Lectures on Random Matrices (Spring 2025)

Leonid Petrov

Spring 2025

Contents

Chapter 1

Moments of random variables and random matrices

1.1 Why study random matrices?

TODO FOR BOOK FORMAT: add ref to git and live sims as below

TODO FOR BOOK FORMAT: fix texorpdfstring for titles

TODO FOR BOOK FORMAT: catch just a small handful of undefined references

On the history. Random matrix theory (RMT) is a fascinating field that studies properties of matrices with randomly generated entries, focusing (at least initially) on the statistical behavior of their eigenvalues. This theory finds its roots in the domain of nuclear physics through the pioneering work of Wigner, Dyson, and others [Wig55], [Dys62a], [Dys62b], who utilized it to analyze the energy levels of complex quantum systems. Other, earlier roots include statistics [Dix05] and classical Lie groups [Hur97]. Today, RMT has evolved to span a wide array of disciplines, from pure mathematics, including areas such as integrable systems and representation theory, to practical applications in fields like data science and engineering.

Classical groups and Lie theory. Random matrices are deeply connected to *classical Lie groups*, particularly the orthogonal, unitary, and symplectic groups. This connection emerges primarily due to the invariance properties of these groups, such as those derived from the Haar measure.

Random matrices significantly impact representation theory, linking to integrals over matrix groups through character expansions. The symmetry classes of random matrix ensembles, like the Gaussian Orthogonal (GOE), Unitary (GUE), and Symplectic (GSE), correspond to respective symmetry groups.

Toolbox. RMT utilizes a broad range of tools ranging across all of mathematics, including probability theory, combinatorics, analysis (classical and modern), algebra, representation theory, and number theory. The theory of random matrices is a rich source of problems and techniques for all of mathematics.

The main content of this course is to explore the toolbox around random matrices, including going into discrete models like dimers and statistical mechanics. Some of this will be included in the lectures, and some other topics will be covered in the reading course component, which is individualized.

Applications. Random matrix theory finds applications across a diverse set of fields. In nuclear physics, random matrix ensembles serve as models for complex quantum Hamiltonians, thereby explaining the statistics of energy levels. In number theory, connections have been drawn between random matrices and the Riemann zeta function, particularly concerning the distribution of zeros on the critical line. Wireless communications benefit from random matrix theory through the analysis of eigenvalue distributions, which helps in understanding channel capacity in multi-antenna (MIMO) systems. In the burgeoning field of machine learning, random weight matrices and their spectra are key to analyzing neural networks and their generalization capabilities. High-dimensional statistics and econometrics also draw on random matrix tools for tasks such as principal component analysis and covariance estimation in large datasets. Additionally, combinatorial random processes exhibit connections to random permutations, random graphs, and partition theory, all through the lens of matrix integrals.

1.2 Recall Central Limit Theorem

1.2.1 Central Limit Theorem and examples

We begin by establishing the necessary groundwork for understanding and proving the Central Limit Theorem. The theorem's power lies in its remarkable universality: it applies to a wide variety of probability distributions under mild conditions.

Definition 1.1. A sequence of random variables $\{X_i\}_{i=1}^{\infty}$ is said to be *independent and identically distributed (iid)* if:

- Each X_i has the same probability distribution as every other X_j, for all i, j.
- The variables are mutually independent, meaning that for any finite subset $\{X_1, X_2, \ldots, X_n\}$, the joint distribution factors as the product of the individual distributions:

$$\mathbb{P}(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = \mathbb{P}(X_1 \le x_1)\, \mathbb{P}(X_2 \le x_2) \cdots \mathbb{P}(X_n \le x_n).$$

Theorem 1.2 (Classical Central Limit Theorem). *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$. Define the normalized sum*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)\,. \tag{1.1}$$

Then, as $n \to \infty$, the distribution of Z_n converges in distribution to a normal random variable with mean 0 and variance σ^2, i.e.,

$$Z_n \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Convergence in distribution means

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le x) = \mathbb{P}(Z \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}}\, dt \qquad \text{for all } x \in \mathbb{R}, \tag{1.2}$$

where $Z \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian random variable.

Remark 1.3. For a general random variable instead of $Z \sim \mathcal{N}(0, \sigma^2)$, the convergence in distribution (1.2) holds only for x at which the cumulative distribution function of Z is continuous. Since the normal distribution is absolutely continuous (has density), the convergence holds for all x.

Example 1.4. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of iid Bernoulli random variables with parameter p, meaning that each X_i takes the value 1 with probability p and 0 with probability $1-p$. The mean and variance of each X_i are given by:

$$\mu = \mathbb{E}[X_i] = p, \quad \sigma^2 = \mathrm{Var}(X_i) = p(1-p).$$

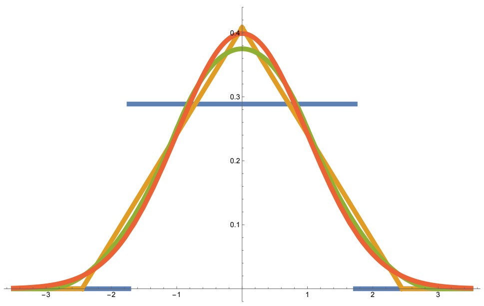

Figure 1.1: Densities of U_1, U_1+U_2, $U_1+U_2+U_3$ (where U_i are iid uniform on $[0,1]$), and $\mathcal{N}(0,1)$, normalized to have the same mean and variance.

We also have the distribution of $X_1+\cdots+X_n$:

$$\mathbb{P}\left(X_1+\cdots+X_n=k\right)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad k=0,1,\ldots,n.$$

Introduce the normalized quantity

$$z=\frac{k-np}{\sqrt{np(1-p)}}, \tag{1.3}$$

and assume that throughout the asymptotic analysis, this quantity stays finite.

Our aim is to show that, for k such that z remains bounded as $n\to\infty$, the following holds:

$$\mathbb{P}(S_n=k)=\frac{1}{\sqrt{2\pi np(1-p)}}\exp\Big(-\frac{z^2}{2}\Big)(1+o(1)).$$

For large n, Stirling's formula gives

$$m!\sim\sqrt{2\pi m}\,m^m e^{-m}, \quad \text{as } m\to\infty.$$

Apply Stirling's approximation to $n!$, $k!$, and $(n-k)!$:

$$n!\sim\sqrt{2\pi n}\,n^n e^{-n}, \quad k!\sim\sqrt{2\pi k}\,k^k e^{-k}, \quad (n-k)!\sim\sqrt{2\pi(n-k)}\,(n-k)^{n-k}e^{-(n-k)}.$$

Thus,

$$\binom{n}{k}\sim\frac{\sqrt{2\pi n}\,n^n e^{-n}}{\sqrt{2\pi k}\,k^k e^{-k}\sqrt{2\pi(n-k)}\,(n-k)^{n-k}e^{-(n-k)}}=\frac{n^n}{k^k(n-k)^{n-k}}\frac{1}{\sqrt{2\pi\,k(n-k)/n}}.$$

More precisely, one often writes

$$\binom{n}{k} \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Big(n \ln n - k \ln k - (n-k)\ln(n-k)\Big),$$

where $p \approx k/n$ thanks to the fact that z (1.3) is assumed to be finite.

We have

$$k = np + z\sqrt{np(1-p)}.$$

Then, consider the second-order Taylor expansion. We have

$$n \ln n - k \ln k - (n-k)\ln(n-k) \sim nH - \frac{z^2}{2},$$

where $H = -[p \ln p + (1-p)\ln(1-p)] + c(z;p)/\sqrt{n}$ (for an explicit function $c(z;p)$) is the "entropy" term which exactly cancels with the prefactors coming from $p^k(1-p)^{n-k}$.

After combining the approximations from the binomial coefficient and the probability weights, one arrives at

$$\mathbb{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{z^2}{2}\right),$$

as desired.

(Note that this is a *local* CLT as opposed to the convergence (1.2) in the classical CLT; but one can get the latter from the local CLT by integration.)

1.2.2 Moments of the normal distribution

Proposition 1.5. *The moments of a random variable $Z \sim \mathcal{N}(0, \sigma^2)$ are given by:*

$$\mathbb{E}[Z^k] = \begin{cases} 0, & \text{if } k \text{ is odd}, \\ \sigma^k (k-1)!! = \sigma^k \cdot (k-1)(k-3)\cdots 1, & \text{if } k \text{ is even}. \end{cases} \tag{1.4}$$

Proof. We just compute the integrals. Assume k is even (for odd, the integral is zero by symmetry). Also assume $\sigma = 1$ for simplicity. Then

$$\mathbb{E}[Z^k] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^k e^{-z^2/2}\, dz.$$

Applying integration by parts (putting $ze^{-z^2/2}$ under d), we get

$$\mathbb{E}[Z^k] = \frac{1}{\sqrt{2\pi}} \left[-z^{k-1}e^{-z^2/2}\right]_{-\infty}^{\infty} + \frac{k-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{k-2} e^{-z^2/2}\, dz.$$

The first term vanishes at infinity (you can verify this using L'Hôpital's rule), leaving us with:

$$\mathbb{E}[Z^k] = (k-1)\,\mathbb{E}[Z^{k-2}].$$

This gives us a recursive formula, and completes the proof. □

1.2.3 Moments of sums of iid random variables

Let us now show the CLT by moments. For example, the source is [Bil95, Section 30] or [Fil10].

Remark 1.6. This proof requires an additional assumption that all moments of the random variables are finite. This is quite a strong assumption, and while the CLT holds without it, this proof by moments is more algebraic, and will translate to random matrices more directly.

Computation of moments

Denote $Y_i = X_i - \mu$, these are also iid, but have mean 0. We consider

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right].$$

Expanding the k-th power using the multinomial theorem, we obtain:

$$\left(\sum_{i=1}^{n} Y_i\right)^k = \sum_{j_1+j_2+\cdots+j_n=k} Y_{j_1}Y_{j_2}\ldots Y_{j_n}.$$

Taking the expectation and using linearity, we have:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = \sum_{j_1+j_2+\cdots+j_n=k} \mathbb{E}\left[Y_{j_1}Y_{j_2}\ldots Y_{j_n}\right].$$

The sum over all $j_1, \ldots, j_n$ with $j_1 + \ldots + j_n = k$ is the number of ways to partition k into n non-negative integers. We can order these integers, and thus obtain the sum over all partitions of k into $\le n$ parts. Since n is large, we simply sum over all partitions of k. For each partition λ of k (where $k = \lambda_1 + \lambda_2 + \ldots + \lambda_n$ and $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n \ge 0$), we must count the number of distinct multisets of indices $(j_1, j_2, \ldots, j_n)$ that yield the same collection $\{\lambda_1, \lambda_2, \ldots\}$. Then,

$$\mathbb{E}\left[Y_{j_1} Y_{j_2} \ldots Y_{j_n}\right] = \mathsf{m}_{\lambda_1} \mathsf{m}_{\lambda_2} \ldots \mathsf{m}_{\lambda_n},$$

where $\mathsf{m}_j = \mathbb{E}[Y^j]$ (recall the identical distribution of Y_i). Note that $\mathsf{m}_0 = 1$ and $\mathsf{m}_1 = 0$. Let us illustrate this with an example.

Example 1.7. For $k = 4$, there are only two partitions which have no parts equal to 1: $\lambda = (4)$ and $\lambda = (2, 2)$. The number of ways to get (4) (so that $\mathbb{E}[Y_{j_1} Y_{j_2} Y_{j_3} Y_{j_4}] = \mathsf{m}_4$) is to just assign one of the j_p to be 4, this can be done in n ways.

The number of ways to get $(2, 2)$ (so that $\mathbb{E}[Y_{j_1} Y_{j_2} Y_{j_3} Y_{j_4}] = \mathsf{m}_2^2$) is to assign two of the j_p to be 2 and the other two to be 0, this can be done in $\binom{n}{2}$ ways. Moreover, there are also 6 permutations of the indices $j_p = (i, j)$ which give the same partition $(2, 2)$: (i, i, j, j), (j, j, i, i), (i, j, i, j), (j, i, j, i), (i, j, j, i), (j, i, i, j). Thus, the total number of ways to get $(2, 2)$ is $6\binom{n}{2} \sim 3n^2$.

So, we see that there is an n-dependent factor, and a "combinatorial" factor for each partition.

n-dependent factor

Consider first the n-dependent factor. In the case k is even and $\lambda = (2, 2, \ldots, 2)$, the power of n is $n^{k/2}$. In the case k is even and λ has at least one part ≥ 3, the power of n is at most $n^{k/2-1}$, which is subleading in the limit $n \to \infty$. When k is odd, the "best" we can do (without parts equal to 1) is going to be $\lambda = (3, 2, \ldots, 2)$ with $(k-1)/2$ parts, so the power of n is $n^{(k-1)/2}$. This is also subleading in the limit $n \to \infty$.

Combinatorial factor

Now, we see that we only need to consider the case when k is even and all parts of λ are 2. Then, the n-dependent factor is $\binom{n}{k/2} \sim n^{k/2}/(k/2)!$. The combinatorial factor is equal to the number of ways to partition k into pairs, which is the double factorial:

$$(k-1)!! = (k-1)(k-3)\ldots 1,$$

times the number of permutations of the $k/2$ indices which are assigned to the pairs, so $(k/2)!$. In particular, for $k = 4$ this is 6.

Putting it all together

We have as $n \to \infty$:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = n^{k/2}\frac{(k-1)!!}{(k/2)!}\cdot(k/2)!\sigma^k + o(n^{k/2}) = n^{k/2}(k-1)!!\sigma^k + o(n^{k/2}).$$

Now, we need to consider the normalization of the sum $\sum_{i=1}^{n} Y_i$ by $\sqrt{n}$:

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i\right)^k\right] = \frac{1}{n^{k/2}}\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^k\right] = (k-1)!!\sigma^k + o(1).$$

Therefore, the moments of Z_n (1.1) converge to the moments of the standard normal distribution.

1.2.4 Convergence in distribution

Is convergence of moments enough to imply convergence in distribution? Not necessarily. First, note that the functions $x \mapsto x^k$ are not even bounded on $\mathbb{R}$.

A sufficient condition for convergence in distribution is found in the classical method of moments in probability theory [Bil95, Theorem 30.2]. This theorem states that if the limiting distribution X is uniquely determined by its moments, then convergence in moments implies convergence in distribution.

The normal distribution is indeed uniquely determined by its moments (Problem 1.4.5), so the CLT holds in this case, provided that the original iid random variables X_i have finite moments of all orders.

1.3 Random matrices and semicircle law

We now turn to random matrices.

1.3.1 Where can randomness in a matrix come from?

The study of random matrices begins with understanding how randomness can be introduced into matrix structures. We consider three primary sources:

1. **iid entries:** The simplest form of randomness comes from filling matrix entries independently with samples from a fixed probability distribution. For an $n \times n$ matrix, this gives us n^2 independent random variables. If we do not impose any additional structure on the matrix, then the eigenvalues will be complex. So, often we consider real symmetric, complex Hermitian, or quaternionic matrices with symplectic symmetry.[1]

2. **Correlated entries:** In many physical systems, especially those modeling local interactions, matrix entries are not independent but show correlation patterns. Common examples include:
   - Band matrices, where entries become negligible far from the diagonal
   - Matrices with correlation decay based on the distance between indices
   - Structured random matrices arising from specific physical models
   - Sparse matrices, where most entries are zero

3. **Haar measure on matrix groups:** Randomness can come from considering matrices sampled according to the Haar measure on a compact matrix group, for example, the orthogonal $O(n)$, unitary $U(n)$, or symplectic group $Sp(n)$.[2] One can think of this as a generalization of the uniform distribution (Lebesgue measure) on the unit circle in $\mathbb{C}$, or a unit sphere in $\mathbb{R}^n$. One can also mix and match: one of the most interesting families of random matrices is the one with constant eigenvalues, but random eigenvectors:

$$A = UD_\lambda U^\dagger, \qquad U \in U(n), \quad U \sim \text{Haar}.$$

[1] Real symmetric means $A^\top = A$, complex Hermitian means $A^\dagger = A$ (conjugate transpose). Let us briefly discuss the quaternionic case. It can be modeled over $\mathbb{C}$. A quaternion $q = a + b\,\mathbf{i} + c\,\mathbf{j} + d\,\mathbf{k}$ can be represented by the complex 2×2 matrix

$$q \;\longmapsto\; \begin{pmatrix} a + \mathbf{i}b & c + \mathbf{i}d \\ -c + \mathbf{i}d & a - \mathbf{i}b \end{pmatrix}.$$

The entries a, b, c, d for the quaternion matrix case must be real, and the matrix A of size $2n \times 2n$ should also be Hermitian in the usual complex sense.

[2] The orthogonal and unitary groups are defined in the usual way, by $OO^\top = O^\top O = I$ and $UU^\dagger = U^\dagger U = I$, respectively. The group $Sp(n)$ is the compact real form of the full symplectic group $Sp(2n, \mathbb{C})$, consisting of $2n \times 2n$ matrices A such that $A^\top JA = J$, where J is the skew-symmetric form.

Here D_λ is a diagonal matrix with constant eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_n)$. The random matrix A is the "uniform" random variable taking values in the set of all Hermitian matrices with fixed real eigenvalues λ. Here we may assume that $\lambda_1 \geq \ldots \geq \lambda_n$, since the unitary conjugation can permute the eigenvalues.

1.3.2 Real Wigner matrices

Definition 1.8 (Real Wigner Matrix)**.** An $n \times n$ random matrix $W = W_n = (X_{ij})_{1 \leq i,j \leq n}$ is called a *real Wigner matrix* if:

1. W is symmetric: $X_{ij} = X_{ji}$ for all i, j;
2. The upper triangular entries $\{X_{ij} : 1 \leq i \leq j \leq n\}$ are independent;
3. The diagonal entries $\{X_{ii}\}$ are iid real random variables with mean 0 and variance σ_d;
4. The upper triangular entries $\{X_{ij} : i < j\}$ are iid (possibly with a distribution different from the diagonal entries) real random variables with mean 0 and variance σ;
5. (optional, but we assume this) All entries have finite moments of all orders.

Example 1.9 (Gaussian Wigner Matrices, Gaussian Orthogonal Ensemble (GOE))**.** Let W be a real Wigner matrix where:

- Diagonal entries $X_{ii} \sim \mathcal{N}(0, 2)$;
- Upper triangular entries $X_{ij} \sim \mathcal{N}(0, 1)$ for $i < j$.

We can model W as $(Y + Y^\top)/\sqrt{2}$, where Y is a matrix with iid Gaussian entries $Y_{ij} \sim \mathcal{N}(0, 1)$. The matrix distribution of W is called the *Gaussian Orthogonal Ensemble* (*GOE*).

Remark 1.10 (Wishart Matrices)**.** There are other ways to define random matrices, most notably, *sample covariance matrices*. Let $A = [a_{i,j}]_{i,j=1}^{n,m}$ be an $n \times m$ matrix ($n \leq m$), where entries are iid real random variables with mean 0 and finite variance. Then $M = AA^\top$ is a positive symmetric random matrix of size $n \times n$. It almost surely has full rank.

1.3.3 Empirical spectral distribution

For an arbitrary random matrix of size $n \times n$ with real eigenvalues, the *empirical spectral distribution* (*ESD*) is defined as the random probability measure on $\mathbb{R}$:

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}, \tag{1.5}$$

which puts point masses of size $1/n$ at the eigenvalues λ_i of the matrix.

If you sample the ESD for a large real Wigner matrix, and take a histogram (to cluster the eigenvalues into boxes), you will see the semi-circular pattern. This pattern does not change over several samples. Hence, one can conjecture that the ESD (1.5) converges to a nonrandom measure, after rescaling.

