z_{n,1}[\mu] = n! \Big(\prod_{i=0}^{(n-3)/2} h_i \Big) \cdot \int_A \mathrm{d}\mu(\lambda)\, p_n(\lambda)\,. \tag{31}$$

The formula for n odd contains an extra factor, and is established by direct computation of the pfaffian of size $n+1$ for the matrix of pairings of p_i's augmented by the last column v/last line $-v$. This also shows that $h_i > 0$ and $\int_A \mathrm{d}\mu(\lambda)\, p_{2i+1}(\lambda) > 0$ for any $i \geq 0$, independently of the choice of the skew-orthogonal polynomials. This fact is not at all obvious from the definitions.

6.4 *Orthogonal and skew-orthogonal polynomials*

Lecture 11 (1h30)
November 11th, 2014

The purpose of this section is to show that the relation between (skew)-orthogonal polynomials and invariant ensembles is deeper: (skew) orthogonal polynomial themselves are (closely related to) the expectation value of the characteristic polynomial of the random matrix M drawn from invariant ensembles.

A remark

Let $f \in L^2(A,\mu)$ be a complex-valued such $Z_{n,\beta}[\mu \cdot f]$ converges absolutely. We can write:

$$\mathbb{E}_{n,\beta}\Big[\prod_{i=1}^{n} f(\lambda_i)\Big] = \frac{Z_{n,\beta}[f \cdot \mu]}{Z_{n,\beta}[\mu]},$$

where the expectation value is computed with respect to the probability measure on A^n:

$$Z_{n,\beta}^{-1}[\mu] \cdot \prod_{i=1}^{n} \mathrm{d}\mu(\lambda_i)\,\big|\Delta(\lambda_1,\ldots,\lambda_n)\big|^{\beta}. \tag{32}$$

The previous results show how to compute the partition functions in terms of any family of monic staggered polynomials. Imagine that we choose this family to be orthogonal (or skew-orthogonal if $\beta = 1,4$) polynomials $(p_i)_{i\geq 0}$ *for the measure* μ.

For $\beta = 2$, we find:

$$\begin{aligned} \mathbb{E}_{n,\beta}\big[\det f(M)\big] &= \mathbb{E}_{n,\beta}\Big[\prod_{i=1}^{n} f(\lambda_i)\Big] \\ &= \Big(\prod_{i=0}^{n-1} h_i^{-1}\Big) \det_{1\leq i,j\leq n}\Big[\int_A f(\lambda)\, p_{i-1}(\lambda) p_{j-1}(\lambda)\Big] \\ &= \det_{1\leq i,j\leq n}\Big[\int_A f(\lambda)\, \hat{p}_{i-1}(\lambda)\hat{p}_{j-1}(\lambda)\Big], \end{aligned} \tag{33}$$

where we have introduced the **orthonormal** polynomials $\hat{p}_i(\lambda) = p_i(\lambda)/h_i$.

This formula has a nice interpretation in terms of truncated operators. $L^2(A,\mu)$ is a Hilbert space when equipped with the scalar product $\langle\cdot,\cdot\rangle_2$. Since we assumed that all moments of μ are finite, $\mathbb{R}_{n-1}[X]$ is a subspace, which has $(\hat{p}_i)_{0\leq i\leq n-1}$ as an orthonormal basis. Let us define:

- $\iota_n\,:\,\mathbb{R}_{n-1}[X] \to L^2(A,\mu)$ the canonical inclusion.
- $\pi_n\,:\,L^2(A,\mu) \to \mathbb{R}_{n-1}[X]$ the orthogonal projection.
- $\mathbf{f}\,:\,L^2(A,\mu) \to L^2(A,\mu)$ the endomorphism consisting in pointwise multiplication by the function f.

We see that:

$$\mathbb{E}_{n,\beta}\big[\det f(M)\big] = \operatorname*{Det}_{\mathbb{R}_{n-1}[X]} \pi_n\,\mathbf{f}\,\iota_n\,,$$

where Det is defined as the determinant of the matrix of the operator in any orthonormal basis – and this does not depend on the choice of orthonormal basis. In other words, the expectation value in the left-hand side is the determinant of the truncation of the operator $\mathbf{f}$ to a finite-dimensional subspace – that is increasing (for the inclusion) with n. Yet another reformulation: we can exchange the order of the determinant and the expectation value, provided we replace the scalar function f by an operator $\mathbf{f}$. This fact makes random matrix models analogous to quantum mechanical systems. In particular, if we choose $f(\lambda) = (z - \lambda)$, we have:

$$\mathbb{E}_{n,\beta}\big[\det(z - M)\big] \quad = \quad \mathrm{Det}_{\mathbb{R}_{n-1}[X]}(z - \mathbf{f}_n), \qquad \mathbf{f}_n = \pi_n\, \mathbf{f}\, \iota_n\,. \tag{34}$$

A similar reformulation can be settled for $\beta = 1$ and 4. It is however important to notice that we want to use the skew-orthogonal polynomials $(p_i)_{i\geq 0}$ for the measure μ, and not for $\mu \cdot f$.

$\beta = 2$: orthogonal polynomials

6.15 THEOREM (Heine). *The orthogonal polynomials are given by:*

$$p_n(z) = \mathbb{E}_{n,2}\big[\det(z - M)\big]\,.$$

As a consequence of (34), the orthogonal polynomial of degree n is the characteristic polynomial of the truncation $\mathbf{f}_n$ of $\mathbf{f} \in \mathcal{L}(L^2(A, \mu))$ to a subspace of dimension n. In particular, the zeroes of the orthogonal polynomials are the eigenvalues[18] of $\mathbf{f}_n$.

Proof. We observe that:

$$\prod_{i=1}^{n}(z - \lambda_i)\Delta(\lambda_1, \ldots, \lambda_n) = \Delta(\lambda_1, \ldots, \lambda_n, z)\,.$$

Therefore, with the convention $\lambda_{n+1} = z$ fixed, we can write:

$$Z_{n,2}[\mu] \cdot \mathbb{E}_{n,2}\big[\det(z - M)\big] = \int_{A^n} \prod_{i=1}^{n} \mathrm{d}\mu(\lambda_i) \det_{1\leq i,j\leq n} p_{i-1}(\lambda_j) \cdot \det_{1\leq i,j\leq n+1} p_{i-1}(\lambda_j)\,.$$

Then, we replace the first determinant by its definition:

$$n! \prod_{i=0}^{n-1} h_i \cdot \mathbb{E}_{n,2}\big[\det(z - M)\big] = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma) \int_{A^n} \prod_{i=1}^{n} p_{\sigma(i)-1}(\lambda_i)\mathrm{d}\mu(\lambda_i) \cdot \det_{1\leq i,j\leq n+1} p_{i-1}(\lambda_j)\,.$$

We can relabel the integration variables $i \to \sigma(i)$, and by antisymmetry of the second determinant, this absorbs the signature $\varepsilon(\sigma)$. We are left with $n!$ equal

[18] From this fact, one can e.g. give a proof that the zeroes of p_i interlace those of p_{i+1}.

terms, and the factorial cancels on both sides:

$$\Big(\prod_{i=0}^{n-1} h_i\Big)\mathbb{E}_{n,2}\big[\det(z-M)\big] = \int_{A^n}\prod_{j=1}^{n} p_{j-1}(\lambda_j)\mathrm{d}\mu(\lambda_j)\cdot \det_{1\leq i,j\leq n+1} p_{j-1}(\lambda_i)\,.$$

By linearity of the determinant with respect to its columns, we can include the first factor and the integration over λ_j in the j-th column of second determinant for $j \in [\![1,n]\!]$:

$$\Big(\prod_{i=0}^{n-1} h_i\Big)\mathbb{E}_{n,2}\big[\det(z-M)\big] = \det_{\substack{1\leq i\leq n+1\\ 1\leq j\leq n}}\left[\begin{array}{c|c}\int_A \mathrm{d}\mu(\lambda_j)\,p_{i-1}(\lambda_j)p_{j-1}(\lambda_j) & \begin{array}{c} p_0(z)\\ \vdots \\ p_n(z)\end{array}\end{array}\right].$$

Since the p_i's are orthogonal, the matrix is upper-triangular, with $h_0,\ldots,h_{n-1},p_n(z)$ on the diagonal. Hence:

$$\Big(\prod_{i=0}^{n-1} h_i\Big)\mathbb{E}_{n,2}\big[\det(z-M)\big] = \prod_{i=0}^{n-1} h_i\cdot p_n(z)\,,$$

which is the desired relation. □

$\beta = 4$ skew-orthogonal polynomials

6.16 PROPOSITION (Eynard, 2001). *Let $(c_n)_{n\geq 0}$ be an arbitrary sequence of real numbers. The following formula define $\beta = 4$ skew-orthogonal polynomials:*

$$p_{2n}(z) = \mathbb{E}_{n,4}\big[\det(z-M)\big],\qquad p_{2n+1}(z) = \mathbb{E}_{n,4}\big[(z+\mathrm{Tr}M + c_n)\det(z-M)\big].$$

Notice that the polynomials of degree $2n$ and $2n+1$ are both given by an expectation value over quaternionic self-dual matrices of size n. We remind that in the representation as $2n\times 2n$ complex matrices, M has n eigenvalues with even multiplicities.

Proof. The right-hand sides define monic staggered polynomials $\tilde{p}_i(z)$ for $i\geq 0$. The claim is equivalent to checking the skew-orthogonality relations:

$$(35)\quad \forall n\geq 0,\qquad \left\{\begin{array}{ll} \forall q\in\mathbb{R}_{2n}[X], & \langle\tilde{p}_{2n},Q\rangle_4 = 0 \\ \forall q\in\mathbb{R}_{2n-1}[X], & \langle\tilde{p}_{2n+1},Q\rangle_4 = 0\end{array}\right. .$$

Let us focus on the first relation. We want to compute the pairing:

$$\langle\tilde{p}_{2n},q\rangle_4 \propto \int_{A^{n+1}}\prod_{i=1}^{n+1}\mathrm{d}\mu(\lambda_i)T(\lambda_1,\ldots,\lambda_{n+1})\,,$$

where:

$$\begin{aligned} T(\lambda_1,\ldots,\lambda_n;\lambda_{n+1}) &= R(\lambda_1,\ldots,\lambda_{n+1})Q'(\lambda_{n+1}) - \partial_{\lambda_{n+1}}R(\lambda_1,\ldots,\lambda_{n+1})\,Q(\lambda_{n+1}) \\ R(\lambda_1,\ldots,\lambda_{n+1}) &= \Delta(\lambda_1,\ldots,\lambda_n)^4\cdot\prod_{i=1}^{n}(\lambda_{n+1}-\lambda_i)^2\,. \end{aligned}$$

The notation $\propto$ means "proportional to": since we want to show that the expression vanishes, we do not care about overall multiplicative prefactors. Following the proof of the confluent Vandermonde (Lemma 6.6), we see that:

$$\begin{aligned} R(\lambda_1,\ldots,\lambda_{n+1}) &= \lim_{\substack{x_j\to\lambda_j \\ 1\leq j\leq n}} \frac{\Delta(\lambda_1,x_1,\ldots,\lambda_n,x_n,\lambda_{n+1})}{\prod_{j=1}^n (x_j-\lambda_j)} \\ &= \det\begin{bmatrix} q_0(\lambda_1) & q_0'(\lambda_1) & \cdots & q_0(\lambda_n) & q_0'(\lambda_n) & q_0(\lambda_{n+1}) \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ q_{2n}(\lambda_1) & q_{2n}'(\lambda_1) & \cdots & q_{2n}(\lambda_n) & q_{2n}'(\lambda_n) & q_{2n}(\lambda_{n+1}) \end{bmatrix}, \end{aligned}$$

where $(q_i)_{i\geq 0}$ is an arbitrary monic staggered family of polynomials. And, since the variable λ_{n+1} only appears in the last column, it is easy to differentiate:

$$\partial_{\lambda_{n+1}} R(\lambda_1,\ldots,\lambda_{n+1}) = \det\begin{bmatrix} q_0(\lambda_1) & q_0'(\lambda_1) & \cdots & q_0(\lambda_n) & q_0'(\lambda_n) & q_0'(\lambda_{n+1}) \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ q_{2n}(\lambda_1) & q_{2n}'(\lambda_1) & \cdots & q_{2n}(\lambda_n) & q_{2n}'(\lambda_n) & q_{2n}'(\lambda_{n+1}) \end{bmatrix}.$$

T is a linear combination of these two determinants of size $(n+1)$, therefore it can be represented as the expansion with respect to the last line of a determinant of size $(n+2)$:

$$\begin{aligned} & T(\lambda_1,\ldots,\lambda_n;\lambda_{n+1}) \\ = \ & \det\begin{bmatrix} q_0(\lambda_1) & q_0'(\lambda_1) & \cdots & q_0(\lambda_n) & q_0'(\lambda_n) & q_0(\lambda_{n+1}) & q_0'(\lambda_{n+1}) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ q_{2n}(\lambda_1) & q_{2n}'(\lambda_1) & \cdots & q_{2n}(\lambda_n) & q_{2n}'(\lambda_n) & q_{2n}(\lambda_{n+1}) & q_{2n}'(\lambda_{n+1}) \\ 0 & 0 & \cdots & 0 & 0 & Q(\lambda_{n+1}) & Q'(\lambda_{n+1}) \end{bmatrix}. \end{aligned}$$

The variable λ_{n+1} plays a special role in T, but since we integrate over all λ_i's, relabeling gives the same result. Thus:

$$(36) \quad \langle \tilde{p}_{2n}, Q\rangle_4 \propto \int_{A^{n+1}} \prod_{i=1}^{n+1} \mathrm{d}\mu(\lambda_i)\, T(\lambda_1,\ldots,\lambda_{i-1},\lambda_{i+1},\ldots,\lambda_{n+1};\lambda_i),$$

and by permutation of rows and columns, we notice that

$$T(\lambda_1,\ldots,\lambda_{i-1},\lambda_{i+1},\ldots,\lambda_{n+1};\lambda_i)$$

can be written as a determinant of a matrix identical to that appearing in $T(\lambda_1,\ldots,\lambda_n;\lambda_{n+1})$, except for the last line which consists in $Q(\lambda_i)$ $Q(\lambda_i)'$ in the $(2i-1)$-th and $2i$-th columns and zero for the other entries. And, by linearity of the determinant with respect to its lines, the sum over i can be included in the last line:

$$\langle \tilde{p}_{2n}, Q\rangle_4 \propto \int_{A^{n+1}} \prod_{i=1}^{n+1} \mathrm{d}\mu(\lambda_i)\, \det\begin{bmatrix} q_0(\lambda_1) & q_0'(\lambda_1) & \cdots & q_0(\lambda_{n+1}) & q_0'(\lambda_{n+1}) \\ \vdots & \vdots & & \vdots & \vdots \\ q_{2n}(\lambda_1) & q_{2n}'(\lambda_1) & \cdots & q_{2n}(\lambda_{n+1}) & q_{2n}'(\lambda_{n+1}) \\ Q(\lambda_1) & Q'(\lambda_1) & \cdots & Q(\lambda_{n+1}) & Q'(\lambda_{n+1}) \end{bmatrix}. \tag{37}$$

Finally, if $\deg Q \leq 2n$, then it must be a linear combination of the staggered family q_i for $i \in [\![0, 2n]\!]$, so the lines of this matrix are linearly dependent: the determinant is 0.

Now consider the second relation in (35). Notice that $\mathrm{Tr} M = \sum_{i=1}^{n} \lambda_i$. So, we have to compute:

$$\langle \tilde{p}_{2n}, q\rangle_4 \propto \int_{A^{n+1}} \prod_{i=1}^{n+1} \mathrm{d}\mu(\lambda_i) \tilde{T}(\lambda_1, \ldots, \lambda_{n+1}),$$

with:

$$\begin{aligned} \tilde{T}(\lambda_1, \ldots, \lambda_{n+1}) &= \tilde{R}(\lambda_1, \ldots, \lambda_{n+1}) Q'(\lambda_{n+1}) - \partial_{\lambda_{n+1}} \tilde{R}(\lambda_1, \ldots, \lambda_{n+2}) Q(\lambda_{n+1}) \\ \tilde{R}(\lambda_1, \ldots, \lambda_{n+1}) &= \Big(\lambda_{n+1} + \sum_{i=1}^{n} \lambda_i + c_n\Big) R(\lambda_1, \ldots, \lambda_n). \end{aligned}$$

This can be written:

$$\begin{aligned} \tilde{T}(\lambda_1, \ldots, \lambda_{n+1}) &= \Big(\sum_{i=1}^{n+1} 2\lambda_i + c_n) T(\lambda_1, \ldots, \lambda_{n+1}) \\ &\quad - \Big(R(\lambda_1, \ldots, \lambda_{n+1}) \tilde{Q}'(\lambda_{n+1}) - \partial_{\lambda_{n+1}} R(\lambda_1, \ldots, \lambda_{n+1}) \tilde{Q}(\lambda_{n+1})\Big). \end{aligned} \tag{38}$$

where $\tilde{Q}(\lambda_{n+1}) = \lambda_{n+1} Q(\lambda_{n+1})$. We can then repeat, for each line, the manipulations (symmetrization) we did in the even case. In the first line, this leads to the integral (37) with an extra factor $(\sum_{i=1}^{n+1} 2\lambda_i + c_n)$ in the integrand, and we have seen the determinant is 0. In the second line, this leads to (37) with $\tilde{Q}$ instead of Q. Since $\deg \tilde{Q} = \deg Q + 1 \leq 2n$, the determinant is again zero. □

$\beta = 1$ skew-orthogonal polynomials

Let us start with an elementary lemma about sign functions:

6.17 lemma. *Let $x_1, \ldots, x_n, y$ be pairwise disjoint elements of a completely ordered set X, and n odd. We have:*

$$\sum_{k=1}^{n} \mathrm{sgn}(y - x_k) \prod_{\substack{i=1 \\ i \neq k}}^{n} \mathrm{sgn}(x_k - x_i) = \prod_{i=1}^{n} \mathrm{sgn}(y - x_i).$$

Proof. Both sides are antisymmetric functions of the x_i's, so we can assume $x_1 > \ldots > x_n$. We add to X two elements: $-\infty$ which smaller than all elements in X ; $+\infty$ which is larger than all elements in X. And we complete our sequence by $x_0 = +\infty$ and $x_{n+1} = -\infty$. Then, there is a unique $\ell \in [\![0, n]\!]$ such that $x_\ell > y > x_{\ell+1}$, and the right-hand side is equal to $(-1)^\ell$, while the left-hand side is:

$$-\sum_{k=1}^{\ell}(-1)^{k-1} + \sum_{k=\ell+1}^{n}(-1)^{k-1} = \Big(\sum_{k=1}^{n+1}(-1)^k\Big) - (-1)^{\ell+1}.$$

This last equality comes from the observation that the jump of sign at $k = \ell + 1$ can be interpreted as a gap at position $\ell + 1$ in a sum of $n + 1$ alternating signs. Since n is odd, the first sum is zero and the total is equal also equal to $(-1)^\ell$. ☐

6.18 proposition (Eynard, 2001). *Let $(c_n)_{n\geq 0}$ be an arbitrary sequence of real numbers. The following formula define $\beta = 1$ skew-orthogonal polynomials:*

$$p_{2n}(z) = \mathbb{E}_{2n,1}\big[\det(x - M)\big], \quad p_{2n+1}(z) = \mathbb{E}_{2n,1}\big[(z + \mathrm{Tr}\, M + c_n)\det(z - M)\big].$$

Proof. As for $\beta = 4$, we need to check that the polynomials $\tilde{p}_i(z)$ in the right-hand sides satisfy the skew-orthogonality relations (35) for the pairing $\langle\cdot,\cdot\rangle_1$. We first consider:

$$\langle \tilde{p}_{2n}, Q\rangle_1 \propto \int_{A^{2n+2}} \prod_{i=1}^{2n+2} \mathrm{d}\mu(\lambda_i) S(\lambda_1, \ldots, \lambda_{2n}) \cdot \mathrm{sgn}(\lambda_{2n+2} - \lambda_{2n+1}) \\ \times \Delta(\lambda_1, \ldots, \lambda_{2n+1})\, Q(\lambda_{2n+2}),$$

where for convenience we introduced:

$$S(\lambda_1, \ldots, \lambda_n) = \prod_{1\leq i<j\leq n} \mathrm{sgn}(\lambda_j - \lambda_i).$$

We can extend the product of signs over pairs to $1 \leq i < j \leq 2n+1$, provided we also multiply by an extra $\prod_{i=1}^{2n+1}\mathrm{sgn}(\lambda_{2n+1} - \lambda_i)$. The variables λ_{2n+1} and λ_{2n+2} plays a special role in the integrand, but we are anyway integrating over all λ_i's. So, we can symmetrize the formula in λ_{2n+1}, and find:

$$\langle \tilde{p}_{2n}, Q\rangle_1 \propto \int_{A^{2n+2}} \prod_{i=1}^{2n+2} \mathrm{d}\mu(\lambda_i)\, \Delta(\lambda_1, \ldots, \lambda_{2n+1}) \cdot S(\lambda_1, \ldots, \lambda_{2n+1}) \\ \times \left\{\sum_{k=1}^{2n+1} \mathrm{sgn}(\lambda_{2n+2} - \lambda_k) \prod_{\substack{i=1 \\ i\neq k}}^{2n+1} \mathrm{sgn}(\lambda_k - \lambda_i)\right\} Q(\lambda_{2n+2}).$$

Now, we use the identity of Lemma 6.17 with $2n+1$ odd and $y = \lambda_{2n+2}$. The sum is replaced by a product of signs that can be incorporated in S in which the variable λ_{2n+2} is added:

$$\langle \tilde{p}_{2n}, Q \rangle_1 \propto \int_{A^{2n+2}} \prod_{i=1}^{2n+2} \mathrm{d}\mu(\lambda_i)\, S(\lambda_1, \ldots, \lambda_{2n+2}) \cdot \Delta(\lambda_1, \ldots, \lambda_{2n+1}) \cdot Q(\lambda_{2n+2}),$$

and we can write part of the integrand as a determinant of size $(2n+2)$ of a block-upper triangular matrix:

$$\Delta(\lambda_1, \ldots, \lambda_{2n+1}) \cdot Q(\lambda_{2n+2}) = \det \begin{bmatrix} q_0(\lambda_1) & \cdots & q_0(\lambda_{2n+1}) & q_0(\lambda_{2n+2}) \\ \vdots & & \vdots & \vdots \\ q_{2n}(\lambda_1) & \cdots & q_{2n}(\lambda_{2n+1}) & q_{2n}(\lambda_{2n+2}) \\ 0 & \cdots & 0 & Q(\lambda_{2n+2}) \end{bmatrix},$$

where $(q_i)_{i \geq 0}$ is any monic staggered family of polynomials. At this stage, we can relabel λ_{2n+2}, then symmetrize as we did for $\beta = 4$ and use the linearity of the determinant with respect to its last line. Because the product of signs in S is completely antisymmetric, this yields:

$$\langle \tilde{p}_{2n}, Q \rangle_1 \propto \int_{A^{2n+2}} \prod_{i=1}^{2n+2} \mathrm{d}\mu(\lambda_i)\, S(\lambda_1, \ldots, \lambda_{2n+2}) \cdot \det \begin{bmatrix} q_0(\lambda_1) & \cdots & q_0(\lambda_{2n+2}) \\ \vdots & & \vdots \\ q_{2n}(\lambda_1) & \cdots & q_{2n}(\lambda_{2n+2}) \\ Q(\lambda_1) & \cdots & Q(\lambda_{2n+2}) \end{bmatrix}.$$

If Q has degree $\leq 2n$, it must be linear combination of q_i for $i \in [\![0, 2n]\!]$, and the determinant vanishes.

We now consider $\langle \tilde{p}_{2n+1}, Q \rangle_1$ for a polynomial Q of degree $\leq 2n-1$:

$$\begin{aligned} \langle \tilde{p}_{2n+1}, Q \rangle_1 \quad \propto \quad & \int_{A^{2n+2}} \prod_{i=1}^{2n+2} \mathrm{d}\mu(\lambda_i)\, S(\lambda_1, \ldots, \lambda_{2n+2}) \cdot \Delta(\lambda_1, \ldots, \lambda_{2n+2}) \cdot Q(\lambda_{2n+2}) \\ & \times \Big(\sum_{i=1}^{2n+1} \lambda_i + c_n \Big) \end{aligned}$$

Following the same steps, we arrive to an integral of:

$$\begin{aligned} & \Big(\sum_{i=1}^{2n+2} \lambda_i + c_n \Big) S(\lambda_1, \ldots, \lambda_{2n+2}) \Delta(\lambda_1, \ldots, \lambda_{2n+2})\, Q(\lambda_{2n+2}) \\ & - S(\lambda_1, \ldots, \lambda_{2n+2}) \Delta(\lambda_1, \ldots, \lambda_{2n+2})\, \tilde{Q}(\lambda_{2n+2}) \end{aligned}$$

with $\tilde{Q}(\lambda_{2n+2}) = \lambda_{2n+2} Q(\lambda_{2n+2}$. Then, we can symmetrize each term with respect to λ_{2n+2} as before (in the first line, the prefactor in bracket is already completely symmetric, thus it remains a factor after the symmetrization). We can apply the same remark concerning the vanishing of the determinant for Q or $\tilde{Q}$ – which both have degree $\leq 2n$ – to conclude. □

7 Invariant ensembles: eigenvalue statistics and kernels

Lecture 12 (1h30)
November 17th, 2014

We harvest the fruits of Chapter 5, relying on the notions on quaternionic determinants (Definition 5.3) and quasi-projectors (Definition 5.8).