We can guess the rescaling by looking at the first two moments of the ESD. The first moment is

$$\int_{\mathbb{R}} x \, \mu_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i = \frac{1}{n} \operatorname{Tr}(W) = \frac{1}{n} \sum_{i=1}^{n} X_{ii}, \tag{1.6}$$

and this sum has mean zero (and small variance), so it converges to zero. The second moment is

$$\int_{\mathbb{R}} x^2 \, \mu_n(dx) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 = \frac{1}{n} \operatorname{Tr}(W^2) = \frac{1}{n} \sum_{i,j=1}^{n} X_{ij}^2. \tag{1.7}$$

This sum has mean $\sim \sigma^2 n^2$, so even normalized by n, it still goes to infinity. But, if we normalize the matrix as $\frac{1}{\sqrt{n}} W$, then the second moment becomes bounded, and one can convince oneself that the ESD of the normalized Wishart matrix has a limit. Indeed, this is the case:

Theorem 1.11 (Wigner's Semicircle Law)**.** *Let W be a real Wigner matrix of size $n \times n$ (with off-diagonal entries having a fixed variance σ^2, independent of n). Then as $n \to \infty$, the ESD of $W/(\sigma\sqrt{n})$ converges in distribution to the semicircular law:*

$$\nu_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \longrightarrow \mu_{\mathrm{sc}}, \tag{1.8}$$

where μ_{sc} is the semicircular distribution with density with respect to the Lebesgue measure:

$$\mu_{\mathrm{sc}}(dx) \coloneqq \frac{1}{2\pi} \sqrt{4 - x^2} \, \mathbf{1}_{|x| \le 2} \, dx. \tag{1.9}$$

Remark 1.12. The convergence in (1.8) may mean either *weakly in probability* or *weakly almost surely*. The first notion, weak convergence in probability, means that for every bounded continuous function f, we have

$$\int_{\mathbb{R}} f(x)\,\nu_n(dx) \longrightarrow \int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx), \qquad n \to \infty, \tag{1.10}$$

where in (1.10) the convergence is in probability. Indeed, the left-hand side of (1.10) is a random variable, so we need to qualify which sense of convergence we mean.

The weakly almost sure convergence means that the convergence in (1.10) holds for almost all realizations of the random matrix W, that is, for every bounded continuous function f, the random variable $\int_{\mathbb{R}} f(x)\,\nu_n(dx)$ converges almost surely to $\int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx)$.

Remark 1.13. There exists a version of the limiting ESD for the Wishart matrices (Remark 1.10). In this case, the limiting distribution is the *Marchenko-Pastur law* [MP67].

1.3.4 Expected moments of traces of random matrices

The main computation in the proof of Theorem 1.11 is the computation of expected moments of the ESD. This computation of moments is somewhat similar to the one in the proof of the CLT by moments, but has its own random matrix flavor.

Definition 1.14 (Normalized Moments)**.** For each $k \geq 1$, the normalized k-th moment of the empirical spectral distribution of $W_n/\sqrt{n}$ is given by

$$m_k^{(n)} = \int_{\mathbb{R}} x^k\,\nu_n(dx) = \frac{1}{n^{k/2+1}}\,\mathrm{Tr}(W^k).$$

Our first goal is to study the asymptotic behavior of $\mathbb{E}[m_k^{(n)}]$ as $n \to \infty$ for each fixed $k \geq 1$, just like we did in (1.6)–(1.7) for $k = 1, 2$:

$$\mathbb{E}[m_1^{(n)}] = 0, \qquad \mathbb{E}[m_2^{(n)}] \to \sigma^2.$$

Note that $\mathbb{E}[m_2^{(n)}]$ is not exactly equal to σ^2 because of the presence of the diagonal elements which have a different distribution. In general, we will see that the contribution of the diagonal elements to the moments is negligible in the limit $n \to \infty$.

Lemma 1.15 (Convergence of Expected Moments)**.** *For each fixed $k \geq 1$, we have*

$$\lim_{n\to\infty} \mathbb{E}[m_k^{(n)}] = \begin{cases} 0 & \textit{if } k \textit{ is odd,} \\ \sigma^k C_{k/2} & \textit{if } k \textit{ is even,} \end{cases}$$

where $C_m = \frac{1}{m+1}\binom{2m}{m}$ is the m-th Catalan number.

The even moments are scaled by powers of σ just as in the case $k = 2$, while the odd moments vanish due to the symmetry of the limiting distribution around zero. As we will see, the appearance of Catalan numbers is not accidental, but it is due to the underlying combinatorics.

Proof of Lemma 1.15. The trace of W^k expands as a sum over all possible index sequences:

$$\mathrm{Tr}(W^k) = \sum_{i_1,\dots,i_k=1}^{n} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}. \tag{1.11}$$

Due to independence and the fact that $\mathbb{E}[X_{ij}] = 0$ for all i, j, the only nonzero contributions come from index sequences where each matrix element appears least twice.

As in the CLT proof, there is a power-n factor and a combinatorial factor.

For k odd, let us count the power of n first. As in the CLT proof, the maximum power comes from index sequences where all matrix elements appear exactly twice except for one which appears three times. Indeed, this corresponds to the maximum freedom of choosing k indices among the large number n of indices, and thus to the maximum power of n. This maximum power of n is $n^{1+\lfloor k/2 \rfloor}$ (note that there is an extra factor n compared to the CLT proof, as now we have $\sim n^2$ random variables in the matrix instead of n). Since this is strictly less than the normalization $n^{k/2+1}$ in $m_k^{(n)}$, the term with odd k vanish in the limit $n \to \infty$.

Assume now that k is even. Then the maximum power of n comes from index sequences where each matrix element appears exactly twice. This power of n is $n^{k/2+1}$, which exactly matches the normalization in $m_k^{(n)}$.

It remains to count the combinatorial factor, assuming that k is even. For each term in the trace expansion, we can represent the sequence of indices $(i_1, \dots, i_k)$ as a directed closed path with vertices $\{1, \dots, n\}$ and edges given by the matrix entries $X_{i_a i_{a+1}}$. For example, if $k = 4$ and we have a term $X_{12}X_{23}X_{34}X_{41}$, this corresponds to the path $1 \to 2 \to 3 \to 4 \to 1$. Recall that our path must have each matrix entry exactly twice (within the symmetry $X_{ij} = X_{ji}$), and the path must be closed. The condition that each edge appears exactly twice means that if we forget the direction of the edges and the multiplicities, we must get a *tree*, with $k/2$ edges and $k/2+1$ vertices. The complete justification of this counting is the problem in Problem 1.4.9.

The n-powers counting implies that the combinatorial factor (for even k) is equal to σ^k times the number of *rooted* (*planar*) *trees* with $k/2$ edges. The rooted condition comes from the fact that we are free to fix the starting point of the path to be 1 (this ambiguity is taken into account by the power-n factor).

In Problem 1.4.10, we show that the number of these rooted trees is the $k/2$-th Catalan number $C_{k/2}$. This completes the proof of Lemma 1.15. □

1.3.5 Immediate next steps

The proof of Theorem 1.11 is continued in the next Chapter 2. Immediate next steps are:

1. Show that the number of rooted trees with $k/2$ edges is the $k/2$-th Catalan number, and give the exact formula for the Catalan numbers.

2. Compute the moments of the semicircular distribution.

3. Make sure that the moment computation suffice to show the weak in probability convergence of the ESD to the semicircular law.

1.4 Problems

Each problem is a subsection (like Problem 1.4.1), and may have several parts.

1.4.1 Normal approximation

1. In Figure 1.1, which color is the normal curve and which is the sum of three uniform random variables?

2. Show that the sum of 12 iid uniform random variables on $[-1,1]$ (without normalization) is approximately standard normal.

3. Find (numerically is okay) the maximum discrepancy between the distribution of the sum of 12 iid uniform random variables on $[-1,1]$ and

the standard normal distribution:

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\sum_{i=1}^{12} U_i \le x \right) - \mathbb{P}\left(Z \le x\right) \right| .$$

1.4.2 Convergence in distribution

Convergence in distribution $X_n \to X$ for real random variables X_n and X means, by definition, that

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$$

for all bounded continuous functions f. Show that convergence in distribution is equivalent to the condition outlined in (1.2):

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x)$$

for all x at which the cumulative distribution function of X is continuous.

1.4.3 Moments of sum justification

Justify the computations of the power of n in Section 1.2.3.

1.4.4 Distribution not determined by moments

Show that the log-normal random variable e^Z (where $Z \sim \mathcal{N}(0,1)$) is not determined by its moments.

1.4.5 Uniqueness of the normal distribution

Show that the normal distribution is uniquely determined by its moments.

1.4.6 Quaternions

Show that the 2×2 matrix representation of a quaternion given in Footnote 1 indeed satisfies the quaternion multiplication rules. Hint: Use linearity and distributive law.

1.4.7 Ensemble $UD_\lambda U^\dagger$

Let U be the random Haar-distributed unitary matrix of size $N \times N$. Let D_λ be the diagonal matrix with constant real eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \geq \ldots \geq \lambda_N$. Let us fix λ to be, say, $\lambda = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, for some proportion of 1's and 0's (you can start with half ones and half zeros).

Use a computer algebra system to sample the eigenvalues of the matrix obtained from $UD_\lambda U^\dagger$ by taking only its top-left corner of size $k \times k$, where $k = 1, 2, \ldots, N$. For a fixed k, let $\lambda_1^{(k)} \geq \ldots \geq \lambda_k^{(k)}$ be the eigenvalues of the top-left corner of size $k \times k$. Plot the two-dimensional array

$$\left\{ (\lambda_i^{(k)}, k) \colon i = 1, \ldots, k,\ k = 1, \ldots, N \right\} \subset \mathbb{R} \times \mathbb{Z}_{\geq 1}.$$

1.4.8 Invariance of the GOE

Show that the distribution of the GOE is invariant under conjugation by orthogonal matrices:

$$\mathbb{P}(OWO^\top \in \mathrm{A}) = \mathbb{P}(W \in \mathrm{A})$$

for all orthogonal matrices O and Borel sets A.

1.4.9 Counting n-powers in the real Wigner matrix

Show that in the expansion of the expected trace of the k-th power of the real Wigner matrix, the maximum power of n is $k/2 + 1$ for even k and less for odd k. For even k, the power $k/2 + 1$ comes from index sequences where each off-diagonal matrix element appears exactly twice, and no diagonal elements are present.

1.4.10 Counting trees

Show that the number of rooted trees with m edges is the m-th Catalan number:

$$C_m = \frac{1}{m+1} \binom{2m}{m}.$$

Chapter 2

Wigner semicircle law

2.1 Recap

We are working on the Wigner semicircle law.

1. Wigner matrices W: real symmetric random matrices with iid entries X_{ij}, $i > j$ (mean 0, variance σ^2); and iid diagonal entries X_{ii} (mean 0, some other variance and distribution).

2. Empirical spectral distribution (ESD)

$$\nu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}},$$

   which is a random probability measure on $\mathbb{R}$.

3. Semicircle distribution μ_{sc}:

$$\mu_{\text{sc}}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2}\, dx, \qquad x \in [-2, 2].$$

4. Computation of expected traces of powers of W (with variance 1). We showed that

$$\int_{\mathbb{R}} x^k \, \nu_n(dx) \to \# \left\{\text{rooted planar trees with } k/2 \text{ edges}\right\}.$$

Remark 2.1. If the off-diagonal elements of the matrix have variance σ^2, then the semicircle distribution should be scaled to be supported on $[-2\sigma, 2\sigma]$. We assume that the variance of the off-diagonal elements is 1 in most arguments throughout the lecture.

2.2 Two computations

First, we finish the combinatorial part, and match the limiting expected traces of powers of W to moments of the semicircle law.

2.2.1 Moments of the semicircle law

We also need to match the Catalan numbers to the moments of the semicircle law. Let $k = 2m$, and we need to compute the integral

$$\int_{-2}^{2} x^{2m} \frac{1}{2\pi} \sqrt{4 - x^2}\, dx.$$

By symmetry, we write:

$$\int_{-2}^{2} x^{2m} \rho(x)\, dx = \frac{2}{\pi} \int_{0}^{2} x^{2m} \sqrt{4 - x^2}\, dx.$$

Using the substitution $x = 2\sin\theta$, we have $dx = 2\cos\theta\, d\theta$. The integral becomes:

$$\frac{2}{\pi} \int_{0}^{\pi/2} (2\sin\theta)^{2m} (2\cos\theta)(2\cos\theta\, d\theta) = \frac{2^{2m+2}}{\pi} \int_{0}^{\pi/2} \sin^{2m}\theta \cos^2\theta\, d\theta.$$

Using $\cos^2\theta = 1 - \sin^2\theta$, we split the integral:

$$\frac{2^{2m+2}}{\pi} \left(\int_{0}^{\pi/2} \sin^{2m}\theta\, d\theta - \int_{0}^{\pi/2} \sin^{2m+2}\theta\, d\theta \right).$$

Using the standard formula (cf. Problem 2.6.1)

$$\int_{0}^{\pi/2} \sin^{2n}\theta\, d\theta = \frac{\pi}{2} \frac{(2n)!}{2^{2n}(n!)^2}, \tag{2.1}$$

we compute each term:

$$\frac{2^{2m+2}}{\pi} \left(\frac{\pi}{2} \frac{(2m)!}{2^{2m}(m!)^2} - \frac{\pi}{2} \frac{(2m+2)!}{2^{2m+2}((m+1)!)^2} \right).$$

After simplification, this becomes C_m, the m-th Catalan number.

2.2.2 Counting trees and Catalan numbers

Throughout this section, for a random matrix trace moment of order k, we use $m = k/2$ as our main parameter. Note that m can be arbitrary (not necessarily even).

Definition 2.2 (Dyck Path)**.** A *Dyck path* of semilength m is a sequence of $2m$ steps in the plane, each step being either $(1,1)$ (up step) or $(1,-1)$ (down step), starting at $(0,0)$ and ending at $(2m,0)$, such that the path never goes below the x-axis. We denote an up step by U and a down step by D.

Definition 2.3 (Rooted Plane Tree)**.** A *rooted plane tree* is a tree with a designated root vertex where the children of each vertex have a fixed left-to-right ordering. The size of such a tree is measured by its number of edges, which we denote by m.

Definition 2.4 (Catalan Numbers)**.** The sequence of *Catalan numbers* $\{C_m\}_{m\geq 0}$ is defined recursively by:

$$C_0 = 1, \quad C_{m+1} = \sum_{j=0}^{m} C_j C_{m-j} \quad \text{for } m \geq 0. \tag{2.2}$$

Alternatively, they have the closed form[1]

$$C_m = \frac{1}{m+1}\binom{2m}{m} = \binom{2m}{m} - \binom{2m}{m+1}. \tag{2.3}$$

These numbers appear naturally in the moments of random matrices, where $m = k/2$ for trace moments of order k.

Lemma 2.5. *Formulas* (2.2) *and* (2.3) *are equivalent.*

Proof. One can check that the closed form satisfies the recurrence relation by direct substitution. The other direction involves generating functions. Namely, (2.2) can be rewritten for the generating function

$$C(z) = \sum_{m=0}^{\infty} C_m z^m$$

as

$$C(z) = 1 + zC(z)^2.$$

[1]See Problem 2.6.4 for a combinatorial proof of the second inequality.

Solving for $C(z)$, we get

$$C(z) = \frac{1 \pm \sqrt{1-4z}}{2z}. \tag{2.4}$$

We need to pick the solution which is nonsingular at $z = 0$, and it corresponds to the minus sign. Taylor expansion of the right-hand side of (2.4) at $z = 0$ gives the closed form. □

Remark 2.6. Catalan numbers enumerate many (too many!) combinatorial objects. For a comprehensive treatment, see [Sta15].

Proposition 2.7 (Dyck Path–Rooted Tree Correspondence)**.** *For any m, there exists a bijection between the set of Dyck paths of semilength m and the set of rooted plane trees with m edges.*

Proof. Given a Dyck path of semilength m, we build the corresponding rooted plane tree as follows (see Figure 2.1 for an illustration):

1. Start with a single root vertex
2. Read the Dyck path from left to right:
   - For each up step (U), add a new child to the current vertex
   - For each down step (D), move back to the parent of the current vertex
3. The order of children is determined by the order of up steps

This is clearly a bijection, and we are done. □

It remains to show that the Dyck paths or rooted plane trees are counted by the Catalan numbers, by verifying the recursion (2.2) for them. By Proposition 2.7, it suffices to consider only Dyck paths.

Proposition 2.8. *The number of Dyck paths of semilength m satisfies the Catalan recurrence* (2.2)*.*

Proof. We need to show that the number of Dyck paths of semilength $m+1$ is given by the sum in the right-hand side of (2.2). Consider a Dyck path of semilength $m+1$, and let the *first* time it returns to zero be at semilength $j+1$, where $j = 0, \ldots, m$. Then the first and the $(2j+1)$-st steps are, respectively, U and D. From 0 to $2j+2$, the path does not return to the x-axis, so we can remove the first and the $(2j+1)$-st steps, and get a proper Dyck path of semilength j. The remainder of the Dyck path is a Dyck path of semilength $m-j$. This yields the desired recurrence. □

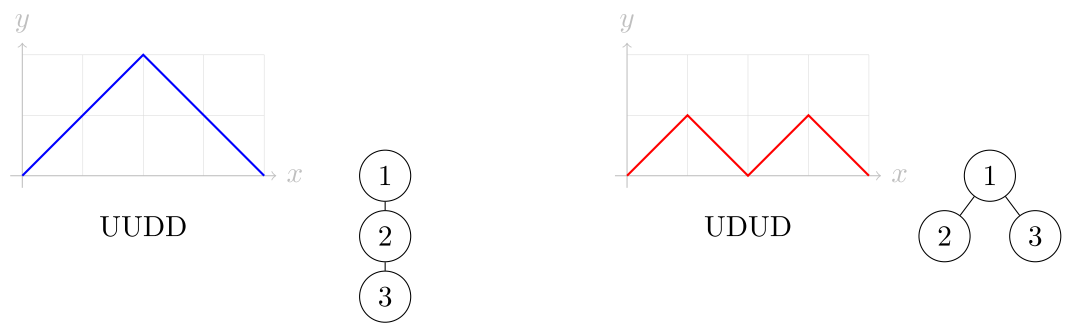

Figure 2.1: The two possible Dyck paths of semilength $m = 2$ and their corresponding rooted plane trees.

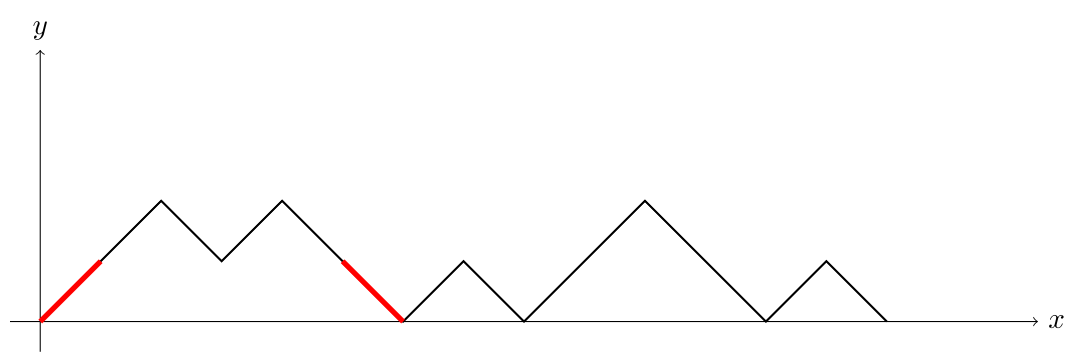

Figure 2.2: Illustration of a Dyck path decomposition for the proof of Proposition 2.8.

2.3 Analysis steps in the proof

We are done with combinatorics, and it remains to justify that the computations lead to the desired semicircle law from Chapter 1.

Let us remember that so far, we showed that

$$\lim_{n\to\infty} \frac{1}{n^{k/2+1}} \mathbb{E}\left[\operatorname{Tr} W^k\right] = \begin{cases} \sigma^{2m} C_m & \text{if } k = 2m \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases}$$

Here, W is real Wigner (unnormalized) with mean 0, where its off-diagonal entries are iid with variance σ^2.