7.1 *Definition of some eigenvalue statistics*

Imagine we have a probability measure $\mathbb{P}_n$ on $\mathbb{R}^n$, and consider the pushforward of $\mathbb{P}_n$ by the projection $\mathbb{R}^n \to \mathbb{R}^n/\mathfrak{S}_n$. $\{\lambda_1, \ldots, \lambda_n\} \in \mathbb{R}^n/\mathfrak{S}_n$ could be e.g. eigenvalues of a random matrix, and our main example is $\mathbb{P}_n$ having a density given by (32), but the notions we will introduce hold in full generality. We would like to ask what is the probability of finding k eigenvalues around given positions in $\mathbb{R}$, to be expressed in terms of $\mathbb{P}_n$ and the corresponding expectation value $\mathbb{E}_n$. Because the eigenvalues are not labeled, we will have to carry combinatorial factors all the way through. Let μ be an (arbitrary) positive measure on $\mathbb{R}$ with finite total mass. We introduce:

7.1 definition. If it exists, the k**-point density correlation** with respect to μ is a function $\rho_{k|n} \in L^1(\mathbb{R}^k, \mu^{\otimes k})$ such that $\rho_{k|n} \geq 0$ is almost everywhere positive, $\rho_{k|n}$ is symmetric in its k variables, and for any bounded, continuous μ-almost everywhere, test function f on $\mathbb{R}^k$, symmetric in its k variables:

$$\mathbb{E}_n\Big[\sum_{\substack{1 \leq i_1, \ldots, i_k \leq n \\ \text{pairwise distinct}}} f(\lambda_{i_1}, \ldots, \lambda_{i_k})\Big] = \int_{\mathbb{R}^k} \rho_{k|n}(x_1, \ldots, x_k)\, f(x_1, \ldots, x_k) \prod_{i=1}^{k} \mathrm{d}\mu(x_i).$$

Clearly, $\rho_{k|n}$ is uniquely characterized by Definition 7.1. Lemma 7.2 below give a formula to compute it. By taking $f \equiv 1$, we observe the normalization:

$$(39)\quad \int_{\mathbb{R}^k} \rho_{k|n}(x_1, \ldots, x_k) \prod_{i=1}^{k} \mathrm{d}\mu(x_i) = \frac{n!}{(n-k)!} = \#\left\{ \begin{array}{l} \text{choosing } k \text{ labeled} \\ \text{elements among } n \end{array} \right\}$$

and in particular:

$$(40)\quad \int_{\mathbb{R}} \rho_{1|n}(x)\, \mathrm{d}\mu(x) = n, \qquad \int_{\mathbb{R}^n} \rho_{n|n}(x_1, \ldots, x_n) \prod_{i=1}^{n} \mathrm{d}\mu(x_i) = n!\,.$$

For this reason, $\rho_{k|n}$ is not the density of a probability measure on $\mathbb{R}^k$, in particular it is not the marginal of a probability measure on $\mathbb{R}^n$. However, it is a marginal up to a combinatorial factor:

7.2 lemma. *Assume that there exists an n-point density correlation $\rho_{n|n}$. Then:*

$$(41)\quad n! \cdot \mathbb{P}_n = \rho_{n|n} \cdot \mu^{\otimes n}$$

for any $k \in [\![1, n]\!]$, the k-point correlation density exists and is given by the formula:

$$\rho_{k|n}(x_1, \ldots, x_k) = \frac{1}{(n-k)!} \int_{\mathbb{R}^{n-k}} \rho_{n|n}(x_1, \ldots, x_n) \prod_{i=k+1}^{n} \mathrm{d}\mu(x_i)\,.$$

The normalization in this formula is compatible with (39)-(40).

Proof. The relation between $\mathbb{P}_n$ and $\rho_{n|n}$ follows from the definition. Let f be a function in k variables which is bounded and continuous μ-almost everywhere. We set:

$$F(x_1, \ldots, x_n) = \sum_{\substack{1 \leq i_1, \ldots, i_k \leq n \\ \text{pairwise distinct}}} f(x_{i_1}, \ldots, x_{i_k}).$$

Using the definition of $\rho_{n|n}$ and its symmetry, we can compute:

$$\begin{aligned}
& \mathbb{E}_n\Big[\sum_{\sigma \in \mathfrak{S}_n} F(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})\Big] \\
= & \int_{\mathbb{R}^n} F(x_1, \ldots, x_n)\,\rho_{n|n}(x_1, \ldots, x_n) \prod_{i=1}^{n} \mathrm{d}\mu(x_i) \\
= & \frac{n!}{(n-k)!} \int_{\mathbb{R}^k} f(x_1, \ldots, x_k) \Big(\int_{\mathbb{R}^{n-k}} \rho_{n|n}(x_1, \ldots, x_n) \prod_{i=k+1}^{n} \mathrm{d}\mu(x_i) \Big) \prod_{j=1}^{k} \mathrm{d}\mu(x_j).
\end{aligned}$$

On the other hand, if $\rho_{k|n}$ exists, we would have:

$$\mathbb{E}_n\Big[\sum_{\sigma \in \mathfrak{S}_n} F(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})\Big] = n! \int_{\mathbb{R}}^{k} f(x_1, \ldots, x_k)\,\rho_{k|n}(x_1, \ldots, x_k) \prod_{i=1}^{k} \mathrm{d}\mu(x_k)$$

The comparison of these two formulas gives the result. □

If $A_1, \ldots, A_k$ are disjoint measurable subsets, the probability that at least one eigenvalue is found in each A_i is evaluated by taking $f(x_1, \ldots, x_k) = \prod_{i=1}^{k} \mathbf{1}_{A_i}(x_i)$ in Definition 7.1:

$$\mathbb{P}_n\Big[\bigcap_{i=1}^{k} \{\mathrm{Sp}\, M_n \cap A_i \neq \emptyset\}\Big] = \int_{A_1} \mathrm{d}\mu(x_1) \cdots \int_{A_k} \mathrm{d}\mu(x_k)\,\rho_{k|n}(x_1, \ldots, x_k).$$

Other characterization and gap probabilities

If g is a continuous bounded function and $t \in \mathbb{R}$, the identity:

$$\begin{aligned}\prod_{i=1}^{n}(1+tg(\lambda_i)) &= \sum_{k=0}^{n} t^k \sum_{1\leq i_1<\ldots<i_k\leq n} \prod_{a=1}^{k} g(\lambda_{i_a}) \\ &= \sum_{k=0}^{n} \frac{t^k}{k!} \sum_{\substack{1\leq i_1,\ldots,i_k\leq n \\ \text{pairwise distinct}}} \prod_{a=1}^{k} g(\lambda_{i_a})\end{aligned}$$

translates into:

$$\mathbb{E}_n\Big[\prod_{i=1}^{n}(1+tg(\lambda_i))\Big] = \sum_{k=0}^{n}\frac{t^k}{k!}\int_{\mathbb{R}^k}\prod_{i=1}^{k}\rho_{k|n}(x_1,\ldots,x_k)\prod_{i=1}^{k}g(x_i)\mathrm{d}\mu(x_i)$$

7.3 LEMMA. *Formula 42 characterizes* $(\rho_{k|n})_{1\leq k\leq n}$.

Proof. Assume that for $k \in [\![1,n]\!]$ we have identified symmetric, non-negative functions of k-variables $\rho_{k|n} \geq 0$ such that (42) holds. Then, for any $g_1,\ldots,g_k$ continuous bounded functions, we can set $t = 1$ and choose $g = \sum_{i=1}^{k} t_i g_i$. Picking up the coefficient of $t_1\cdots t_k$ in both sides of the equation, we find:

$$\mathbb{E}\Big[\sum_{\substack{1\leq i_1,\ldots,i_k\leq n \\ \text{pairwise distinct}}}\prod_{a=1}^{k}g(\lambda_{i_a})\Big] = \int_{\mathbb{R}^k}\rho_{k|n}(x_1,\ldots,x_k)\prod_{i=1}^{k}\mathrm{d}\mu(x_i)\,.$$

Then, the span of functions of the $\prod_{i=1}^{k} g_i(x_i)$ is dense for the sup norm on any compact in the space continuous bounded functions of k variables. Therefore, $\rho_{k|n}$ can be identified with k-point density correlation. ☐

If A is a measurable subset of $\mathbb{R}$, choosing $g = \mathbf{1}_A$ and $t = -1$ gives:

$$\mathbb{G}_n(A) := \mathbb{E}_n\Big[\prod_{i=1}^{n}\mathbf{1}_{A^c}(\lambda_i)\Big] = \sum_{k=0}^{n}\frac{(-1)^k}{k!}\int_{A^k}\prod_{i=1}^{k}\rho_{k|n}(x_1,\ldots,x_k)\prod_{i=1}^{k}\mathrm{d}\mu(x_i)\,.$$

7.4 DEFINITION. $\mathbb{G}_n(A)$ is the probability that no eigenvalue is found in A, it is called the *gap probability*.

For instance:

$$\mathbb{P}_n\big[\lambda_{\max} \leq s\big] = \mathbb{G}_n\big([s,+\infty]\big).$$

7.2 *Kernel for beta = 2*

We put ourselves in the framework of the master integration formula for $\beta = 2$ (Lemma 6.1) and the use of orthogonal functions. Let $(\phi_i)_{i\geq 0}$ be a sequence of real-valued functions in $L^2(A,\mu)$, which are orthonormal for the scalar product:

$$\forall i,j \geq 1, \qquad \langle \phi_i, \phi_j \rangle_2 = \int_A \phi_i(x)\phi_j(x)\mathrm{d}\mu(x) = \delta_{i,j}\,.$$

We consider a model with the n-point density correlation is:

$$(42) \quad \rho^{(2)}_{n|n}(x_1,\ldots,x_n) = \det_{1\leq i,j\leq n} \phi_i(x_j) \cdot \det_{1\leq i,j\leq n} \psi_i(x_j)\,.$$

Using $\det A \cdot \det B = \det A^T B$, we can put the n-point density correlation in the form:

$$\rho^{(2)}_{n|n}(x_1,\ldots,x_n) = \det_{1\leq i,j\leq n} K^{(2)}_n(x_i,x_j), \qquad K^{(2)}_n(x,y) = \sum_{k=0}^{n-1} \phi_k(x)\psi_k(y)\,.$$

In this chapter, we will see that for $\beta = 1$ or 4, we can also find a kernel $K^{(\beta)}_n$ – although quaternionic – such that $\rho_{n|n}$ is an $n \times n$ quaternionic determinant of $K^{(\beta)}_n(x_i,x_j)$, and which is a quasi projector in the sense of Definition 5.8. The computation of all k-point densities and gap probabilities will follow from Corollary 5.10.

7.3 *Kernel for beta = 4*

We put ourselves in the framework of the master integration lemma for $\beta = 4$ (Lemma 6.5) and the use of skew-orthogonal functions. Namely, let $(\phi_i)_{i\geq 1}$ and $(\psi_i)_{i\geq 1}$ be two sequences of real or complex-valued function in $L^2(A,\mu)$. We denote $V \subseteq L^2(A,\mu)$ (resp. $W \subseteq L^2(A,\mu)$) be the Hilbert space generated by the ϕ's (resp. the ψ's). Let $\hat{u} : V \to W$ be the linear operator sending ϕ_i to ψ_i for any $i \geq 1$. We introduce on V the skew-symmetric bilinear form:

$$\langle f, g\rangle_4 = \int_{\mathbb{R}} \big[f(x)(\hat{u}g)(x) - g(x)(\hat{u}f)(x)\big]\mathrm{d}\mu(x).$$

We assume that ϕ's are skew-orthonormal in the sense that:

$$\forall i,j \geq 1, \qquad \left\{ \begin{array}{l} \langle \phi_{2i}, \phi_{2j}\rangle_4 = \langle \phi_{2i-1}, \phi_{2j-1}\rangle_4 = 0 \\ \langle \phi_{2i-1}, \phi_{2j}\rangle_4 = -\langle \phi_{2i}, \phi_{2j-1}\rangle_4 = \delta_{i,j} \end{array}\right. ,$$

and consider a model where the n-point density correlation is:

$$(43) \quad \rho^{(4)}_{n|n}(x_1,\ldots,x_n) = \det_{\substack{1\leq i\leq 2n \\ 1\leq j\leq n}} \big[\phi_i(x_j) \;\; \psi_i(x_j)\big]\,.$$

The $2n \times 2n$ matrix in the determinant – after simultaneous line and column permutations – can always be written $\Theta_n(A)$ for the quaternionic matrix $A = \big[\overline{\chi}_i(x_j)\big]_{ij}$ of size $n \times n$, with:

$$\Theta_1[\overline{\chi}_i(x)] = \begin{pmatrix} \psi_{2i}(x) & \psi_{2i-1}(x) \\ \phi_{2i}(x) & \phi_{2i-1}(x) \end{pmatrix}$$

and its quaternion dual:

$$\Theta_1[\chi_i(x)] = \begin{pmatrix} \phi_{2i-1}(x) & -\psi_{2i-1}(x) \\ -\phi_{2i}(x) & \psi_{2i}(x) \end{pmatrix}.$$

Then, using the properties of the quaternionic determinant (Lemma 5.5):

$$\rho^{(4)}_{n|n}(x_1, \ldots, x_n) = \det \Theta_n(A) = \det_{\mathbb{H}} A\overline{A} = \det_{\mathbb{H}}\,_{1 \leq i,j \leq n} K^{(4)}_n(x_i, x_j)\,, \tag{44}$$

with:

$$K^{(4)}_n(x, y) = \sum_{k=1}^{n} \overline{\chi}_k(x)\chi_k(y)\,.$$

7.5 LEMMA. *K is a projector of trace n.*

Proof. We compute:

$$\begin{aligned} \Theta_1(\overline{\chi}_k\chi_\ell) &= \begin{pmatrix} \psi_{2k}\phi_{2\ell-1} - \psi_{2k-1}\phi_{2\ell} & \psi_{2k-1}\psi_{2\ell} - \psi_{2k}\psi_{2\ell-1} \\ \phi_{2k}\phi_{2\ell-1} - \phi_{2k-1}\phi_{2\ell} & \phi_{2k-1}\psi_{2\ell} - \phi_{2k}\psi_{2\ell-1} \end{pmatrix} \\ \Theta_1(\chi_k\overline{\chi}_\ell) &= \begin{pmatrix} \phi_{2k-1}\psi_{2\ell} - \phi_{2k}\psi_{2\ell-1} & \phi_{2k-1}\psi_{2\ell-1} - \psi_{2k-1}\phi_{2\ell-1} \\ \phi_{2k}\psi_{2\ell} - \psi_{2k}\phi_{2\ell} & \psi_{2k}\phi_{2\ell-1} - \phi_{2k}\psi_{2\ell-1} \end{pmatrix}. \end{aligned}$$

Let us first look at the trace:

$$\mathrm{Tr}\, K^{(4)}_n = \sum_{k=1}^{n} \int_A \big(\overline{\chi}_k(z)\chi_k(z)\big)^{(1)}\, \mathrm{d}\mu(z).$$

From the expression of $\overline{\chi}_k(z)\chi_k(z)$, the off-diagonal terms are 0, while the integration of the diagonal terms against $\mathrm{d}\mu(z)$ reconstructs the skew-products: we find

$$\Theta_1\Big(\int_A \overline{\chi}_k(z)\chi_k(z)\mathrm{d}\mu(z)\Big) = \begin{pmatrix} \langle \phi_{2k-1}, \phi_{2k}\rangle_4 & 0 \\ 0 & \langle \phi_{2k-1}, \phi_{2k}\rangle_4 \end{pmatrix} = \mathbf{1}_2\,,$$

hence $\mathrm{Tr}\, K^{(4)}_n = n$. Then, we compute the kernel of the operator $K^{(4)}_n \circ K^{(4)}_n$:

$$(K^{(4)}_n \circ K^{(4)}_n)(x, y) = \sum_{k,\ell=1}^{n} \overline{\chi}_k(x)\Big(\int_A \chi_k(z)\overline{\chi}_\ell(z)\mathrm{d}\mu(z)\Big)\chi_\ell(y)\,.$$

By skew-orthogonality, the integral yields $\delta_{k,\ell}$ (times the identity quaternion), hence $K \circ K = K$. ☐

7.4 Kernel for beta = 1

Lecture 13 (1h30)
November 18th, 2014

Lecture 14 (1h30)
December 1st, 2014

We put ourselves in the framework of the master integration lemma for $\beta = 1$ (Lemma 6.10) and the use of skew-orthogonal functions. Namely, let $s : A \times A \to \mathbb{R}$ be an antisymmetric function, and $(\phi_i)_{i \geq 1}$ a sequence of real or complex-valued function in $L^2(A,\mu)$. We introduce:

$$(\hat{s}f)(y) = \int_A s(y,x)\, f(x)\mathrm{d}\mu(x)$$

and the skew-symmetric bilinear form:

$$\langle f,g\rangle_1 = \langle \hat{s}f,g\rangle_2 = \int_A s(y,x)\, f(x)g(y)\mathrm{d}\mu(x)\mathrm{d}\mu(y)\,.$$

We assume that ϕ's are skew-orthonormal:

$$\forall i,j \geq 1, \qquad \begin{cases} \langle \phi_{2i},\phi_{2j}\rangle_1 = \langle \phi_{2i-1},\phi_{2j-1}\rangle_1 = 0 \\ \langle \phi_{2i-1},\phi_{2j}\rangle_1 = -\langle \phi_{2i},\phi_{2j-1}\rangle_1 = \delta_{i,j} \end{cases},$$

and we consider a model where the n-point density correlation is:

$$\rho^{(1)}_{n|n}(x_1,\ldots,x_n) = \mathop{\mathrm{pf}}_{1\leq i,j\leq n} s(x_j,x_i)\cdot \det_{1\leq i,j\leq n} \phi_i(x_j)\,. \tag{45}$$

For simplicity, we will assume $n = 2m$ is even. The computations for n odd are slightly more involved, but also lead to the construction of a kernel which is a quasi-projector, and we shall give the result in the summary of § 7.5.

Let us define:

$$\Theta_1(\overline{\chi}_i(x)) = \begin{pmatrix} \phi_{2i}(x) & \phi_{2i-1}(x) \\ \hat{s}\phi_{2i}(x) & \hat{s}\phi_{2i-1}(x) \end{pmatrix}.$$

Its quaternion dual is:

$$\Theta_1(\chi_i(x)) = \begin{pmatrix} \hat{s}\phi_{2i-1}(x) & -\phi_{2i-1}(x) \\ -\hat{s}\phi_{2i}(x) & \phi_{2i}(x) \end{pmatrix}.$$

We introduce the kernel:

$$K^{(1)}_n(x,y) = \sum_{k=0}^{m-1} \overline{\chi}_k(x)\chi_k(y) - s(x,y)E_{21}\,,$$

where E_{ab} is the quaternion represented by the elementary 2×2 matrix with zero entries except for a 1 at position (a,b). $K^{(1)}_n$ is clearly self-dual and trace-class.

7.6 lemma. *Assume $\phi_i(x_j)$ invertible for $(x_1,\ldots,x_n)$ $\mu^{\otimes n}$-almost everywhere. Then:*

$$\rho_{n|n}(x_1,\ldots,x_n) = \mathop{\mathrm{det}_{\mathbb{H}}}_{1\leq i,j\leq n} K^{(1)}_n(x_i,x_j)\,,$$

Proof. Let us introduce the $n\times n$ matrices $A = (\phi_i(x_j))_{ij}$ and $S = (s(x_j - x_i))_{ij}$.

Using the properties of the pfaffian, in particular Lemma 5.2, we can write:

$$\mathrm{pf}\, S \cdot \det A = (-1)^m\, \mathrm{pf}(-S) \cdot \mathrm{pf}\, A J_N A^T = (-1)^m \underset{2n}{\mathrm{pf}}\, \tilde{A}\,,$$

with:

$$\tilde{A} := \begin{pmatrix} A & A\alpha^T \\ \alpha A & \alpha A \alpha^T - S \end{pmatrix},$$

and for any matrix $\alpha \in \mathscr{M}_n(\mathbb{C})$. Let P be the matrix of the permutation of $\mathfrak{S}_{2n}$ sending $2i-1$ to i, and $2i$ to $i+n$ for $i \in [\![1,n]\!]$. Its signature is $\det P = (-1)^m$. Therefore:

$$\mathrm{pf}\, S \cdot \det A = \det P \cdot \mathrm{pf}\, \tilde{B} = \mathrm{pf}\, P^T \tilde{A} P = \mathrm{det}_{\mathbb{H}} C, \qquad \Theta_n(C) = J_n^{-1} \tilde{A}\,.$$

The self-dual quaternionic matrix C of size $n \times n$ indeed takes the form $C_{ij} = \tilde{K}(x_i, x_j; \alpha)$ for a quaternionic-valued function $\tilde{K}$ depending on α. We now describe a clever choice of α, for which skew-orthogonality of the ϕ's will translate for the kernel into the property of being a quasi-projector. For this purpose, we want that ϕ_i and $\hat{s}\phi_i$ both appear in $\tilde{K}$. If we define the $n \times n$ matrix $B = [\hat{s}\phi_i(x_j)]_{ij}$, we choose:

$$\alpha = -(A^{-1}B)^T.$$

This is well-defined for $(x_1, \ldots, x_n)$ $\mu^{\otimes n}$-almost everywhere since our assumption implies almost-everywhere invertibility of A. Tracking back all transformations from A to C, we find with this choice:

$$C_{ij} = K_n^{(1)}(x_i, x_j)$$

with the announced expression. □

7.7 lemma. *$K_n^{(1)}$ is a quasi-projector of trace n:*

$$K_n^{(1)} \circ K_n^{(1)} = K_n^{(1)} + [\lambda, K_n^{(1)}], \qquad \Theta_1(\lambda) = \begin{pmatrix} 1/2 & 0 \\ -1/2 & 0 \end{pmatrix}.$$

Proof. We compute the products of quaternionic functions:

$$\begin{aligned} \Theta_1(\overline{\chi}_k \chi_\ell) &= \begin{pmatrix} \phi_{2k} \cdot \hat{s}\phi_{2\ell-1} - \phi_{2k-1} \cdot \hat{s}\phi_{2\ell} & \phi_{2k-1} \cdot \phi_{2\ell} - \phi_{2k} \cdot \phi_{2\ell-1} \\ \hat{s}\phi_{2k} \cdot \hat{s}\phi_{2\ell-1} - \hat{s}\phi_{2k-1} \hat{s}\phi_{2\ell} & \hat{s}\phi_{2k-1} \cdot \phi_{2\ell} - \hat{s}\phi_{2k} \cdot \phi_{2\ell-1} \end{pmatrix} \\ \Theta_1(\chi_k \overline{\chi}_\ell) &= \begin{pmatrix} \hat{s}\phi_{2k-1} \cdot \phi_{2\ell} - \phi_{2k-1} \cdot \hat{s}\phi_{2\ell} & \hat{s}\phi_{2k-1} \cdot \phi_{2\ell-1} - \phi_{2k-1} \cdot \hat{s}\phi_{2\ell-1} \\ \phi_{2k} \cdot \hat{s}\phi_{2\ell} - \hat{s}\phi_{2k} \cdot \phi_{2\ell} & \phi_{2k} \cdot \hat{s}\phi_{2\ell-1} - \hat{s}\phi_{2k} \cdot \phi_{2\ell-1} \end{pmatrix}. \end{aligned}$$

Integrating the first equality over $\mathrm{d}\mu(z)$ and taking the scalar part gives:

$$\mathrm{Tr}\, K_n^{(1)} = \sum_{k=1}^{n/2} \int_A \mathrm{d}\mu(z) \big(\overline{\chi}_k(z) \chi_k(z)\big)^{(1)} = 2\langle \phi_{2k-1}, \phi_{2k} \rangle_1 = n.$$

Let us compute the kernel of $K_n^{(1)} \circ K_n^{(1)}$:

$$(K_n^{(1)} \circ K_n^{(1)})(x,y) = \sum_{k,\ell=1}^{n/2} \overline{\chi}_k(x)\Big(\int_A \chi_k(z)\overline{\chi}_\ell(z)\mathrm{d}\mu(z)\Big)\chi_\ell(y) \tag{46}$$
$$= -\sum_{\ell=1}^{n/2}\Big(\int_A s(x,z)E_{21}\overline{\chi}_\ell(z)\mathrm{d}\mu(z)\Big)\chi_\ell(y)$$
$$= -\sum_{k=1}^{n/2}\overline{\chi}_k(x)\Big(\int_A s(z,y)\,\chi_k(z)E_{21}\mathrm{d}\mu(z)\Big),$$

where we have introduce the quaternion E_{ab} represented by the elementary 2×2 matrix with zero entries except for a 1 in position (a,b). The integral in the first term is equal to the scalar quaternion $2\delta_{k,l}$ by skew-orthogonality of the ϕ's. We compute:

$$\Theta_1\Big(\int_A s(x,z)\,E_{21}\overline{\chi}_\ell(z)\mathrm{d}\mu(z)\Big) = \begin{pmatrix} 0 & 0 \\ \hat{s}\phi_{2\ell}(x) & \hat{s}\phi_{2\ell-1}(x)\end{pmatrix} = \Theta_1(E_{22}\overline{\chi}_\ell(x))$$
$$\Theta_1\Big(\int_A s(z,y)E_{21}\chi_k(z)\mathrm{d}\mu(z)\Big) = \begin{pmatrix} \hat{s}\phi_{2k-1}(y) & 0 \\ -\hat{s}\phi_{2k}(y) & 0\end{pmatrix} = \Theta_1(\chi_k(y)E_{11}).$$

Therefore:

$$(K_n^{(1)} \circ K_n^{(1)})(x,y) = 2A - E_{22}A - E_{11}, \qquad A = \sum_{k=1}^{n}\overline{\chi}_k(x)\chi_k(y).$$

We also remark that:

$$2E_{21} - E_{22}E_{21} - E_{21}E_{11} = 0.$$

So we can reconstitute $\tilde{K}_n^{(1)}(x,y) = A - s(x,y)E_{21} = \tilde{A}$ in the previous equation. And we conclude noticing that:

$$2\tilde{A} - E_{22}\tilde{A} - E_{21}\tilde{A} = \tilde{A} + \Big[\frac{E_{11}-E_{22}}{2}, \tilde{A}\Big].$$

□

In this case, we could construct a quasi-projector, but not a projector.