2.3.1 The semicircle distribution is determined by its moments

We use (without proof) the known Carleman's criterion for the uniqueness of a distribution by its moments.

Proposition 2.9 (Carleman's criterion [ST43, Theorem 1.10], [Akh65])**.** *Let X be a real-valued random variable with moments $m_k = \mathbb{E}[X^k]$ of all orders. If*

$$\sum_{k=1}^{\infty} (m_{2k})^{-1/(2k)} = \infty, \tag{2.5}$$

then the distribution of X is uniquely determined by its moments $(m_k)_{k\geq 1}$.

Remark 2.10. Note that we do not assume that the measure is symmetric, but use only even moments for the Carleman criterion. Indeed, in determining uniqueness, the decisive aspect is how the distribution mass "escapes" to $\pm\infty$. Since $\int |x|^n d\mu(x)$ can be bounded by twice $\int x^{2\lfloor n/2 \rfloor} d\mu(x)$ (roughly speaking), controlling $\int x^{2n} d\mu(x)$ also controls $\int |x|^n d\mu(x)$. Thus, one does not need to worry about positive or negative signs in x; the even powers handle both sides of the real line at once.

Moreover, the convergence of (2.5), as for any infinite series, is only determined by arbitrarily large moments, for the same reason.

Remark 2.11. By the Stone-Wierstrass theorem, the semicircle distribution on $[-2, 2]$ is unique among distributions with an arbitrary, but fixed compact support with the moments $\sigma^{2k} C_k$. However, we need to guarantee that there are no distributions on $\mathbb{R}$ with the same moments.

Now, the moments satisfy the asymptotics

$$m_{2k} = C_k \sigma^{2k} \sim \frac{4^k}{k^{3/2}\sqrt{\pi}} \sigma^{2k},$$

so

$$\sum_{k=1}^{\infty} (m_{2k})^{-1/(2k)} \sim \sum_{k=1}^{\infty} \left(\frac{k^{3/2}\sqrt{\pi}}{4^k} \right)^{1/2k} \sigma^{-1}.$$

The k-th summands converges to $1/(2\sigma)$, so the series diverges.

Remark 2.12. See also Problem A.4 from Chapter 1 on an example of a distribution not determined by its moments.

2.3.2 Convergence to the semicircle law

Recall [Bil95, Theorem 30.2] that convergence of random variables in moments plus the fact that the limiting distribution is uniquely determined by its moments implies convergence in distribution. However, we need weak convergence in probability or almost surely (see the previous Chapter 1). which deals with random variables

$$\int_{\mathbb{R}} f(x)\, \nu_n(dx), \qquad f \in C_b(\mathbb{R}),$$

and we did not compute the moments of these random variables.

To complete the argument, let us show that for each fixed integer $k \geq 1$, we have almost sure convergence of the moments (of a random distribution, so that the $Y_{n,k}$'s are random variables):

$$Y_{n,k} := \int_{\mathbb{R}} x^k\, \nu_n(dx) \xrightarrow[n\to\infty]{\text{a.s.}} m_k, \qquad n \to \infty,$$

where m_k are the moments of the semicircle distribution, and ν_n is the ESD corresponding to the scaling of the eigenvalues as $\lambda_i/\sqrt{n}$.

As typical in asymptotic probability, we not only need the expectation of $Y_{n,k}$, but also their variances, to control the almost sure convergence. Recall that we showed $\mathbb{E}(Y_{n,k}) \to m_k$. Let us assume the following:

Proposition 2.13 (Variance bound)**.** *For each fixed integer $k \geq 1$ and large enough n, we have*

$$\operatorname{Var}(Y_{n,k}) \leq \frac{m_k}{n^2}.$$

We will prove Proposition 2.13 in Section 2.4 below. Let us finish the proof of convergence to the semicircle law modulo Proposition 2.13.

A concentration bound and the Borel–Cantelli lemma

From Chebyshev's inequality,

$$\mathbb{P}\Big(\big|Y_{n,k} - \mathbb{E}[Y_{n,k}]\big| \geq n^{-\frac{1}{4}}\Big) \leq \mathrm{Var}[Y_{n,k}]\sqrt{n} = O(n^{-\frac{3}{2}}),$$

where in the last step we used Proposition 2.13.

Hence the probability that $|Y_{n,k} - \mathbb{E}[Y_{n,k}]| > n^{-\frac{1}{4}}$ is summable in n. By the Borel–Cantelli lemma, with probability 1 only finitely many of these events occur. Since $\mathbb{E}[Y_{n,k}] \to m_k$, we conclude

$$\big|Y_{n,k} - m_k\big| \ \leq\ \big|Y_{n,k} - \mathbb{E}[Y_{n,k}]\big| + \big|\mathbb{E}[Y_{n,k}] - m_k\big| \ \xrightarrow[n\to\infty]{}\ 0 \quad \text{almost surely.}$$

Tightness of $\{\nu_n\}$ and subsequential limits

Since $|Y_{n,k}| = \big|\int x^k \, \nu_n(dx)\big|$ stays almost surely bounded for each k, one readily checks (Problem 2.6.5) that almost surely, for each fixed k,

$$\nu_n\big(\{x : |x| > M\}\big) \ \leq\ \frac{C}{M^k}. \tag{2.6}$$

Here, C may depend on k, but its growth is at most exponential in k due to the Catalan number moments. By choosing k large, we see that ν_n puts arbitrarily little mass outside any interval $[-M, M]$ for sufficiently large M. Thus, the sequence of probability measures $\{\nu_n\}$ is *tight*. By Prokhorov's theorem [Bil95, Theorem 25.10], there exists a subsequence ν_{n_j} converging weakly to some probability measure ν^*. We will now characterize all subsequential limits ν^* of ν_n.

Characterizing the limit measure

We claim that $\nu^* = \mu_{\mathrm{sc}}$, the semicircle distribution (and in particular, this measure is not random). Indeed, fix k. Since x^k is a bounded function on a sufficiently large interval, and $\nu_{n_j} \to \nu^*$ weakly, we have

$$\int_{\mathbb{R}} x^k \, \nu_{n_j}(dx) \ \to\ \int_{\mathbb{R}} x^k \, \nu^*(dx).$$

On the other hand, we have already shown

$$\int_{\mathbb{R}} x^k \, \nu_{n_j}(dx) \ =\ Y_{n_j,k} \ \xrightarrow[j\to\infty]{\text{a.s.}} m_k \ =\ \int_{\mathbb{R}} x^k \, \mu_{\mathrm{sc}}(dx).$$

Thus

$$\int_{\mathbb{R}} x^k \,\nu^*(dx) \;=\; m_k \;=\; \int_{\mathbb{R}} x^k \,\mu_{\mathrm{sc}}(dx) \qquad \text{for all } k \geq 1.$$

By Proposition 2.9, the measure ν^* is uniquely determined by its moments. Hence ν^* must coincide with μ_{sc}.

Remark 2.14. In Section 2.3.2 and **??** we tacitly assumed that we choose an elementary outcome ω, and view ν_n as measures depending on ω. Then, since the convergence of moments is almost sure, ω belongs to a set of full probability. The limiting measure ν^* must coincide with μ_{sc} for this ω, and thus, ν^* is almost surely nonrandom.

Any subsequence of $\{\nu_n\}$ has a further sub-subsequence convergent to ν. By a standard diagonal argument, this forces $\nu_n \to \nu$ in the weak topology (almost surely). This completes the proof that the ESD of our Wigner matrix (rescaled by $\sqrt{n}$) converges to the semicircle distribution weakly almost surely, modulo Proposition 2.13. (See also Problem 2.6.6 for the weakly in probability convergence.)

2.4 Proof of Proposition 2.13: bounding the variance

There is one more "combinatorial" step in the proof of the semicircle law: we need to show that the variance of the moments of the ESD is bounded by m_k/n^2.

Recall that

$$Y_{n,k} = \int_{\mathbb{R}} x^k \,\nu_n(dx) = \frac{1}{n^{1+\frac{k}{2}}} \sum_{i_1,\ldots,i_k=1}^{n} X_I, \quad \text{where} \quad X_I = X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}.$$

Here we use the notation I for the multi-index $(i_1, \ldots, i_k)$, and throughout the computation below, we use the notation $I \in [n]^k$, where $[n] = \{1, \ldots, n\}$. We have

$$\operatorname{Var}\bigl(Y_{n,k}\bigr) = \frac{1}{n^{2+k}} \operatorname{Var}\Bigl(\sum_{I \in [n]^k} X_I\Bigr) = \frac{1}{n^{2+k}} \sum_{I,J \in [n]^k} \operatorname{Cov}\bigl(X_I, X_J\bigr).$$

We claim that the sum of all covariances is bounded by a constant times n^k, which then implies $\operatorname{Var}\bigl(Y_{n,k}\bigr) \leq \text{const} \cdot n^k / n^{2+k} = O\bigl(\frac{1}{n^2}\bigr)$.

Step 1. Identifying when $\mathrm{Cov}(X_I, X_J)$ **can be nonzero.** For each k-tuple $I = (i_1, i_2, \ldots, i_k) \in [n]^k$, the product

$$X_I = X_{i_1 i_2} X_{i_2 i_3} \ldots X_{i_k i_1}$$

is the product of the entries of our Wigner matrix corresponding to the directed "edges" $(i_1 \to i_2), (i_2 \to i_3), \ldots, (i_k \to i_1)$. Similarly, X_J is determined by the edges of another closed directed walk J.

1. If I and J use disjoint collections of matrix entries, then X_I and X_J are independent, and hence $\mathrm{Cov}(X_I, X_J) = 0$.

2. If there is an edge (say, $X_{i_1 i_2}$) which appears *only once* in exactly one of I or J but not both, then that edge factor is independent and forces $\mathrm{Cov}(X_I, X_J) = 0$ since $\mathbb{E}[X_{i_1 i_2}] = 0$. Indeed, for example if $X_{i_1 i_2}$ appears only in X_I, then

   $$\mathbb{E}[X_I] = \mathbb{E}[X_{i_1 i_2}] \cdot \mathbb{E}[\text{other factors}] = 0, \qquad \mathbb{E}\big[X_I X_J\big] = \mathbb{E}[X_{i_1 i_2}] \cdot \mathbb{E}\big[\text{other factors}\big] = 0.$$

Thus, the only way we could get a nonzero covariance is if *every* edge that appears in $I \cup J$ appears at least twice overall. Graphically, let us represent each k-tuple I by a directed closed walk in the complete graph on $[n]$. The union $I \cup J$ must be a connected subgraph in which every directed edge has total multiplicity ≥ 2.

Step 2. Counting the contributions to the sum. Denote by $q = |V(I \cup J)|$ the number of distinct vertices involved in the union $I \cup J$. In principle, there are $O(n^q)$ ways to choose q vertices from $[n]$. Then we need to specify how the edges form two closed walks of length k.

We split into two cases:

1. $\boldsymbol{q \leq k}$**.** Then the n-power in the sum over I, J is at most n^k, which yields the overall contribution $O(n^{-2})$, as desired.

2. $\boldsymbol{q \geq k + 1}$**.** Ignoring directions and multiplicities, we see that the subgraph corresponding to $I \cup J$ contains at most k edges. Since $q \geq k+1$, we must have $q = k + 1$ (by connectedness). Thus, $I \cup J$ is a double tree. Since I and J are subsets of this double tree and $q = k + 1$, they also must be double trees. Thus, there exists an edge which appears in both I and J, and at least twice in I and twice in J, so four times in $I \cup J$. This contradicts the assumption that $I \cup J$ is a double tree.

   This implies that there are no leading contributions to the sum when $q \geq k + 1$.

Combining these two cases, we conclude that the total number of pairs (I, J) with nonzero covariance is of order at most n^k, This yields the desired bound on the variance, and completes the proof of Proposition 2.13.

With that, we are done with the Wigner semicircle law proof for real Wigner matrices (with weakly almost sure convergence; see Chapter 1 for the definitions).

Also, see Problem 2.6.7 for the complex case of the Wigner semicircle law.

2.5 Remark: Variants of the semicircle law

Let us briefly outline a few examples of the semicircle law for real/complex Wigner matrices which relax the iid conditions and the conditions that all moments of the entries must be finite. This list is not comprehensive, it is presented as an illustration of the universality / robustness of the semicircle law.

Theorem 2.15 (Gaussian β-Ensembles [Joh98], [For10])**.** *Let $\beta > 0$, and consider an $n \times n$ random matrix ensemble with joint eigenvalue density:*

$$p_n(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_{n,\beta}} \exp\left(-\frac{\beta}{4} \sum_{i=1}^{n} \lambda_i^2\right) \prod_{1 \le i < j \le n} |\lambda_i - \lambda_j|^\beta \tag{2.7}$$

where $Z_{n,\beta}$ is the normalization constant.[2] *Then the ESD of the normalized eigenvalues $\lambda_i/\sqrt{n}$ converges weakly almost surely to the semicircle law.*

Theorem 2.16 (Correlated entries [SSB05])**.** *Let $W_n = \left(\frac{1}{\sqrt{n}} X_{pq}\right)_{1 \le p,q \le n}$ be a sequence of $n \times n$ Hermitian random matrices where:*

1. *The entries X_{pq} are complex random variables that are:*
   - *Centered:* $\mathbb{E}[X_{pq}] = 0$,
   - *Unit variance:* $\mathbb{E}[|X_{pq}|^2] = 1$,
   - *Moment bound:* $\sup_n \max_{p,q=1,\ldots,n} \mathbb{E}\left[|X_{pq}|^k\right] < \infty$ *for all* $k \in \mathbb{N}$.

[2] For $\beta = 1, 2, 4$, this is the joint eigenvalue density of the Gaussian Orthogonal, Unitary, and Symplectic Ensembles, respectively. For general β, there is no invariant random matrix distribution (while the eigenvalue density (2.7) makes sense), and we can still treat all the β cases in a unified manner.

2. *There exists an equivalence relation* $\sim_n$ *on pairs of indices* (p, q) *in* $\{1, \ldots, n\}^2$ *such that:*

   - *Entries* $X_{p_1 q_1}, \ldots, X_{p_j q_j}$ *are independent when* $(p_1, q_1), \ldots, (p_j, q_j)$ *belong to distinct equivalence classes.*
   - *The relation satisfies the following bounds:*

     (a) $\max_p \#\big\{(q, p', q') \in \{1, \ldots, n\}^3 \mid (p, q) \sim_n (p', q')\big\} = o(n^2)$,

     (b) $\max_{p,q,p'} \#\big\{q' \in \{1, \ldots, n\} \mid (p, q) \sim_n (p', q')\big\} \leq B$ *for some constant* B,

     (c) $\#\big\{(p, q, p') \in \{1, \ldots, n\}^3 \mid (p, q) \sim_n (q, p') \text{ and } p \neq p'\big\} = o(n^2)$.

3. *The matrices are Hermitian:* $X_{pq} = \overline{X_{qp}}$. *In particular,* $(p, q) \sim_n (q, p)$, *and this is consistent with the conditions on the equivalence relation.*

Then, as $n \to \infty$, *the ESD of* W_n *converges to the semicircle law.*

There are variants of this theorem without the assumption that all moments of the entries are finite.

Theorem 2.17 ([BGK16])**.** *Let* $M_n = [X_{ij}]_{i,j=1}^n$ *be a symmetric* $n \times n$ *matrix with random entries such that:*

- *The off-diagonal elements* X_{ij}, *for* $i < j$, *are i.i.d. random variables with* $\mathbb{E}[X_{ij}] = 0$ *and* $\mathbb{E}[X_{ij}^2] = 1$.

- *The diagonal elements* X_{ii} *are i.i.d. random variables with* $\mathbb{E}[X_{ii}] = 0$ *and a finite second moment,* $\mathbb{E}[X_{ii}^2] < \infty$, *for* $1 \leq i \leq n$.

Then the ESD of M_n, *normalized by* $\sqrt{n}$, *converges to the semicircle law.*

Theorem 2.18. *For each* $n \in \mathbb{Z}_+$, *let* $M_n = [X_{ij}]_{i,j=1}^n$ *be a symmetric* $n \times n$ *matrix with real random entries satisfying the following conditions:*

- *The entries* X_{ij} *are independent (but not necessarily identically distributed) random variables with* $\mathbb{E}[X_{ij}] = 0$ *and* $\mathbb{E}[X_{ij}^2] = 1$.

- *There exists a constant* C *such that* $\sup_{i,j,n} \mathbb{E}\left[|X_{ij}|^4\right] < C$.

Then the ESD of M_n, *normalized by* $\sqrt{n}$, *converges to the semicircle law almost surely. The second condition can also be replaced by a uniform integrability condition on the variances.*

Theorem 2.19 (For example, see [SB95]). *Let $M_n = [X_{ij}]_{i,j=1}^n$ be a symmetric $n \times n$ matrix with random entries. Assume that the expected matrix $\mathbb{E}[M_n]$ has rank $r(n)$, where*

$$\lim_{n\to\infty} \frac{r(n)}{n} = 0.$$

Additionally, suppose $\mathbb{E}[X_{ij}] = 0$, $\mathrm{Var}(X_{ij}) = 1$, and

$$\sup_{i,j,n} \mathbb{E}\left[|X_{ij} - \mathbb{E}[X_{ij}]|^4\right] < \infty.$$

Then the ESD of M_n, normalized by $\sqrt{n}$, converges to the semicircle law almost surely.

2.6 Problems

2.6.1 Standard formula

Prove formula (2.1):

$$\int_0^{\pi/2} \sin^{2n}\theta \, d\theta = \frac{\pi}{2} \frac{(2n)!}{2^{2n}(n!)^2}.$$

2.6.2 Tree profiles

Show that the expected height of a uniformly random Dyck path of semilength m is of order $\sqrt{m}$.

2.6.3 Ballot problem

Suppose candidate A receives p votes and candidate B receives q votes, where $p > q \geq 0$. In how many ways can these votes be counted such that A is always strictly ahead of B in partial tallies?

2.6.4 Reflection principle

Show the equality

$$C_m = \binom{2m}{m} - \binom{2m}{m-1},$$

where C_m counts the number of lattice paths from $(0,0)$ to $(2m,0)$ with steps $(1,1)$ and $(1,-1)$ that never go below the x-axis, and binomial coefficients count arbitrary lattice paths from $(0,0)$ to $(2m,0)$ or to $(2m,2)$ with steps $(1,1)$ and $(1,-1)$. In other words, show that the difference between the number of paths to $(2m,0)$ and to $(2m,2)$ is C_m, the number of paths that never go below the x-axis.

2.6.5 Bounding probability in the proof

Show inequality (2.6).

2.6.6 Almost sure convergence and convergence in probability

Show that in Wigner's semicircle law, the weakly almost sure convergence of random measures ν_n to μ_{sc} implies weak convergence in probability.

2.6.7 Wigner's semicircle law for complex Wigner matrices

Complex Wigner matrices are Hermitian symmetric, with iid complex off-diagonal entries, and real iid diagonal entries (all mean zero). Each complex random variable has independent real and imaginary parts.

1. Compute the expected trace of powers of a complex Wigner matrix.

2. Outline the remaining steps in the proof of Wigner's semicircle law for complex Wigner matrices.

2.6.8 Semicircle law without the moment condition

Prove Theorem 2.17.

Chapter 3

Gaussian and tridiagonal matrices

3.1 Recap

We have established the semicircle law for real Wigner random matrices. If W is an $n \times n$ real symmetric matrix with independent entries X_{ij} above the main diagonal (mean zero, variance 1), and mean zero diagonal entries, then the empirical spectral distribution of $W/\sqrt{n}$ converges to the semicircle law as $n \to \infty$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} = \mu_{\mathrm{sc}}, \tag{3.1}$$

where

$$\mu_{\mathrm{sc}}(dx) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}\,dx, & \text{if } |x| \le 2, \\ 0, & \text{otherwise.} \end{cases}$$

The convergence in (3.1) is weakly almost sure. The way we got the result is by expanding $\mathbb{E}\,\mathrm{Tr}(W^k)$ and counting trees, plus analytic lemmas which ensure that the convergence of expected powers of traces is enough to conclude the convergence (3.1) of the empirical spectral measures.