7.5 *Summary: kernels and consequences*

We have seen that the n-point density in the models (42), (43) and (45) can be put in the form:

$$\rho_{n|n}^{(\beta)}(x_1,\ldots,x_n) = \det_{\substack{\mathbb{H}\\ 1\leq i,j\leq n}} K_n^{(\beta)}(x_i,x_j).$$

The kernel $K_n^{(\beta)}$ is always a quasi-projector of trace n, which is quaternionic for $\beta = 1,4$, but scalar for $\beta = 2$. Corollary 5.10 yields:

7.8 proposition. *The k-point density correlation in those models reads:*

$$\rho_{k|n}^{(\beta)}(x_1,\ldots,x_n) = \underset{1\leq i,j\leq k}{\mathrm{det}_{\mathbb{H}}}\, K_n^{(\beta)}(x_i,x_j),$$

and the gap probabilities are:

$$\mathbb{G}_n^{(\beta)}(J) = \mathrm{Det}_{\mathbb{H}}(I - K_n^{(\beta)})_{L^2(\mu,J)}\,.$$

We stress that the kernel depends on n, and that $\rho_{k|n}$ is a $k\times k$ quaternionic determinant: for $\beta = 2$, this is a $k\times k$ determinant, and for $\beta = 1,4$, this is a $2k\times 2k$ pfaffian. Therefore, if we understand the $n\to\infty$ limit of $K_n^{(\beta)}$, we understand the limit $n\to\infty$ of all density correlations and of the gap probabilities.

Let us recollect the expressions for the kernels:

Case beta = 2

$$K_n^{(2)}(x,y) = \sum_{k=1}^{n} \phi_k(x)\psi_k(y).$$

In the case of unitary invariant ensembles, we have with respect to the Lebesgue measure:

$$\rho_{n|n}^{(2)}(x_1,\ldots,x_n) = \frac{1}{\prod_{i=0}^{n-1} h_i}\big(\Delta(x_1,\ldots,x_n)\big)^2\prod_{i=1}^{n} w(x_i),$$

and we have seen in § 6.1 that it can be put in the form (7.5) with:

$$\phi_{i+1}(x) = \psi_{i+1}(x) = \sqrt{\frac{w(x)}{h_i}}\,p_i(x).$$

Here, $(p_i)_{i\geq 0}$ are the orthogonal polynomials associated to the measure $w(x)\mathrm{d}x$. Then, $K_n^{(2)}$ is the **Christoffel-Darboux kernel** and can be readily computed:

7.9 lemma. *For any $n\geq 1$, we have the Christoffel-Darboux formula:*

$$\begin{aligned} K_n^{(2)}(x,y) &:= w^{1/2}(x)w^{1/2}(y)\sum_{k=0}^{n-1}\frac{p_k(x)p_k(y)}{h_k} \\ &= \frac{w^{1/2}(x)w^{1/2}(y)}{h_{n-1}}\,\frac{p_n(x)p_{n-1}(y)-p_{n-1}(x)p_n(y)}{x-y} \end{aligned}$$

The second equality is very useful, since it only requires the knowledge of two consecutive orthogonal polynomials to compute the kernel. So, if one wishes to study the limit $n\to\infty$ of $K_n^{(2)}$, it is enough to study the limit of $p_n(x)$. In general, there is no simple formula for p_n, so computing this asymptotics is a difficult problem. It can be attacked with the so-called **Riemann-Hilbert methods**, see e.g. the book of Deift for an introduction.

Case $\beta = 4$

The kernel:

$$(47)\quad \Theta_1\big(K_n^{(4)}(x,y)\big) = \begin{pmatrix} \hat{u}_x L_n(x,y) & -\hat{u}_x\hat{u}_y L_n(x,y) \\ L_n(x,y) & -\hat{u}_y L_n(x,y) \end{pmatrix},$$

can be deduced from the scalar function:

$$L_n(x,y) = \sum_{k=1}^{n} \phi_{2k}(x)\phi_{2k-1}(y) - \phi_{2k-1}(x)\phi_{2k}(y).$$

The operator $\hat{u}_x$ replaces $\phi_i(x)$ by $\psi_i(x)$.

In the case of quaternionic unitary invariant ensembles, we have with respect to Lebesgue measure:

$$\rho_{n|n}^{(4)}(x_1,\ldots,x_n) = \frac{1}{\prod_{i=0}^{n-1} h_i}\big(\Delta(x_1,\ldots,x_n)\big)^4 \prod_{i=1}^{n} w(x_i).$$

We have seen in § 6.2 that it can be brought to the form (7.5), with:

$$\phi_{i+1}(x) = \sqrt{\frac{w(x)}{h_i}}\, p_i(x), \qquad \psi_{i+1}(x) = \sqrt{\frac{w(x)}{h_i}}\, p_i'(x).$$

where p_i are the skew-orthogonal polynomials associated to:

$$\begin{aligned} \langle f, g\rangle_4 &= \int_{\mathbb{R}} \big(f(x)g'(x) - f'(x)g(x)\big)w(x)\mathrm{d}x \\ &= \int_{\mathbb{R}} \big(f(x)\cdot\hat{u}g(x) - \hat{u}f(x)\cdot g(x)\big)w(x)\mathrm{d}x \\ &= \int_{\mathbb{R}} \big(f(x)\cdot\hat{D}g(x) - \hat{D}f(x)\cdot g(x)\big)w(x)\mathrm{d}x. \end{aligned}$$

where we have introduced the operator $\hat{D} = w^{-1}\hat{u}w$:

$$(48)\quad \hat{D}f(x) = w^{-1/2}(x)\partial_x\big\{w^{1/2}(x)f(x)\big\}.$$

Because of the equality (48), following the derivation (44), we see that we can equally replace $\hat{u}$ by $\hat{D}$ to define our kernel. Unlike the case of orthogonal polynomials and the Christoffel-Darboux formula, there is no simple and general expression of $K_n^{(4)}$ in terms of a few consecutive skew-orthogonal polynomials.

Case beta = 1

When n even, the kernel takes the form:

$$(49)\quad \Theta_1\big(K_n^{(1)}(x,y)\big) = \begin{pmatrix} \hat{s}_y R_n(x,y) & -R_n(x,y) \\ \hat{s}_x\hat{s}_y R_n(x,y) - s(x,y) & -\hat{s}_x R_n(x,y) \end{pmatrix}.$$

It be deduced from the scalar function:

$$R_n(x,y) = \sum_{k=1}^{\lfloor n/2 \rfloor} \phi_{2k}(x)\phi_{2k-1}(y) - \phi_{2k-1}(x)\phi_{2k}(y).$$

The operator $\hat{s}_x$ replaces $f(x)$ by:

$$\hat{s}f(x) = \int_A s(x,z)f(z)\,\mathrm{d}\mu(z).$$

When n is odd, the kernel contains an additional term:

$$\Theta_1\big(K^{(1)}_{n,\mathrm{odd}}(x,y)\big) = \begin{pmatrix} \phi_{n+1}(x) & 0 \\ (\hat{s}\phi_{n+1})(y) - (\hat{s}\phi_{n+1})(x) & \phi_{n+1}(y) \end{pmatrix},$$

where ϕ_{n+1} completes the skew-orthonormal family $\phi_1, \ldots, \phi_n$, i.e.

$$\forall k \in [\![1,n]\!], \qquad \langle \phi_k, \phi_{n+1} \rangle_1 = \delta_{k,n},$$

and must be chosen such that:

$$\int_A \phi_{n+1}(x)\mathrm{d}\mu(x) = 1.$$

In the case of orthogonal invariant ensembles, we have with respect to Lebesgue measure:

$$\rho^{(1)}_{n|n}(x_1,\ldots,x_n) = \frac{1}{h_{n,\mathrm{odd}} \cdot \prod_{i=0}^{n/2-1} h_i} |\Delta(x_1,\ldots,x_n)| \prod_{i=1}^n w(x_i),$$

where the factor $h_{n,\mathrm{odd}}$ only occurs when n is odd. We have seen in § 6.3 that it can be written in the form (45) with $s(x,y) = \mathrm{sgn}(x-y)$ and:

$$\forall i \in [\![0,n-1]\!], \qquad \phi_{i+1}(x) = \sqrt{\frac{w(x)}{h_{\lfloor i/2 \rfloor}}}\, p_i(x).$$

These expressions involve the skew-orthogonal polynomials $(p_i)_{i\geq 0}$ for the skew-bilinear form:

$$\langle f,g \rangle_4 = \int_{\mathbb{R}^2} \mathrm{sgn}(y-x)\, f(x)g(y)\mathrm{d}\mu(x)\mathrm{d}\mu(y).$$

When n is odd, we also have:

$$\phi_n(x) = \frac{\sqrt{w(x)}}{h_{n,\mathrm{odd}}}\, p_n(x), \qquad h_{n,\mathrm{odd}} = \int_{\mathbb{R}} p_{2n}(z)\, w(z)\mathrm{d}z.$$

The operator $\hat{s}$ is:

$$\hat{s}f(x) = \int_{\mathbb{R}} \operatorname{sgn}(x-z)\, f(z) w(z) \mathrm{d}z = 2 \int_{-\infty}^{x} f(z)\, w(z) \mathrm{d}z - \int_{\mathbb{R}} f(z)\, w(z) \mathrm{d}z,$$

therefore:

$$\partial_x \hat{s}_x f = f(x) w(x).$$

In other words, $\hat{s}_x$ is a right inverse for $w(x)^{-1}\partial_x$. As for the quaternionic unitary case and unlike the unitary case, there is no simple and general expression for $K_n^{(1)}$ in terms of a few consecutive skew-orthogonal polynomials.

8 Gaussian invariant ensembles

Lecture 15 (1h30)
December 2nd, 2014

In this chapter, we focus on the Gaussian ensemble of random matrices $M_n \in \mathscr{H}_{n,\beta}$. These are invariant ensembles for which the n-point density correlation of the eigenvalues – with respect to Lebesgue measure – is:

$$\rho_{n|n}^{(\beta)}(x_1, \ldots, x_n) \propto \big|\Delta(x_1, \ldots, x_n)\big|^{\beta} \prod_{i=1}^{n} w_\beta(x_i),$$

and we choose the normalisation:

$$w_2(x) = w_4(x) := w(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad w_1(x) = w^{1/2}(x) = \frac{e^{-x^2/4}}{(2\pi)^{1/4}}.$$

The Gaussian ensembles form of the few instances where the (skew)-orthogonal polynomials can be computed in a simple way. We then study the kernels, and their $n \to \infty$ limit.

8.1 *Hermite polynomials*

The orthogonal polynomials for the measure:

$$\mathrm{d}\nu(x) = w(x)\mathrm{d}x, \qquad w(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

are called **Hermite polynomials**. We denote them $(H_k(x))_{k \geq 0}$. Let us summarize their properties:

8.1 LEMMA. *For any $n \geq 0$, with the convention $H_{-1} = 0$:*

- *H_n has parity $(-1)^n$.*
- *$H_n'(x) = nH_{n-1}(x)$.*
- *The norm of H_n is $h_n = n!$.*
- *$H_0 = 1$, and $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$.*
- *$H_n''(x) - xH_n'(x) + nH_n(x) = 0$.*
- *We have the formula $H_n(x) = (-1)^n e^{x^2/2}\, \partial_x^n\big(e^{-x^2/2}\big)$.*

Proof. Since μ is even, H_n must have parity $(-1)^n$. H_n' is a degree $n-1$ polynomial, so decomposes on the basis $(H_k)_{0 \leq k \leq n-1}$. Integration by parts yields:

$$(H_n'|H_m) + (H_n|H_m') - (H_n|xH_m) = 0.$$

For $m \leq n-2$, the two last term are 0 by orthogonality. For $m = n-1$, the second is zero, while the third term is $(H_n|xH_{n-1}) = (H_n|H_n + \cdots)$, where the $\cdots$ represent a polynomial of degree $\leq n-1$. By orthogonality,

$(H_n|xH_{n-1}) = h_n$. Thus:

$$H'_n(x) = \frac{h_n}{h_{n-1}} H_{n-1}(x).$$

and comparing the coefficient of the leading term, we find $h_n = nh_{n-1}$. Since $H_0 = 1$ and the measure μ has total mass 1, we have $h_0 = 1$ and by recursion:

$$H'_n(x) = nH_{n-1}, \qquad h_n = n!\,.$$

The 3-term recurrence relation is then given by Lemma 9.1 in next Chapter, using the fact that $\beta_n = (xH_n|H_n) = 0$ for parity reasons. Using $H'_n = nH_{n-1}$ then translates the 3-term recurrence relation into a differential equation of order 2 for H_n. It is straightforward to check that the given expression for H_n is the unique solution the 3-term recurrence relation with initial condition $H_0(x) = 1$ and $H_1(x) = x$. $\square$

It is convenient to define the **Hermite wave functions**:

$$\forall n \geq 0, \qquad \varphi_n(x) = w^{1/2}(x)\,\frac{H_n(x)}{\sqrt{h_n}} = \frac{e^{-x^2/4}H_n(x)}{(2\pi)^{1/4}\sqrt{n!}},$$

which form an orthonormal Hilbert basis on $L^2(\mathbb{R})$ equipped with the Lebesgue measure:

$$\int_{\mathbb{R}} \varphi_n(x)\varphi_m(x)\,\mathrm{d}x = 0.$$

delta_{n,m}

Their properties can be directly established from Lemma 8.1.

8.2 lemma. *For any $n \geq 0$:*

- $x\varphi_n(x) = \sqrt{n+1}\varphi_{n+1}(x) + \sqrt{n-1}\varphi_{n-1}(x).$
- $(-\partial_x^2 + x^2/4)\varphi_n(x) = (n+1/2)\varphi_n(x).$
- *We have the integral representation:*

$$\varphi_n(x) = \frac{\mathrm{i}^n\, e^{x^2/4}}{(2\pi)^{1/4}\sqrt{n!}} \int_{\mathbb{R}} \mathrm{d}z\, z^n\, e^{-z^2/2-\mathrm{i}xz}\,.$$

Remark. As we observe on the differential equation, $\varphi_n(x)$ are the eigenvectors of the self-adjoint operator $\mathcal{H} = -\partial_x^2 + x^2/4$ on $L^2(\mathbb{R})$, and the corresponding eigenvalues are $(n+1/2)$, indexed by $n \geq 0$. In quantum mechanics, $\mathcal{H}$ is the hamiltonian of the harmonic oscillator. It turns out that, after separation of the angular variables in the hamiltonian of the hydrogen atom, one is left with an hamiltonian of the form $\mathcal{H}$ governing the radial part of the electron wave function, so the Hermite wave functions also appear in this problem.

8.2 *Moments of Gaussian hermitian matrices*

We focus on $\beta = 2$. The Christoffel-Darboux kernel is:

$$K_n^{(2)}(x,y) = \frac{e^{-(x^2+y^2)/4}}{\sqrt{2\pi}}\,\frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(n-1)!\,(x-y)}\,.$$

It can be written in terms of the Hermite wave functions:

$$K_n^{(2)}(x,y) = \sqrt{n}\,\frac{\varphi_n(x)\varphi_{n-1}(y) - \varphi_n(x)\varphi_{n-1}(y)}{x-y}\,.$$

The generating series of moments of the eigenvalues can be evaluated explicitly.

8.3 PROPOSITION. *For any $n \geq 1$, we have:*

$$\frac{1}{n}\mathbb{E}_n\big[\mathrm{Tr}\,e^{sM}\big] = e^{s^2/2}\Big(\sum_{k=0}^{n-1}\mathrm{Cat}(k)\,\frac{n(n-1)\cdots(n-k)}{n^k}\,\frac{(s\sqrt{n})^{2k}}{(2k)!}\Big).$$

Proof. Denote $A_n(s)$ the right-hand side. We have by definition of the 1-point density correlation (see Definition 7.1):

$$A_n(s) = \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} e^{s\lambda_i}\Big] = \frac{1}{n}\int_{\mathbb{R}}\rho_{1|n}^{(2)}(x)\,e^{sx}\mathrm{d}x\,. \tag{50}$$

Although $f(x) = e^{sx}$ is not bounded, the integral is absolutely convergent due to the gaussian tails of $\rho_{1|n}^{(2)}(x)$, so we can use the approximation $f_m(x) = \min(e^{sx}, m)$ which is bounded continuous for any m, and converges pointwise to $f(x) = e^{sx}$, and dominated convergence to write (50). The general formula in $\beta = 2$ invariant ensembles gives:

$$\rho_{1|n}^{(2)} = K_n^{(2)}(x,x) = \sqrt{n}\big(\varphi_n'(x)\varphi_{n-1}(x) - \varphi_{n-1}'(x)\varphi_n(x)\big).$$

With integration by parts:

$$A_n(s) = \frac{1}{s\sqrt{n}}\int_{\mathbb{R}}\mathrm{d}x e^{sx}\big(-\varphi_n''(x)\varphi_{n-1}(x) + \varphi_{n-1}''(x), \varphi_n(x)\big)$$

and using the differential equation for Hermite wave functions:

$$\begin{aligned} A_n(s) &= \frac{1}{s\sqrt{n}}\int_{\mathbb{R}} e^{sx}\varphi_n(x)\varphi_{n-1}(x) = \frac{1}{s\,n!}\int_{\mathbb{R}} e^{-x^2/2+sx}\,H_n(x)H_{n-1}(x)\,\frac{\mathrm{d}x}{\sqrt{2\pi}} \\ &= \frac{e^{s^2/2}}{s\,n!}\int_{\mathbb{R}} H_n(x+s)H_{n-1}(x+s)\,w(x)\mathrm{d}x\,. \end{aligned}$$

Using the explicit formula for the Hermite polynomials, it is easy to show:

$$H_n(x+s) = \sum_{k=0}^{n} \binom{n}{k} H_{n-k}(x)\, s^k \,.$$

Therefore, we can compute $A_n(s)$ using orthogonality of the Hermite polynomials with respect to $w(x)\mathrm{d}x$:

$$A_n(s) = \frac{e^{s^2/2}}{s\, n!} \Big\{ \sum_{k=1}^{n} \binom{n}{k} \binom{n-1}{k-1} s^{2k-1}\,(n-k)! \Big\},$$

which can be rewritten in the announced form. □

It is easy to take the large n limit of this formula, and we deduce a proof of Wigner theorem in the case of hermitian Gaussian random matrices. Remind the definition of the empirical measure:

$$L^{(M_n)} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i^{(M_n)}}\,,$$

where $\lambda_i := \lambda_i^{(M_n)}$ are the eigenvalues of A.

8.4 corollary. *$L^{(\tilde{M}_n)}$ converges weakly in expectation to the semi-circle law.*

Proof. Let $\tilde{M}_n = n^{-1/2}M_n$. The Fourier transform of the empirical measure of $\tilde{M}_n$ is:

$$\mathbb{E}_n\Big[\int e^{\mathrm{i}sx}\mathrm{d}L^{(\tilde{M}_n)}(x)\Big] = \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{n} e^{s\lambda_i/\sqrt{n}}\Big] = A_n(s/\sqrt{n})\,.$$

From Lemma 8.3, we deduce:

$$\lim_{n\to\infty} A_n(s/\sqrt{n}) = \sum_{k=0}^{\infty} \mathrm{Cat}(k)\,\frac{s^{2k}}{(2k)!}\,,\qquad \mathrm{Cat}(k) = \frac{1}{k+1}\binom{2k}{k+1}\,.$$

We recognize in the coefficients the moments of the semi-circle law (Lemma 3.2):

$$\lim_{n\to\infty} A_n(s/\sqrt{n}) = \int_{\mathbb{R}} e^{\mathrm{i}sx}\mathrm{d}\mu_{\mathrm{sc}}(x),\qquad \mathrm{d}\mu_{\mathrm{sc}}(x) = \frac{\mathrm{d}x}{2\pi}\sqrt{4-x^2}\cdot \mathbf{1}_{[-2,2]}(x)\,.$$

Pointwise convergence of the expectation value of the Fourier transform of $L^{(\tilde{M}_n)}$ to that of μ_{sc} implies weak convergence in expectation of $L^{(\tilde{M}_n)}$ towards μ_{sc}. □

This convergence, as we already pointed out in § 3.7, does not give the convergence in probability $n^{-1/2}\lambda_{\max} \to 2$. But the exact evaluation of moments provides us much more precise information that Wigner's theorem:

8.5 proposition (Ledoux bound). *There exist constants $c_1, c_2 > 0$ independent of* *n such that, for any $\eta > 0$,*

$$\mathbb{P}_n\Big[\frac{\lambda_{\max}}{2\sqrt{n}} \geq e^{n^{-2/3}t}\Big] \leq c_1 t^{-3/4}\, e^{-c_2 t^{3/2}}\,.$$

In particular, $n^{-1/2}\lambda_{\max} \to 2$ in probability when $n \to \infty$.

This suggests that $\lambda_{\max} = \sqrt{n}\big[2 + O(n^{-2/3})\big] = 2\sqrt{n} + O(n^{-1/6})$ where the $O(\cdots)$ represent fluctuations. We shall see later that this is indeed the correct scale, and we shall characterise the distribution of these fluctuations.

Proof. Since we already know the k-th moments converge to $\mathrm{Cat}(k)$, let us decompose:

$$\frac{1}{n}\,\mathbb{E}\big[\,\mathrm{Tr}\,\tilde{M}_n^{2k}\big] = \mathrm{Cat}(k)\, m_n(k)\,.$$

We have by construction of the exponential generating series:

$$A_n(s/\sqrt{n}) = \sum_{k=0}^{\infty} \mathrm{Cat}(k)\, m_n(k)\, \frac{s^{2k}}{(2k)!}\,.$$

and the $m_n(k)$ are positive and be explicitly by collecting the powers of s^2 in Lemma 8.3. One finds – this is an exercise left to the reader – that they satisfy the recursion:

$$m_n(k+1) = m_n(k) + \frac{k(k+1)}{4n^2} m_n(k-1).$$

In particular, we deduce that $m_n(k)$ is increasing, and:

$$m_n(k) \leq m_n(k+1) \leq m_n(k)\Big(1 + \frac{k(k+1)}{4n^2}\Big),$$

and thus:

$$\begin{aligned} m_n(k+1) \quad & \leq \quad \prod_{\ell=0}^{k}\Big(1 + \frac{\ell(\ell+1)}{4n^2}\Big) \\ & \leq \quad \exp\Big\{\sum_{\ell=0}^{k} \frac{\ell(\ell+1)}{4n^2}\Big\} = \exp\Big(\frac{k(k+1)(k+2)}{12n^2}\Big) \leq e^{c_0 k^3/n^2} \end{aligned}$$

for some constant $c_0 > 0$, e.g. $c_0 = 1/2$. Now, we can use Markov inequality to write, for any $k \geq 1$:

$$\begin{aligned} \mathbb{P}_n\Big[\frac{\lambda_{\max}}{2\sqrt{n}} \geq e^u\Big] \quad & \leq \quad e^{-2ku}\,\mathbb{E}_n\big[(\lambda_{\max}/2\sqrt{n})^{2k}\big] \leq n e^{-2ku} 4^{-k} \cdot \frac{1}{n}\,\mathbb{E}\big[\,\mathrm{Tr}\,\tilde{M}_n^{2k}\big] \\ & \leq \quad n e^{-2ku}\, 4^{-k}\mathrm{Cat}(k)\, e^{c_0 k^3/n^2}. \end{aligned}$$

When k is large, we have $4^{-k}\mathrm{Cat}(k) \sim (\pi k)^{-1/2}$. If n is large, we see that the bound remains non trivial in the regime where $k = \kappa n^{2/3}$ and $u = n^{-2/3}t$ with

t and κ finite. There exists $c > 0$ such that:

$$\mathbb{P}_n\Big[\frac{\lambda_{\max}}{2\sqrt{n}} \geq e^{n^{-2/3}t}\Big] \leq c\,\kappa^{-3/2}\,e^{-2\kappa t + c_0\kappa^3}.$$

We can choose for instance to optimize the term in the exponential, by choosing $\kappa := n^{-2/3}\lfloor n^{2/3}t^{1/2}\rfloor \sim t$ when $n \to \infty$. Thus, we obtain an upper bound of the form:

$$\mathbb{P}_n\Big[\frac{\lambda_{\max}}{2\sqrt{n}} \geq e^{n^{-2/3}t}\Big] \leq c_1\,t^{-3/4}\,e^{-c_2 t^{3/2}}$$

for constants $c_1, c_2 > 0$ independent of n. It remains the justify the convergence in probability. If $\epsilon > 0$, we take $t_n(\epsilon) = n^{2/3}\ln(1+\epsilon)$ in the previous bound. Since $t_n(\epsilon) \to +\infty$, the right-hand side goes to 0 when $n \to \infty$ and we have:

$$\lim_{n\to\infty} \mathbb{P}_n\Big[\frac{\lambda_{\max}}{2\sqrt{n}} \geq 1+\epsilon\Big] = 0.$$

For the lower bound, we can repeat the argument of Lemma 3.22 using the version of Wigner theorem we proved here in Lemma 8.4. ☐

8.3 *Skew-orthogonal polynomials for beta = 4*

We shall construct the skew-orthogonal polynomials for:

$$\langle f, g\rangle_4 = \int_{\mathbb{R}} \big(f(x)g'(x) - f'(x)g(x)\big)w(x)\mathrm{d}x, \qquad w(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

8.6 lemma. *Let $\tilde{H}_{2n+1} = H_{2n+1}$, and define:*

$$\tilde{H}_{2n} = \sum_{k=0}^{n} \frac{n!}{k!}\,2^{n-k}\,H_{2k}.$$

$(\tilde{H}_n)_{n\geq 0}$ is a family of skew-orthogonal polynomials, with $\langle \tilde{H}_{2n}, \tilde{H}_{2n+1}\rangle_4 = \tilde{h}_n = (2n+1)!$.

Proof. For parity reasons, we have $\langle H_{2n}, H_{2m}\rangle_4 = \langle H_{2n+1}, H_{2m+1}\rangle_4 = 0$. And, using $H_n' = nH_{n-1}$, we compute:

$$\begin{aligned}\langle H_{2n}, H_{2m+1}\rangle_4 &= (2m+1)\langle H_{2n}, H_{2m}\rangle_2 - 2n\langle H_{2n-1}, H_{2m+1}\rangle_2 \\ &= (2n+1)!\delta_{n,m} - (2n)!\delta_{n,m+1}.\end{aligned}$$

Using the definition of $\tilde{H}$'s, we compute:

$$\begin{aligned}\langle \tilde{H}_{2n}, \tilde{H}_{2m+1}\rangle_4 &= \sum_{k=0}^{n} \frac{n!}{k!}\,2^{n-k}\langle H_{2k}, H_{2m+1}\rangle_4 \\ &= \sum_{k=0}^{n} \frac{n!}{k!}\,2^{n-k}\big(\delta_{k,m}(2k+1)! - (2k)!\delta_{k,m+1}\big).\end{aligned}$$

It vanishes for $m > n$. For $n = m$, we find:

$$\tilde{h}_n := \langle \tilde{H}_{2n}, \tilde{H}_{2n+1} \rangle_4 = (2n+1)!$$

while for $m < n$:

$$\langle \tilde{H}_{2n}, \tilde{H}_{2m+1} \rangle_4 = n!\, 2^n \Big(\frac{(2m+1)!}{m! 2^m} - \frac{(2m+2)!}{(m+1)! 2^{m+1}} \Big) = 0.$$

□

It will be convenient to have an alternative representation for $\tilde{H}_{2n}$:

8.7 lemma. *For any $n \geq 0$:*

$$\tilde{H}_{2n} = \frac{e^{x^2/4}}{2} \int_x^\infty e^{-y^2/4}\, H_{2n+1}(y).$$

Proof. Let us define the two integrals:

$$I_n(x) := \frac{e^{x^2/4}}{2} \int_x^\infty e^{-y^2/4}\, H_{2n+1}(y)\, \mathrm{d}y, \qquad J_n(x) := \frac{e^{x^2/4}}{2} \int_x^\infty e^{-y^2/4}\, y H_{2n}(y)\, \mathrm{d}y.$$

We have by integration by parts:

$$J_n(x) = e^{x^2/4} \big[-e^{-y^2/4} H_{2n} \big]_x^\infty + e^{x^2/4} \int_x^\infty e^{-y^2/4}\, H'_{2n}(y)\, \mathrm{d}y = H_{2n}(x) + 4n\, I_{n-1}(x),$$

using $H'_{2n} = 2n H_{2n-1}$. And, using the 3-term recurrence relation for Hermite polynomials:

$$I_n(x) = J_n(x) - 2n\, I_{n-1}(x).$$

Combining these two relations gives:

$$I_n(x) = H_{2n}(x) + 2n\, I_{n-1}(x),$$

and we have the initial condition:

$$I_0(x) = \frac{e^{x^2/4}}{2} \int_x^\infty e^{-y^2/4} y\, \mathrm{d}y = 1.$$

This is the same initial condition and recurrence relation that is satisfied by $\tilde{H}_{2n}$, thus $I_n = \tilde{H}_{2n}$ for any $n \geq 0$. □

We can now address the computation of the kernel $K_n^{(4)}$. From § 7.5, we learned that its matrix elements can all be deduced from each other. For instance, it is enough to compute the $(1,1)$ entry:

$$\hat{u}_x L(x,y) = \frac{e^{-(x^2+y^2)/4}}{\sqrt{2\pi}} \sum_{k=0}^{n-1} \frac{\hat{D}\tilde{H}_{2k+1}(x)\cdot\tilde{H}_{2k}(y) - \hat{u}\tilde{H}_{2k}(x)\cdot\tilde{H}_{2k+1}(y)}{\tilde{h}_k},$$

where $\hat{D} = w^{1/2}\hat{u}w^{-1/2}$ is the operator:

$$\hat{D}f(x) = w^{1/2}\partial_x(w^{-1/2}f_(x)) = e^{x^2/4}\partial_x\{e^{-x^2/4}f(x)\}.$$

8.8 proposition. *The* $(1,1)$ *entry of* $K_n^{(4)}(x,y)$ *is:*

$$\hat{u}_x L(x,y) = \frac{K_{2n}^{(2)}(x,y)}{2} - \frac{e^{-x^2/4}}{\sqrt{2\pi}}\,\frac{H_{2n}(x)}{4}\int_x^{\infty} e^{-y^2/4}\,H_{2n-1}(y)\,\mathrm{d}y.$$

Proof. Lemma 8.7 tells us:

$$\hat{D}\tilde{H}_{2n} = -\frac{H_{2n+1}}{2}. \tag{51}$$

We compute directly, using the 3-term recurrence relation:

$$\hat{D}\tilde{H}_{2n+1} = (2n+1)H_{2n} - \frac{x}{2}\,H_{2n+1} = \frac{1}{2}\big(-H_{2n+2} + (2n+1)H_{2n}\big), \tag{52}$$

and we remind the recurrence relation obvious from Lemma 8.6:

$$\tilde{H}_{2n} = H_{2n} + 2n\tilde{H}_{2n-2}.$$

Therefore:

$$\begin{aligned}\hat{D}_x L(x,y) &= \frac{w^{1/2}(x)w^{1/2}(y)}{2} \\ &\quad\times\Big\{\sum_{k=0}^{n-1}\frac{H_{2k}(x)\big[H_{2k}(y)+2k\tilde{H}_{2k-2}(y)\big]}{(2k)!} - \frac{H_{2k+2}(x)\cdot\tilde{H}_{2k}(y)}{(2k+1)!} \\ &\quad+\frac{H_{2k+1}(x)\cdot H_{2k+1}(y)}{(2k+1)!}\Big\}.\end{aligned}$$

We recognize in the first and third term half the Christoffel-Darboux kernel of size $2n$, and remains a telescopic sum:

$$\begin{aligned}\hat{D}_x L(x,y) &= \frac{K_{2n}^{(2)}(x,y)}{2} + \frac{w^{1/2}(x)w^{1/2}(y)}{2}\Big\{\sum_{k=0}^{n-1}\frac{H_{2k}(x)\tilde{H}_{2k-2}(y)}{(2k-1)!} - \frac{H_{2k+2}(x)\tilde{H}_{2k}(y)}{(2k+1)!}\Big\} \\ &= \frac{K_{2n}^{(2)}(x,y)}{2} - \frac{w^{1/2}(x)w^{1/2}(y)}{2\,(2n-1)!}\,H_{2n}(x)\cdot\tilde{H}_{2n-2}(y)\,,\end{aligned}$$

hence the result. □

It is then easy to reconstruct the other entries of $K_n^{(4)}$ using (51)-(52).

8.4 *Skew-orthogonal polynomials for beta = 1*

Lecture 16 (1h30)
December 8th, 2014

Here, we choose another normalization for the Gaussian weight, namely:

$$w_1(x) = w^{1/2}(x) = \frac{e^{-x^2/4}}{(2\pi)^{1/4}}.$$

The reason is that we find that skew-orthogonal polynomials for $w_1(x)\mathrm{d}x$ are related to $H_n(x)$ – whereas they would be related to $2^{-n/2}H_n(\sqrt{2}x)$ for the weight $w(x)$.

The skew-bilinear form is:

$$\langle f, g\rangle_1 = \int_{\mathbb{R}} \mathrm{sgn}(y-x) f(x)g(y)\, w_1(x)w_1(y)\mathrm{d}x\mathrm{d}y = \langle w_1^{-1}\cdot \hat{s}f, g\rangle_2.$$

It can be rewritten in terms of the operator $\hat{s}$ defined such that:

$$\frac{\hat{s}f(x)}{w_1(x)} := \int_{\mathbb{R}} \mathrm{sgn}(x-z)f(z)\,\frac{w_1(z)}{w_1(x)}\mathrm{d}z = -2e^{x^2/4}\int_x^{\infty} f(z)\,e^{-z^2/4}\mathrm{d}z + \int_{\mathbb{R}} f(z)\,\frac{w_1(z)}{w_1(x)}\mathrm{d}z,$$

and the usual scalar product:

$$\langle f, g\rangle_2 = \int_{\mathbb{R}} f(x)g(x)\,w(x)\mathrm{d}x.$$

In particular:

$$\frac{\hat{s}H_{2n+1}(x)}{w_1(x)} = -4\tilde{H}_{2n}(x). \tag{53}$$

We have used that the total integral over $\mathbb{R}$ vanishes by parity, and recognized the function $\tilde{H}_{2n}(x)$ met in Lemma 8.7.

8.9 lemma. *Define $\check{H}_{2n} = H_{2n}$ and $\check{H}_{2n+1} = H_{2n+1} - 2nH_{2n-1}$. Then, $(\check{H}_n)_{n\geq 0}$ is a family of skew-orthogonal polynomials with:*

$$\check{h}_n = \langle \check{H}_{2n}, \check{H}_{2n+1}\rangle_1 = 4\,(2n)!\,.$$

Proof. For parity reasons,

$$\forall n, m \geq 0, \qquad \langle H_{2n}, H_{2m}\rangle_1 = \langle H_{2n+1}, H_{2m+1}\rangle_1 = 0\,.$$

We compute using Lemma 8.6 to represent $\tilde{H}_{2n}$:

$$\begin{aligned} \langle H_{2n}, H_{2m+1}\rangle_1 &= -\langle H_{2m+1}, H_{2n}\rangle_1 = 4\langle \tilde{H}_{2m}, H_{2n}\rangle_2 \\ &= \frac{2^{m+2}\,m!}{2^n\,n!}\,(2n)!\,. \end{aligned} \tag{54}$$

if $n \leq m$, and 0 otherwise. Then, defining $\check{H}_{2n} = H_{2n}$ and $\check{H}_{2n+1} = H_{2n+1} -$

$2nH_{2n-1}$ satisfy the skew-orthogonality relations. And:

$$\check{h}_n = \langle \check{H}_{2n}, \check{H}_{2n+1} \rangle_1 = 4\,(2n)!\,.$$

□

As we have seen in § 7.5, the kernel $K_n^{(1)}$ can be deduced from its $(1,1)$ matrix element:

$$\hat{s}_y R(x,y) = w_1^{1/2}(x) w_1^{1/2}(y) \sum_{k=0}^{\lfloor n/2 \rfloor - 1} \frac{\check{H}_{2k+1}(x)\cdot \hat{s}\check{H}_{2k}(y) - \check{H}_{2k}(x)\cdot \hat{s}\check{H}_{2k+1}(y)}{\check{h}_k}\,.$$

8.10 proposition. *Let $m = \lfloor n/2 \rfloor$. We have:*

$$\hat{s}_y R(x,y) = w_1^{-1/2}(x) w_1^{1/2}(y)\, K_{2m}^{(2)}(x,y) + \frac{w_1^{1/2}(x) w_1^{1/2}(y)}{4\,(2m-1)!}\, H_{2m-1}(x)\cdot \hat{s}H_{2m}(y)\,.$$

Proof. From Equation (53), the definition of $\check{H}$'s and Lemma 8.6 for the $\tilde{H}$'s:

$$\frac{\hat{s}\check{H}_{2k+1}(y)}{w_1(y)} = -4\big(\tilde{H}_{2k}(y) - 2k\tilde{H}_{2k-2}(y)\big) = -4H_{2k}(y)\,. \tag{55}$$

We also need to compute $\hat{s}\check{H}_{2k}(y) = \hat{s}H_{2k}(y)$. For this, we remark:

$$\hat{D}\Big\{\frac{\hat{s}\check{H}_{2k}(y)}{w_1(y)}\Big\} = 2H_{2k}(y)\,,$$

with the operator $\hat{D} = w_1^{-1}\partial_x w_1$ defined in (48). It can be compared to the formula 52:

$$\hat{D}H_{2k+1} = \frac{1}{2}\Big(-H_{2k+2} + (2k+1)H_{2k}\Big)\,.$$

Therefore, there exists a constant c_k such that:

$$(2k+1)\frac{\hat{s}H_{2k}(y)}{w_1(y)} - \frac{\hat{s}H_{2k+2}}{w_1(y)} = 4H_{2k+1}(y)\,. \tag{56}$$

This identity is obtained by integration of $w^{-1}(x)\hat{u}_x = \partial_x$, and the integration constant here is 0 since $\hat{s}H_{2k}(0) = 0$ by parity of H_{2k}. Now, we can compute:

$$\begin{aligned}
\hat{s}_y R(x,y) &= w_1^{1/2}(x) w_1^{1/2}(y) \sum_{k=0}^{m/2-1} \frac{\big[H_{2k+1}(x) - 2kH_{2k-1}(x)\big]\cdot \hat{s}H_{2k}(y) - H_{2k}(x)\cdot \hat{s}\check{H}_{2k+1}(y)}{4\,(2k)!} \\
&= w_1^{1/2}(x) w_1^{3/2}(y) \sum_{k=0}^{m/2-1} \frac{H_{2k+1}(x)H_{2k+1}(y)}{4\,(2k+1)!} + \frac{H_{2k+1}(x)\cdot \hat{s}H_{2k+2}(y)}{4\,(2k+1)!\,w_1(y)} \\
&\quad - \frac{H_{2k-1}(x)\hat{s}H_{2k}(y)}{4\,(2k-1)!\,w_1(y)} + \frac{H_{2k}(x)H_{2k}(y)}{(2k)!}\,.
\end{aligned}$$

We recognize the Hermite Christoffel-Darboux kernel, namely if we recall

$w_1(x) = w^{1/2}(x)$:

$$\hat{s}_y R(x,y) = \sqrt{\frac{w_1(y)}{w_1(x)}}\, K_{2m}^{(2)}(x,y) + \frac{w_1^{1/2}(x) w_1^{1/2}(y)}{4\,(2m-1)!}\, H_{2m-1}(x)\cdot \hat{s} H_{2m}(y)\,,$$

which is the announced formula. □

It is then easy to reconstruct the other entries of $K_n^{(1)}$ using (55)-(56), as well as:

$$\hat{D}_x\Big\{\frac{\hat{s}f(x)}{w_1(x)}\Big\} = f(x)\,.$$

8.5 *Asymptotics of Hermite polynomials*

The integral formula of Lemma 8.2 for the Hermite polynomial is convenient to derive the large n asymptotics of $H_n(x)$. First, we want to argue that a priori, three regimes of x will occur. Recall Heine formula (Lemma 6.15):

$$H_n(x) = \mathbb{E}_n\Big[\prod_{i=1}^{n}(x-\lambda_i)\Big]\,.$$

We know from Wigner theorem that the λ_i are typically distributed, in the large n limit, in a segment $[-2\sqrt{n}, 2\sqrt{n}]$. So, the function $\prod_{i=1}^n (x-\lambda_i)$ is typically of constant sign when $|x| \gg 2\sqrt{n}$. This suggests that most of the n zeroes of $H_n(x)$ accumulate[19] in $[-2\sqrt{n}, 2\sqrt{n}]$, and we can expect then to be smoothly spaced, i.e. the typical spacing should be $O(n^{-1/2})$. So, in the regime $n \to \infty$, we expect the asymptotics of $H_n(x)$ to be:

- of oscillatory nature, with oscillations at scale $n^{-1/2}$, whenever $|x| < 2\sqrt{n}$ far away from the edges. This is the **bulk regime**.
- of exponential nature (without zeroes at leading order) when $|x| \gg 2\sqrt{n}$.

and there should exist a crossover between the two regimes x approaches $\pm 2\sqrt{n}$ at a rate depending on n: this is the **edge regime**. We will indeed find this dichotomy by computation. The regime $|x| \gg 2\sqrt{n}$ is less interesting. Indeed, it probes eigenvalues statistics in regions where only a finite number of them can be typically, and they typically behave like independent r.v. For this reason, we focus on the bulk and edge regime.

We remind that the Airy function is the solution of the differential equation $f''(x) = xf(x)$ expressed as:

$$\mathrm{Ai}(x) = \int_{\mathbb{R}} \frac{\mathrm{d}z}{2\pi}\, e^{-\mathrm{i}z^3/3 - \mathrm{i}zx}\,.$$

[19] Actually, one can prove that the distribution of zeroes of the orthogonal polynomials has the same limit law – the so-called equilibrium measure described in Chapter 10 – as those of eigenvalues. For such results, see the book of Deift.

8.11 proposition (Plancherel-Rotach asymptotics). *Let s be an integer, and X be a real number independent of n. In the bulk: we fix $x_0 \in]-2,2[$ and we have*

$$(57) \quad \varphi_{n+s}(n^{1/2}x_0 + n^{-1/2}X) = \frac{2\cos\big[\theta_n(x_0,X,s)\big]}{n^{1/4}\sqrt{2\pi}(4-x_0^2)^{1/4}} + O(n^{-3/4})\,,$$

with the phase:

$$\theta_n(x_0,X,s) = (n+s+1)\arcsin(x_0/2) - \frac{\pi(n+s)}{2} + \frac{nx_0\sqrt{4-x_0^2}}{4} + \frac{X\sqrt{4-x_0^2}}{2}\,.$$

At the edge, we have:

$$(58) \quad \varphi_{n+s}(2\sqrt{n} + n^{-1/6}X) = n^{-1/12}\mathrm{Ai}(X) + O(n^{-5/12})\,.$$

These asymptotics are uniform with respect to differentiation.

Proof. We start from the integral representation in Lemma 8.2:

$$(59) \quad \varphi_{n+s}(n^{1/2}x) = \frac{\mathrm{i}^n\,e^{nx^2/4}}{(2\pi)^{1/4}\sqrt{n!}}\,\tilde{\varphi}_{n+s}(n^{1/2}x)\,,$$

with focus on the integral:

$$(60) \quad \tilde{\varphi}_{n+s}(n^{1/2}x) = \int_{\mathbb{R}} \mathrm{d}z\,z^s\,e^{\mathrm{i}\sqrt{n}(x_0-x)z}\,e^{-S_n(z)}\,,$$

with:

$$S(z) = -n\ln z + z^2/2 + \mathrm{i}\sqrt{n}x_0 z.$$

At this stage, we consider $x = x_0 + o(1)$ when $n \to \infty$, for some $x_0 \in \mathbb{R}$. The strategy to derive the large n asymptotics of such integrals is called "steepest descent analysis". Firstly, we look for "approximate" critical points of S, i.e. points $z \in \mathbb{C}$ such that $S_n'(z) = -(n+s)/z + z + \mathrm{i}\sqrt{n}x = 0$. Dropping the term proportional to s and replacing x by x_0 at leading order, we find two approximate critical points at $z_\pm$ with:

$$z_\pm := \frac{\sqrt{n}}{2}\Big(-\mathrm{i}x_0 \pm \sqrt{4-x_0^2}\Big).$$

The game consists in moving the contour (here $\mathbb{R}$) in the complex plane to pass through critical points and follow "steepest descent contours" to join ∞, i.e. in a direction when $\mathrm{Re}\,S_n(x;z)$ decreases the most. The constraint in doing so is that the asymptotic direction of the contour at ∞ should remain in a region of convergence of the integral – here $|\arg z| < \pi/4$ or $|\pi - \arg z| < \pi/4$. We must discuss several cases:

- For $|x_0| < 2$, the steepest descent contour homologous to $\mathbb{R}$ is the union of two paths γ_- and γ_+, passing through $z_\pm$ and 0 (Figure 10).
- For $|x_0| = 2$, it passes through $z_+ = z_-$ (Figure 11).

- For $|x_0| > 2$, it passes through $z_+ = -\mathrm{i}\frac{\sqrt{n}}{2}(x_0 - \sqrt{x_0^2 - 4})$ (Figure 12).

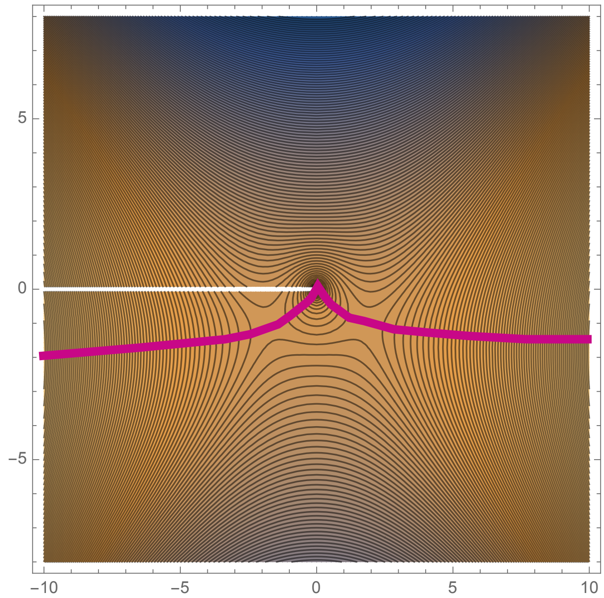

Figure 10: Level lines of S for $x_0 = 1$ and $n = 4$. We recognize the two saddle points $z_\pm$, and the steepest descent contour passes through the saddles and is normal to the level lines at every point.

We shall treat in detail the cases $|x_0| < 2$ and $x_0 = 2$. The case $x_0 = -2$ follows by parity, and the regime $|x_0| > 2$ is technically similar to $|x_0| < 2$ and left to the reader. Then, it is convenient to set $x_0 = 2\sin\theta_0$ for $\theta_0 \in]-\pi/2, \pi/2]$. This leads to:

$$z_\varepsilon = \sqrt{n}\varepsilon\, e^{-\mathrm{i}\varepsilon\theta_0}, \qquad \varepsilon \in \{-1, 1\}.$$

To analyze the local behavior of S at the critical point, we compute the Hessian:

$$S''(z_\varepsilon) = \frac{n}{z_\varepsilon^2} + 1 = \frac{\sqrt{4 - x_0^2}}{2}\Big(\sqrt{4 - x_0^2} + \mathrm{i}\varepsilon x_0\Big) = \sqrt{4 - x_0^2}\, e^{\mathrm{i}\varepsilon\theta_0}.$$

Case $|\mathbf{x_0}| < \mathbf{2}$. $S''(z_\varepsilon)$ does not vanish and is of order 1. This suggests, on the integration path $\gamma_\pm$, to use the new variable $z = z_{n,\pm} + \zeta$. We then collect all terms in $S_{n,x}(z)$ that contribute up to $o(1)$ when $n \to \infty$. This also suggests the natural scaling

$$x - x_0 = X/n$$

in (60). For ζ in a compact independent of n, the third derivative of S is $O(n^{-1/2})$, and the extra term $\sqrt{n}(x - x_0)\zeta$ is also $O(n^{-1/2})$. Therefore, by

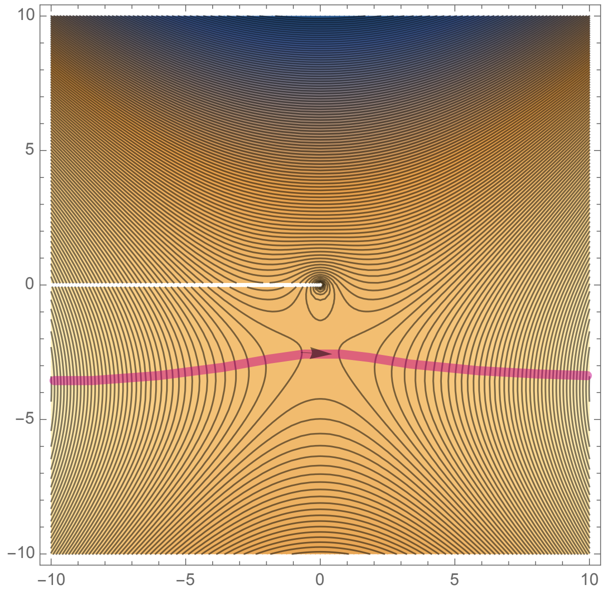

Figure 11: Level lines of S and steepest descent contour for $x_0 = 2$, $n = 4$. The saddle point of order 3 arises from the collision between the two saddles when $x_0 \to 2$.

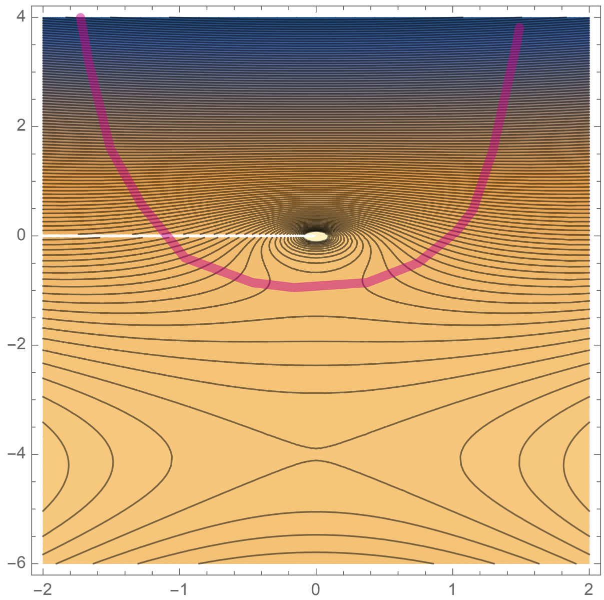

Figure 12: Level lines of S and steepest descent contour for $x_0 = 2.5$, $n = 4$. The second saddle does not contribute at leading order.

Taylor expansion with integral remainder at third order for S, we obtain:

$$\begin{aligned} z^s\, e^{-S(z)+\mathrm{i}\sqrt{n}(x_0-x)z} &= \big(\sqrt{n}\varepsilon\, e^{-\mathrm{i}\varepsilon\theta_0}\big)^{n+s} \exp\Big[\mathrm{i}\varepsilon\, e^{-\mathrm{i}\varepsilon\theta_0} X - n + \frac{n e^{-2\mathrm{i}\varepsilon\theta_0}}{2}\Big] \\ &\quad \times e^{-S''(z_\varepsilon)\zeta^2/2-\mathrm{i}X\zeta}\big(1+O(n^{-1/2})\big), \end{aligned}$$

and the remainder term is integrable when $\zeta \to \infty$ along the contour – because the prefactor is decaying exponentially. Therefore, we obtain:

$$\begin{aligned} \tilde{\varphi}_{n+s}(\sqrt{n}x) &= \Big\{ \sum_{\varepsilon=\pm 1} \big(\sqrt{n}\varepsilon\, e^{-\mathrm{i}\varepsilon\theta_0}\big)^{n+s} \sqrt{\frac{\pi}{S''(z_\varepsilon)}}\, e^{-X^2/2S''(z_\varepsilon)} \\ &\quad \times \exp\Big[\mathrm{i}\varepsilon\, e^{-\mathrm{i}\varepsilon\theta_0} X - n + \frac{n e^{-2\mathrm{i}\varepsilon\theta_0}}{2}\Big]\Big\} \cdot \big(1+O(n^{-1/2})\big). \end{aligned}$$

Taking into account the prefactor in (60) and using Stirling formula, we arrive to the announced result (57) after a tedious algebra.