Today, we are going to focus on Gaussian ensembles. The plan is:

- Definition and spectral density for real symmetric Gaussian matrices (GOE).
- Other random matrix ensembles with explicit eigenvalue densities: Wishart (Laguerre) and Jacobi (MANOVA/CCA) ensembles.
- Tridiagonalization and general beta ensemble.

- (next week, not today) Wigner's semicircle law via tridiagonalization.

3.2 Gaussian ensembles

3.2.1 Definitions

Recall that a real Wigner matrix W can be modeled as

$$W = \frac{Y + Y^\top}{\sqrt{2}},$$

where Y is an $n \times n$ matrix with independent entries Y_{ij}, $1 \le i, j \le n$, such that Y_{ij} are mean zero, variance 1. Then for $1 \le i < j \le n$, we have for the matrix $W = (X_{ij})$:

$$\mathrm{Var}\,(X_{ii}) = \mathrm{Var}(\sqrt{2}Y_{ii}) = 2, \qquad \mathrm{Var}\,(X_{ij}) = \mathrm{Var}\left(\frac{Y_{ij} + Y_{ji}}{\sqrt{2}}\right) = 1.$$

If, in addition, we assume that Y_{ij} are standard Gaussian $\mathcal{N}(0,1)$, then the distribution of W is called the *Gaussian Orthogonal Ensemble* (GOE).

For the complex case, we have the *standard complex Gaussian random variable*

$$Z = \frac{1}{\sqrt{2}}\left(Z^R + \mathbf{i}Z^I\right), \qquad \mathbb{E}(Z) = 0, \qquad \mathrm{Var}_{\mathbb{C}}(Z) := \mathbb{E}(|Z|^2) = \frac{\mathbb{E}(|Z^R|^2) + \mathbb{E}(|Z^I|^2)}{2} = 1,$$

where Z^R and Z^I are independent standard Gaussian real random variables $\mathcal{N}(0,1)$.

If we take Y to be an $n \times n$ matrix with independent entries Y_{ij}, $1 \le i, j \le n$ distributed as Z, then the random matrix[1]

$$W = \frac{Y + Y^\dagger}{\sqrt{2}}$$

is said to have the *Gaussian Unitary Ensemble* (GUE) distribution. For the GUE matrix $W = (X_{ij})$, we have for $1 \le i < j \le n$:

$$\mathrm{Var}_{\mathbb{C}}(X_{ii}) = 1, \qquad \mathrm{Var}_{\mathbb{C}}(X_{ij}) = \frac{1}{4}\Big[\mathbb{E}(Z^R_{ij} + Z^R_{ji})^2 + \mathbb{E}(Z^I_{ij} + Z^I_{ji})^2\Big] = 1.$$

Both GOE and GUE have real eigenvalues $\lambda_1 \ge \ldots \ge \lambda_n$. We are going to describe the joint distribution of these eigenvalues. Despite the fact that

[1] $Y^\dagger$ denotes the transpose of Y combined with complex conjugation.

the map from a matrix to its eigenvalues is quite complicated and nonlinear (you need to solve an equation of degree n), the distribution of eigenvalues in the Gaussian cases is fully explicit.

See Problem 3.6.1 for invariance of GOE/GUE under orthogonal/unitary conjugation (this is where the names "orthogonal" and "unitary" come from).

Remark 3.1. There is a third player in the game, the *Gaussian Symplectic Ensemble* (GSE),which we will mainly ignore in this course due to its less intuitive quaternionic nature.

3.2.2 Joint eigenvalue distribution for GOE

In this section, we give a derivation of the joint probability density for the GOE.

Theorem 3.2 (GOE Joint Eigenvalue Density)**.** *Let W be an $n \times n$ real symmetric matrix with the GOE distribution (Section 3.2.1). Then its ordered real eigenvalues $\lambda_1 \le \cdots \le \lambda_n$ of $W/\sqrt{2}$ have a joint probability density function on $\mathbb{R}^n$ given by:*

$$p(\lambda_1, \ldots, \lambda_n) \;=\; \frac{1}{Z_n} \prod_{1 \le i < j \le n} \left|\lambda_i - \lambda_j\right| \, \exp\Big(-\frac{1}{2}\sum_{k=1}^{n} \lambda_k^2\Big),$$

where Z_n is a constant (depending on n but not on λ_i) ensuring the density integrates to 1:

$$Z_n = Z_n^{GOE} = \frac{(2\pi)^{n/2}}{n!} \prod_{j=0}^{n-1} \frac{\Gamma(1 + (j+1)\beta/2)}{\Gamma(1+\beta/2)}, \qquad \beta = 1.$$

Remark 3.3. We renormalized the GOE by a factor of $\sqrt{2}$ to make the Gaussian part of the density, $\exp(-\frac{1}{2}\sum_{k=1}^{n} \lambda_k^2)$, standard. In the GUE case, no normalization is required.

We break the proof into four major steps, considered in Section 3.2.3 and **??????** below.

3.2.3 Step A. Joint density of matrix entries

Let us label all independent entries of $W/\sqrt{2}$:

$$\{\underbrace{X_{12}, X_{13}, \ldots, X_{23}, \ldots}_{\text{above diag}}, \underbrace{X_{22}, X_{33}, \ldots}_{\text{diag}}\}.$$

There are $\frac{n(n-1)}{2}$ off-diagonal entries with variance 1/2, and n diagonal entries with variance 1. The joint density of these entries (ignoring normalization for a moment) is proportional to

$$f(x_{12}, x_{13}, \ldots, x_{22}, x_{33}, \ldots) \propto \exp\Big(-\sum_{i<j} x_{ij}^2 - \frac{1}{2}\sum_{i=1}^{n} x_{ii}^2\Big) = \exp\Big(-\frac{1}{2}\sum_{i,j=1}^{n} x_{ij}^2\Big), \tag{3.2}$$

where in the right-hand side, we have $x_{ij} = x_{ji}$ for $i \neq j$. We then recognize

$$\sum_{i,j=1}^{n} x_{ij}^2 = \mathrm{Tr}(W^2) = \sum_{k=1}^{n} \lambda_k^2.$$

Including the normalization for Gaussians, one arrives at the density on $\mathbb{R}^{n(n+1)/2}$:

$$f(W)\, dW \;=\; \pi^{-\frac{n(n-1)}{4}} \big(2\pi\big)^{-\frac{n}{4}} \exp\Big(-\tfrac{1}{2}\,\mathrm{Tr}(W^2)\Big)\, dW,$$

where dW is the product measure over the $\frac{n(n+1)}{2}$ independent entries.

3.2.4 Step B. Spectral decomposition

Since W is real symmetric, it can be orthogonally diagonalized:

$$W = Q\,\Lambda\,Q^\top, \quad Q \in O(n),$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ has the eigenvalues. Then, as we saw before, we have

$$\mathrm{Tr}(W^2) = \mathrm{Tr}\big(Q\,\Lambda\,Q^\top Q\,\Lambda\,Q^\top\big) = \mathrm{Tr}(\Lambda^2) = \sum_{k=1}^{n} \lambda_k^2.$$

The map from W to (Λ, Q) is not one-to one, but in case W has distinct eigenvalues, the preimage of (Λ, Q) contains 2^n elements. See Problems 3.6.2 and 3.6.3.

It remains to make the change of variables from W to Λ, which involves the Jacobian.

3.2.5 Step C. Jacobian

We now examine how the measure dW in the space of real symmetric matrices factors into a piece depending on $\{\lambda_i\}$ and a piece depending on Q. Formally,

$$dW = \Big|\det\big(\tfrac{\partial W}{\partial(\Lambda, Q)}\big)\Big|\, d\Lambda\, dQ,$$

where dQ is the Haar measure[2] on $O(n)$, and $d\Lambda$ is the Lebesgue measure on $\mathbb{R}^n$. The Lebesgue measure later needs to be restricted to the "Weyl chamber" $\lambda_1 \le \cdots \le \lambda_n$ if we want an ordering, this introduces the simple factor $n!$ in the final density.

Lemma 3.4 (Jacobian for Spectral Decomposition). *For real symmetric $W = Q\Lambda Q^\top$, one has*

$$\left|\det\left(\tfrac{\partial W}{\partial(\Lambda,Q)}\right)\right| \;=\; \text{const} \prod_{1\le i<j\le n} \left|\lambda_i - \lambda_j\right|,$$

where the constant is independent of the λ_i's and depends only on n.

Remark 3.5. Equivalently, one often writes

$$dW \;=\; \left|\Delta(\lambda_1,\dots,\lambda_n)\right| \, d\Lambda \, dQ, \quad \text{where } \Delta(\lambda_1,\dots,\lambda_n) = \prod_{i<j}(\lambda_j - \lambda_i)$$

is the *Vandermonde determinant.*

We prove Lemma 3.4 in the rest of this subsection.

Consider small perturbations of Λ and Q. Write

$$W = Q\,\Lambda\,Q^\top, \quad \Lambda = \operatorname{diag}(\lambda_1,\dots,\lambda_n).$$

Let δW be an infinitesimal change in W. We want to see how δW depends on $\delta\Lambda$ and δQ.

Parametrizing δQ. Since $Q \in O(n)$, any small variation of Q can be expressed as

$$Q\exp(B) \approx Q(I+B),$$

where B is an infinitesimal skew-symmetric matrix ($B^\top = -B$). Indeed, $\exp(B)$ must be orthogonal, so $\exp(B)^\top\exp(B) = I$. Thus, we have

$$(I+B)^\top(I+B) = I, \qquad \text{or} \qquad B^\top + B = 0.$$

Note that $\exp(B)$ is the matrix exponential of B, which is defined by the usual power series. Note also that the dimension of $O(n)$ is $\dim(O(n)) = \frac{n(n-1)}{2}$, which matches the dimension of the space of skew-symmetric matrices.

[2] Recall that the Haar measure on $O(n)$ is the unique (up to a constant factor) measure that is invariant under group shifts (in this situation, both left and right shifts work). In probabilistic terms, if a random orthogonal matrix Q is Haar-distributed, then QR and RQ are also Haar-distributed for any fixed orthogonal matrix R.

Computing δW. Under an infinitesimal change, say,

$$Q \mapsto Q\,(I+B), \quad \Lambda \mapsto \Lambda + \delta\Lambda,$$

we have

$$W = Q\Lambda Q^\top \implies Q^\top \delta W Q = \delta\Lambda + B\Lambda - \Lambda B,$$

to first order in small quantities. Here we used the orthogonality of Q and the skew-symmetry of B.

Local structure of the map. We see that the map $W \mapsto (\Lambda, Q)$ in a neighborhood of (Λ, Q) determined by $\delta\Lambda$ and B locally translates by $Q^\top \delta\Lambda Q$, which implies the Lebesgue factor $d\lambda_1 \ldots d\lambda_n$ in δW. Indeed, the Lebesgue measure on $\mathbb{R}^n$ is invariant under orthogonal transformations.

The next terms, the commutator $[B, \Lambda]$, has the form (recall that B is infinitesimally small and Λ is diagonal):

$$\begin{aligned}
B\Lambda - \Lambda B &= \begin{pmatrix} 0 & b_{12} & \cdots \\ -b_{12} & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} - \begin{pmatrix} \lambda_1 & 0 & \cdots \\ 0 & \lambda_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} 0 & b_{12} & \cdots \\ -b_{12} & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&= \begin{pmatrix} 0 & b_{12}\lambda_2 & \cdots \\ -b_{12}\lambda_1 & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} - \begin{pmatrix} 0 & b_{12}\lambda_1 & \cdots \\ b_{12}\lambda_2 & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \\
&= \begin{pmatrix} 0 & b_{12}(\lambda_2 - \lambda_1) & \cdots \\ b_{12}(\lambda_1 - \lambda_2) & 0 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.
\end{aligned}$$

Thus, this action locally means that the infinitesimal b_{ij} is multiplied by $\lambda_i - \lambda_j$, for all $1 \le i < j \le n$. This is a scalar factor that does not depend on the orthogonal component Q, but only on the eigenvalues. Therefore, this factor is the same in $Q^\top \delta W Q$.

This completes the proof of Lemma 3.4. See also Problem 3.6.5 for the GUE Jacobian.

3.2.6 Step D. Final Form of the density

Putting Steps A–C together, we find:

$$dW = \text{const} \cdot \prod_{i<j} |\lambda_i - \lambda_j|\, d\Lambda \underbrace{(\text{Haar measure on } O(n))}_{\text{does not depend on } \lambda_i}.$$

Hence, the joint density of $\{\lambda_1, \ldots, \lambda_n\}$ is, up to normalization depending only on n, equal to

$$\prod_{i<j} |\lambda_i - \lambda_j| \ \exp\Big(-\tfrac{1}{2} \sum_{k=1}^{n} \lambda_k^2\Big). \tag{3.3}$$

We leave the computation of the normalization constant in Theorem 3.2 as Problem 3.6.6.

Remark 3.6. We emphasize that in the GOE case, the normalization $W/\sqrt{2}$ for (3.3) is so that the variance is 1 on the diagonal and $\frac{1}{2}$ off the diagonal.

3.3 Other classical ensembles with explicit eigenvalue densities

Let us briefly discuss other classical ensembles with explicit eigenvalue densities, which are not necessarily Gaussian, but are related to other classical structures like orthogonal polynomials. These ensembles also have a built-in parameter β (and in the cases $\beta = 1, 2, 4$, they have invariance under orthogonal/unitary/symplectic conjugation).

3.3.1 Wishart (Laguerre) ensemble

In this subsection, we describe another classical family of random matrices whose eigenvalues form a fundamental example of a β-ensemble with a "logarithmic" pairwise interaction. These are called the *Wishart* or *Laguerre* ensembles. Their importance arises in statistics (covariance estimation, principal component analysis), signal processing, and many other areas.

Definition via SVD

Let X be an $n\times m$ random matrix with iid entries drawn from a real/complex/quaternionic normal distribution. We assume $n \le m$. We can perform the *singular value decomposition* (SVD) of X:

$$X = U \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix} V^\dagger,$$

where U, V are orthogonal/unitary/symplectic matrices (depending on β), $s_1, \ldots, s_n \ge 0$ are the singular values of X, and $\dagger$ means the corresponding conjugation. For example, in the real case, $s_1, \dots, s_n$ are the square roots of the eigenvalues of $XX^\top$.

Moreover, let $W = XX^\dagger$; this is called the Wishart random matrix ensemble. We have

$$\lambda_i = s_i^2, \qquad i = 1, \dots, n; \qquad \lambda_1 \geq \cdots \geq \lambda_n \geq 0.$$

These eigenvalues admit a closed-form joint probability density function (pdf) in complete analogy with the GOE/GUE calculations from previous subsections.

Joint density of eigenvalues

Theorem 3.7 (Wishart eigenvalue density)**.** *The ordered eigenvalues $\lambda_1, \dots, \lambda_n \geq 0$ of the $n \times n$ Wishart matrix W have the joint density on $\{\lambda_i \geq 0\}$ proportional to*

$$\prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^\beta \prod_{i=1}^n \lambda_i^{\frac{\beta}{2}(m-n+1)-1} \exp\Big(-\frac{\lambda_i}{2}\Big),$$

where $\beta = 1, 2, 4$ corresponds to the real, complex, or quaternionic case, respectively.

Idea of proof (sketch). The proof is a variant of the derivation for the joint eigenvalue density in the GOE/GUE case (see Section 3.2.2). One writes down the joint distribution of all entries of X, changes variables to singular values and orthogonal/unitary transformations, and identifies the Jacobian factor as $\prod_{i<j} |s_i^2 - s_j^2|^\beta = \prod_{i<j} |\lambda_i - \lambda_j|^\beta$. The extra factors in front arise from the powers of λ_i (i.e. from $\prod_i s_i$) and the Gaussian exponential $\exp\big(-\frac{1}{2}\sum s_i^2\big)$ when reshaped to $\exp\big(-\frac{1}{2}\sum \lambda_i\big)$. □

Remark 3.8. The exponent of λ_i in the product is often written as $\alpha = \frac{\beta}{2}(m - n + 1) - 1$. One also sees the name *multivariate Gamma distribution* in statistics. For $\beta = 1$ the ensemble is sometimes called the *real Wishart* (or *Laguerre Orthogonal*) ensemble; for $\beta = 2$ it is the *complex Wishart* (or *Laguerre Unitary*) ensemble; and $\beta = 4$ (not discussed in detail here) is the *symplectic version.* In point processes, the case $\beta = 2$ is also referred to as the *Laguerre orthogonal polynomial ensemble.*

3.3.2 Jacobi (MANOVA/CCA) ensemble

The *Jacobi* (sometimes called *MANOVA* or *CCA*) ensemble arises when one looks at the interaction between two independent rectangular Gaussian matrices that share the same number of columns. Statistically, this corresponds to questions of canonical correlations or multivariate Beta distributions. In random matrix theory, it appears as yet another fundamental example of a β-ensemble with an explicit eigenvalue density.

Setup

Let X be an $n \times t$ real (or complex) matrix and Y be a $k \times t$ matrix, with $n \leq k \leq t$. Assume X and Y have iid Gaussian entries (real or complex) of mean 0 and variance 1 and are independent of each other.

Definition 3.9 (Projectors and canonical correlations)**.** Denote by

$$P_X \;=\; X^\top (X\,X^\top)^{-1} X \quad \bigl(\text{or } X^\dagger (X\,X^\dagger)^{-1} X\bigr),$$

the orthogonal (unitary) projector onto the row span of X. Similarly, define

$$P_Y \;=\; Y^\top (Y\,Y^\top)^{-1} Y.$$

These are $t \times t$ projection matrices of ranks n and k, respectively, embedded in a space of dimension t. One checks that P_X and P_Y commute if and only if the row spaces of X and Y are aligned in a certain way. The *canonical correlations* between these two subspaces are the singular values of $P_X P_Y$. Equivalently, the *squared* canonical correlations are the nonzero eigenvalues of $P_X P_Y$.

Since $\operatorname{rank}(P_X P_Y) \leq \min(n, k)$, there are at most $\min(n, k)$ nonzero eigenvalues of $P_X P_Y$. In fact, generically (when the subspaces are in "general position"), there are exactly $\min(n, k)$ nonzero eigenvalues.

Example 3.10. For $n = k = 1$, we have

$$P_X P_Y = \frac{\langle X, Y\rangle}{\langle X, X\rangle \langle Y, X\rangle} X^\top Y,$$

which is a rank one matrix with the only nonzero singular eigenvalue $\langle X, Y\rangle$. Therefore, the singular value is exactly the sample correlation coefficient between X and Y.

Jacobi ensemble

Theorem 3.11 (Jacobi/MANOVA/CCA Distribution)**.** *Let X and Y be as above, each having iid (real or complex) Gaussian entries of size $n \times t$ and $k \times t$, respectively, with $n \leq k \leq t$. Assume further that X and Y are independent of each other (this is the null hypothesis in statistics).*

Then the nonzero eigenvalues $\lambda_1, \ldots, \lambda_n$ of the matrix $P_X P_Y$ lie in the interval $[0,1]$ and have the joint density function of the form

$$\prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^{n} \lambda_i^{\frac{\beta}{2}(k-n+1)-1} \big(1-\lambda_i\big)^{\frac{\beta}{2}(t-n-k+1)-1},$$

up to a normalization constant that depends on n, k, t (but not on $\{\lambda_i\}$). Here again $\beta = 1$ for the real case and $\beta = 2$ for the complex case.