Case $\mathbf{x_0} = \mathbf{2}$. The two critical points have coalesced $z_+ = z_- = z_0 := -\mathrm{i}\sqrt{n}$, thus leading to $S''(z_0) = 0$. The leading contribution in Taylor formula now comes from the third order:

$$S'''(z_0) = \frac{2\mathrm{i}}{\sqrt{n}}.$$

This suggest to define $z = z_0 + n^{1/6}\zeta$ so that the main contribution to the integral comes from finite range in the new variable ζ. Then, the fourth order remainder in Taylor formula is of order $(n^{1/6}\zeta)^4 \cdot S''''(z_0 + n^{1/6}\zeta) \in O(n^{-1/3})$, uniformly for ζ in any compact independent of n. This also suggests to set

$$x - x_0 = n^{-2/3} X$$

for X independent of n. Then:

$$\begin{aligned} \mathrm{d}z\, z^s\, e^{-S(z)+\mathrm{i}\sqrt{n}(x_0-x)z} &= \mathrm{d}\zeta\, n^{1/6}\, (-\mathrm{i}\sqrt{n})^{n+s}\, e^{-3n/2}\, e^{-\frac{\mathrm{i}\zeta^3}{3}-\mathrm{i}X\zeta} \\ &\quad \times \big(1+O(n^{-1/3})\big). \end{aligned}$$

Therefore:

$$\tilde{\varphi}_{n+s}(\sqrt{n}x) = n^{1/6}\, (-\mathrm{i}\sqrt{n})^{n+s}\, e^{-3n/2}\, \mathrm{Ai}(X) \cdot \big(1+O(n^{-1/3})\big),$$

and (58) follows from a (slightly less) tedious but straightforward algebra.

□

8.6 *Limit laws*

Thanks to the asymptotics of the Hermite Christoffel-Darboux kernel, we can compute the laws governing the statistics of eigenvalues in the bulk at the edge. Though these expressions are computed in the Gaussian ensemble, they are universal, i.e. hold in a large class of invariant ensemble, as well as Wigner matrices. These universal law depend only on β, i.e. the symmetry class considered.

Spectral density

First, we have a new proof of Wigner theorem for Gaussian matrices with $\beta = 1, 2, 4$ by direct computation. A much simpler proof will be given in Chapter 10.

8.12 PROPOSITION. *The* 1*-point density correlation – also called* ***spectral densities*** *– in the Gaussian β ensembles for $\beta \in \{1, 2, 4\}$ converges to the semi-circle law:*

$$\rho_{\rm sc}(x_0) = \frac{\sqrt{4 - x_0^2}}{2\pi} \cdot \mathbf{1}_{[-2,2]}(x_0)$$

after suitable normalisation. Namely:

$$\begin{aligned} \lim_{n\to\infty} n^{-1/2} \rho_{1|n}^{(2)}(n^{1/2}x_0) &= \rho_{\rm sc}(x_0), \\ \lim_{n\to\infty} (2n)^{-1/2} \rho_{1|n}^{(4)}((2n)^{1/2}x_0) &= \rho_{\rm sc}(x_0), \\ \lim_{n\to\infty} n^{-1/2} \rho_{1|n}^{(1)}(n^{1/2}x_0) &= \rho_{\rm sc}(x_0). \end{aligned}$$

Proof. ($\beta = 2$) For $\beta = 2$, its expression is

$$\rho_n^{(2)}(x) = K_n^{(2)}(x,x) = \sqrt{n}\big[\varphi_n'(x)\varphi_{n-1}(x) - \varphi_{n-1}'(x)\varphi_n(x)\big].$$

If we set $x = n^{1/2}x_0$, the differentiation is represented by $\partial_x = n^{-1/2}\partial_{x_0}$. The Plancherel-Rotach asymptotics give:

$$\varphi_{n+s}'(x) = \frac{2n^{-3/4}}{\sqrt{2\pi}(4-x_0^2)^{1/4}}\bigg\{ -\frac{x_0\cos\big[\theta_n(x_0,0,s)\big]}{2(4-x_0^2)} - \partial_{x_0}\theta_n(x_0,0,s)\,\sin\big[\theta_n(x_0,0,s)\big]\bigg\}.$$

We observe that:

$$\begin{aligned} \theta_n(x_0,0,s) &= n\bigg(-\arccos(x_0/2) + \frac{x_0\sqrt{4-x_0^2}}{4}\bigg) + O(1) \\ \partial_{x_0}\theta_n(x_0,0,s) &= \frac{n}{2}\sqrt{4-x_0^2} + O(1) \end{aligned}$$

is independent of s to leading order. Thus:

$$\partial_{x_0}\theta_n(x_0,0,s)\cdot\varphi_{n+s}'(x) = -\frac{n^{1/4}(4-x_0^2)^{1/4}}{\sqrt{2\pi}}\Big\{\sin\big[\theta_n(x_0,0,s)\big] + O(1/n)\Big\},$$

and:

$$\begin{aligned} K_n^{(2)}(x,x) &= \frac{2n^{1/2}}{2\pi}\Big\{\cos\big[\theta_n(x_0,0,-1)\big]\sin\big[\theta_n(x_0,0,0)\big] \\ &\quad -\cos\big[\theta_n(x_0,0,0)\big]\sin\big[\theta_n(x_0,0,-1)\big]+O(n^{-1/2})\Big\} \\ &= \frac{2n^{1/2}}{2\pi}\sin\big[\theta_n(x_0,0,0)-\theta_n(x_0,0,-1)\big]+O(1) \\ &= \frac{2n^{1/2}}{2\pi}\sin\big[\arccos(x_0/2)\big]+O(1) \\ &= \frac{n^{1/2}}{2\pi}\sqrt{4-x_0^2}+O(1)\,. \end{aligned}$$

□

We will only sketch the computations needed for the cases $\beta=1$ and 4.

Proof. ($\beta=4$). We use the expression (47) for the kernel, and our recent computation of the diagonal matrix elements – the $(1,1)$ and the $(2,2)$ elements are identical by antisymmetry of $L(x,y)$ – in Lemma 8.8:

$$\begin{aligned} \rho_{1|n}^{(4)}(x) &= \frac{1}{2}\,\mathrm{Tr}\big[\Theta_1(K_n^{(4)}(x,x))\big] \\ &= \frac{1}{2}\Big(K_{2n}^{(2})(x,x)-\frac{e^{-x^2/4}}{2\pi}\,H_{2n}(x)\int_x^\infty e^{-y^2/4}\,H_{2n}(y)\,\mathrm{d}y\Big) \\ &= \frac{1}{2}\Big(\rho_{1|2n}^{(1)}(x)-\frac{e^{-x^2/4}}{2\pi\,(2n-1)!}\,H_{2n}(x)\int_x^\infty e^{-y^2/4}\,H_{2n}(y)\,\mathrm{d}y\Big)\,. \end{aligned} \tag{61}$$

In the regime $x=(2n)^{1/2}x_0$ with $x_0\in]-2,2[$, one can check that the second term is $O(1/n)$ – in particular one should perform the change of variable $y=n^{1/2}x_0+n^{-1/2}Y$, and use the fact that H_{2n} is decaying exponentially fast outside of $[-2\sqrt{n},2\sqrt{n}]$. Since we have seen for $\beta=2$ that the first term is $O(n^{-1/2})$, we deduce:

$$\rho_{1|n}^{(4)}(x)=(2n)^{1/2}\,\rho_{\mathrm{sc}}(x_0)+O(1)\,.$$

□

Proof. ($\beta=1$). We use the expression (49) for the kernel, and our recent computation of the diagonal matrix elements – the $(1,1)$ and the $(2,2)$ are identical by antisymmetry of $R(x,y)$ – in Lemma 8.10. Let $m=\lfloor n/2\rfloor$.

$$\begin{aligned} \rho_{1|n}^{(1)} &= \frac{1}{2}\,\mathrm{Tr}\big[\Theta_1(K_n^{(1)}(x,x))\big] \\ &= K_{2m}(x,x)+\frac{e^{-x^2/8-y^2/8}}{4\,(2\pi)^{1/4}\,(2m-1)!}\,H_{2m-1}(x)\cdot\hat{s}H_{2m}(x)\,. \end{aligned} \tag{62}$$

Once again, we find that the second term is $O(1/m)$. And, for n odd, the extra term that should be added to the kernel is also $O(1/m)$. Therefore:

$$\rho^{(1)}_{1|n}(x) = n^{1/2}\rho^{(2)}_{1|n}(x_0) + O(1)$$

independently of the parity of n. □

Eigenvalue statistics for the GβE

We can also deduce the asymptotics of the Christoffel-Darboux kernel. Let us define the sine kernel and the Airy kernel:

$$K_{\sin}(x,y) = \frac{\sin\pi(x-y)}{\pi(x-y)}, \qquad K_{\mathrm{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}\,.$$

For $\beta = 4$, we define the quaternionic kernels:

$$\begin{aligned}
\Theta_1\big[K^{(4)}_{\sin}(x,y)\big] &= \frac{1}{2}\begin{pmatrix} K_{\sin}(x,y) & -\partial_y K_{\sin}(x,y) \\ \int_y^x K_{\sin}(z,y)\mathrm{d}z & K_{\sin}(x,y)\end{pmatrix} \\
\Theta_1\big[K^{(4)}_{\mathrm{Airy}}(x,y)\big] &= \frac{1}{2}\begin{pmatrix} K_{\mathrm{Airy}}(x,y) & -\partial_y K_{\mathrm{Airy}}(x,y) \\ \int_y^x K_{\mathrm{Airy}}(z,y)\mathrm{d}z & K_{\mathrm{Airy}}(x,y)\end{pmatrix} \\
&\quad -\frac{1}{4}\begin{pmatrix} \mathrm{Ai}(x)\int_y^\infty \mathrm{Ai}(z)\mathrm{d}z & \mathrm{Ai}(x)\mathrm{Ai}(y) \\ \big(\int_y^x \mathrm{Ai}(z)\mathrm{d}z\big)\big(\int_y^\infty \mathrm{Ai}(z')\mathrm{d}z'\big) & \mathrm{Ai}(y)\int_x^\infty \mathrm{Ai}(z)\mathrm{d}z\end{pmatrix}. \qquad (63)
\end{aligned}$$

And $\beta = 1$, we define the quaternionic kernels:

$$\begin{aligned}
\Theta_1\big[K^{(1)}_{\sin}(x,y)\big] &= \begin{pmatrix} K_{\sin}(x,y) & -\partial_y K_{\sin}(x,y) \\ \frac{1}{2}\mathrm{sgn}(y-x) + \int_y^x K_{\sin}(z,y)\mathrm{d}z & K_{\sin}(x,y)\end{pmatrix} \\
\Theta_1\big[K^{(1)}_{\mathrm{Airy}}(x,y)\big] &= \begin{pmatrix} K_{\mathrm{Airy}}(x,y) & -\partial_y K_{\mathrm{Airy}}(x,y) \\ \int_y^x K_{\mathrm{Airy}}(z,y)\mathrm{d}z & K_{\mathrm{Airy}}(x,y)\end{pmatrix} \\
&\quad +\frac{1}{2}\begin{pmatrix} \mathrm{Ai}(x)\big(1-\int_\infty^y \mathrm{Ai}(z)\mathrm{d}z\big) & -\mathrm{Ai}(x)\mathrm{Ai}(y) \\ \big(\int_y^x \mathrm{Ai}(z)\mathrm{d}z\big)\big(1-\int_y^\infty \mathrm{Ai}(z')\mathrm{d}z'\big) & \mathrm{Ai}(y)\big(1-\int_x^\infty \mathrm{Ai}(z')\mathrm{d}z'\big)\end{pmatrix}.
\end{aligned}$$

Eventually, we define the effective degrees of freedom n_β, sometimes called **Fermi number**:

$$n_1 = n_2 = n, \qquad n_4 = 2n\,.$$

8.13 proposition (Bulk). *If $x_0 \in]-2,2[$, we have:*

$$\lim_{n\to\infty} \frac{K_n^{(\beta)}\Big(n_\beta^{1/2}x_0 + \frac{X}{\rho^{(\beta)}_{1|n}(n_\beta^{1/2}x_0)}, n_\beta^{1/2}x_0 + \frac{Y}{\rho^{(\beta)}_{1|n}(n_\beta^{1/2}x_0)}\Big)}{\rho^{(\beta)}_{1|n}(n_\beta^{1/2}x_0)} = K^{(\beta)}_{\sin}(X,Y)$$

with uniform convergence for (X,Y) in any compact of $\mathbb{R}^2$. Hence, for any fixed

$k \geq 1$, the k-point density correlations in the bulk are:

$$\lim_{n\to\infty} \frac{\rho_{k|n}^{(\beta)}\Big(n_\beta^{1/2}x_0 + \frac{X_1}{\rho_{1|n}^{(\beta)}(n_\beta^{1/2}x_0)}, \ldots, n_\beta^{1/2}x_0 + \frac{X_k}{\rho_{1|n}^{(\beta)}(n_\beta^{1/2}x_0)}\Big)}{\big(\rho_{1|n}^{(\beta)}(n_\beta^{1/2}x_0)\big)^k} = \det_{\substack{\mathbb{H} \\ 1\leq i,j\leq k}} K_{\sin}^{(\beta)}(X_i, X_j)\,.$$

For any compact $A \subseteq \mathbb{R}$, the gap probabilities are:

$$\lim_{n\to\infty} \mathbb{P}_n^{(\beta)}\left[\mathrm{Sp}\, M \cap \left(n_\beta^{1/2}x_0 + \frac{A}{\rho_{1|n}^{(\beta)}(n_\beta^{1/2}x_0)}\right) = \emptyset\right] = \mathrm{Det}_{\mathbb{H}}(1 - K_{\sin}^{(\beta)})_{L^2(A)}\,.$$

The rescaling by the local spectral density at $\rho_{1|n}^{(\beta)}(n_\beta^{1/2}x_0)$, means that measured in the variable X, the local density of eigenvalues is ~ 1. Such a normalisation is necessary before comparing eigenvalue statistics in two different ensembles.

8.14 proposition (Edge). *We have:*

$$\lim_{n\to\infty} \frac{K_n^{(\beta)}(2n_\beta^{1/2} + n_\beta^{-1/6}X, 2n_\beta^{1/2} + n_\beta^{-1/6}Y)}{n_\beta^{1/6}} = K_{\mathrm{Airy}}^{(\beta)}(X, Y)$$

with uniform convergence for (X, Y) bounded from below, i.e. in any semi-infinite segment of the form $[m, +\infty[$. So, the k-point density correlations are:

$$\lim_{n\to\infty} \frac{\rho_{k|n}^{(\beta)}\big(2n_\beta^{1/2} + n_\beta^{-1/6}X_1, \ldots, 2n_\beta^{1/2} + n_\beta^{-1/6}X_k\big)}{n_\beta^{k/6}} = \det_{\substack{\mathbb{H} \\ 1\leq i,j\leq k}} K_{\mathrm{Airy}}^{(\beta)}(X_i, X_j)\,.$$

And, for any A bounded from below:

$$\lim_{n\to\infty} \mathbb{P}_n^{(\beta)}\big[\mathrm{Sp}\, M \cap (2n_\beta^{1/2} + n_\beta^{-1/6}A) = \emptyset\big] = \mathrm{Det}_{\mathbb{H}}(1 - K_{\mathrm{Airy}}^{(\beta)})_{L^2(A)}\,.$$

In particular, by choosing $A = [s, +\infty[$, we have access to the law of fluctuations of the maximum eigenvalue:

$$\mathrm{TW}_\beta(s) := \lim_{n\to\infty} \mathbb{P}_n^{(\beta)}\big[\lambda_{\max} \leq 2n_\beta^{1/2} + n_\beta^{-1/6}s\big] = \mathrm{Det}(1 - K_{\mathrm{Airy}}^{(\beta)})_{L^2([s,+\infty[)}\,.$$

These are the **Tracy-Widom laws** introduced in § 1[redacted]

We give the details of the proof only for $\beta = 2$. For $\beta = 1, 4$, we have already gathered all the ingredients so $\beta = 2$ and some extra but straightforwards computations with the Plancherel-Rotach formula give the results.

Proof. Bulk: let $x_0 \in]-2,2[$. We have

$$\begin{aligned}
& K_n^{(2)}\big(n^{1/2}x_0 + n^{-1/2}\tilde{X}, n^{1/2}x_0 + n^{-1/2}\tilde{Y}\big) \\
= & \frac{4n^{-1/2}}{2\pi\sqrt{4-x_0^2}}\Big\{\cos\big[\theta_n(x_0,\tilde{X},0)\big]\cos\big[\theta_n(x_0,\tilde{Y},-1)\big] - \cos\big[\theta_n(x_0,\tilde{X},-1)\big]\cos\big[\theta_n(x_0,\tilde{Y},0)\big]\Big\} \\
& + O(1/n)
\end{aligned}$$

and using the expression of the phase θ_n given in Proposition 8.11 and trigonometric formulas:

$$\begin{aligned}
& K_n^{(2)}\big(n^{1/2}x_0 + n^{-1/2}\tilde{X}, n^{1/2}x_0 + n^{-1/2}\tilde{Y}\big) \\
= & \frac{4n^{-1/2}}{2\pi\sqrt{4-x_0^2}}\, 2\sin\big[\arccos(x_0/2)\big]\cdot\sin\Big[(\tilde{X}-\tilde{Y})\sqrt{4-x_0^2}/2\Big] + O(1/n).
\end{aligned}$$

Therefore, if we rescale:

$$n^{-1/2}\tilde{X} = \frac{X}{\rho_{1|n}^{(2)}(n^{1/2}x_0)} = n^{-1/2}\,\frac{2\pi X}{\sqrt{4-x_0^2}} + O(1/n)$$

and similarly for Y, we obtain the result in the announced form. □

Proof. (*Edge*) Since the leading asymptotics of φ_{n+s} at the edge does not depend on s, the leading contribution to the kernel $K_n^{(2)}$ at the edge comes from the subleading terms in φ_{n+s}. But we can rearrange slightly the kernel to avoid computing further coefficients in the asymptotics. We use the the exact equation:

$$\varphi_n'(x) = -\frac{x}{2}\varphi_n(x) + \sqrt{n}\varphi_{n-1}(x)$$

to write:

$$K_n^{(2)}(x,y) = -\frac{\varphi_n(x)\varphi_n(y)}{2} + \frac{\varphi_n(x)\varphi_{n-1}'(y) - \varphi_n'(x)\varphi_{n-1}(y)}{x-y}\,.$$

When x and y are chosen in the edge regime $(x-y)$ is of order $n^{-1/6}$, so the first term is negligible compared to the last one, and we find taking into account $\partial_x = n^{1/6}\partial_X$:

$$\begin{aligned}
K_n^{(2)}\big(2\sqrt{n}+n^{-1/6}X, 2\sqrt{n}+n^{-1/6}Y\big) & = n^{1/6}\,\frac{\mathrm{Ai}(X)\mathrm{Ai}'(Y) - \mathrm{Ai}'(x)\mathrm{Ai}(Y)}{X-Y} \\
& \quad + O(n^{-1/6})\,.
\end{aligned}$$

□

9 Orthogonal polynomials and integrable systems

Lecture 17 (1h30)
December 9th, 2014

Let μ_0 be a reference positive measure on the real line. We want to consider the system of orthogonal polynomials $(p_n)_{n\geq 0}$ associated to the new measure:

$$\mu_t = \mu_0 \cdot e^{-V(x)}, \qquad V(x) = \sum_{j\geq 1} \frac{t_j\, x^j}{j}\,.$$

The polynomial V is called the **potential**, and the coefficients t_j the **times**. We assume that μ_0 and V is chosen such that all the moments of μ are finite, thus ensuring existence of $(p_n)_{n\geq 0}$. We consider the scalar product on $\mathbb{R}[X]$;

$$(f|g) = \int_{\mathbb{R}} f(x)g(x)\,\mathrm{d}\mu_t(x)\,.$$

Let us recollect some notations:

$$\left\{\begin{array}{lcl} h_n & = & (p_n|p_n) \\ \beta_n & = & (xp_n|p_n) \end{array}\right., \qquad \left\{\begin{array}{lcl} u_n & = & \ln h_n \\ v_n & = & \beta_n/h_n \end{array}\right..$$

The partition function of the $\beta = 2$ invariant ensemble associated to μ can be computed in terms of the norms:

$$Z_n := \int_{\mathbb{R}^n} \prod_{i=1}^{n} \mathrm{d}\mu_t(\lambda_i) \prod_{1\leq i<j\leq n} (\lambda_j - \lambda_i)^2 = n! \prod_{m=0}^{n-1} h_m\,.$$

So, u_n is a discrete derivative of the free energy $F_n = \ln(Z_n/n!)$, namely $u_n = F_n - F_{n-1}$.

9.1 *Operator formalism*

It is a basic fact that orthogonal polynomials satisfy:

9.1 lemma. *The* 3*-term recurrence relation:*

$$\forall n \geq 0, \qquad xp_n(x) = p_{n+1}(x) + v_n p_n(x) + e^{u_n - u_{n-1}} p_{n-1}(x)\,.$$

where it is understood that quantities with negative index are 0*.*

Proof. $xp_n(x) - p_{n+1}(x)$ is a polynomial of degree $\leq n$, and we can decompose it in the basis of p_k indexed by $k \in [\![0, n]\!]$. We have $(xp_n|p_n) = \beta_n = v_n(p_n|p_n)$, and:

$$(xp_n|p_{n-1}) = (p_n|xp_{n-1}) = (p_n|p_n + \ldots) = h_n = e^{u_n - u_{n-1}}(p_{n-1}|p_{n-1})\,.$$

The $\ldots$ is a polynomial of degree $\leq n-1$: it is orthogonal to p_n whence the last equality. And, for $m \leq n-2$, we have $(xp_n|p_m) = (p_n|xp_m)$ which is equal to 0 since xp_m has degree $m + 1 \leq n - 1$. □

Position operator

The recurrence relation can be reformulated in terms of the operator $\hat{Q}$: $\mathbb{R}[X] \to \mathbb{R}[X]$ of "multiplication by X": its matrix in the basis $(p_n)_{n\geq 0}$ is a (semi-infinite) band matrix of width 1:

$$\hat{Q} = \begin{pmatrix} v_0 & h_0 & & & & \\ h_0 & v_1 & \frac{h_1}{h_0} & & & \\ & \frac{h_1}{h_0} & v_2 & \frac{h_2}{h_1} & & \\ & & \frac{h_2}{h_1} & v_3 & \frac{h_3}{h_2} & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}$$

and we have:

$$x p_n(x) = \sum_{m\geq 0} \hat{Q}_{nm} p_m(x)\,.$$

We thus have:

$$\hat{Q}_{nn+1} = 1, \qquad \hat{Q}_{nn} = v_n, \qquad \hat{Q}_{nn-1} = e^{u_n - u_{n-1}}\,.$$

Differentiation with respect to times

We would like to compute $j\partial_{t_j} p_n$ via its decomposition on the basis of orthogonal polynomials. Since $p_n = x^n + \ldots$, the polynomial $j\partial_{t_j} p_n$ has degree $\leq n-1$, i.e. the matrix representing $j\partial_{t_j}$ is strictly lower triangular. Differentiating the scalar product:

$$j\partial_{t_j} h_n\, \delta_{n,m} = (j\partial_{t_j} p_n | p_m) + (p_n | j\partial_{t_j} p_m) - (\hat{Q}^j p_n | p_m)\,.$$

For $n = m$, the two first terms vanish par orthogonality, and we find:

$$j\partial_{t_j} h_n = \hat{Q}^j_{nn}\,.$$

For $m < n$, the second term vanishes by orthogonality, and we find:

$$(j\partial_{t_j} p_n | p_m) = (\hat{Q}^j p_n | p_m)\,.$$

Therefore:

$$j\partial_{t_j} p_n(x) = \sum_{m\geq 0} (\hat{Q}^j)^-_{nm} p_m(x),$$

where $(\cdots)^-$ denotes the strict lower triangular part of the matrix $\cdots$.

The matrices $\hat{Q}^k$ themselves depend on the times. We shall now examine their evolution. Let us differentiate, for any $n \geq 0$:

$$\begin{aligned} j\partial_{t_j}\big(\hat{Q}^k p_n(x)\big) &= \sum_{\ell\geq 0} (j\partial_{t_j} \hat{Q}^k)_{n\ell}\, p_\ell(x) + \sum_{m\geq 0} \hat{Q}^k_{nm}\, j\partial_{t_j} p_m(x) \\ &= \sum_{m\geq 0} (j\partial_{t_j} \hat{Q}^k_{nm})\, p_m(x) + \sum_{m,\ell\geq 0} \hat{Q}^k_{nm}\, (Q^j)^-_{m\ell}\, p_\ell(x)\,. \end{aligned} \tag{64}$$

But this is also equal to:

$$
\begin{aligned}
j\partial_{t_j}(x^k p_n(x)) &= x^k\, j\partial_{t_j} p_n(x) = x^k \sum_{m\geq 0} (\hat{Q}^j)^-_{nm} p_m(x) \\
&= \sum_{m,\ell\geq 0} (\hat{Q}^j)^-_{nm} \hat{Q}^k_{m\ell}\, p_\ell(x)\,.
\end{aligned}
\tag{65}
$$

Comparing the two expressions, we find:

$$\forall j,k\geq 1, \qquad j\partial_{t_j}\hat{Q}^k = \left[(\hat{Q}^j)^-, \hat{Q}^k\right]. \tag{66}$$

9.2 *Impulsion operator*

We can also define the operator $\hat{P} : \mathbb{R}[X] \to \mathbb{R}[X]$ of multiplication by X. It satisfies the commutation relation:

$$[\hat{P},\hat{Q}] = 1.$$

We denote $(\hat{P}_{nm})_{m,n\geq 0}$ its matrix in the basis of $(p_n)_{n\geq 0}$:

$$\partial_x p_n(x) = \sum_{m=0}^{n-1} \hat{P}_{nm} p_m(x).$$

This matrix is strictly lower triangular, and identifying the leading coefficient: $\hat{P}_{nn-1} = n$.

In the special case where μ_0 is the Lebesgue measure, i.e. $\mu = e^{-V(x)}\mathrm{d}x$, we can compute $\hat{P}$ by using integration by parts in the scalar product:

$$(\hat{P}p_n|p_m) + (p_n|\hat{P}p_m) = (V'(\hat{Q})p_n|p_n), \tag{67}$$

where we identify $V'(\hat{Q})$ as the operator of multiplication by the polynomial $V'(X)$. Using that $\hat{P}$ is strict lower triangular, we get for $m < n$ the equation $\hat{P}_{nm} = V'(\hat{Q})_{nm}$. In other words:

$$\hat{P} = [V'(\hat{Q})]^-.$$

The previous formula for $m = n-1$, and (67) for $n = m$ give two constraints on the operator $\hat{Q}$, called **discrete string equation**:

$$V'(\hat{Q})_{nn} = 0, \qquad V'(\hat{Q})_{nn-1} = n.$$

Folding

The 3-term recurrence relation can be written as a recurrence on the column vector:

$$\Psi_{n+1}(x) = A_n(x)\Psi_n(x),$$

with:

$$\Psi_n(x) = \begin{pmatrix} p_n(x) \\ p_{n-1}(x) \end{pmatrix}, \qquad A_n(x) = \begin{pmatrix} x - v_n & -\frac{h_n}{h_{n-1}} \\ 1 & 0 \end{pmatrix}.$$

Then, the differential equations with respect to x and t_j can be written in terms of $\Psi_n(x)$ only.

9.2 lemma. *There exist 2×2 matrices B_n and $C_n^{(j)}$, whose entries are polynomials in x and depend on the t's, such that, for any $n \geq 1$:*

$$\partial_x \Psi_n(x) = B_n(x), \qquad \forall j \geq 1, \quad j\partial_{t_j} \Psi_n(x) = C_n^{(j)}(x).$$

It is implicit that all the matrices here depend on the times. In the case $\mu(x) = e^{-V(x)}\mathrm{d}x$, There exist explicit formulas involving $V'(\hat{Q})$ for B_n and Q^j for $C_n^{(j)}$, which we will not state here.

Proof. Since $\Psi_1(x) = (x\ 1)^T$, we can take:

$$B_1(x) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad C_1^{(j)} = 0.$$

Assume we have constructed B_n and $C_n^{(j)}$. Note that the matrix $A_n(x)$ is invertible, and its inverse:

$$A_n^{-1}(x) = \begin{pmatrix} 0 & 1 \\ -\frac{h_{n-1}}{h_n} & \frac{h_{n-1}}{h_n}(x - v_n) \end{pmatrix}$$

is a polynomial in x. Then, we have:

$$\begin{aligned} \partial_x \Psi_{n+1}(x) &= \partial_x\big(A_n(x)\Psi_n(x)\big) = (\partial_x A_n(x)) \cdot \Psi_n(x) + A_n(x)\partial_x \Psi_n(x) \\ &= \partial_x A_n(x) A_n^{-1}(x) + A_n(x) B_n(x) A_n^{-1}(x), \end{aligned} \tag{68}$$

so we can take:

$$B_{n+1}(x) = \partial_x A_n(x) \cdot A_n^{-1}(x) + A_n(x) B_n(x) A_n^{-1}(x). \tag{69}$$

Similarly, we can set:

$$C_{n+1}^{(j)} = j\partial_{t_j} A_n(x) \cdot A_n^{-1}(x) + A_n(x) C_n^{(j)} A_n^{-1}(x). \tag{70}$$

By induction, B's and $C^{(j)}$ exist and have polynomial entries in x. □

9.3 *Introduction to integrable system: the Toda chain*

The vector of consecutive orthogonal polynomials $\Psi_n(x)$ exists, and is solution to many ODE's:

$$\big(\partial_x - A_n(x)\big)\Psi_n(x) = 0, \quad \big(j\partial_{t_j} - C_n^{(j)}(x)\big)\Psi_n(x) = 0\,.$$

So, the coefficients of those ODEs – which are polynomials in x and functions of all the t's – cannot be arbitrary. They must satisfy the **zero-curvature** conditions[20]:

$$(71)\quad \forall j,k,n \geq 1,\ \ [\partial_x - A_n(x), j\partial_{t_j} - C_n^{(j)}] = 0\ \ [j\partial_{t_j} - C_n^{(j)}, k\partial_{t_k} - C_n^{(k)}] = 0\,.$$

These are nonlinear PDE's with respect to the t_j's. They are called **integrable** because they arise as compatibility conditions of linear ODEs. The set of linear ODE's is called **associated linear system**. Only very special nonlinear PDEs are integrable, and proving that a given collection of nonlinear PDE's is integrable has often been done by guessing a rewriting as compatibility equations.

Part of Equation 71 are actually a reformulation – after folding to dimension 2 – of the evolution equations:

$$(72)\quad \forall j \geq 1, \qquad j\partial_{t_j}\hat{Q} = \big[(\hat{Q}^j)^-, \hat{Q}\big]\,.$$

These are called the Toda hierarchy, and ∂_{t_j} generates the ***j*-th Toda flow**. Equations 72 expresses the compatibility of the semi-infinite system:

$$x p_n(x) = \sum_{m\geq 0}^{n} \hat{Q}_{nm} p_m(x), \qquad j\partial_{t_j} p_n(x) = \sum_{m\geq 0} (\hat{Q}^j)_{mn} p_m(x)\,,$$

so the Toda hierarchy is integrable according to our definition.

A system of equations of the form:

$$\partial_{t_j} A = [L, A]$$

for an unknown matrix or operator A is called a **Lax system**. If implies that the eigenvalues of A are conserved quantities under the evolution of all t_j's. Equivalently, the spectral invariants $(\mathrm{Tr}\, A^k)_{k\geq 1}$ are conserved quantities:

9.3 LEMMA. *The Toda hierarchy has the property to be* ***hamiltonian****, i.e. the j-th equation can be written in Hamilton-Jacobi form:*

$$(73)\quad \forall j \geq 1, \qquad \partial_{t_j} u_n = \frac{\partial H_j}{\partial v_n}, \qquad \partial_{t_j} v_n = -\frac{\partial H_j}{\partial u_n},$$

with hamiltonians $H_j = -\frac{\mathrm{Tr}\,\hat{Q}^{j+1}}{j(j+1)}$.

The proof is left as exercise. Formally, the H_j are conserved quantities. We

[20]Equations (69)-(70) actually expresses the compatibility of the difference equation – with respect to the discrete variable n – of the ∂_x and ∂_{t_j} systems.

shall describe below this structure for $j=1$ and 2 only.

There exist actually several notions of "integrable systems", not necessarily equivalent. It turns out that for almost all known integrable PDE's, a representation in Lax form has been found, thus giving a systematic way to produce a countable set of independent conserved quantities. Integrable hamiltonian PDE's form a (large) subclass of integrable PDE's, to which the Toda hierarchy belongs.

First Toda equation

Let us compute the differential equation with respect to t_1 satisfied by the coefficients of $\hat{Q}$. Since $\hat{Q}$ is a band matrix of width 1 and $\hat{Q}^-$ is strict lower triangular, and $\hat{Q}_{nn+1}=1$ anyway, we have two non trivial equations:

$$\left\{\begin{array}{lcl} \partial_{t_1}\hat{Q}_{nn} & = & \hat{Q}_{nn-1}\hat{Q}_{n-1n} - \hat{Q}_{nn+1}\hat{Q}_{n+1n} \\ \partial_{t_1}\hat{Q}_{nn-1} & = & \hat{Q}_{nn-1}\hat{Q}_{n-1n-1} - \hat{Q}_{nn}\hat{Q}_{nn-1} \end{array}\right. . \tag{74}$$

In terms of u's and v's they become:

$$\left\{\begin{array}{l} \partial_{t_1} v_n = e^{u_n-u_{n-1}} - e^{u_{n+1}-u_n} \\ \partial_{t_1}(u_n-u_{n-1}) = v_{n-1}-v_n \end{array}\right. .$$

By summing the second equation over n and using the fact that (74) is also valid for $n=0$ with the convention $u_{-j}=v_{-j}=0$ for $j>0$, we obtain:

$$\left\{\begin{array}{lcl} \partial_{t_1} v_n & = & e^{u_n-u_{n-1}} - e^{u_{n+1}-u_n} \\ \partial_{t_1} u_n & = & -v_n \end{array}\right. .$$

These are nonlinear difference-differential equations, which take the Hamilton-Jacobi form (73), with Hamiltonian:

$$H_1 = -\sum_{n\geq 0}\left\{\frac{v_n^2}{2} + e^{u_n-u_{n-1}}\right\}.$$

We indeed remark that:

$$H_1 = -\frac{1}{2}\,\mathrm{Tr}\,\hat{Q}^2$$

Although the sum in H_1 is semi-infinite and there is an issue of convergence, its derivative with respect to its variables u_n and v_n are finite sums, so the Hamilton-Jacobi equations are unambiguously defined.

Second Toda equation

To write down the evolution equation (72) with respect to t_2, we need to compute $(Q^2)^-$. Its non-zero terms are:

$$\begin{array}{lcl} (\hat{Q}^2)_{nn-1} & = & \hat{Q}_{nn}\hat{Q}_{nn-1} + \hat{Q}_{nn-1}\hat{Q}_{n-1n-1} = e^{u_n-u_{n-1}}(v_n+v_{n-1}) \\ (\hat{Q}^2)_{nn-2} & = & \hat{Q}_{nn-1}\hat{Q}_{n-1n-2} = e^{u_n-u_{n-2}}. \end{array}$$

Then:

$$\left\{\begin{array}{l} 2\partial_{t_2} v_n = e^{u_n - u_{n-1}}(v_n + v_{n-1}) - e^{u_{n+1}-u_n}(v_{n+1} + v_n) \\ 2\partial_{t_2}(u_n - u_{n-1}) = v_{n-1}^2 - v_n^2 + (e^{u_{n-1}-u_{n-2}} - e^{u_n - u_{n-1}}) - (e^{u_{n+1}-u_n} - e^{u_n - u_{n-1}}) \end{array}\right. .$$

Once again, summing the second equation over n, we obtain:

$$\left\{\begin{array}{lcl} 2\partial_{t_2} v_n & = & -v_{n+1}e^{u_{n+1}-u_n} + v_n(e^{u_n-u_{n-1}} - e^{u_{n+1}-u_n}) + v_{n-1}e^{u_n-u_{n-1}} \\ 2\partial_{t_2} u_n & = & -v_n^2 - e^{u_n - u_{n-1}} - e^{u_{n+1}-u_n} \end{array}\right. .$$

This can again be put in Hamilton-Jacobi form (73), with Hamiltonian:

$$H_2 = -\frac{1}{2}\sum_{n\geq 0}\left\{\frac{v_n^3}{3} + v_n(e^{u_n-u_{n-1}} + e^{u_{n+1}-u_n})\right\}.$$

We remark that:

$$H_2 = -\frac{1}{6}\mathrm{Tr}\,\hat{Q}^3$$

again in agreement with Lemma 9.3.

Reduction: Volterra equation

If the measure μ is even, we have $v_n = 0$. Then, the second Toda equation (72) reduces to an equation on $R_n := \hat{Q}_{nn-1} = h_n/h_{n-1}$:

$$2\partial_{t_2} R_n = R_n(R_{n-1} - R_{n+1}),$$

which is called the Volterra equation. Studying the continuum limit of this difference equation, one can find in a suitable regime the Korteweg-de Vries equation:

$$\partial_\tau f = \partial_{\xi^3} f + 6f\partial_\xi f\,.$$

The latter is one of the first nonlinear PDEs which have been shown to be integrable in the late 60s, first by finding a countable number of conserved quantities, and later by representing it as the compatibility of a Lax system.

10 Beta ensembles and potential theory

The β ensembles are defined by the probability measure on $\mathbb{R}^N$:

$$\text{(75)}\quad \mathrm{d}\mu_N(\lambda_1,\ldots,\lambda_N) = \frac{1}{Z_N}\prod_{1\leq i<j\leq N}|\lambda_i-\lambda_j|^{\beta}\prod_{i=1}^{N} e^{-N(\beta/2)V(\lambda_i)}\mathrm{d}\lambda_i\,.$$

$V:\mathbb{R}\to\mathbb{R}$ is a continuous function called **potential**, or sometimes **external field**. We assume it is **confining**:

$$\text{(76)}\quad \liminf_{|x|\to\infty}\frac{V(x)}{2\ln|x|} > 1\,.$$

This assumption guarantees that V grows fast enough at infinity for the integral to be absolutely convergent, and but also – as we will see – that most of the interesting behavior when $N\to\infty$ arise from configurations of λ_i's in a region bounded uniformly in N. We wish to obtain the large N behavior of the partition function:

$$Z_N = \int_{\mathbb{R}^N}\prod_{1\leq i<j\leq N}|\lambda_i-\lambda_j|^{\beta}\prod_{i=1}^{N}e^{-N(\beta/2)V(\lambda_i)}\mathrm{d}\lambda_i = \int_{\mathbb{R}^N} e^{-N^2(\beta/2)\mathcal{E}_\Delta(L_N^{(\lambda)})}\prod_{i=1}^{N}\mathrm{d}\lambda_i$$

and study the convergence when of the random probability measure $L_N^{(\lambda)}$. This will be achieved in § 10.4.

The configuration of λ_i's which minimize the quantity:

$$\text{(77)}\quad \begin{aligned}\mathcal{E}_\Delta(L_N^{(\lambda)}) &= \sum_{1\leq i<j\leq N}-\ln|\lambda_i-\lambda_j| + \sum_{i=1}^{N}V(\lambda_i)\\ &= \iint_{x\neq y}-\ln|x-y|\mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y) + \int V(x)\mathrm{d}L_N^{(\lambda)}(x)\\ &= \iint_{x\neq y}\Big(-\ln|x-y| + \frac{V(x)+V(y)}{2}\Big)\,\mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y)\end{aligned}$$

should give the dominant contribution, when $N\to\infty$, to Z_N. For a fixed N, the configurations $(\lambda_1^{(N)},\ldots,\lambda_N^{(N)})$ minimizing (77) are called the **Fekete points**. They play an important role in approximation theory, because they are in some sense "optimal" interpolation points. We will not study Fekete points here, but since we are interesting in the limit $N\to\infty$, we will study the auxiliary problem of minimizing the **energy functional**:

$$\mathcal{E}(\mu) := \iint E(x,y)\,\mathrm{d}\mu(x)\mathrm{d}\mu(y),\qquad E(x,y) := -\ln|x-y| + \frac{1}{2}\big(V(x)+V(y)\big)$$

over all probability measures $\mu\in\mathcal{M}_1(\mathbb{R})$. This is the continuous version of the minimizing problem for Fekete points. Most of the information about the large N limit of the model can be extracted from the energy functional.

10.1 *The energy functional*

Since the logarithm singularity is integrable, $\mathcal{E}(\mu)$ is finite at least when μ has compact support and is Lebesgue continuous. But we have to justify that it is well-defined for any probability measure μ. We also establish the basic properties of this functional:

10.1 LEMMA. *For any probability measure μ, $\mathcal{E}(\mu)$ is well-defined in $\mathbb{R} \cup \{+\infty\}$. Besides, for any $M \in \mathbb{R}$, its level sets:*

$$H_M = \{\mu \in \mathcal{M}_1(\mathbb{R}), \quad \mathcal{E}(\mu) \leq M\}$$

are compact. In particular, they are closed, i.e. $\mathcal{E}$ is lower semi-continuous.

Proof. Let us regularize the logarithmic singularity by defining:

$$\begin{aligned} \mathcal{E}_{(\epsilon)}(\mu) &:= \iint\limits_{[-1/\epsilon,1/\epsilon]^2} E_{(\epsilon)}(x,y)\,\mathrm{d}\mu(x)\mathrm{d}\mu(y) \\ E^{(\epsilon)}(x,y) &:= -\frac{1}{2}\ln((x-y)^2+\epsilon^2) + \frac{1}{2}(V(x)+V(y)). \end{aligned}$$

$\mathcal{E}_{(\epsilon)}(\mu)$ is well-defined for any $\epsilon > 0$, and is a decreasing function of ϵ. Therefore, the following limit exists and defines $\mathcal{E}(\mu)$:

$$\mathcal{E}(\mu) := \lim_{\epsilon \to 0} \mathcal{E}_{(\epsilon)}(\mu) \in \mathbb{R} \cup \{+\infty\}\,.$$

Since $E_{(\epsilon)}$ is continuous and bounded on $[-1/\epsilon, 1/\epsilon]^2$, the functional $\mathcal{E}_{(\epsilon)}$ is continuous over $\mathcal{M}_1(\mathbb{R})$ for the weak topology. Then, $\mathcal{E}$ is lower semi-continuous as the supremum of continuous functionals. This means that, for any sequence $(\mu_n)_n$ converging to μ, we have the inequality:

$$\liminf_{n\to\infty} \mathcal{E}(\mu_n) \geq \mathcal{E}(\mu)\,.$$

This is equivalent to the property that, for any M, the level set H_M is closed. We will now check that H_M is tight: this imply compactness by Prokhorov theorem. Since we assumed that V is confining (Equation (76)), $\lim_{|x|\to\infty} V(x) = +\infty$ and there exist $c > 0$ and $c' \in \mathbb{R}$ such that[21]:

$$\forall x,y \in \mathbb{R}, \qquad E(x,y) \geq c\big(|V(x)| + |V(y)|\big) + c'\,.$$

The terms proportional to c are nonnegative, so we have a lower bound:

$$\mathcal{E}(\mu) \geq c v_m\, \mu([-m,m]^c) + c'$$

with $v_m = \inf_{x\in[-m,m]^c}|V|$ going to $+\infty$ when $m \to +\infty$. The left-hand side has a uniform upper bound by M when $\mu \in H_M$. Therefore, $\mu([-m,m]^c)$ converges to 0 uniformly for $\mu \in H_M$ when $m \to \infty$, i.e. H_M is tight, and we conclude the proof. ☐

[21] The left-hand side is $+\infty$ when $x = y$, but this does not spoil the lower bound.

We now address the problem of minimizing the energy functional.

10.2 LEMMA. *$\mathcal{E}$ is strictly convex.*

Proof. Let $\mu_0, \mu_1 \in \mathcal{M}_1(\mathbb{R})$. For any $t \in [0,1]$, $\mu_t = (1-t)\mu_0 + t\mu_1$ is a probability measure, and we decompose:

$$\mu_t \otimes \mu_t = (1-t)\mu_0^{\otimes 2} + t\mu_1^{\otimes 2} - t(1-t)(\mu_0 - \mu_1)^{\otimes 2}\,. \tag{78}$$

We want to prove the convexity inequality:

$$\mathcal{E}(\mu_t) \leq (1-t)\mathcal{E}(\mu_0) + t\mathcal{E}(\mu_1)\,,$$

with equality iff $\mu_1 = \mu_0$. We can assume $\mathcal{E}(\mu_0)$ and $\mathcal{E}(\mu_1)$ finite, since the bound is trivial otherwise. It makes sense to integrate (78) against $E(x,y)$ term by term, since the two first terms have finite integrals, while the last integral belongs a priori to $\mathbb{R} \cup \{-\infty\}$. $\nu = \mu_0 - \mu_1$ is a finite, signed measure of mass zero. Therefore:

$$\begin{aligned} \mathcal{Q}(\nu) &:= \iint_{\mathbb{R}^2} \mathrm{d}\nu(x)\mathrm{d}\nu(y)\Big(-\ln|x-y| + \frac{V(x)+V(y)}{2}\Big) \\ &= \lim_{\epsilon\to 0} \int_{\mathbb{R}^2} -\frac{1}{2}\ln((x-y)^2+\epsilon^2)\,\mathrm{d}\nu(x)\mathrm{d}\nu(y)\,. \end{aligned} \tag{79}$$

At this point, we use an integral representation of the logarithm:

$$-\frac{1}{2}\ln(u^2+\epsilon^2) = \mathrm{Re}\Big(\int_0^\infty \frac{e^{\mathrm{i}us}-1}{s}\,e^{-\epsilon s}\mathrm{d}s\Big) - \ln\epsilon$$

to find:

$$\begin{aligned} \mathcal{Q}(\nu) &= \lim_{\epsilon\to 0}\mathrm{Re}\Big(\iint_{\mathbb{R}^2}\mathrm{d}\nu(x)\mathrm{d}\nu(y)\int_0^\infty \frac{e^{\mathrm{i(x-y)s}}}{s}\,e^{-\epsilon s}\mathrm{d}s\Big) \\ &= \lim_{\epsilon\to 0}\int_0^\infty \frac{|\widehat{\nu}(s)|^2}{s}\,e^{-\epsilon s}\,\mathrm{d}s = \int_0^\infty \frac{|\widehat{\nu}(s)|^2}{s}\,\mathrm{d}s\,. \end{aligned} \tag{80}$$

We have used Fubini theorem to go from the first line to the second line, since the integral for any fixed $\epsilon > 0$ is absolutely convergent. In the second line, we have recognized the Fourier transform $\widehat{\nu}(s)$ of the measure ν. The last integral obviously belongs to $[0,+\infty]$, with equality iff $\widehat{\nu}(s) = 0$ almost everywhere, i.e. $\nu \equiv 0$. So, we have:

$$\mathcal{E}(\mu_t) = (1-t)\mathcal{E}(\mu_0) + t\mathcal{E}(\mu_1) - t(1-t)\mathcal{Q}(\mu_0-\mu_1) \geq (1-t)\mathcal{E}(\mu_0) + t\mathcal{E}(\mu_0)\,,$$

with equality iff $\mu_0 = \mu_1$. ☐

10.3 COROLLARY. *$\mathcal{E}$ has a unique minimizer over $\mathcal{M}_1(\mathbb{R})$, denoted μ_{eq}. It is called* *the **equilibrium measure**.*

Proof. Since $\mathcal{E}(\mu)$ is finite for the Lebesgue measure supported on a compact, the level set H_M is non empty for some M large enough. Since the level set H_M is compact (Lemma 10.1), $\mathcal{E}$ achieves its minimum. The strict convexity of $\mathcal{E}$ (Lemma 10.2) guarantees that there is a unique minimizer. □

10.2 *The equilibrium measure*

We define the **effective potential**:

$$V_{\text{eff}}(x) := V(x) - 2\int \ln|x-y|\mathrm{d}\mu_{\text{eq}}(y)\,.$$

10.4 THEOREM. *μ_{eq} is characterized by the following properties: there exists a finite constant C_{eff} such that:*

(i) $V_{\text{eff}}(x) = C_{\text{eff}}$ for x μ_{eq}-almost everywhere.

(ii) $V_{\text{eff}}(x) \geq C_{\text{eff}}$ for x (Lebesgue)-almost everywhere.

The minimization of $\mathcal{E}$ over $\mathcal{M}_1(\mathbb{R})$ can be seen as a minimization over the vector space of finite measures under the constraints that the total mass is 1, and the measure is positive. This respectively explains the presence of the Lagrange multiplier C_{eff}, and the inequality (ii).

We will see later in Corollary 10.14 that the empirical measure $L_N^{(\lambda)}$ converges to the equilibrium measure. It is therefore important to describe the properties of μ_{eq}, and hopefully compute it. Heuristically, the effective potential $V_{\text{eff}}(\lambda_i)$ can be interpreted as the potential felt by one eigenvalue in (75): it takes into account the term $NV(\lambda_i)$, as well as the collective effect of repulsion by all the other eigenvalues, which is approximated by

$$\int 2N\ln|x-y|\mathrm{d}L_N^{(\lambda)} \approx \int 2\ln|x-y|\mathrm{d}\mu_{\text{eq}}(y)$$

when N is large. Theorem 10.4 characterizes the equilibrium measure, as the one producing a constant – and minimum – effective potential on the support of μ_{eq}, i.e. the locus where the eigenvalues are typically expected to be.

10.5 REMARK. Theorem 10.4 tells us that V_{eff} has a minimum value C_{eff} on $\operatorname{supp}\mu_{\text{eq}}$, but nothing prevents it to reach its minimum value C_{eff} outside $\operatorname{supp}\mu_{\text{eq}}$ as well.

Proof. Let f, g be two measurable functions with compact support, such that $\inf f(x)\mathrm{d}\mu_{\text{eq}}(x) = -\int g(x)\mathrm{d}x$ and $g \geq 0$. For $t \geq 0$ small enough

$$\mu_t = \mu_{\text{eq}} + t(f\mu_{\text{eq}} + g\mathrm{d}x)$$

is a probability measure. So, we have $\mathcal{E}(\mu_{\text{eq}} + tv) \geq \mathcal{E}(\mu_{\text{eq}})$ by the minimization property. This is a quadratic polynomial in t:

$$\mathcal{E}(\mu_{\text{eq}} + t\nu) = \mathcal{E}(\mu_{\text{eq}}) + t \int V_{\text{eff}}(x)\, \mathrm{d}\nu(x) + t^2 \mathcal{Q}(\nu)$$

so the coefficient of t must be nonnegative. This reads:

$$\int V_{\text{eff}}(x) \left[f(x) \mathrm{d}\mu_{\text{eq}} + g(x) \mathrm{d}x \right] \geq 0\,.$$

Firstly, we choose $g = 0$, and apply this inequality for f and $-f$. We thus obtain the equality:

$$\int V_{\text{eff}}(x)\, f(x)\, \mathrm{d}\mu_{\text{eq}}(x) = 0$$

for any measurable f with compact support and such that $\int f(x) \mathrm{d}\mu_{\text{eq}}(x) = 0$. This implies for any measurable f with compact support:

$$\int (V_{\text{eff}}(x) - C_{\text{eff}})\, f(x)\, \mathrm{d}\mu_{\text{eq}}(x) = 0, \qquad C_{\text{eff}} = \int V_{\text{eff}}(x)\, \mathrm{d}\mu_{\text{eq}}(x)\,,$$

hence the equality $V_{\text{eff}}(x) = C_{\text{eff}}$ for x μ_{eq}-almost everywhere. Secondly, we choose an arbitrary measurable, positive, compactly supported function g, and take f to be the function equal to the constant $-\int g(x)\mathrm{d}x$ on the support of μ_{eq}, and 0 otherwise. This gives:

$$\int V_{\text{eff}}(x) \Big(g(x)\, \mathrm{d}x - \mathrm{d}\mu_{\text{eq}}(x) \int \mathrm{d}y\, g(y) \Big) = \int (V_{\text{eff}}(x) - C_{\text{eff}}) g(x)\, \mathrm{d}x \geq 0\,.$$

Therefore, $V_{\text{eff}}(x) - C_{\text{eff}} \geq 0$ for x Lebesgue almost everywhere. Conversely, let μ be a probability measure satisfying (i) and (ii). We can integrate the inequality $V(x) - 2\int \ln|x-y|\mathrm{d}\mu(y) \geq C$ against the measure $(\mu_{\text{eq}} - \mu)$ to find back:

$$\mathcal{E}[\mu + t(\mu_{\text{eq}} - \mu)] \geq \mathcal{E}[\mu] + t^2 \mathcal{Q}[\mu - \mu_{\text{eq}}] \geq \mathcal{E}(\mu)\,.$$

But convexity also gives an upper bound for the left-hand side:

$$(1-t)\mathcal{E}(\mu) + t\mathcal{E}(\mu_{\text{eq}}) \geq \mathcal{E}(\mu)\,.$$

Hence $\mathcal{E}(\mu_{\text{eq}}) \geq \mathcal{E}(\mu)$, thus μ is a minimizer. By uniqueness of the minimizer, $\mu = \mu_{\text{eq}}$. ☐

10.6 lemma. *μ_{eq} has compact support.*

Proof. We can write:

$$V_{\text{eff}}(x) = \int \big(V(x) - 2\ln|x-y| \big) \mathrm{d}\mu_{\text{eq}}(y)\,.$$

Since V is confining (Equation (76)), we know that for any y, $V(x) - 2\ln|x-y|$ goes to $+\infty$ when $|x| \to \infty$. But V_{eff} must be constant on the support of μ_{eq}. Therefore, this support must be compact. ☐

If the support of μ_{eq} is known a priori, the characterization of Theorem 10.4 gives μ_{eq} as a solution of linear equation. In practice, one often makes an assumption on the shape of the support (e.g. a union of segments), solves the linear equation, then find consistency equations that the support should satisfy. This usually leads to a finite number of possibilities. Among those, if one finds a solution which is a positive measure of mass 1, by uniqueness it must be the equilibrium measure, and this validates the assumption made initially on the support.