This distribution is called the *Jacobi* (or *MANOVA*, or *CCA*) ensemble, and it is also sometimes called the *multivariate Beta distribution.* In point processes, the $\beta = 2$ case is often referred to as the *Jacobi orthogonal polynomial ensemble.*

Remark 3.12. The derivation is again parallel to that in the GOE/GUE context, but one now keeps track of the row spaces and the relevant rectangular dimensions. The matrix $(X\,X^\top)$ (or $(X\,X^\dagger)$) is invertible with high probability whenever $n \le t$ and X is in general position. The distribution above reflects the geometry of overlapping projectors in a higher-dimensional space $\mathbb{R}^t$ (or $\mathbb{C}^t$).

3.3.3 General Pattern and β-Ensembles

We have now seen three classical examples:

- *Wigner (Gaussian) ensembles* (real/complex/quaternionic),
- *Wishart/Laguerre ensembles* $W = XX^\top$,
- *Jacobi/MANOVA/CCA ensembles.*

Their eigenvalue densities (ordered or unordered) always display the same building blocks:

$$\prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^\beta \;\times\; \prod_{i=1}^{n} V(\lambda_i),$$

where β indicates the real ($\beta = 1$), complex ($\beta = 2$), or symplectic ($\beta = 4$) symmetry class, and $V(\lambda)$ is a single-variable potential function. Such distributions are often referred to as *β-ensembles* or *log-gases*, reflecting that the factor $\prod_{i<j} |\lambda_i - \lambda_j|^\beta$ can be interpreted as the Boltzmann weight for charges with a logarithmic pairwise repulsion.

Remark 3.13. Beyond these three classical families, there are many other *matrix models* and *discrete distributions* whose eigenvalues produce similar log-gas structures but with different potentials $V(\lambda)$. These share many of the same techniques and phenomena (e.g. local eigenvalue statistics, largest-eigenvalue asymptotics, etc.) that appear throughout modern random matrix theory.

Remark 3.14. For $\beta = 2$, the connection to orthogonal polynomials suggests discrete models of log-gases, which are powered by most known orthogonal polynomials in one variable from the (q-)*Askey scheme* [KS96]. For example, the model of (uniformly random) lozenge tilings of the hexagon is connected to Hahn orthogonal polynomials [Gor21] whose orthogonality weight is the classical hypergeometric distribution from probability theory.

3.4 Tridiagonal form for real symmetric matrices

Any real symmetric matrix can be orthogonally transformed into a tridiagonal matrix. This fact is standard in numerical linear algebra (the "Householder reduction") and also central in random matrix theory—notably in the Dumitriu–Edelman approach [DE02] for Gaussian ensembles.

Theorem 3.15. *Any real symmetric matrix $W \in \mathbb{R}^{n\times n}$ can be represented as*

$$W \;=\; Q^\top T\, Q, \quad Q \in O(n),$$

where T is real symmetric tridiagonal. Concretely, T has nonzero entries only on the main diagonal and the first super-/sub-diagonals:

$$T \;=\; \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & d_2 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & d_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \alpha_{n-1} \\ 0 & 0 & \cdots & \alpha_{n-1} & d_n \end{pmatrix}.$$

Definition 3.16 (Householder reflection)**.** A *Householder reflection* in $\mathbb{R}^n$ is a matrix H of the form

$$H \;=\; I \;-\; 2\,\frac{v\,v^\top}{\|v\|^2}, \quad v \in \mathbb{R}^n \text{ nonzero column vector.}$$

One checks that $H^\top = H$, $H^2 = I$, and H is orthogonal (i.e. $H^\top H = I$). Geometrically, H is the reflection across the hyperplane orthogonal to v.

Proof of Theorem 3.15. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. We will show how to orthogonally conjugate A into a tridiagonal matrix T.

Step 1: Zeroing out subdiagonal entries in the first column. Write A in block form as

$$A = \begin{pmatrix} a_{11} & r^\top \\ r & B \end{pmatrix},$$

where $r \in \mathbb{R}^{n-1}$ is the rest of the first column below a_{11}, and B is $(n-1)\times(n-1)$. We seek an orthogonal matrix H_1 acting on $\mathbb{R}^{n-1}$ (and in the full space $\mathbb{R}^n$ it preserves the first basis vector e_1 and its orthogonal complement) that "annihilates" the part of this first column below the subdiagonal. Specifically, H_1 is a Householder reflection chosen so that H_1 when acting in the $(n-1)$-dimensional subspace spanned by r zeroes out all but the first entry of r. In the ambient space $\mathbb{R}^n$, H_1 has a block form, so that it does not touch the 11-entry of the matrix A. Since A is symmetric, conjugating A by H_1 also zeroes out the corresponding superdiagonal entries in the first row. Concretely,

$$H_1 A H_1^\top = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & * & * & \cdots & * \end{pmatrix}.$$

This is always possible because Householder reflections can exchange any two given unit vectors. Note also that $\alpha_1 = \|r\|$.

Step 2: Inductive reduction on the trailing principal submatrix. Next, we restrict attention to rows 2 through n and columns 2 through n. Let H_2 be a second Householder reflection that acts as the identity on the first row and column, and zeroes out the subdiagonal entries of the *second* column (viewed within that trailing $(n-1)\times(n-1)$ block). Conjugate again:

$$H_2\left(H_1 A H_1^\top\right) H_2^\top = \left(H_2 H_1\right) A \left(H_1^\top H_2^\top\right).$$

Now the first two columns (and rows) are in the desired form.

Step 3: Repeat for columns (and rows) 3, 4, ... By repeating this procedure for each successive column (and row, by symmetry), we eventually force all off-diagonal entries outside the main and first super-/subdiagonals to be zero. After $n-2$ steps, the resulting matrix

$$T \;=\; Q^\top A\, Q, \qquad Q \;=\; H_1\, H_2\, \cdots\, H_{n-2},$$

is *tridiagonal*, and Q is orthogonal because it is a product of orthogonal (Householder) transformations.

Since each H_k is orthogonal, none of these transformations change the eigenvalues of A. Thus T has the same spectrum as A. This completes the tridiagonalization argument. □

Remark 3.17. This Householder procedure is also used in practical numerical methods for eigenvalue computations: once a real symmetric matrix is reduced to tridiagonal form, specialized algorithms (such as the QR algorithm) can then be applied more efficiently. Overall, computations with tridiagonal matrices are much simpler and with better numerical stability than with general dense matrices.

3.5 Tridiagonalization of random matrices

Here we discuss the tridiagonal form of the GOE random matrices, and extend it to the general beta case.

3.5.1 Dumitriu–Edelman tridiagonal model for GOE

Theorem 3.18. *Let W be an $n \times n$ GOE matrix (real symmetric) with variances chosen so that each off-diagonal entry has variance $1/2$ and each diagonal entry has variance 1. Then there exists an orthogonal matrix Q such that*

$$W \;=\; Q^\top\, T\, Q,$$

where T is a real symmetric tridiagonal matrix of the special form

$$T \;=\; \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots \\ \alpha_1 & d_2 & \alpha_2 & \ddots \\ 0 & \alpha_2 & d_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

and the random variables $\{d_i, \alpha_j\}_{1\le i\le n,\, 1\le j\le n-1}$ are mutually independent, with

$$d_i \sim \mathcal{N}(0,1), \qquad \alpha_j \;=\; \sqrt{\frac{\chi^2_{n-j}}{2}},$$

where χ^2_ν *is a chi-square distribution with* ν *degrees of freedom.*

Remark 3.19 (Chi-square distributions)**.** The *chi-square distribution* with ν degrees of freedom, denoted by χ^2_ν, is a fundamental distribution in statistics and probability theory. It arises naturally as the distribution of the sum of the squares of ν independent standard normal random variables. Formally, if $Z_1, Z_2, \ldots, Z_\nu$ are independent random variables with $Z_i \sim \mathcal{N}(0, 1)$, then the random variable

$$Q = \sum_{i=1}^{\nu} Z_i^2$$

follows a chi-square distribution with ν degrees of freedom, i.e., $Q \sim \chi^2_\nu$. In the context of the Dumitriu–Edelman tridiagonal model (Theorem 3.18), the subdiagonal entries α_j are defined as $\alpha_j = \sqrt{\frac{\chi^2_{n-j}}{2}}$. One can call this a *chi random variable*, as this is a square root of a chi-square variable.

The parameter ν does not need to be an integer, and the chi-square distribution is well defined for any positive real ν, by continuation of the density formula.

Idea of proof of Theorem 3.18. This construction is essentially a specialized version of the Householder reduction in Section 3.4, set up so that each step matches precisely the distributions $\alpha_j \sim \sqrt{\frac{\chi^2_{n-j}}{2}}$ and $d_i \sim \mathcal{N}(0, 1)$. One uses the rotational invariance of Gaussian matrices to ensure at each step that the "residual vector" is isotropic (i.e., its distribution is invariant under orthogonal transformations). The norm of that vector yields the χ^2-type variables. $\square$

Thus, to study the eigenvalues of a GOE matrix W, one can equivalently study the (much sparser) random tridiagonal matrix T.

3.5.2 Generalization to β-ensembles

The tridiagonal GOE construction (Theorem 3.18) extends to a whole family of ensembles, parametrized by $\beta > 0$. In particular, for $\beta = 1, 2, 4$ we get the classical Orthogonal, Unitary, and Symplectic (GOE/GUE/GSE) ensembles, respectively. The general β case is known as the β-ensemble; outside of the classical cases $\beta = 1, 2, 4$, there is no matrix ensemble interpretation with iid entries, but the tridiagonal form model still works.

We saw that the β-ensembles arise naturally as *log-gases* in physics, with density proportional to

$$\exp\Bigl(-\sum_{i=1}^{n} V(\lambda_i)\Bigr) \prod_{1\le i<j\le n} \bigl|\lambda_i-\lambda_j\bigr|^{\beta}$$

for some potential V. The simplest choice, $V(\lambda) = \frac{1}{2}\lambda^2$, corresponds to Gaussian β-ensembles, which in the classical cases reproduce GOE/GUE/GSE.

Remark 3.20 (Tridiagonal Construction for General β)**.** A breakthrough [DE02] showed that the Gaussian β-ensembles (for *any* $\beta > 0$) can be represented as eigenvalues of real symmetric *tridiagonal* matrices whose entries are independent (but not identically distributed), and have Gaussian and chi distributions:

- The diagonal entries are iid standard normal random variables $\mathcal{N}(0,1)$.
- The subdiagonal entries are $\alpha_j = \sqrt{\frac{\chi^2_{(n-j)\beta}}{2}}$, where χ^2_ν is a chi-square distribution with ν degrees of freedom. Here we use the fact that the parameter ν in the chi-square distribution does not need to be an integer.
- The superdiagonal entries are determined by symmetry.

In the next lecture, we will see how the tridiagonal form allows to prove the Wigner's semicircle law for the Gaussian β-ensembles.

3.6 Problems

3.6.1 Invariance of GOE and GUE

Show that the distribution of the GOE and GUE is invariant under, respectively, orthogonal and unitary conjugation. For GOE, this means that if W is a random GOE matrix and Q is a fixed orthogonal matrix of order n, then the distribution of $QWQ^\top$ is the same as the distribution of W. (Similarly for GUE.)

Hint: write the joint density of all entries of GOE/GUE (for instance, GOE is determined by $n(n+1)/2$ real random independent variables) in a coordinate-free way.

3.6.2 Preimage size for spectral decomposition

Show that for a real symmetric matrix W with distinct eigenvalues, if $W = Q\Lambda Q^\top$ is its spectral decomposition where Q is orthogonal and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal with $(\lambda_1 \geq \cdots \geq \lambda_n)$, then there are exactly 2^n different choices of Q that give the same matrix W.

3.6.3 Distinct eigenvalues

Show that under GOE and GUE, almost surely, all eigenvalues are distinct.

3.6.4 Testing distinctness of eigenvalues via rank-1 perturbations

Suppose λ is an eigenvalue of a fixed matrix W with multiplicity ℓ. Consider the rank-1 perturbation

$$W_\varepsilon \;=\; W \;+\; \alpha\, u\, u^\top, \quad \alpha \sim \mathcal{N}(0, \varepsilon),$$

where $u \in \mathbb{R}^n$ is fixed. Prove that with probability one (in α), the eigenvalue λ *splits* into ℓ distinct eigenvalues of W_ε.

Hint: Write the characteristic polynomial of W_ε as $\det(W_\varepsilon - \mu I)$. Show that the infinitesimal change in α moves the roots in a non-degenerate way, splitting a repeated root.

3.6.5 Jacobian for GUE

Arguing similarly to Section 3.2.5, show that the Jacobian for the spectral decomposition of a complex Hermitian matrix is proportional to

$$\prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^2.$$

In particular, make sure you understand where the factor 2 comes from in the complex case.

3.6.6 Normalization for GOE

Compute the n-dimensional integral (in the ordered on unordered form):

$$\int_{\lambda_1 < \ldots < \lambda_n} \prod_{i<j} (\lambda_i - \lambda_j) \exp\Big(-\tfrac{1}{2} \sum_{k=1}^{n} \lambda_k^2\Big)\, d\lambda_1 \cdots d\lambda_n.$$

$$= \frac{1}{n!} \int_{\mathbb{R}^n} \prod_{i<j} |\lambda_i - \lambda_j| \exp\Bigl(-\tfrac{1}{2} \sum_{k=1}^{n} \lambda_k^2\Bigr) \, d\lambda_1 \cdots d\lambda_n.$$

Hint: The following identity might be useful:

$$\int_{-\infty}^{\infty} x^{2m} e^{-x^2/2} \, dx = 2^{m+1/2} \Gamma\left(m + \frac{1}{2}\right).$$

3.6.7 Wishart eigenvalue density

Prove Theorem 3.7 (in the real case $\beta = 1$) by using the singular value decomposition of X and the properties of the Wishart ensemble.

3.6.8 Householder reflection properties

Show that the Householder reflection $H = I - 2\, v\, v^\top / \|v\|^2$ has the following properties:

1. H is orthogonal, i.e., $H^\top H = I$.
2. H is symmetric, i.e., $H^\top = H$.
3. H is idempotent, i.e., $H^2 = I$.
4. H is a reflection across the hyperplane orthogonal to v.

3.6.9 Distribution of the Householder vector in random tridiagonalization

Consider the first step of the Householder tridiagonalization of a GOE matrix W. Denote the first column by $x \in \mathbb{R}^n$, and let

$$v \;=\; x \;+\; \alpha\, e_1, \quad \alpha \;=\; \pm\, \|x\|.$$

Then the first Householder reflection is given by

$$H_1 \;=\; I \;-\; 2\, \frac{v\, v^\top}{\langle v, v \rangle}.$$

Prove that:

1. $\|v\|^2$ follows a χ^2_ν distribution with ν degrees of freedom (determine ν in terms of n).

2. The direction $v/\|v\|$ is uniformly distributed on the unit sphere $\mathbb{S}^{n-1}$ and is independent of $\|v\|$.

Hint: View x as a Gaussian vector in $\mathbb{R}^n$, using the fact that the first column of a GOE matrix (including its diagonal entry) is an isotropic normal vector (up to small adjustments for the diagonal). Orthogonal invariance of the underlying distribution ensures the direction is uniform on $\mathbb{S}^{n-1}$.

3.6.10 Householder reflection for GUE

Modify the tridiagonalization procedure which was discussed for the GOE case, and show that the GUE random matrix can be transformed (by a unitary conjugation) into

$$\begin{pmatrix} \mathcal{N}(0,1) & \chi_{2(n-1)}/\sqrt{2} & 0 & 0 & \cdots \\ \chi_{2(n-1)}/\sqrt{2} & \mathcal{N}(0,1) & \chi_{2(n-2)}/\sqrt{2} & 0 & \cdots \\ 0 & \chi_{2(n-2)}/\sqrt{2} & \mathcal{N}(0,1) & \chi_{2(n-3)}/\sqrt{2} & \cdots \\ 0 & 0 & \chi_{2(n-3)}/\sqrt{2} & \mathcal{N}(0,1) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

(this matrix is symmetric, and in the entries, we list the distributions).

3.6.11 Jacobi ensemble is related to two Wisharts

Let X be an $n \times m$ and Y be a $k \times m$ real Gaussian matrices with iid $\mathcal{N}(0,1)$ entries, independent of each other, and assume $n \le k \le m$. Consider the matrix

$$\left(X\,X^\top + Y\,Y^\top\right)^{-1}\left(X\,X^\top\right) \quad \in \mathbb{R}^{n\times n}.$$

1. Prove that it is well-defined (invertible denominator) with probability 1, and that it is symmetric and diagonalizable in $\mathbb{R}^n$.

2. Show that its eigenvalues lie in $[0,1]$ and follow a Jacobi (MANOVA) distribution of parameters $\beta = 1$ and $\big(n,k,m\big)$.

3. Identify explicitly how these parameters match the shape parameters in the standard multivariate Beta / Jacobi pdf

$$\prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^{n} \lambda_i^{\alpha}\,(1-\lambda_i)^{\gamma},$$

with appropriate α, γ in terms of n, k, m.

Hint: Use that $X\,X^\top$ and $Y\,Y^\top$ are (independent) Wishart matrices. Rewrite

$$\big(X\,X^\top + Y\,Y^\top\big)^{-1} X\,X^\top$$

via block-inversion or projector-based arguments to see it is related to the product of two orthogonal projectors in $\mathbb{R}^m$. The Jacobi distribution then emerges from the overlapping subspace geometry.

Chapter 4

Semicircle law for GβE via tridiagonalization. Beginning determiantal processes

4.1 Recap

Note: I did some live random matrix simulations here and here — check them out. More simulations to come.

4.1.1 Gaussian ensembles

We introduced Gaussian ensembles, and for GOE ($\beta = 1$) we computed the joint eigenvalue density. The normalization is so that the off-diagonal elements have variance $\frac{1}{2}$ and the diagonal elements have variance 1. Then the joint eigenvalue density is

$$p(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_n} \prod_{i=1}^{n} e^{-\frac{1}{2}\lambda_i^2} \prod_{1 \le i < j \le n} (\lambda_i - \lambda_j), \qquad \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n.$$

4.1.2 Tridiagonalization

We showed that any real symmetric matrix A can be tridiagonalized by an orthogonal transformation Q:

$$Q^\top A\, Q \;=\; T,$$

where T is real symmetric tridiagonal, having nonzero entries only on the main diagonal and the first super-/subdiagonals:

$$T = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots & 0 \\ \alpha_1 & d_2 & \alpha_2 & \cdots & 0 \\ 0 & \alpha_2 & d_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \alpha_{n-1} \\ 0 & 0 & \cdots & \alpha_{n-1} & d_n \end{pmatrix}.$$

In the proof, each time we need to act in the orthogonal complement to the subspace $e_1, \ldots, e_{k-1}$ (starting from e_1), and apply a Householder reflection to zero out everything strictly below the subdiagonal. (We apply the transformations like $A \mapsto HAH^\top$, so that the first row transforms in the same way as the first column of A).

4.2 Tridiagonal random matrices

4.2.1 Distribution of the tridiagonal form of the GOE

Applying the tridiagonalization to GOE, we obtain the following random matrix model.

Theorem 4.1. *Let W be an $n \times n$ GOE matrix (real symmetric) with variances chosen so that each off-diagonal entry has variance $1/2$ and each diagonal entry has variance 1. Then there exists an orthogonal matrix Q such that*

$$W = Q^\top T Q,$$

where T is a real symmetric tridiagonal matrix

$$T = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots \\ \alpha_1 & d_2 & \alpha_2 & \ddots \\ 0 & \alpha_2 & d_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \tag{4.1}$$

and the random variables $\{d_i, \alpha_j\}_{1 \le i \le n,\, 1 \le j \le n-1}$ are mutually independent, with

$$d_i \sim \mathcal{N}(0,1), \qquad \alpha_j = \sqrt{\frac{\chi^2_{n-j}}{2}},$$

where χ^2_ν is a chi-square distribution with ν degrees of freedom.