In general, it is difficult to compute completely μ_{eq}, and one can only rely on the characterization to establish qualitative properties of μ_{eq}. But in the situation where we know a priori that μ_{eq} is supported on a segment (the **one-cut regime**), the problem is amenable to a fairly explicit solution, presented in Theorem 10.8 or **??** below. We first give a sufficient – but not necessary – condition to be in the one-cut regime:

10.7 lemma. *If V is $\mathcal{C}^1$ and convex, then μ_{eq} is supported on a segment.*

Proof. Let S_{eq} be the support of μ_{eq}. For any y, $x \mapsto V(x) - 2\ln|x-y|$ is a convex function, which is strictly convex on the domain $x \neq y$. Therefore, integrating against $\mathrm{d}\mu_{\mathrm{eq}}(y)$, we deduce that $x \mapsto V_{\mathrm{eff}}(x)$ is convex, and strictly convex in S_{eq}^c. Then, the set S of values where it achieves its minimum is connected. Assume there is a segment $]x_-, x_+[\subseteq S$ that does not belong to S_{eq} but $x_\pm \in S_{\mathrm{eq}}$, with $x_- < x_+$. Then, differentiating the equality (i) and sending $x \in S_{\mathrm{eq}}$ to x_- or x_+, we obtain:

$$(81)\qquad \left\{\begin{array}{lcl} V'(x_-) - 2\partial_{x=x_-}\Big(\int \ln|x-y|\mathrm{d}\mu_{\mathrm{eq}}(y)\Big) & = & 0 \\ V'(x_+) - 2\partial_{x=x_+}\Big(\int \ln|x-y|\mathrm{d}\mu_{\mathrm{eq}}(y)\Big) & = & 0 \end{array}\right. .$$

By convexity, $V'(x_-) \geq V'(x_+)$, and by strict convexity of $\int \ln|x-y|\mathrm{d}\mu_{\mathrm{eq}}(y)$ on $x \in]x_0, x_1[$, we also have:

$$\partial_{x=x_-}\Big(\int \ln|x-y|\mathrm{d}\mu_{\mathrm{eq}}(y)\Big) < \partial_{x=x_+}\Big(\int \ln|x-y|\mathrm{d}\mu_{\mathrm{eq}}(y)\Big),$$

which is a contradiction with the equality of the two lines in (81). Therefore, we must have $S = S_{\mathrm{eq}}$, which is a segment. ☐

10.8 theorem (Tricomi formula). *Assume V $\mathcal{C}^2$. If the support of μ_{eq} is a segment $[b,a]$ with $b < a$, μ_{eq} is continuous with respect to Lebesgue, and its density reads:*

$$\frac{\mathrm{d}\mu_{\mathrm{eq}}}{\mathrm{d}x}(x) = \int_b^a \frac{V'(x) - V'(\xi)}{x-\xi}\sqrt{\frac{(a-x)(x-b)}{(a-\xi)(\xi-b)}}\,\frac{\mathrm{d}\xi}{2\pi}\,.$$

The endpoints a and b must satisfy the constraints:

$$(82)\qquad \int_b^a \frac{V'(\xi)\mathrm{d}\xi}{\sqrt{(a-\xi)(b-\xi)}} = 0, \qquad \frac{(a+b)}{2}\int_b^a \frac{\xi\, V'(\xi)\,\mathrm{d}\xi}{\sqrt{(a-\xi)(\xi-b)}} = 1\,.$$

In the course of the proof, we accept that the Stieltjes transform of μ_{eq} is continuous when $z \to a$ or b in $\mathbb{C} \setminus [b, a]$, and has continuous boundary values when $z \to 0$ with $\pm \text{Im}\, z > 0$. This fact will be justified in Chapter 11 using Schwinger-Dyson equations, see Corollary 11.2.

Proof. We prove the result for V polynomial, and it can be justified in general by approximations. We can differentiate the equality (i) for $x \in]b, a[$:

$$2\text{p.v} \int \frac{\mathrm{d}\mu_{\text{eq}}(y)}{x - y} = V'(x),$$

where p.v. denotes Cauchy principal value. Let us introduce the Stieltjes transform:

$$W(z) = \int \frac{\mathrm{d}\mu_{\text{eq}}(y)}{z - y}.$$

This is a holomorphic function of $z \in \mathbb{C} \setminus [b, a]$, which behaves like $W(z) \sim 1/z$ when $z \to \infty$. By definition of the principal value integral:

$$\forall x \in]b, a[, \qquad \lim_{\epsilon \to 0} W(x + \mathrm{i}\epsilon) + W(x - \mathrm{i}\epsilon) = V'(x).$$

We accepted that $W(z)$ has continuous boundary values on $[b, a]$, and is continuous when $z \to a, b$ in $\mathbb{C} \setminus [b, a]$, so this equation actually holds for $x \in [b, a]$. $V'(x)/2$ is a particular solution to this equation, so we can set:

$$W_{(0)}(z) = W(z) - \frac{V'(z)}{2},$$

which now satisfies:

$$\forall x \in [b, a], \qquad W_{(0)}(x + \mathrm{i}\epsilon) + W_{(0)}(x - \mathrm{i}\epsilon) = 0,$$

i.e. $W_{(0)}(z)$ takes a minus sign when z crosses the cut $[b, a]$. The function:

$$\sigma(z) = \sqrt{(z - a)(z - b)}$$

has the same behavior. σ is defined as the unique holomorphic function on $\mathbb{C} \setminus [b, a]$ which behaves like $\sigma(z) \sim z$ when $z \to \infty$. So, the function:

$$W_{(1)}(z) = \frac{W_{(0)}(z)}{\sigma(z)}$$

is now continuous when z crosses the cut $]b, a[$. A priori, it behaves like $O((z - a)^{-1/2})$ when $z \to a$, but this continuity implies that the coefficient in front of $(z - a)^{-1/2}$ vanishes, and it must actually be a $O(1)$, and similarly when $z \to b$. Therefore, $W_{(1)}$ is an entire function. Since V is a polynomial, $W_{(1)}$ has polynomial growth, so we can conclude by Liouville theorem that it is a polynomial. It just remains to compute the polynomial that makes our initial $W(z)$ be $1/z + o(1/z)$ when $z \to \infty$. This can be done elegantly with contour integrals. We write:

$$\begin{aligned} W_{(1)}(z) &= \operatorname*{Res}_{\zeta\to z} \frac{W_{(1)}(\zeta)\,\mathrm{d}\zeta}{\zeta - z} = -\operatorname*{Res}_{\zeta\to\infty} \frac{W_{(1)}(\zeta)\,\mathrm{d}\zeta}{\zeta - z} \\ &= \operatorname*{Res}_{\zeta\to\infty} \frac{\big[V'(\zeta)/2 + O(1/\zeta)\big]\,\mathrm{d}\zeta}{\sigma(\zeta)(\zeta - z)}\,. \end{aligned} \tag{83}$$

The $O(1/\zeta)$ does not contribute to the residue at ∞ because it comes with a prefactor which is already $O(1/\zeta)$. Then, moving back the contour to surround the cut of σ on $[b,a]$, we also pick up the residue at the simple pole $\zeta = z$:

$$W_{(1)}(z) = -\frac{V'(z)}{2\sigma(z)} + \oint_{[b,a]} \frac{\mathrm{d}\zeta}{4\mathrm{i}\pi} \frac{V'(\zeta)\,\mathrm{d}\zeta}{\sigma(\zeta)(z-\zeta)}\,,$$

and coming back to the Stieltjes transform itself:

$$W(z) = \oint_{[b,a]} \frac{\mathrm{d}\zeta}{4\mathrm{i}\pi} \frac{\sigma(z)}{\sigma(\zeta)} \frac{V'(\zeta)}{z-\zeta}\,. \tag{84}$$

Since $\sigma(z) = z - (a+b)/2 + O(1/z)$, we identify:

$$W(z) = c_0 + c_1/z + O(1/z^2)\,,$$

when $z \to \infty$ with:

$$c_0 = \oint_{[b,a]} \frac{\mathrm{d}\zeta}{4\mathrm{i}\pi} \frac{V'(\zeta)}{\sigma(\zeta)}, \qquad c_1 = -\frac{a+b}{2} \oint_{[a,b]} \frac{\mathrm{d}\zeta}{4\mathrm{i}\pi} \frac{\zeta\, V'(\zeta)}{\sigma(\zeta)}\,,$$

and we have to satisfy the constraints $c_0 = 0$ and $c_1 = 1$. Squeezing the contour integral to the segment $[b,a]$, and using that the signs of the integrand are opposite on the upper and lower side of $[b,a]$, we obtain the constraints (82). The density of μ_{eq} is obtained from the discontinuity of $W(z)$ when z crosses $[b,a]$. Since in (84) the point z is outside the contour, we can move it back inside the contour by writing:

$$W(z) = \frac{V'(z)}{2} + \sigma(z) R(z)\,,$$

with:

$$R(z) = \oint_{[b,a]} \frac{\mathrm{d}\zeta}{4\mathrm{i}\pi} \frac{V'(\zeta) - V'(z)}{\sigma(\zeta)(\zeta - z)} = \frac{1}{2\pi} \int_b^a \frac{\mathrm{d}\zeta}{\sqrt{(a-\zeta)(\zeta-b)}} \frac{V'(\zeta) - V'(z)}{\zeta - z}\,.$$

Then, it is easy to compute the discontinuity, since it only comes from the prefactor $\sigma(z)$, whose discontinuity at x is equal to $2\sigma(x)$:

$$\begin{aligned}\frac{\mathrm{d}\mu_{\mathrm{eq}}}{\mathrm{d}x}(x) &= \frac{1}{2\mathrm{i}\pi}\lim_{\epsilon\to 0^+}\big[W(x-\mathrm{i}\epsilon)-W(x+\mathrm{i}\epsilon)\big] \\ &= \frac{2\sigma(x)R(x)}{2\mathrm{i}\pi} = \frac{\sqrt{(a-x)(x-b)}}{\pi}\,R(x)\,. \end{aligned} \tag{85}$$

We observe that, if V is convex on $[b,a]$, then $R(x)\geq 0$, ensuring that μ_{eq} is positive measure. When V' is not convex, if the constraints (82) have several solutions (a,b), there must be unique one for which formula (85) gives a positive measure. ☐

10.3 *Large deviation theory: principles*

Lecture 18 (1h30)
January 12th, 2015

X denotes a Polish space, for instance $\mathbb{R}$ or $\mathcal{M}_1(\mathbb{R})$. We describe the basics of large deviation theory. A sequence $(\mu_n)_n$ of probability measures on X satisfies a **large deviation principle** (LDP) if:

- there exists a sequence of positive real numbers $(\alpha_n)_n$ such that
$$\lim_{n\to\infty}\alpha_n = +\infty.$$
- there exists a lower semi-continuous function $J\,:\,X\to[0,+\infty]$
- for any open set $\Omega\subseteq X$:
$$\liminf_{n\to\infty}\alpha_n^{-1}\ln\mu_n[\Omega]\geq -\inf_{x\in\Omega} I(x)\,.$$
- for any closed set $F\subseteq X$:
$$\limsup_{n\to\infty}\alpha_n^{-1}\ln\mu_n[F]\leq -\inf_{x\in F} I(x)\,.$$

α_n is called the **speed**, and I the **rate** of the LDP. In particular, by taking $F=X$ in the last inequality, one obtains that $\inf_{x\in X} I = 0$. I is a **good rate** function if for any M, the level set $I_M = I^{-1}([0,M])$ is compact. By the previous remark, if I is a good rate function, there must exist $x\in X$ such that $I(x)=0$.

$(\mu_n)_n$ satisfies a **weak LDP** if one requires only "F compact" in the last point. $(\mu_n)_n$ is **exponentially tight** if there exists an increasing sequence of compacts $(K_M)_{M\geq 0}$ such that:

$$\lim_{M\to\infty}\limsup_{n\to\infty}\alpha_n^{-1}\ln\mu_n[K_M^c] = -\infty\,.$$

10.9 lemma. *If $(\mu_n)_n$ is exponentially tight, a weak LDP implies a LDP.*

Proof. Let F is a closed set, and K_M a compact such that

$$\limsup_{n\to\infty}\alpha_n^{-1}\ln\mu_n[K_M^c] = -M.$$

Then:

$$\begin{aligned}\limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[F] &\leq \limsup_{n\to\infty} \alpha_n^{-1} \ln \left(\mu_n[F\cap K_M] + \mu_n[K_M^c]\right)\\ &\leq \max\left\{\limsup_{n\to\infty} \alpha_n^{-1} \ln \left(2\mu_n[F\cap K_M]\right), \limsup_{n\to\infty} \alpha_n^{-1} \ln \left(2\mu_n[K_M^c]\right)\right\}.\end{aligned}$$

We then use the LDP estimate on the compact $F \cap K_M$, and the definition of K_M:

$$\limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[F] \leq \max\left[-\inf_{x\in F\cap K_M} I(x), -M\right] \leq \max\left[-\inf_{x\in F} I(x), -M\right].$$

Taking the limit $M \to \infty$, we find:

$$\limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[F] \leq -\inf_{x\in F} I(x)\,.$$

□

10.10 lemma. *Assume that, for any $x \in X$:*

$$-I(x) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[B(x,\epsilon)] = \lim_{\epsilon\to 0} \liminf_{n\to\infty} \alpha_n^{-1} \ln \mu_n[B(x,\epsilon)]\,.$$

Then, $(\mu_n)_n$ satisfies a weak LDP.

Proof. Let Ω be an open set. If Ω is empty the inequality trivially holds, so we assume Ω non-empty. If $x \in \Omega$, the open ball $B(x,\epsilon)$ is contained in Ω for ϵ small enough. Thus:

$$\liminf_{n\to\infty} \alpha_n^{-1} \ln \mu_n[\Omega] \geq \lim_{\epsilon\to 0} \liminf_{n\to\infty} \alpha_n^{-1} \ln \mu_n[B(x,\epsilon)] = -I(x)\,.$$

Optimizing over $x \in \Omega$, we deduce:

$$\liminf_{n\to\infty} \alpha_n^{-1} \ln \mu_n[\Omega] \geq -\inf_{x\in\Omega} I(x).$$

Let F be a compact set. Since X is a Polish space, for any integer $m > 0$, there exists a finite covering $F \subseteq \bigcup_{i=1}^{k(\epsilon)} B(x_i^{(\epsilon)}, \epsilon)$ by balls with centers $x_i^{(\epsilon)} \in F$. Using an obvious union upper bound for the probability of F:

$$\begin{aligned}\limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[F] &\leq \lim_{\epsilon\to 0} \limsup_{n\to\infty} \alpha_n^{-1} \ln \Big(\sum_{i=1}^{k(\epsilon)} \mu_n[B(x_i^{(\epsilon)},\epsilon)]\Big)\\ &\leq \lim_{\epsilon\to 0} \max_{1\leq i\leq k(\epsilon)} \limsup_{n\to\infty} \alpha_n^{-1} \ln \left(k(\epsilon)\, \mu_n[B(x_i^{(\epsilon)},\epsilon)]\right)\\ &\leq \max_{\substack{\epsilon'>0\\ 1\leq i\leq k(\epsilon')}} \left\{\lim_{\epsilon\to 0} \limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[B(x_i^{(\epsilon')},\epsilon)]\right\}. \qquad (86)\end{aligned}$$

Using the assumption, we arrive to:

$$\limsup_{n\to\infty} \alpha_n^{-1} \ln \mu_n[F] \leq \max_{\substack{\epsilon'>0 \\ 1\leq i\leq k(\epsilon')}} \big\{ -I(x_i^{(\epsilon')})\big\} \leq -\inf_{x\in F} I(x)\,.$$

$\square$

Therefore, to derive an LDP, it is enough to establish exponential tightness, and an upper and a lower bound for probabilities of small balls in X. Indeed, Lemma 10.9 upgrades the latter to large deviation estimates for probabilities in compact sets (weak LDP), and with exponential tightness, Lemma 10.9 concludes for any closed sets (LDP).

A fundamental result in the theory of large deviations is:

10.11 lemma (Varadhan). *Assume that $(\nu_n)_n$ satisfies a LDP with good rate function I and speed α_n. Let $J : X \to \mathbb{R}$ continuous bounded. Then:*

$$\lim_{n\to\infty} \alpha_n^{-1} \nu_n\big[e^{\alpha_n J(x)}\big] = \sup_{x\in X} \big\{J(x) - I(x)\big\}\,.$$

This lemma gives the free energy of a statistical mechanics system in presence of a continuous bounded source J. In physically interesting models, one often face the problem of extending such results to sources J that do not satisfy these assumptions, and it usually requires some non-trivial work to reduce it to the use of Varadhan's lemma. As usual, singularities make all the richness of physics. We will see how it can be done in beta ensembles, where we typically want J to contain logarithmic singularities.

Proof. We prove the existence, and compute the limit by showing an upper and a lower bound which coincide. We start with the lower bound estimate. Fix $\epsilon > 0$. For any $x \in X$, since J is lower semicontinuous, there exists an open neighborhood $U_{x,\epsilon}$ of x such that

$$\inf_{y\in U_{x,\epsilon}} J(y) \geq J(x) - \epsilon\,.$$

We can write:

$$\begin{aligned}
\liminf_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_X \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big] \quad & \geq \quad \lim_{\epsilon\to 0} \liminf_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_{U_{x,\epsilon}} \mathrm{d}\nu(y)\, e^{\alpha_n (J(x)-\epsilon)}\Big] \\
& \geq \quad \lim_{\epsilon\to 0} \big\{J(x) - \epsilon - \inf_{y\in U_{x,\epsilon}} I(y)\big\} \\
& \geq \quad \lim_{\epsilon\to 0} \big\{J(x) - I(x) - \epsilon\big\} = J(x) - I(x)\,.
\end{aligned}$$

Optimizing over $x \in X$, we find:

$$\liminf_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_X \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big] \geq \sup_{x\in X} \big\{J(x) - I(x)\big\}\,.$$

Now, we turn to the upper bound estimate. Using the lower semi-continuity of I and the upper semi-continuity of J, we denote $V_{x,\epsilon}$ a closed neighborhood of $x \in X$ such that:

$$(87) \quad \inf_{y \in V_{x,\epsilon}} I(y) \geq I(x) - \epsilon, \qquad \sup_{y \in V_{x,\epsilon}} J(y) \leq J(x) + \epsilon\,.$$

Fix a real number M. By assumption the level set I_M is compact. So, there exists a finite covering $X = \bigcup_{i=1}^{k(M)} V_{x_i,\epsilon} \cap \overline{I_M^c}$. With the maximum bound already used in the proof of the previous lemma:

$$\begin{aligned}
&\limsup_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_X \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big] \\
&\leq \max\Big\{\max_{1\leq i\leq k(M)} \limsup_{n\to\infty} \alpha_n^{-1}\ln\Big[\int_{V_{x_i,\epsilon}} \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big], \limsup_{n\to\infty}\alpha_n^{-1}\ln\Big[\int_{\overline{I_M^c}} \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big]\Big\}.
\end{aligned}$$

For the integrals over $V_{x_i,\epsilon}$, we use the inequality (87) for J, the LDP for ν_n, and then the inequality (87) for I. For the integral over I_M^c, we use the boundedness of J, the LDP for ν_n and the definition of the level set. All in all:

$$\limsup_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_X \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big] \leq \max\Big\{\max_{1\leq i\leq k(M)} J(x_i) - I(x_i) + 2\epsilon\,;\, \sup_{x\in X} J(x) - M\Big\}.$$

The innermost maximum is bounded by the supremum of $J(x) - I(x)$ when x runs over X, and then in the limit $M \to +\infty$:

$$\limsup_{n\to\infty} \alpha_n^{-1} \ln\Big[\int_X \mathrm{d}\nu_n(y)\, e^{\alpha_n J(y)}\Big] \leq \sup_{x\in X}\big\{J(x) - I(x)\big\}\,.$$

Consequently, $\alpha_n^{-1}\nu_n[e^{\alpha_n J}]$ has a limit given by $\sup_{x\in X}\{J(x) - I(x)\}$. □

10.4 *Large deviations of the empirical measure*

We consider the β ensemble:

$$\mathrm{d}\mu_N(\lambda_1,\ldots,\lambda_N) = \frac{1}{Z_N}\prod_{i=1}^N \mathrm{d}\lambda_i\, e^{-N(\beta/2)V(\lambda_i)} \prod_{1\leq i<j\leq N} |\lambda_i - \lambda_j|^\beta\,,$$

with a potential V which is continuous and confining (Equation (76)). This is a probability measure on $\mathbb{R}^N$, that can be rewritten:

$$\mathrm{d}\mu_N(\lambda_1,\ldots,\lambda_N) = \frac{1}{Z_N}\prod_{i=1}^N \mathrm{d}\lambda_i\, e^{-(\beta/2)V(\lambda_i)}\, \exp\Big(-\frac{\beta N^2}{2}\iint_{x\neq y} E(x,y)\,\mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y)\Big)\,,$$

with:

$$E(x,y) = -\ln|x-y| + \frac{1}{2}\big(V(x) + V(y)\big)\,.$$

Notice the absence of the factor N in $e^{-(\beta/2)V(x)}$. It is due to the fact that the integral of $V(x)\mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y)$ avoiding the diagonal $x = y$ produces $\frac{(N-1)}{N^2}\sum_{i=1}^{N} V(\lambda_i)$. We denote $\mathbb{P}_N$ the probability measure on $X = \mathcal{M}_1(\mathbb{R})$, which is the direct image of the probability measure μ_n on $\mathbb{R}^n$, by the map:

$$
\begin{array}{rccc} (88) & \mathbb{R}^n & \longrightarrow & \mathcal{M}_1(\mathbb{R}) \\ & (\lambda_1,\ldots,\lambda_n) & \longmapsto & L_N^{(\lambda)} = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}\,. \end{array}
$$

In practice, it means that the probability of a measurable subset $A \subseteq \mathcal{M}_1(\mathbb{R})$ is computed by the formula:

$$
\mathbb{P}_N[A] = \int_{\{\lambda\in\mathbb{R}^N\,:\,L_N^{(\lambda)}\in A\}} \mathrm{d}\mu_N(\lambda_1,\ldots,\lambda_N)\,.
$$

We have introduced in Chapter 10 the energy functional:

$$
\mathcal{E}(\mu) = \iint E(x,y)\,\mathrm{d}\mu(x)\mathrm{d}\mu(y) \in \mathbb{R}\cup\{+\infty\}\,,
$$

and we have shown that it is lower semi-continuous, has compact level sets, and admits a unique minimizer μ_{eq}. Therefore:

$$
(89)\quad I(\mu) := \frac{\beta}{2}\big[\mathcal{E}(\mu) - \mathcal{E}(\mu_{\mathrm{eq}})\big] \in [0,+\infty]
$$

is a good rate function on $\mathcal{M}_1(\mathbb{R})$. The main results we shall prove are:

10.12 theorem. *We have:*

$$
\lim_{N\to\infty}\frac{\ln Z_N}{N^2} = -\frac{\beta}{2}\,\mathcal{E}(\mu_{\mathrm{eq}})\,.
$$

10.13 theorem. $(\mathbb{P}_N)_{N\geq 1}$ *satisfies a LDP with speed* N^2 *and good rate function* I *given by* (89).

10.14 corollary. *The empirical measure* $L_N^{(\lambda)}$ *– which is a* $\mathcal{M}^1(\mathbb{R})$ *random variable with law given by* $\mathbb{P}_N$ *– converges almost surely (for the weak topology) to the equilibrium measure* μ_{eq}.

$\mathcal{M}$^1 valued

Proof. (of the Corollary) It is enough to prove that, for any $\epsilon > 0$, the series $\sum_{N\geq 1}\mathbb{P}_N[\mathfrak{d}(L_N,\mu_{\mathrm{eq}}) \geq \epsilon]$ converges. $\mathfrak{d}$ is the Vasershtein distance, see Chapter 0. But the upper bound in the LDP implies, for N large enough:

$$
\mathbb{P}_N \leq e^{-N^2 c(\epsilon)}, \qquad c(\epsilon) = \frac{1}{2}\inf_{d(\mu,\mu_{\mathrm{eq}})\geq\epsilon}\mathcal{E}(\mu)\,.
$$

Since the minimizer of $\mathcal{E}$ is unique, $c(\epsilon)$ is positive, so the series converges. □

For the Gaussian ensembles, μ_{eq} is the semi-circle law and we retrieve Wigner theorem for $\beta = 1,2,4$. We actually obtain here a stronger version of

Wigner theorem, because it ensures almost sure convergence – and not only convergence in expectation.

Theorem 10.12 will be a side consequence of the proof of Theorem 10.13. As explained in § 10.3, we will prove exponential tightness, and estimate the probability of small balls to prove the LDP. We denote $\overline{\mathbb{P}_N} = Z_N \cdot \mathbb{P}_N$ the measure on $\mathcal{M}_1(\mathbb{R})$ of total mass Z_N.

Exponential tightness

Our candidate is $K_M = \{\mu \in \mathcal{M}_1(\mathbb{R}),\ \int |V(x)| \mathrm{d}L_N(x) \leq M\}$. First of all, since V is proper, K_M is indeed compact. And, since V is confining, there exist $c > 0$ and $c' \in \mathbb{R}$ such that:

$$E(x,y) \geq c\big(|V(x)| + |V(y)|\big) + c'\,.$$

Therefore:

$$\mathbb{P}_N[K_M^c] \leq \frac{1}{Z_N} \int_{K_M^c} \exp\Big[-\beta N^2\Big(c\int |V(x)|\mathrm{d}L_N(x) + c'/2\Big)\Big] \leq \frac{e^{-\beta N^2(cM + c'/2)}}{Z_N}\,.$$

To conclude, we need a rough lower bound on the partition function. It can be obtained for instance by using Jensen inequality with respect to the probability measure:

$$\text{(90)}\quad \mathrm{d}\nu_N(\lambda_1,\ldots,\lambda_N) := \prod_{i=1}^{N} \frac{\mathrm{d}\lambda_i\, e^{-(\beta/2)V(\lambda_i)}}{m}, \qquad m := \int_{\mathbb{R}} \mathrm{d}x\, e^{-(\beta/2)V(x)}\,.$$

It indeed gives:

$$\ln\Big(\frac{Z_N}{m^N}\Big) \geq -\frac{\beta N^2}{2} \int_{\mathbb{R}^N} \mathrm{d}\nu_N(\lambda_1,\ldots,\lambda_N) \int_{x\neq y} \mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y)\, E(x,y) \geq -CN^2$$

for some finite constant C since the logarithmic singularity is integrated against the Lebesgue continuous measure $\mathrm{d}\nu_N$. Thus:

$$\limsup_{N\to\infty} N^{-2} \ln \mathbb{P}_N[K_M^c] \leq C - \beta(cM + c'/2)\,,$$

and:

$$\lim_{M\to\infty} \limsup_{N\to\infty} N^{-2} \ln \mathbb{P}_N[K_M^c] = -\infty\,.$$

Upper bound for probability of small balls

If $M > 0$, we denote $E_M := \min(E, M)$, so that we always have $E_M \leq E$, and the functional $\mathcal{E}_M(\nu) := \iint E_M(x,y)\, \mathrm{d}\nu(x)\mathrm{d}\nu(y)$ is continuous. We can estimate:

$$\overline{\mathbb{P}_N}[B(\nu,\epsilon)] \leq \int_{\mathfrak{d}(L_N^{(\lambda)},\nu)\leq\epsilon} e^{-N^2(\beta/2)\mathcal{E}_M(L_N^{(\lambda)})} \prod_{i=1}^{N} e^{-(\beta/2)V(\lambda_i)}\mathrm{d}\lambda_i\,.$$

Since E_M is regular on the diagonal, the functional is continuous. Repeating the steps of the proof of Varadhan lemma (Lemma 10.11), one can prove for any fixed M:

$$\lim_{N\to\infty} N^{-2}\ln\Bigg(\int_{\mathfrak{d}(L_N^{(\lambda)},\nu)\leq\epsilon} e^{-N^2(\beta/2)\mathcal{E}_M(L_N^{(\lambda)})} \prod_{i=1}^{N} e^{-(\beta/2)V(\lambda_i)}\mathrm{d}\lambda_i\Bigg) = -\inf_{\mathfrak{d}(\nu',\nu)\leq\epsilon}\mathcal{E}_M(\nu')\,.$$

Since $\mathcal{E}_M$ is continuous, we can take the limit $\epsilon\to 0$:

$$\lim_{\epsilon\to 0}\limsup_{N\to\infty} N^{-2}\,\ln\overline{\mathbb{P}_N}[B(\nu,\epsilon)] \leq -\mathcal{E}_M(\nu)\,.$$

We can then take the limit $M\to\infty$ in the right-hand side by monotone convergence. Hence:

$$\limsup_{N\to\infty} N^{-2}\,\ln\overline{\mathbb{P}_N}[B(\nu,\epsilon)] \leq -\inf_{\mathfrak{d}(\nu',\nu)\leq\epsilon}(\beta/2)\mathcal{E}(\nu')\,.$$

Similarly, if we integrate not over the ball of radius ϵ but over all the space $\mathcal{M}_1(\mathbb{R})$:

$$\limsup_{N\to\infty} N^{-2}\,\ln Z_N \leq -\inf_{\nu'\in\mathcal{M}_1(\mathbb{R})}(\beta/2)\mathcal{E}(\nu') = -(\beta/2)\mathcal{E}(\mu_{\mathrm{eq}})\,. \tag{91}$$

Lower bound for probability of small balls

We want to prove that, for all $\nu\in\mathcal{M}_1(\mathbb{R})$:

$$\lim_{\epsilon\to 0}\liminf_{N\to\infty} N^{-2}\ln\overline{\mathbb{P}_N}[B(\nu,\epsilon)] \geq -\mathcal{E}(\nu)\,. \tag{92}$$

We can assume that ν is such that $\mathcal{E}(\nu)<+\infty$, otherwise the inequality holds trivially. This means in particular that ν has no atoms. One can also assume that ν has compact support, since the result for general ν can be deduced by approximation.

We define the **semiclassical positions** of ν as $\xi_{N;i}$ for $i\in[\![1,N]\!]$:

$$\xi_{N;i} = \inf\Big\{x\in\mathbb{R},\quad \nu\big(]-\infty,x]\big)\geq\frac{i}{N+1}\Big\}\,.$$

We remark that, since ν is assumed to have compact support, $\xi_{N;i}$ all belong to a compact K that is independent of i and N.

By approximation of the integral, for any $\epsilon > 0$, there exists N_ϵ such that:

$$\forall N \geq N_\epsilon, \qquad \mathfrak{d}\Big(\tilde{\nu}_N, \nu\Big) < \frac{\epsilon}{2}, \qquad \tilde{\nu}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{N;i}}\,,$$

and this implies:

$$\forall N \geq N_\epsilon, \quad B(\tilde{\nu}_N, \epsilon/2) \subseteq B(\nu, \epsilon)\,.$$

Further, let:

$$\Lambda_{N,\epsilon} := \big\{\lambda \in \mathbb{R}^N\,, \quad 0 \leq \lambda_1 < \ldots < \lambda_N \leq \epsilon/2\big\}.$$

Reminding that the Vasershtein distance is the supremum over a class of 1-Lipschitz functions, we have:

$$(93) \quad \big\{\mu \in \mathcal{M}_1(\mathbb{R}), \quad \exists \lambda \in \Lambda_{N,\epsilon},\ \mu = L_N^{(\lambda)}\big\} \subseteq B(\tilde{\nu}_N, \epsilon/2).$$

We then integrate on the smaller event to write a lower bound:

$$\lim_{\epsilon \to 0} \liminf_{N \to \infty} N^{-2} \ln \overline{\mathbb{P}_N}[B(\nu,\epsilon)] \geq \lim_{\epsilon \to 0} \liminf_{N \to \infty} N^{-2} \ln Z_{N,\epsilon}(\nu)$$

with:

$$Z_{N,\epsilon}(\nu) := \int_{0 \leq x_1 < \ldots < x_N \leq \epsilon/2} \prod_{i=1}^{N} \mathrm{d}x_i\, e^{-N(\beta/2)V(\xi_{N;i}+x_i)} \prod_{1 \leq i < j \leq N} |\xi_{N;i} - \xi_{N;j} + x_i - x_j|^\beta\,.$$

Since both $\xi_{N;i}$ and x_i are increasing with i, the interaction term always satisfies the lower bound:

$$\begin{aligned} |\xi_{N;i} - \xi_{N;j} + x_i - x_j| & \geq & |\xi_{N;i} - \xi_{N;j}| \\ & \geq & |x_i - x_j| \\ & \geq & |\xi_{N;i} - \xi_{N;j}|^{1/2} \cdot |x_i - x_j|^{1/2}\,. \end{aligned}$$

It is convenient, as we will see later in the proof of Lemma 10.4, to use the third bound for nearest neighbors (i.e. $i+1 = j$), and the first bound for all other pairs $\{i,j\}$. Our lower bound for the integral splits into:

$$(94) \quad Z_{N,\epsilon}(\nu) \geq Z_{N,\epsilon}^{(\mathrm{int})}(\nu) \cdot Z_N^{(0)}(\nu)\,,$$

with:

$$\begin{aligned} Z_{N,\epsilon}^{(\mathrm{int})}(\nu) & = & \int_{0 \leq x_1 < \ldots < x_N \leq \epsilon/2} \prod_{i=1}^{N-1} |x_{i+1} - x_i|^{\beta/2} \prod_{i=1}^{N} e^{-N(\beta/2)(V(\xi_{N;i}+x_i) - V(\xi_{N;i}))} \mathrm{d}x_i \\ Z_N^{(0)}(\nu) & = & \prod_{1 \leq i < j \leq N-1} |\xi_{N;i} - \xi_{N;j+1}|^\beta \prod_{i=1}^{N-1} |\xi_{N;i} - \xi_{N;i+1}|^{\beta/2} \prod_{i=1}^{N} e^{-N(\beta/2)V(\xi_{N;i})}\,. \end{aligned}$$

We recognize in the last factor an approximation to the answer we look for:

10.15 lemma.

$$\liminf_{N\to\infty} N^{-2}\,\ln Z_N^{(0)}(\nu) \geq -\mathcal{E}(\nu)\,.$$

Proof. It follows from the definition of the semiclassical positions that:

$$\int_{\xi_{N;i}}^{\xi_{N;i+1}} \int_{\xi_{N;j}}^{\xi_{N;j+1}} \mathrm{d}\nu(x)\mathrm{d}\nu(y)\mathbf{1}_{x<y} = \begin{cases} \frac{1}{(N+1)^2} & \text{if } i<j \\ \frac{1}{2(N+1)^2} & \text{if } i=j \end{cases}\,.$$

Therefore, we can write:

$$\begin{aligned} \ln Z_N^{(0)}(\nu) &= \beta(N+1)^2 \sum_{1\leq i\leq j\leq N-1} \ln(\xi_{N;j+1}-\xi_{N;i}) \int_{\xi_{N;i}}^{\xi_{N;i+1}} \int_{\xi_{N;j}}^{\xi_{N;j+1}} \mathrm{d}\nu(x)\mathrm{d}\nu(y)\mathbf{1}_{x<y} \\ &\quad -N^2(\beta/2)\int V(x)\,\mathrm{d}\tilde{\nu}_N(x) \\ &\geq \beta(N+1)^2 \iint_{[\xi_{N;1},\xi_{N;N}]^2} \mathbf{1}_{x<y}\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) - N^2(\beta/2)\int V(x)\mathrm{d}\tilde{\nu}_N(x)\,. \end{aligned}$$

We clearly have:

$$\lim_{N\to\infty} \iint_{[\xi_{N;1},\xi_{N;N}]^2} \ln|x-y|\,\mathbf{1}_{x<y}\,\mathrm{d}\nu(x)\mathrm{d}\nu(y) = -\frac{1}{2}\iint \ln|x-y|\mathrm{d}\nu(x)\mathrm{d}\nu(y)\,.$$

Since $\tilde{\nu}_N$ converges to ν and V is continuous hence bounded on the compact K uniformly containing the support of $\tilde{\nu}_N$ and ν, we have for the second term:

$$\lim_{N\to\infty} V(x)\mathrm{d}\tilde{\nu}_N(x) = \int V(x)\,\mathrm{d}\nu(x)\,.$$

Therefore, we conclude that:

$$\liminf_{N\to\infty} Z_N^{(0)}(\nu) \geq \frac{\beta}{2}\Big(\iint \ln|x-y|\mathrm{d}\nu(x)\mathrm{d}\nu(y) - \int V(x)\,\mathrm{d}\nu(x)\Big) = -(\beta/2)\mathcal{E}(\nu)\,.$$

□

It remains to show that the second factor in (94) leads does not spoil the lower bound:

10.16 lemma.

$$\lim_{\epsilon\to 0}\liminf_{N\to\infty} N^{-2}\ln Z_{N,\epsilon}^{(\mathrm{int})}(\nu) \geq 0\,.$$

Proof. We first observe that, since $\xi_{N;i}$ belongs to the compact K and V is uniformly continuous on K:

$$\limsup_{N\to\infty} \sup_{\substack{|x|\leq\epsilon/2 \\ 1\leq i\leq N}} \big|V(\xi_{N;i}+x) - V(\xi_{N;i})\big| = c(\epsilon)\,,$$

with $c(\epsilon) \to 0$ when $\epsilon \to 0$. Then, we use the change of variables:

$$y_1 = x_1, \qquad y_i = x_i - x_{i-1},\ i \in [\![2, N]\!] .$$

For a lower bound, we can restrict the integration to the domain $y_i \in [0, \epsilon/2N]$:

$$\liminf_{N\to\infty} N^{-2} \ln Z_{N,\epsilon}^{(\mathrm{int})}(\nu) \geq -(\beta/2)c(\epsilon) + \liminf_{N\to\infty} N^{-2} \ln \Bigg(\int_{[0,\epsilon/2N]^N} \mathrm{d}y_1 \prod_{i=2}^{n} y_i^{\beta/2} \mathrm{d}y_i \Bigg) .$$

The integral can be explicitly computed and is $O(e^{c' N \ln N})$, does not contribute to the limit. We conclude by sending ϵ to 0. ☐

Putting the two lemmas together, we have proved (92). We remark further that we also get a lower bound for the partition function. Indeed, for any $\nu \in \mathcal{M}_1(\mathbb{R})$, we can write a trivial lower bound

(95)
$$\liminf_{N\to\infty} N^{-2} \ln Z_N \geq \liminf_{\epsilon\to 0} \liminf_{N\to\infty} N^{-2} \ln \overline{\mathbb{P}_N}[B(\mu_{\mathrm{eq}}, \epsilon)] \geq -(\beta/2)\mathcal{E}(\mu_{\mathrm{eq}}) ,$$

and the last inequality is the result we just proved.

Conclusion

For the partition function: we have obtained identical upper bound (91) and lower bound (95), so we deduce Theorem 10.12:

$$\lim_{N\to\infty} N^{-2} \ln Z_N = -(\beta/2)\mathcal{E}(\mu_{\mathrm{eq}}) .$$

For the probability of small balls, we have obtained identical upper bound for the limsup and lower bound for the liminf, so the two limits must be equal:

$$-(\beta/2)\mathcal{E}(\mu) = \lim_{\epsilon\to 0} \liminf_{N\to\infty} N^{-2} \ln \overline{\mathbb{P}_N}[B(\nu,\epsilon)] = \lim_{\epsilon\to 0} \limsup_{N\to\infty} N^{-2} \ln \overline{\mathbb{P}_N}[B(\nu,\epsilon)] .$$

Hence, we can come back to the normalized measure $\mathbb{P}_N = Z_N^{-1}\overline{\mathbb{P}_N}$ using our result for the partition function:

$$-I(\mu) = \lim_{\epsilon\to 0} \liminf_{N\to\infty} N^{-2} \ln \mathbb{P}_N[B(\nu,\epsilon)] = \lim_{\epsilon\to 0} \limsup_{N\to\infty} N^{-2} \ln \mathbb{P}_N[B(\nu,\epsilon)]$$

with the good rate function:

$$I(\mu) = (\beta/2)\big[\mathcal{E}(\mu) - \mathcal{E}(\mu_{\mathrm{eq}})\big] .$$

According to Lemma 10.10, this implies a weak LDP for $(\mathbb{P}_N)_{N\geq 1}$. Thanks the exponential tightness – established in § 10.4 – and Lemma 10.9, it is upgraded to an LDP, and we obtain Theorem 10.13.

11 Schwinger-Dyson equations

Lecture 19 (1h30)
January 19th, 2015

We consider again the β ensembles with confining potential V, that we furthermore assume $\mathcal{C}^1$. The partition function and the correlation functions are expressed as integrals over $\mathbb{R}^N$. Performing a change of variable does not change the integral. We can exploit this freedom to derive exact relations between expectation values of various observables in the model. These are called the Schwinger-Dyson equations, or also Ward identities, loop equations, Pastur equations. In the context of matrix models, they were first written down by Migdal.

11.1 *Derivation of the first Schwinger-Dyson equation*

For instance, if f is any diffeomorphism of $\mathbb{R}$ to itself, we have:

$$\begin{aligned} Z_N &= \int_{\mathbb{R}^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i \, e^{-N(\beta/2)V(\lambda_i)} \prod_{1\leq i<j\leq N} |\lambda_i - \lambda_j|^\beta \\ &= \int_{\mathbb{R}^N} \prod_{i=1}^{N} \mathrm{d}\lambda_i \, f'(\lambda_i)\, e^{-N(\beta/2)V(f(\lambda_i))} \prod_{1\leq i<j\leq N} |f(\lambda_i) - f(\lambda_j)|^\beta \,. \end{aligned} \tag{96}$$

Let $h : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^1$ function such that h and h' are bounded on $\mathbb{R}$. Then, $f_t(x) = x + th(x)$ is a diffeomorphism of $\mathbb{R}$ for t small enough. The invariance of the integral under this differentiable family of change of variables implies that the coefficient of t in (96) with $f = f_t$ vanishes. To compute it, we remark:

$$\begin{aligned} & \partial_{t=0}\Big(\frac{\mathrm{d}\mu_N(f_t(\lambda_1),\ldots,f_t(\lambda_N))}{\mathrm{d}\mu_N(\lambda_1,\ldots,\lambda_N)}\Big) \\ = & \sum_{i=1}^{N} \Big(h'(\lambda_i) - N(\beta/2)\, V'(\lambda_i)\, h(\lambda_i)\Big) + \sum_{1\leq i<j\leq N} \beta\, \frac{h(\lambda_i) - h(\lambda_j)}{\lambda_i - \lambda_j}\,, \end{aligned}$$

where $\mathrm{d}\mu_N$ is the usual measure of the β ensemble. So, the vanishing of the coefficient of $O(t)$ after integration can be written as:

$$\mathbb{E}\Big[\sum_{i=1}^{N} \Big(h'(\lambda_i) - N(\beta/2)\, V'(\lambda_i) h(\lambda_i)\Big) + \sum_{1\leq i<j\leq N} \beta\, \frac{h(\lambda_i) - h(\lambda_j)}{\lambda_i - \lambda_j}\Big] = 0\,. \tag{97}$$

Equivalently, we can symmetrize the sum over (i,j) in (97), and add and subtract the diagonal terms. Multiplying by $2/\beta$, this yields:

$$\mathbb{E}\Big[(2/\beta - 1)\sum_{i=1}^{N} h'(\lambda_i) - N\sum_{i=1}^{N} V'(\lambda_i)\, h(\lambda_i) + \sum_{1\leq i,j\leq N} \frac{h(\lambda_i) - h(\lambda_j)}{\lambda_i - \lambda_j}\Big] = 0\,.$$

We remark that for the case of hermitian random matrices, $\beta = 2$ and the first term disappears. The equation is then somewhat simpler because there is no derivative of h. The combination:

$$\mathcal{D}[h](x,y) = \frac{h(x)-h(y)}{x-y} = \int_0^1 h'(xu + y(1-u))\mathrm{d}u$$

is called the **non-commutative derivative** of h.

11.2 *Higher Schwinger-Dyson equations*

If we remind the Definition 7.1 of the k-point density correlations $\rho_{k|N}$, we see that it gives a functional relation between $\rho_{2|N}$ and $\rho_{1|N}$. The fact that it involves statistics of pairs of eigenvalues is due to the pairwise interaction contained in the Vandermonde. Writing down the first Schwinger-Dyson equation for the model with potential $V(t) = V - (2/\beta N)\sum_{a=1}^{k} t_a\, h_a(x)$, and picking up the coefficient of $O(t_1 \cdots t_k)$, we obtain the $(k+1)$-th Schwinger-Dyson equation

$$\begin{aligned}&\mathbb{E}\Big[\Big\{(2/\beta - 1)\sum_{i=1}^{N} h'(\lambda_i) - N\sum_{i=1}^{N} V'(\lambda_i)h(\lambda_i)\\
&+ \sum_{1\leq i<j\leq N} \mathcal{D}[h](\lambda_i,\lambda_j)\Big\}\Big(\prod_{a=1}^{k}\sum_{i_a=1}^{N} h(\lambda_{i_a})\Big)\\
&+\frac{2}{\beta}\sum_{a=1}^{k}\sum_{i_a=1}^{N} h'(\lambda_{i_a})\Big(\prod_{b\neq a}\sum_{i_b=1}^{N} h(\lambda_{i_b})\Big)\Big] = 0.\end{aligned}$$

It is a functional relation involving $\rho_{k+1|N},\ldots,\rho_{1|N}$.

In general, the collection of k-th Schwinger-Dyson equations for $k = 1, 2, \ldots$ do not allow to solve for the density correlations. Written in this form, they do not close – we need to know $\rho_{k+1|N}$ to get $\rho_{k|N}$ for the k-th equation – and actually they have many solutions. However, these equations can be used effectively to establish asymptotic expansions of the partition function and expectation values of various observables when $N \to \infty$. And under certain assumptions – for instance, a slightly stronger version of the one-cut property – they do allow a recursive computation of the coefficients of these expansions.

11.3 *Stieltjes transform of the equilibrium measure*

Yet another way to rewrite the equation is in terms of the empirical measure:

$$\begin{aligned}&\mathbb{E}\Big[\frac{2/\beta - 1}{N}\int h'(x)\,\mathrm{d}L_N^{(\lambda)}(x) - \int V'(x)h(x)\,\mathrm{d}L_N^{(\lambda)}(x)\\
&+\iint \mathcal{D}[h](x,y)\,\mathrm{d}L_N^{(\lambda)}(x)\mathrm{d}L_N^{(\lambda)}(y)\Big] = 0.\end{aligned}$$

This shows that the first term involving h' is negligible in front of the two last terms when $N \to \infty$. As a consequence of Corollary 10.14, L_N converges to

μ_{eq} in expectation, so this equation gives in the large N limit:

$$(98) \quad \int V'(x)h(x)\mathrm{d}\mu_{\mathrm{eq}}(x) = \iint \mathcal{D}[h](x,y)\mathrm{d}\mu_{\mathrm{eq}}(x)\mathrm{d}\mu_{\mathrm{eq}}(y)\,.$$

Since the Schwinger-Dyson equation is linear in h, it is also valid for $h : \mathbb{R} \to \mathbb{C}$. For $z \in \mathbb{C} \setminus \mathbb{R}$ and outside the support of μ_{eq}, let us choose:

$$h(x) = h_z(x) := \frac{1}{z-x}\,.$$

It is indeed a bounded function with bounded derivative for $x \in \mathbb{R}$. We have:

$$\mathcal{D}[h_z](x,y) = \frac{1}{(z-x)(z-y)}\,.$$

So, in terms of the Stieltjes transform $W(z)$ of μ_{eq}, we have:

$$\iint \mathcal{D}[h_z](x,y)\mathrm{d}\mu_{\mathrm{eq}}(x)\mathrm{d}\mu_{\mathrm{eq}}(y) = W(z)^2\,.$$

For simplicity, let us assume that V is a polynomial. We then have:

$$\begin{aligned}\int V'(x)h_z(x)\mathrm{d}\mu_{\mathrm{eq}}(x) &= V'(z)\int h_z(x)\mathrm{d}\mu_{\mathrm{eq}}(x) - \int \frac{V'(z)-V'(x)}{z-x}\,\mathrm{d}\mu_{\mathrm{eq}}(x) \\ &= V'(z)W(z) - P(z)\,,\end{aligned}$$

and since $(V'(x)-V'(z))/(x-z)$ is a polynomial, P is also a polynomial. By continuity, we can actually put z on the real line and outside the support S_{eq} of μ_{eq} since $W(z)$ is defined on $\mathbb{C} \setminus S_{\mathrm{eq}}$.

11.1 lemma. *If V is a polynomial, the Stieltjes transform of μ_{eq} satisfies the equation:*

$$W^2(z) - V'(z)W(z) + P(z) = 0, \qquad P(z) := \int \frac{V'(z)-V'(x)}{z-x}\,\mathrm{d}\mu_{\mathrm{eq}}(x).$$

Although the polynomial P is a priori unknown, this equation is very useful to study μ_{eq}.

11.2 corollary. *If V is a polynomial, $W(z)$ is uniformly bounded when $z \in \mathbb{C} \setminus S_{\mathrm{eq}}$, and the limit:*

$$\lim_{\epsilon \to 0^+} W(x \pm \mathrm{i}0)$$

exist and are continuous functions of x. Besides, S_{eq} is a finite union of segments, and μ_{eq} is continuous with respect to Lebesgue measure.

Proof. The solution of the quadratic equation is:

$$W(z) = \frac{V'(z) - \sqrt{V'(z)^2 - 4P(z)}}{2}.$$

Since P is a polynomial, we see that $W(x)$ remains bounded in a compact neighborhood of S_{eq}. Then, since $W(x) \in O(1/x)$ when $x \to \infty$, $W(x)$ is uniformly bounded over $\mathbb{C}$. The boundary values of $W(x)$ only differ by the sign given to the squareroot, and they are clearly continuous functions of x, and the density of μ_{eq} is:

$$\frac{\mathrm{d}\mu_{\mathrm{eq}}}{\mathrm{d}x}(x) = \frac{\sqrt{4P(x) - V'(x)^2}}{2\pi}.$$

With this formula, we identify:

$$S_{\mathrm{eq}} = \big\{x \in \mathbb{R}, \quad 4P(x) - V'(x)^2 \geq 0\big\}.$$

Since $4P(x) - V'(x)^2$ is a polynomial, this is a finite union of segments. □