Remark 4.2 (Chi-square distributions)**.** The *chi-square distribution* with ν degrees of freedom, denoted by χ^2_ν, is a fundamental distribution in statistics and probability theory. It arises naturally as the distribution of the sum of the squares of ν independent standard normal random variables. Formally, if $Z_1, Z_2, \ldots, Z_\nu$ are independent random variables with $Z_i \sim \mathcal{N}(0,1)$, then the random variable

$$Q = \sum_{i=1}^{\nu} Z_i^2$$

follows a chi-square distribution with ν degrees of freedom, i.e., $Q \sim \chi^2_\nu$. In the context of Theorem 4.1, the α_j's can be called *chi random variables.*

The parameter ν does not need to be an integer, and the chi-square distribution is well defined for any positive real ν, for example, by continuation of the density formula. The probability density is

$$f(x) \;=\; \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, x^{\nu/2-1}\, e^{-x/2}, \qquad x \geq 0.$$

Proof of Theorem 4.1. In the process of tridiagonalization, we apply Householder reflections. Note that the diagonal entries stay fixed, and we only change the off-diagonal entries. Let us consider these off-diagonal entries.

In the first step, we apply the reflection in $\mathbb{R}^{n-1}$ to turn the column vector $(a_{2,1}, a_{3,1}, \ldots, a_{n,1})$ into a vector parallel to $(1, 0, \ldots, 0) \in \mathbb{R}^{n-1}$. Since the Householder reflection is orthogonal, it preserves lengths. So,

$$\alpha_1 = \sqrt{a_{21}^2 + a_{31}^2 + \cdots + a_{n1}^2}, \qquad a_{i1} \sim \mathcal{N}(0, \frac{1}{2}).$$

This implies that α_1 has the desired chi distribution. The distribution of the other entries is obtained similarly by the recursive application of the Householder reflections.

Note that α_j's and d_i's depend on nonintersecting subsets of the matrix entries, so they are independent. This completes the proof. □

4.2.2 Dumitriu–Edelman GβE tridiagonal random matrices

Let us define a general β extension of the tridiagonal model for the GOE.

Definition 4.3. Let $\beta > 0$ be a parameter. The tridiagonal GβE is a random $n \times n$ tridiagonal real symmetric matrix T as in (4.1), where $d_i \sim \mathcal{N}(0,1)$ are independent standard Gaussians, and

$$\alpha_j \sim \frac{1}{\sqrt{2}} \chi_{\beta(n-j)}, \qquad 1 \leq j \leq n-1,$$

are chi-distributed random variables.

We showed that for $\beta = 1$, the GβE is the tridiagonal form of the GOE random matrix model. The same holds for the two other classical betas:

Proposition 4.4 (Without proof). *For $\beta = 2$, the GβE is the tridiagonal form of the GUE random matrix model, which is the random complex Hermitian matrix with Gaussian entries and maximal independence. Similarly, for $\beta = 4$, the GβE is the tridiagonal form of the GSE random matrix model.*

Moreover, for all β, the joint eigenvalue density of GβE is explicit:

Theorem 4.5 ([DE02]). *Let T be a GβE matrix as in Definition 4.3. Then the joint eigenvalue density is given by*

$$p(\lambda_1, \ldots, \lambda_n) = \frac{1}{Z_{n,\beta}} e^{-\frac{1}{2}\sum_{i=1}^n \lambda_i^2} \prod_{1 \le i < j \le n} |\lambda_i - \lambda_j|^\beta, \qquad \lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_n.$$

This theorem is also given without proof. The proof involves linear algebra and computation of the Jacobians of the change of variables from the matrix entries to the eigenvalues in the tridiagonal setting. It can be found in the original paper [DE02].

4.2.3 The case $\beta = 2$

For many questions involving *local eigenvalue statistics*, the case $\beta = 2$ (the GUE, Gaussian Unitary Ensemble) is the most tractable. This is because the joint density of the eigenvalues admits a determinantal structure coming from a *square* Vandermonde factor $\prod_{i<j}(\lambda_i - \lambda_j)^2$ and the Gaussian exponential $\exp\bigl(-\frac{1}{2}\sum \lambda_j^2\bigr)$. Moreover, for $\beta = 2$, the random matrix model and its correlation functions can be expressed explicitly through determinants involving *orthogonal polynomials*, namely, the *Hermite polynomials*.

Proposition 4.6 (Joint density for GUE and orthogonal polynomials). *Consider the GUE (Gaussian Unitary Ensemble) random matrix model, i.e. an $n \times n$ complex Hermitian matrix whose entries are i.i.d. up to the Hermitian condition, with each off-diagonal entry distributed as $\mathcal{N}(0, \frac{1}{2}) + \mathrm{i}\,\mathcal{N}(0, \frac{1}{2})$ and each diagonal entry $\mathcal{N}(0, 1)$. The ordered eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$ (or, without ordering, thought of as an unordered set) satisfy the joint probability density*

$$p(\lambda_1, \ldots, \lambda_n) \;=\; \frac{1}{Z_{n,2}} \prod_{j=1}^n e^{-\frac{1}{2}\lambda_j^2} \prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)^2, \tag{4.2}$$

where $Z_{n,2}$ is a normalization constant.

Moreover, if $\{\psi_k(\lambda)\}_{k=0}^{\infty}$ is the family of Hermite polynomials, orthonormal with respect to the measure $w(\lambda)\,d\lambda = e^{-\lambda^2/2}\,d\lambda$ on $\mathbb{R}$ (i.e., $\int_{-\infty}^{\infty} \psi_k(\lambda)\,\psi_\ell(\lambda)\,w(\lambda)\,d\lambda = \mathbf{1}_{k=\ell}$), then one can also write

$$p(\lambda_1, \ldots, \lambda_n) \;=\; \mathrm{const} \cdot \det\Big[\psi_{j-1}(\lambda_k)e^{-\frac{\lambda_k^2}{4}}\Big]_{j,k=1}^{n} \;\det\Big[\psi_{j-1}(\lambda_k)e^{-\frac{\lambda_k^2}{4}}\Big]_{j,k=1}^{n} \tag{4.3}$$

(the two determinants are identical, but let us keep this notation for future convenience).

The *square determinant* structure is extremely useful. It is precisely the $\beta = 2$ counterpart of the squared Vandermonde factor $\prod_{i<j}(\lambda_i - \lambda_j)^2$.

Remark 4.7 (Hermite polynomials)**.** There are various normalizations of Hermite polynomials. In random matrix theory for the Gaussian ensembles, we often use the *probabilists' Hermite polynomials* (sometimes called He_k, but we use the notation H_k). There are various normalizations due to the factor in the exponent of x^2.

A convenient definition for use with the weight $e^{-x^2/2}$ is:

$$H_k(x) \;=\; (-1)^k\, e^{\frac{x^2}{2}}\, \frac{d^k}{dx^k}\Big(e^{-\frac{x^2}{2}}\Big), \qquad k = 0, 1, \ldots, \tag{4.4}$$

whose leading term is x^k. Polynomials with the leading coeffient 1 are called *monic*. The first few monic Hermite polynomials are

$$H_0(x) = 1, \qquad H_1(x) = x, \qquad H_2(x) = x^2 - 1, \qquad H_3(x) = x^3 - 3x, \qquad H_4(x) = x^4 - 6x^2 + 3.$$

The difference between H_k and ψ_k entering Proposition 4.6 is in a constant normalization, since H_k are monic but not orthonormal, while ψ_k are orthonormal but not monic.

Sketch of the determinantal representation. In brief, one observes that the factor $\prod_{i<j}(\lambda_i - \lambda_j)$ is exactly the Vandermonde determinant $\Delta(\lambda_1, \ldots, \lambda_n) = \det\big[\lambda_k^{j-1}\big]_{j,k=1}^{n}$. Next, the Vandermonde determinant is also equal to the determinant built out of any monic family of polynomials of the corresponding degrees (by linear transformations), and so we get the desired representation. □

We will work with Hermite polynomials and the determinantal structure in Proposition 4.6 in the next Chapter 5).

4.3 Wigner semicircle law via tridiagonalization

If W is an $n \times n$ real Wigner matrix with entries of mean zero and variance 1 on the off-diagonal, then as $n \to \infty$, the empirical spectral distribution (ESD) of $W/\sqrt{n}$ converges weakly almost surely to the Wigner semicircle distribution:

$$\mu_{\mathrm{sc}}(dx) \;=\; \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{|x|\le 2}\,dx.$$

We already derived this in Chapter 2 by a direct combinatorial argument on the trace. Now we present another proof by using the tridiagonal form of W. The argument is conceptually simpler in some steps, because the matrix is sparser (only tridiagonal). At the same time, we will establish the Wigner semicircle law for the general GβE case (but only Gaussian), and thus it will apply to GUE and GSE.

4.3.1 Moments for tridiagonal matrices

Consider the rescaled GβE matrix $T/\sqrt{n}$:

$$\frac{T}{\sqrt{n}} \;=\; \begin{pmatrix} d_1/\sqrt{n} & \alpha_1/\sqrt{n} & 0 & \cdots \\ \alpha_1/\sqrt{n} & d_2/\sqrt{n} & \alpha_2/\sqrt{n} & \ddots \\ 0 & \alpha_2/\sqrt{n} & d_3/\sqrt{n} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

where $d_i \sim \mathcal{N}(0,1)$ and $\alpha_j \sim \frac{1}{\sqrt{2}}\chi_{\beta(n-j)}$. We want to show that the ESD of $T/\sqrt{n}$ converges to the semicircle law. We will mostly consider expected traces of powers, and leave the analytic parts of the argument to the reader.

The k-th (random) moment of the ESD $\frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i/\sqrt{n}}$ is

$$\frac{1}{n}\,\mathrm{Tr}\Big(\tfrac{T}{\sqrt{n}}\Big)^k \;=\; \frac{1}{n^{1+\frac{k}{2}}}\sum_{i_1,\ldots,i_k=1}^{n} t_{i_1,i_2}\cdots t_{i_k,i_1}, \tag{4.5}$$

where t_{ij} are the non-rescaled entries of T. But now t_{ij} is nonzero only if $|i-j|\le 1$, i.e. the (i,j) entry is on the main or first super-/subdiagonal. In a closed product $t_{i_1 i_2}\cdots t_{i_k i_1}$, we thus get a *closed walk* in a linear graph on the vertex set $\{1,2,\ldots,n\}$ with edges only between consecutive indices.

The relevant combinatorial objects encoding these walks are lattice walks in $\mathbb{Z}^2_{\ge 0}$ starting at $(0,m)$, ending at (k,m), and consisting of steps $(1,0)$,

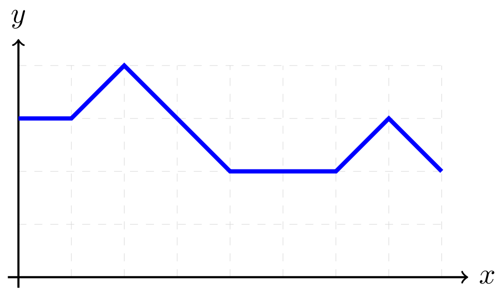

Figure 4.1: Example of a lattice path starting at height 3.

$(1,1)$, and $(1,-1)$. The steps $(1,0)$ correspond to picking the diagonal element; steps $(1,1)$ correspond to picking $i_{\ell+1} = i_\ell + 1$, and steps $(1,-1)$ correspond to $i_{\ell+1} = i_\ell - 1$. See Figure 4.1 for an illustration of a path.

Now, each term in the sum in (4.5) corresponds to a path. Moreover, for each path shape, there are $O(n)$ summands corresponding to it. The number of paths of length k starting from a fixed m is finite (independent of n for $m \gg 1$), so we need to look more closely at the asymptotics of the product in (4.5). This product involves chi random variables which depend on n, too.

4.3.2 Asymptotics of chi random variables

One additional technical point in analyzing $T/\sqrt{n}$ is to note that α_j is roughly $\sqrt{\beta(n-j)/2}$ for large n. Indeed, we have

$$\chi_\nu^2 = \sum_{i=1}^{\nu} Z_i^2, \qquad \mathbb{E}[\chi_\nu^2] = \nu, \qquad \mathrm{Var}[\chi_\nu^2] = 2\nu.$$

Now, since we are dividing by $\sqrt{n}$, we have

$$\frac{\alpha_j}{\sqrt{n}} \sim \sqrt{\frac{\beta}{2}}\sqrt{1-\theta}, \qquad \theta = \frac{j}{n} \in [0,1].$$

This estimate is valid in the "bulk" region, that is, when θ is strictly between 0 and 1.

Let us make these estimates more precise. We have:

Proposition 4.8 (Pointwise asymptotics in the bulk)**.** *Fix small $\delta > 0$, and let j range so that $\theta_j := j/n \in [\delta, 1-\delta]$. Then for each such j, we have*[1]

$$\frac{\alpha_j}{\sqrt{n}} = \sqrt{\frac{\beta}{2}\Big(1-\frac{j}{n}\Big)} + O_p\Big(\frac{1}{\sqrt{n}}\Big),$$

In particular,

$$\lim_{n\to\infty} \frac{\alpha_j}{\sqrt{n}} = \sqrt{\frac{\beta}{2}(1-\theta_j)} \quad \textit{in probability.}$$

Remark 4.9. Outside the bulk region (i.e. very close to $j=0$ or $j=n$), one would need a different statement to handle the case $\beta(n-j)$ is not large. In our application, we only need the bulk behavior. See also Problem 4.7.3.

Meanwhile, on the diagonal, $d_i/\sqrt{n}$ almost surely vanishes in the limit as $n\to\infty$, because d_i is standard Gaussian and does not depend on n.

4.3.3 Completing the proof: global semicircle behavior

Putting the above pieces together, we see that

$$\frac{T}{\sqrt{n}} = \frac{1}{n} \sum_{i_1,\ldots,i_k=1}^{n} \prod_{\ell=1}^{k} \frac{t_{i_\ell i_{\ell+1}}}{\sqrt{n}}, \quad i_{k+1} = i_1 \text{ by agreement.} \tag{4.6}$$

The terms in the sum have all i_ℓ's close together (there are k indices, and they differ by ±1 from each other). We may think that they are close to some θn, where $\theta \in [0,1]$. We can consider only the case when $\delta < \theta < 1-\delta$ for some fixed small $\delta > 0$; the case of edges does not contribute (see Problem 4.7.3).

If at least one of the t_{ij}'s in (4.6) is on the diagonal, the term vanishes in the limit. Therefore, it suffices to consider only the off-diagonal α_j's. The number of length k walks starting from $m = \theta n$ for $\theta > \delta$ is just the number of lattice walks with steps $(1, \pm1)$. This number is $\binom{k}{k/2}$.[2] (From now on till the end of the section, we assume that k is even — the moments become zero for odd k).

Fixing the starting location $\theta = \frac{i_\ell}{n} \in (\delta, 1-\delta)$, we have

$$\prod_{\ell=1}^{k} \frac{t_{i_\ell i_{\ell+1}}}{\sqrt{n}} \to (\beta/2)^{k/2}(1-\theta)^{k/2}.$$

[1]Here and below, $O_p(\cdot)$ denotes a term that is stochastically bounded at the indicated order as $n\to\infty$. That is, $X_n = O_p(a_n)$ means that for any $\epsilon > 0$, there exists $M > 0$ such that $\mathbb{P}(|X_n/a_n| > M) < \epsilon$ for all sufficiently large n.

[2]Not Catalan yet!

There is an extra factor $1/n$ in front in (4.6), which is interpreted as transforming the sum over $i_1, \ldots, i_k$ into an integral in θ. We thus see that the moments converge to

$$(\beta/2)^{k/2}\binom{k}{k/2}\int_0^1 (1-\theta)^{k/2}\, d\theta = (\beta/2)^{k/2}\binom{k}{k/2}\cdot\frac{1}{1+k/2},$$

and we recover our favorite Catalan moments of the semicircle distribution.

This completes the proof.

Remark 4.10 (The factor $(\beta/2)^{k/2}$)**.** Note that the factor $\beta^{k/2}$ refers just to the scaling of the Wigner semicircle law, and does not affect the semicircle shape. More precisely, the limiting semicircle distribution lies from $[-\sqrt{2\beta}, \sqrt{2\beta}]$.

The density of the semicircle distribution on $[-\sqrt{2\beta}, \sqrt{2\beta}]$ is

$$\frac{\sqrt{2-\frac{x^2}{\beta}}}{\pi\sqrt{\beta}}, \qquad |x| < \sqrt{2\beta},$$

and the moments are precisely $(\beta/2)^{k/2}C_{k/2}$ (for even k).

4.4 Wigner semicircle law via Stieltjes transform

Let us stay in the tridiagonal setting, and explore a more analytic method to derive the Wigner semicircle law.

4.4.1 Tridiagonal structure and characteristic polynomials

We let

$$T - \lambda I = \begin{pmatrix} d_1-\lambda & \alpha_1 & 0 & \cdots \\ \alpha_1 & d_2-\lambda & \alpha_2 & \ddots \\ 0 & \alpha_2 & d_3-\lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

We want to understand eigenvalues, that is, zeros of the characteristic polynomial $\det(T-\lambda I)$.

Three-term recurrence for the characteristic polynomial

As a warm-up, let us consider the characteristic polynomial of a tridiagonal matrix.

For each $k = 1, \dots, n$, denote by T_k the top-left $k \times k$ submatrix of T. Define the *characteristic polynomial* of that block:

$$p_k(\lambda) \;=\; \det\bigl(T_k - \lambda I_k\bigr).$$

By convention, set $p_0(\lambda) := 1$. Then a determinant expansion argument along the first column gives the following three-term recurrence relation:

Lemma 4.11 (Three-Term Recurrence)**.** *The characteristic polynomial $p_k(\lambda)$ of the $k \times k$ tridiagonal matrix T_k satisfies the three-term recurrence*

$$p_{k+1}(\lambda) \;=\; (d_{k+1} - \lambda)\, p_k(\lambda) - \alpha_k^2\, p_{k-1}(\lambda), \quad k = 1, \dots, n-1,$$

μ

See also Problem 4.7.4.

Spectral connection and eigenvalues

The eigenvalues $\lambda_1, \dots, \lambda_n$ of T are exactly the roots of $p_n(\lambda)$. For any $\lambda \in \mathbb{C}$, if λ is not an eigenvalue, then $\bigl(T - \lambda I\bigr)$ is invertible.

When λ is close to a real eigenvalue, the behavior of the resolvent $\bigl(T - \lambda I\bigr)^{-1}$ becomes large. Tracking these poles in the complex plane is the key to the resolvent or Stieltjes transform approach.

4.4.2 Stieltjes transform / resolvent

Recall that for a matrix A with real eigenvalues $\lambda_1, \dots, \lambda_n$, the *Stieltjes transform* (or Green's function, or resolvent trace) is

$$G_n(z) \;=\; \frac{1}{n}\,\mathrm{Tr}\bigl[(A - zI)^{-1}\bigr], \quad z \in \mathbb{C} \setminus \mathbb{R}.$$

If $z = x + \mathrm{i}y$ is in the upper half-plane ($y > 0$), this $G_n(z)$ can be seen as

$$G_n(z) \;=\; \int_{\mathbb{R}} \frac{d\mu_n(\lambda)}{\lambda - z},$$

where $\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k}$ is the empirical spectral measure. Equivalently, $\mathrm{Im}\, G_n(x + \mathrm{i}0^+)$ encodes the density of eigenvalues around x. Thus, understanding $G_n(z)$ for large n pinpoints the limiting spectral distribution.

Let us apply this to $A = T/\sqrt{n}$ (an $n \times n$ tridiagonal matrix). We want to investigate

$$G_n(z) := \frac{1}{n}\operatorname{Tr}\bigl(T/\sqrt{n} - zI\bigr)^{-1},$$

for complex z. Since $T/\sqrt{n}$ has nonzero entries only on the main and first off-diagonals, one can write down a linear recurrence for the entries R_{ij} of the resolvent $R(z) = (T/\sqrt{n} - zI)^{-1}$, from the equation

$$\sum_k \bigl(T/\sqrt{n} - zI\bigr)_{ik} R_{kj} = \mathbf{1}_{i=j}.$$

We have

$$\left(\frac{d_i}{\sqrt{n}} - z\right) R_{ij} + \frac{\alpha_i}{\sqrt{n}} R_{i+1,j} + \frac{\alpha_{i-1}}{\sqrt{n}} R_{i-1,j} = \mathbf{1}_{i=j}.$$

Let $f_u(\theta) := R_{\lfloor n\theta\rfloor,\lfloor nu\rfloor}$. Then the above equation becomes

$$\left(\frac{d_{\lfloor n\theta\rfloor}}{\sqrt{n}} - z\right) f_u(\theta) + \frac{\alpha_{\lfloor n\theta\rfloor}}{\sqrt{n}} f_u(\theta + 1/n) + \frac{\alpha_{\lfloor n\theta\rfloor - 1}}{\sqrt{n}} f_u(\theta - 1/n) = \mathbf{1}_{\theta=u}.$$

Scaling with n (and ignoring the boundary conditions and convergence issues), we get a differential equation for $f_u(\theta)$:

$$-zf_u(\theta) + \sqrt{\frac{\beta(1-\theta)}{2}}\left[f_u''(\theta) + 2f_u(\theta)\right] = \delta(\theta - u). \tag{4.7}$$

The resolvent trace (the Stieltjes transform) is then the integral of the solution:

$$\frac{1}{n}\sum_{i=1}^{n} R_{ii} \sim G(z) := \int_0^1 f_\theta(\theta)\, d\theta.$$

At this point (2025-01-30), I am stuck on how to pass from (4.7) to the Stieltjes transform $G(z)$. This would be an excellent topic to explore for a presentation. See Problem 4.7.7.

Update 2025-02-05: Probably, the limit of $\alpha_j/\sqrt{n}$ should be taken as 1 and not as a function of τ. At least this is what is done in the next approach in Section 4.4.3.

4.4.3 Approach via continued fractions

We derive the Wigner semicircle law using the continued fraction representation of the Stieltjes transform (or Green's function) associated with a tridiagonal (Jacobi) matrix. In the Dumitriu–Edelman model for the GUE (let us assume $\beta = 2$ for simplicity) after appropriate rescaling, the matrix's diagonal entries vanish and the off-diagonal entries become essentially constant in the bulk. This leads to a homogeneous three-term recurrence for the corresponding monic orthogonal polynomials. We then show that the Stieltjes transform of the limiting measure may be written as an infinite continued fraction, which yields a quadratic self–consistent equation. Solving that equation and applying the Stieltjes inversion formula recovers the semicircle density.

A real symmetric tridiagonal matrix (a *Jacobi matrix*) has the form

$$J = \begin{pmatrix} a_0 & b_1 & 0 & \cdots & 0 \\ b_1 & a_1 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \cdots & 0 & b_{n-1} & a_{n-1} \end{pmatrix},$$

with $b_j > 0$. Associated with J is a sequence of monic polynomials $\{p_n(z)\}_{n\geq 0}$ defined by the three–term recurrence

$$\begin{aligned} p_0(z) &= 1, \\ p_1(z) &= z - a_0, \\ p_{n+1}(z) &= (z - a_n)p_n(z) - b_n^2\, p_{n-1}(z), \quad n \geq 1. \end{aligned} \tag{4.8}$$

It is well known that there exists a probability measure μ on $\mathbb{R}$ such that the polynomials $\{p_n(z)\}$ are orthogonal with respect to μ.

In the Dumitriu–Edelman tridiagonal model for the GUE (with $\beta = 2$) the matrix is constructed so that, after rescaling by $\sqrt{n}$, one obtains

$$\frac{T}{\sqrt{n}} = \begin{pmatrix} d_1/\sqrt{n} & \alpha_1/\sqrt{n} & 0 & \cdots \\ \alpha_1/\sqrt{n} & d_2/\sqrt{n} & \alpha_2/\sqrt{n} & \ddots \\ 0 & \alpha_2/\sqrt{n} & d_3/\sqrt{n} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix},$$

with

$$d_i \sim \mathcal{N}(0,1), \qquad \alpha_j \sim \frac{1}{\sqrt{2}}\chi_{2(n-j)}.$$

In the large n limit, the diagonal entries $d_i/\sqrt{n}$ vanish and (in the bulk) one has

$$\frac{\alpha_j^2}{n} \to 1.$$

Thus, in the limit the recurrence coefficients become

$$a_n = 0, \quad b_n = 1,$$

for all n.

Note 2025-02-05: This is probably the correct way to approach the global asymptotic behavior of T's spectrum in connection with the Stieltjes transform. This should be justified; however, this idea should help to unstick the argument in Section 4.4.2.

In this homogeneous case the three-term recurrence (4.8) reduces to

$$p_0(z) = 1, \quad p_1(z) = z, \quad p_{n+1}(z) = z\,p_n(z) - p_{n-1}(z).$$

The *Stieltjes transform* of the measure μ is defined by

$$m(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{z-x}, \qquad z \in \mathbb{C}\setminus\mathbb{R}.$$

A **classical result in the theory of orthogonal polynomials** (e.g., see [Sok20]) is that $m(z)$ may be written as the continued fraction

$$m(z) = \cfrac{1}{z - a_0 - \cfrac{b_1^2}{z - a_1 - \cfrac{b_2^2}{z - a_2 - \cfrac{b_3^2}{z - a_3 - \cdots}}}}. \tag{4.9}$$

In our case, since $a_n = 0$ for all n and $b_n = 1$ for all n, this simplifies to

$$m(z) = \cfrac{1}{z - \cfrac{1}{z - \cfrac{1}{z - \cfrac{1}{\ddots}}}}. \tag{4.10}$$

Observe that the infinite continued fraction in (4.10) is self–similar; that is, if we denote the entire continued fraction by $m(z)$, then the tail of the continued fraction is again $m(z)$. Thus we have the relation

$$m(z) = \frac{1}{z - m(z)}.$$

Multiplying both sides by the denominator yields

$$m(z)\Big(z - m(z)\Big) = 1.$$

Expanding the left–hand side we obtain the quadratic equation

$$m(z)^2 - z\,m(z) + 1 = 0. \tag{4.11}$$

The quadratic (4.11) has the solutions

$$m(z) = \frac{z \pm \sqrt{z^2 - 4}}{2}.$$

To determine the correct branch, recall that for z in the upper half–plane ($\mathrm{Im}(z) > 0$) we must have $\mathrm{Im}\; m(z) > 0$. The proper solution is

$$m(z) = \frac{z - \sqrt{z^2 - 4}}{2}, \tag{4.12}$$

where the square root is defined so that $\sqrt{z^2 - 4} \sim z$ as $z \to \infty$ and $\mathrm{Im}\sqrt{z^2 - 4} > 0$ when $\mathrm{Im}(z) > 0$.

The density $\rho(x)$ of the measure μ is recovered from the Stieltjes transform via the inversion formula:

$$\rho(x) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \mathrm{Im}\; m(x + i\epsilon).$$

For x in the interval $(-2, 2)$ one computes that

$$\sqrt{(x + i\epsilon)^2 - 4} \xrightarrow[\epsilon \to 0^+]{} i\sqrt{4 - x^2}.$$

Thus, from (4.12) we have, for $x \in (-2, 2)$,

$$m(x + i0) = \frac{x - \mathrm{i}\sqrt{4 - x^2}}{2}.$$

Taking the imaginary part gives

$$\mathrm{Im}\; m(x + i0) = \frac{\sqrt{4 - x^2}}{2},$$

so that

$$\rho(x) = \frac{1}{\pi}\,\mathrm{Im}\; m(x + i0) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad x \in (-2, 2).$$

This is precisely the celebrated Wigner semicircle law.

4.5 Determinantal point processes (discrete)

We are now going to start the discussion of the local eigenvalue behavior at $\beta = 2$, started in Section 4.2.3. We begin with a general discussion of *determinantal point processes* (DPPs), starting in discrete world. The continuous world is going to be considered in the next Chapter 5.

In this section, we introduce *determinantal point processes (DPPs)* over a discrete state space and explore some of their properties. Our main reference is [Bor11].

Setup. Let $\mathfrak{X}$ be a (finite or countably infinite) discrete set endowed with the counting measure μ. A *point configuration* on $\mathfrak{X}$ is any subset $X \subset \mathfrak{X}$, finite or infinite, with no repeated points.[3] We write $Conf(\mathfrak{X})$ for the set of all point configurations, which carries the natural σ-algebra generated by the functions $\mathbf{1}_{\{x\in X\}}$, $x \in \mathfrak{X}$. A *random point process* P on $\mathfrak{X}$ is a probability measure on $Conf(\mathfrak{X})$.

Definition 4.12 (Determinantal point process)**.** A random point process P on a discrete set $\mathfrak{X}$ is *determinantal* if there exists a kernel function $K : \mathfrak{X} \times \mathfrak{X} \to \mathbb{C}$ such that for every finite collection of pairwise distinct points $x_1, \ldots, x_n \in \mathfrak{X}$,

$$\mathbb{P}\{x_1, \ldots, x_n \in X\} = \det\big[K(x_i, x_j)\big]_{i,j=1}^{n}. \tag{4.13}$$

That is, all finite-dimensional distributions of P take a determinantal form. The function K is called a *correlation kernel* for P.

Correlation functions and the kernel. The condition (4.13) captures all finite-dimensional distributions of P. Equivalently, let

$$\rho_n(x_1, \ldots, x_n) \coloneqq \mathbb{P}\{\text{there is a particle at each } x_i\}$$

for distinct $x_1, \ldots, x_n$. In the discrete setting, ρ_n is sometimes called the *(unordered) correlation function.* The process is determinantal if and only if

$$\rho_n(x_1, \ldots, x_n) = \det\big[K(x_i, x_j)\big]_{i,j=1}^{n} \quad \text{for each } n \geq 1.$$

Basic properties. If P is a DPP with correlation kernel $K\colon \mathfrak{X} \times \mathfrak{X} \to \mathbb{C}$, then for any subset $I \subset \mathfrak{X}$,

$$\mathbb{P}\{X \cap I = \varnothing\} = \det\big[\mathbf{1} - K_I\big], \tag{4.14}$$

[3] Some texts allow multiplicities, but we disallow them here.

where K_I is the operator $\big[K(x,y)\big]_{x,y\in I}$ (viewed as a matrix if $\mathfrak{X}$ is finite, or an infinite matrix if $\mathfrak{X}$ is countably infinite with convergent sums). More generally, if $I_1,\ldots,I_m \subset \mathfrak{X}$ are disjoint subsets, then the joint event $\{|X \cap I_k| = n_k \text{ for } 1 \le k \le m\}$ can be expressed via the determinant $\det\big[\mathbf{1} - \sum_{k=1}^m z_k K_{I_k}\big]$ and its derivatives.

Remark 4.13. For any function $\phi : \mathfrak{X} \to \mathbb{C}$ such that the operator $\big[(1 - \phi(x))K(x,y)\big]_{x,y\in\mathfrak{X}}$ is trace class, the exponential generating function for ϕ is

$$\mathbb{E}\Big[\prod_{x\in X}\phi(x)\Big] = \det\big[\mathbf{1} - (1-\phi)K\big].$$

This identity makes determinantal point processes more tractable than general processes.

4.6 Application of determinantal processes to random matrices at $\beta = 2$

In this final section of the lecture, we illustrate how the theory of determinantal point processes (DPPs) introduced in Section 4.5 applies to the study of local eigenvalue statistics of random matrices. We concentrate on the $\beta = 2$ setting, where DPPs typically govern the joint behavior of eigenvalues at microscopic (local) scales in the *bulk* and at the *edge* of the spectrum. We also include a simpler example of a Poisson process to highlight the role of correlation functions.

4.6.1 Local eigenvalue statistics (bulk and edge scaling limits)

Given an $n \times n$ random Hermitian matrix W whose eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ are real, we often want to study the *local arrangement* of the eigenvalues:

- *Bulk regime:* eigenvalues near some interior point α of the limiting (global) spectral support, rescaled so that we see "microscopic" spacing on the order of $O(\frac{1}{n})$. For Wigner or Gaussian ensembles, one typically looks at a point α in the interior $(-2,2)$ of the semicircle support and then rescales eigenvalues around α by the typical local spacing $1/(n\rho(\alpha))$. Here $\rho(\alpha)$ is the density of eigenvalues at α, which is semicircle density in the Wigner case.

- *Edge regime:* eigenvalues near an endpoint of the support (for instance, near $x = 2$ for the semicircle distribution). One then uses a rescaling of order $n^{2/3}$ (in many classical models) to see nontrivial statistics describing how eigenvalues "peel off" near the boundary.

In both cases, one replaces the original sequence of eigenvalues $\{\lambda_i\}$ by a *point process* on $\mathbb{R}$. The *bulk scaling* leads to the sine-kernel process (e.g. $\sin(\pi(x-y))/(\pi(x-y))$ in the GUE) or more generally to other determinantal processes. The *edge scaling* typically leads to the Airy-kernel process. For Gaussian ensembles at $\beta = 2$, these processes are determinantal, and one can explicitly write correlation kernels involving special functions (sine, Airy, and more generally Hermite polynomials).

4.6.2 Correlation functions and densities

We recall from Section 4.5 (in the discrete setting) that a point process $\mathcal{X}$ on a space $\mathfrak{X}$ can be described by its *correlation functions* $\{\rho_k\}_{k=1}^{\infty}$. In the continuous setting (e.g. $\mathfrak{X} = \mathbb{R}$ or an interval), these are defined so that

$$\rho_k(x_1, \ldots, x_k)\, dx_1 \cdots dx_k \;=\; (\text{probability that there is a particle in each small set } dx_i \text{ near } x_i, \text{ for } 1 \leq \tag{4.15}$$

Equivalently, ρ_k is the k-th *(unordered) joint density* of the process. In particular,

$$\rho_1(x)\, dx \;=\; \text{expected number of particles in a small interval of length } dx \text{ near } x.$$

For a *determinantal* point process in the continuous setting, there is a kernel $K(x, y)$ such that

$$\rho_k(x_1, \ldots, x_k) \;=\; \det\big[K(x_i, x_j)\big]_{i,j=1}^{k} \quad \text{for each } k \geq 1. \tag{4.16}$$

The simplest example is the *Poisson process* (see Section 4.6.3).

4.6.3 Poisson process example

A *Poisson point process with intensity* $\lambda > 0$ on $\mathbb{R}$ is defined by:

- Particles are scattered independently over real line,
- The expected number of particles in an interval $I \subset \mathbb{R}$ is $\lambda|I|$.

Equivalently, one often states that the number of points in any interval I follows a Poisson($\lambda|I|$) distribution, and disjoint intervals are filled independently. One can also check that the correlation functions factorize completely:

$$\rho_k(x_1, \ldots, x_k) \ = \ \lambda^k.$$

Hence, in the Poisson process, there is no “interaction” or “repulsion” between points: the position of one particle does not affect the probability of having other particles nearby. In contrast, a determinantal point process typically exhibits *repulsion*: if you know a particle is present near x, it lowers the density of particles nearby. This effect is crucial in random matrix ensembles at $\beta = 2$.

4.7 Problems

4.7.1 Eigenvalue density of GβE

Read and understand the main principles of the proof of Theorem 4.5 in [DE02].

4.7.2 Chi-square mean and variance

Let X be a random variable with χ^2_ν distribution. Compute the mean and variance of X. (If ν is an integer, you can use the fact that χ^2_ν is a sum of ν independent squares of standard normal random variables. How to extend this to non-integer ν?)

4.7.3 Edge contributions in the tridiagonal moment computation

Show that the cases when the i_ℓ's are close to the edge ($\theta = 0$ or 1) in (4.6) do not contribute to the limit of the moments.

4.7.4 Hermite polynomials and three-term recurrence

Show that the monic Hermite polynomials $H_k(x)$ (4.4) satisfy the three-term recurrence relation

$$H_k(x) = xH_{k-1}(x) - (k-1)H_{k-2}(x).$$

4.7.5

Compute the determinant

$$\det \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix}.$$

4.7.6 Gap probabilities

1. Prove identity (4.14) for DPPs.

2. Prove the generalization computing $\{|X \cap I_k| = n_k \text{ for } 1 \le k \le m\}$.

4.7.7 Stieltjes transform approach for tridiagonal matrices

Complete the derivation from Section 4.4.2 to obtain the limiting Stieltjes transform $G(z)$ for the tridiagonal matrix $T/\sqrt{n}$.

Remark 4.14. This is more of a literature search. It is extensive, and would make an excellent topic for a presentation.

Chapter 5

Determinantal Point Processes and the GUE

5.1 Recap

In Chapter 4 we discussed global spectral behavior of tridiagonal GβE random matrices, and obtained the Wigert semicircle law for the eigenvalue density.

In this lecture we shift our focus to another powerful technique in random matrix theory: the theory of *determinantal point processes* (DPPs). In the $\beta = 2$ (GUE) case the joint eigenvalue distributions can be written in determinantal form. We begin by discussing the discrete version of determinantal processes, and then derive the correlation kernel for the GUE using orthogonal polynomial methods. Finally, we show how the Christoffel–Darboux formula yields a compact representation of the kernel and indicate how one may represent it as a double contour integral—an expression well suited for steepest descent analysis in the large-n limit.

5.2 Discrete determinantal point processes

5.2.1 Definition and basic properties

Let $\mathfrak{X}$ be a (finite or countably infinite) discrete set. A *point configuration* on $\mathfrak{X}$ is any subset $X \subset \mathfrak{X}$ (with no repeated points). A random point process is a probability measure on the space of such configurations.

Definition 5.1 (Determinantal Point Process)**.** A random point process P on $\mathfrak{X}$ is called *determinantal* if there exists a function (the *correlation* *kernel*) $K : \mathfrak{X} \times \mathfrak{X} \to \mathbb{C}$ such that for any n and every finite collection of distinct points $x_1, \ldots, x_n \in \mathfrak{X}$, the joint probability that these points belong to the random configuration is

$$\mathbb{P}\{x_1, \ldots, x_n \in X\} = \det\Big[K(x_i, x_j)\Big]_{i,j=1}^{n}.$$

Determinantal processes are very useful in probability theory and random matrices. They are a natural extension of Poisson processes, and have some parallel properties. Many properties of determinantal processes can be derived from "linear algebra" (broadly understood) applied to the kernel K. There are a few surveys on them: [Sos00], [HKPV06], [Bor11], [KT12]. Let us just mention two useful properties.

Proposition 5.2 (Gap Probability)**.** *If $I \subset \mathfrak{X}$ is a subset, then*

$$\mathbb{P}\{X \cap I = \varnothing\} = \det\Big[I - K_I\Big],$$

where K_I is the restriction of the kernel to I. If I is infinite, then the determinant is understood as a Fredholm determinant.

Remark 5.3. The Fredholm determinant might "diverge" (equal to 0 or 1).

Proposition 5.4 (Generating functions)**.** *Let $f : \mathfrak{X} \to \mathbb{C}$ be a function such that the support of $f - 1$ is finite. Then the generating function of the multiplicative statistics of the determinantal point process is given by*

$$\mathbb{E}\left[\prod_{x \in X} f(x)\right] = \det\Big[I + (\Delta_f - I)K\Big],$$

where the expectation is over the random point configuration $X \subseteq \mathfrak{X}$, Δ_f denotes the operator of multiplication by f (i.e., $(\Delta_f g)(x) = f(x)g(x)$) and the determinant is interpreted as a Fredholm determinant if $\mathfrak{X}$ is infinite.

Remark 5.5 (Fredholm Determinant — Series Definition)**.** The Fredholm determinant of an operator A on $\ell^2(\mathfrak{X})$ is given by the series

$$\det(I + A) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{x_1, \ldots, x_n \in \mathfrak{X}} \det\big[A(x_i, x_j)\big]_{i,j=1}^{n},$$

where the term corresponding to $n = 0$ is defined to be 1.

5.3 Determinantal structure in the GUE

5.3.1 Correlation functions as densities with respect to Lebesgue measure

In the discrete setting discussed above the joint probabilities of finding points in specified subsets of $\mathfrak{X}$ are given by determinants of the kernel evaluated at those points. When the underlying space is continuous (typically a subset of $\mathbb{R}$ or $\mathbb{R}^d$), one works instead with correlation functions which serve as densities with respect to the Lebesgue measure.

Let $X \subset \mathbb{R}$ be a random point configuration. The *n-point correlation function* $\rho_n(x_1, \ldots, x_n)$ is defined by the relation

$$\begin{aligned}\mathbb{P}\{\text{there is a point in each of the infinitesimal intervals } [x_i, x_i{+}dx_i],\ i = 1, \ldots, n\} \\ = \rho_n(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.\end{aligned}$$

For a determinantal point process the correlation functions take a determinantal form:

$$\rho_k(x_1, \ldots, x_k) = \det\Big[K(x_i, x_j)\Big]_{i,j=1}^{k}.$$

Remark 5.6. The reference measure does not necessarily have to be the Lebesgue measure. For example, in the discrete setting, we can also talk about the reference measure, it is the counting measure. The correlation kernel $K(x, y)$ is better understood not as a function of two variables, but as an operator on the Hilbert space $L^2(\mathfrak{X}, d\mu)$, where μ is the reference measure. One can also write $K(x, y)\,\mu(dy)$ or $K(x, y)\sqrt{\mu(dx)\,\mu(dy)}$ to emphasize this structure.

This formulation is particularly useful in the continuous setting, as it allows one to express statistical properties of the point process in terms of integrals over the kernel. For example, the expected number of points in a measurable set $A \subset \mathbb{R}$ is given by

$$\mathbb{E}[\#(X \cap A)] = \int_A \rho_1(x)\, dx,$$

while higher order joint intensities provide information about correlations between points.

5.3.2 The GUE eigenvalues as DPP

Setup

We start from the joint eigenvalue density for the Gaussian Unitary Ensemble (GUE)

$$p(x_1, \dots, x_n)\, dx_1 \cdots dx_n = \frac{1}{Z_{n,2}} \prod_{j=1}^{n} e^{-x_j^2/2} \prod_{1 \le i < j \le n} (x_i - x_j)^2 \, dx_1 \cdots dx_n. \tag{5.1}$$

We will show step by step why this is a determinantal point process,

$$\rho_k(x_1, \dots, x_k) = \det\Big[K_n(x_i, x_j)\Big]_{i,j=1}^{k}, \qquad k \ge 1,$$

with the kernel defined as

$$K_n(x, y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y),$$

where the functions

$$\psi_j(x) = \frac{1}{\sqrt{h_j}}\, p_j(x)\sqrt{w(x)}, \qquad w(x) = e^{-x^2/2},$$

are constructed from the monic Hermite polynomials $\{p_j(x)\}$ which are orthogonal with respect to the weight $w(x)$:

$$\int_{-\infty}^{\infty} p_j(x) p_k(x) e^{-x^2/2}\, dx = h_j\, \delta_{jk}.$$

Recall that "monic" means that the leading coefficient of $p_j(x)$ is 1, and we divide by the norm to make the polynomials orthonormal.

Writing the Vandermonde as a determinant

The product

$$\prod_{1 \le i < j \le n} (x_i - x_j)^2$$

is the square of the Vandermonde determinant. Recall that the Vandermonde determinant is given by

$$\Delta(x_1, \dots, x_n) = \prod_{1 \le i < j \le n} (x_j - x_i) = \det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix}.$$

Thus, we have

$$\prod_{1\le i<j\le n} (x_i - x_j)^2 = \left(\det\left[x_i^{j-1}\right]_{i,j=1}^n\right)^2.$$

Orthogonalization by linear operations

Since determinants are invariant under elementary row or column operations, we can replace the monomials x^{j-1} by any sequence of monic polynomials of degree $j-1$. In particular, we choose the monic Hermite polynomials $p_{j-1}(x)$ and obtain

$$\det\left[x_i^{j-1}\right]_{i,j=1}^n = \det\left[p_{j-1}(x_i)\right]_{i,j=1}^n.$$

The first few monic Hermite polynomials are

$$p_0(x) = 1, \qquad p_1(x) = x, \qquad p_2(x) = x^2-1, \qquad p_3(x) = x^3-3x, \qquad p_4(x) = x^4-6x^2+3.$$

The orthogonality condition for these polynomials is

$$\int_{-\infty}^{\infty} p_j(x)p_k(x)e^{-x^2/2}\,dx = h_j\,\delta_{jk}.$$

We define the functions

$$\phi_j(x) = p_j(x)e^{-x^2/4}, \tag{5.2}$$

and then introduce the orthonormal functions

$$\psi_j(x) = \frac{1}{\sqrt{h_j}}\phi_j(x) = \frac{1}{\sqrt{h_j}}\,p_j(x)e^{-x^2/4}. \tag{5.3}$$

Note that here the weight splits as $e^{-x^2/2} = e^{-x^2/4}e^{-x^2/4}$, which is useful in the next step. The functions ψ_j form an orhtonormal basis of the Hilbert space $L^2(\mathbb{R}, dx)$:

$$\int_{-\infty}^{\infty} \psi_j(x)\psi_k(x)\,dx = \delta_{jk}, \qquad j,k = 0,1,\ldots.$$

Rewriting the density in determinantal form

Substituting the determinant form into the joint density (5.1), we have

$$p(x_1, \ldots, x_n) = \frac{1}{Z_{n,2}} \prod_{j=1}^{n} e^{-x_j^2/2} \Big[\det\Big[p_{j-1}(x_i)\Big]_{i,j=1}^{n}\Big]^2.$$

Incorporate the weight factors into the determinant by writing

$$\prod_{i=1}^{n} e^{-x_i^2/2} = \prod_{i=1}^{n} \left(e^{-x_i^2/4} \cdot e^{-x_i^2/4}\right),$$

so that

$$\prod_{i=1}^{n} e^{-x_i^2/4} \det\Big[p_{j-1}(x_i)\Big]_{i,j=1}^{n} = \det\Big[\phi_{j-1}(x_i)\Big]_{i,j=1}^{n}.$$

Thus, the joint density becomes

$$p(x_1, \ldots, x_n) = \frac{1}{\tilde{Z}_{n,2}} \Big[\det\Big[\phi_{j-1}(x_i)\Big]_{i,j=1}^{n}\Big]^2.$$

This squared-determinant structure is characteristic of determinantal point processes.

We now compute the k-point correlation function by integrating out the remaining $n-k$ variables:

$$\rho_k(x_1, \ldots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \ldots, x_n)\, dx_{k+1} \cdots dx_n. \tag{5.4}$$

Remark 5.7. When defining the k-point correlation function, one might initially expect a combinatorial factor corresponding to the number of ways of choosing k variables out of n, namely $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The absence of an extra $k!$ in the denominator is due to the fact that $x_1, \ldots, x_k$ are fixed, and we are not integrating over all permutations of these variables.

Theorem 5.8 (Determinantal structure for squared-determinant densities)**.** *We have*

$$\rho_k(x_1, \ldots, x_k) = \det\Big[K_n(x_i, x_j)\Big]_{i,j=1}^{k},$$

with the correlation kernel given by

$$K_n(x, y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y).$$

Proof. We begin by writing the joint density as

$$p(x_1, \ldots, x_n) = \frac{1}{\tilde{Z}_{n,2}} \left[\det\Big[\phi_{j-1}(x_i)\Big]_{i,j=1}^{n}\right]^2.$$

Expanding the square of the determinant, we have

$$\left[\det\Big[\phi_{j-1}(x_i)\Big]_{i,j=1}^{n}\right]^2 = \sum_{\sigma,\tau\in S_n} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \prod_{i=1}^{n} \phi_{\sigma(i)-1}(x_i)\phi_{\tau(i)-1}(x_i),$$

where S_n denotes the symmetric group on n elements.

Next, to obtain the k-point correlation function $\rho_k(x_1, \ldots, x_k)$, we integrate out the remaining $n-k$ variables using (5.4). Substituting the expansion of the squared determinant into the expression for ρ_k, we have

$$\rho_k(x_1, \ldots, x_k) = \frac{n!}{(n-k)!\,\tilde{Z}_{n,2}} \sum_{\sigma,\tau\in S_n} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$$
$$\left\{\prod_{i=1}^{k} \phi_{\sigma(i)-1}(x_i)\phi_{\tau(i)-1}(x_i) \prod_{j=k+1}^{n} \int_{\mathbb{R}} \phi_{\sigma(j)-1}(x)\phi_{\tau(j)-1}(x)\,dx\right\}. \quad (5.5)$$

Now, change the functions $\phi_j(x)$ to the orthonormal functions $\psi_j(x)$ using the relation

$$\phi_j(x) = \sqrt{h_j}\,\psi_j(x).$$

This substitution yields

$$\int_{\mathbb{R}} \phi_{\sigma(j)-1}(x)\phi_{\tau(j)-1}(x)\,dx = \sqrt{h_{\sigma(j)-1}h_{\tau(j)-1}} \int_{\mathbb{R}} \psi_{\sigma(j)-1}(x)\psi_{\tau(j)-1}(x)\,dx.$$

By the orthonormality of the ψ_j's, we have

$$\int_{\mathbb{R}} \psi_{\sigma(j)-1}(x)\psi_{\tau(j)-1}(x)\,dx = \delta_{\sigma(j),\tau(j)}.$$

Therefore, for the indices $j = k+1, \ldots, n$, the integrals enforce the condition $\sigma(j) = \tau(j)$. As a result, the double sum over σ and τ reduces to a single sum over permutations on the first k indices, and the factors for the remaining indices simply contribute to the normalization constant.

Let us add more details here. In (5.5), we get, using the symmetry over $x_1, \ldots, x_k$:

$$\rho_k(x_1, \ldots, x_k) = \frac{1}{(n-k)!\,\widehat{Z}_{n,2}} \sum_{\substack{\sigma,\tau\in S_n \\ \sigma(k+1)=\tau(k+1),\ldots,\sigma(n)=\tau(n)}} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \prod_{i=1}^{k} \psi_{\sigma(i)-1}(x_i)\psi_{\tau(i)-1}(x_i). \quad (5.6)$$

Indeed, here we integrated over $x_{k+1}, \ldots, x_n$, and passed from the functions $\phi_0, \phi_1, \ldots, \phi_{n-1}$ to $\psi_0, \psi_1, \ldots, \psi_{n-1}$. The passage to the orthonormal functions only introduces the constant $h_0 h_1 \ldots h_{n-1}$ (by symmetry), and together with $n!$, we include it into the normalization $\widehat{Z}_{n,2}$. The normalization constant does not depend on k, and we later will show that the final normalization becomes 1.

To continue with (5.6), we need two general lemmas.

Lemma 5.9 (Cauchy–Binet formula)**.** *Let A_{ij} and B_{ij} be rectangular matrices of size $m \times p$ and $p \times m$, respectively, with $m \leq p$. Then*

$$\det \left[\sum_{\ell=1}^{p} A_{i\ell} B_{\ell j} \right]_{i,j=1}^{m} = \sum_{\ell_1 < \ell_2 < \cdots < \ell_p} \det \Big[A_{i,\ell_j} \Big]_{i,j=1}^{m} \det \Big[B_{\ell_i,j} \Big]_{j=1}^{m}.$$

Proof. For any $1 \leq k \leq p$, the coefficient of z^{p-k} in the polynomial $\det(zI_p + X)$ is the sum of the $k \times k$ principal minors of X. If $m \leq p$ and A is an $m \times p$ matrix and B is an $p \times m$ matrix, then

$$\det(zI_p + BA) = z^{p-m} \det(zI_m + AB). \tag{5.7}$$

If we compare the coefficient of z^{p-m} in (5.7), the left hand side will give the sum of the principal minors of BA while the right hand side will give the constant term of $\det(zI_m + AB)$, which is simply $\det(AB)$. This yields the desired result. $\square$

Lemma 5.10 (Andreief identity)**.** *Let $f_i(x), g_i(x) \in L^1(\mathbb{R})$ for $i = 1, \ldots, n$. Then*

$$\int_{\mathbb{R}^n} \det[f_i(x_j)]_{i,j=1}^{n} \det[g_i(x_j)]_{i,j=1}^{n} dx_1 \cdots dx_n = n! \det \Big[\int_{\mathbb{R}} f_i(x) g_j(x)\, dx \Big]_{i,j=1}^{n}.$$

Proof. We have by expanding the determinants in the left-hand side:

$$\int_{\mathbb{R}^n} \sum_{\sigma,\tau \in S_n} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) \prod_{i=1}^{n} f_{\sigma(i)}(x_i) g_{\tau(i)}(x_i) dx_1 \cdots dx_n.$$

Now, we can sum over $\sigma\tau^{-1}$, and use the fact that the operation of integration over $\mathbb{R}^n$ is symmetric in the variables $x_1, \ldots, x_n$. We thus need to integrate the products of $f_{(\sigma\tau^{-1})(i)}(x_i)$, yielding the desired determinant in the right-hand side. The factor $n!$ comes from the fact that for each fixed $\sigma\tau^{-1}$, there are $n!$ different pairs (σ, τ). This completes the proof. $\square$

Let us now continue with (5.6), and finish the proof of Theorem 5.8. To sum over σ, τ, let us denote $I = \{\sigma(1), \dots, \sigma(k)\} \subseteq [n] = \{1, \dots, n\}$. The set $[n] \setminus I$ can be ordered in $(n-k)!$ ways, and since σ and τ must coincide on $[n] \setminus I$, the product of their (partial) signs is $+1$ there. Thus, we have

$$(5.6) = \mathrm{const}_n \sum_{I \subseteq [n],\, |I|=k} \sum_{\sigma', \tau' \in S(I)} \operatorname{sgn}(\sigma') \operatorname{sgn}(\tau') \prod_{i=1}^{k} \psi_{\sigma'(i)-1}(x_i) \psi_{\tau'(i)-1}(x_i).$$

where $S(I)$ is the set of all permutations of I. The sum over σ', τ' is actually a product of two sums over two independent permutations, and thus we get the product of two determinants:

$$\det\Big[\psi_{\ell_i - 1}(x_j)\Big]_{i=1}^{k} \det\Big[\psi_{\ell_i - 1}(x_j)\Big]_{i=1}^{k}, \qquad I = \{\ell_1 < \ell_2 < \cdots < \ell_k\}.$$

By Lemma 5.9, we can rewrite the sum (over I) of products of two determinants as a single determinant of the sum. Thus, we have

$$\rho_k(x_1, \dots, x_k) = \mathrm{const} \cdot \det\Big[K_n(x_i, x_j)\Big]_{i,j=1}^{k}, \tag{5.8}$$

where the kernel is given by

$$K_n(x, y) = \sum_{j=0}^{n-1} \psi_j(x) \psi_j(y).$$

The fact that the normalization constant in (5.8) is indeed 1 follows from Lemma 5.10. Indeed, once the integral of ρ_n over $\mathbb{R}^n$ is equal to $n!$, the integral over $x_1 > \cdots > x_n$ becomes 1 by symmetry, as it should be. This completes the proof of Theorem 5.8. □

5.3.3 Christoffel–Darboux formula

Theorem 5.11 (Christoffel–Darboux Formula)**.** *Let $\{p_j(x)\}_{j \ge 0}$ be a family of* monic *orthogonal polynomials with respect to a weight function $w(x)$ on an interval $I \subset \mathbb{R}$. Their squared norms are given by*

$$\int_I p_j(x)\, p_k(x)\, w(x)\, dx = h_j\, \delta_{jk}.$$

Define the orthonormal functions

$$\psi_j(x) = \frac{1}{\sqrt{h_j}}\, p_j(x) \sqrt{w(x)}.$$

Then the kernel

$$K_n(x,y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y) = \sqrt{w(x)w(y)} \sum_{j=0}^{n-1} \frac{p_j(x)p_j(y)}{h_j},$$

admits the closed-form representation

$$K_n(x,y) = \sqrt{w(x)w(y)}\, \frac{1}{h_{n-1}} \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x-y}, \tag{5.9}$$

with the obvious continuous extension when $x = y$.

Proof. Define

$$S_n(x,y) = \sum_{j=0}^{n-1} \frac{p_j(x)p_j(y)}{h_j},$$

so that

$$K_n(x,y) = \sqrt{w(x)w(y)}\, S_n(x,y).$$

Our goal is to prove that

$$(x-y)S_n(x,y) = \frac{1}{h_{n-1}} \Big[p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y) \Big]. \tag{5.10}$$

Since the polynomials are monic and orthogonal, they satisfy the three-term recurrence relation

$$x\, p_j(x) = p_{j+1}(x) + \alpha_j\, p_j(x) + \beta_j\, p_{j-1}(x), \quad j \geq 0,$$

with the convention $p_{-1}(x) = 0$ and where $\beta_j = \frac{h_j}{h_{j-1}}$. This recurrence comes from the three facts:

1. The polynomials are orthogonal with respect to the weight function $w(x)$ supported on the real line;
2. The operator of multiplication by x is self-adjoint with respect to the inner product induced by $w(x)$.
3. The multiplication by x of p_j gives p_{j+1} plus a correction of degree $\leq j$.

Writing the recurrence for both $p_j(x)$ and $p_j(y)$ yields:

$$x\, p_j(x) = p_{j+1}(x) + \alpha_j\, p_j(x) + \beta_j\, p_{j-1}(x),$$
$$y\, p_j(y) = p_{j+1}(y) + \alpha_j\, p_j(y) + \beta_j\, p_{j-1}(y).$$

Multiplying the first equation by $p_j(y)$ and the second by $p_j(x)$, and then subtracting, we obtain:

$$(x-y)p_j(x)p_j(y) = p_{j+1}(x)p_j(y)-p_j(x)p_{j+1}(y)+\beta_j\Big[p_{j-1}(x)p_j(y)-p_j(x)p_{j-1}(y)\Big].$$

Dividing by h_j and summing over $j = 0, \dots, n-1$ gives:

$$(x-y)S_n(x,y) = \sum_{j=0}^{n-1}\frac{1}{h_j}\Big[p_{j+1}(x)p_j(y)-p_j(x)p_{j+1}(y)\Big]+\sum_{j=0}^{n-1}\frac{\beta_j}{h_j}\Big[p_{j-1}(x)p_j(y)-p_j(x)p_{j-1}(y)\Big].$$

A reindexing of the sums shows that the series telescopes, leaving only the boundary terms. In particular, one finds

$$(x-y)S_n(x,y) = \frac{1}{h_{n-1}}\Big[p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)\Big].$$

This establishes (5.10), and hence the representation (5.9) for $K_n(x,y)$.

The continuous extension to $x = y$ is obtained via l'Hôpital's rule. □

5.4 Problems

5.4.1 Gap Probability for Discrete DPPs

Let $\mathfrak{X}$ be a (finite or countably infinite) discrete set and suppose that a point process on $\mathfrak{X}$ is determinantal with kernel

$$K : \mathfrak{X} \times \mathfrak{X} \to \mathbb{C},$$

so that for any finite collection of distinct points $x_1, \dots, x_n \in \mathfrak{X}$ the joint probability that these points belong to the configuration is

$$\mathbb{P}\{x_1, \dots, x_n \in X\} = \det\Big[K(x_i,x_j)\Big]_{i,j=1}^{n}.$$

Show that for any subset $I \subset \mathfrak{X}$ (finite or such that the Fredholm determinant makes sense) the gap probability

$$\mathbb{P}\{X \cap I = \varnothing\} = \det\Big[I - K_I\Big],$$

where K_I is the restriction of K to $I \times I$.

5.4.2 Generating Functions for Multiplicative Statistics

Let $f : \mathfrak{X} \to \mathbb{C}$ be a function such that the support of $f - 1$ is finite. Prove that for a determinantal point process on $\mathfrak{X}$ with kernel K the generating function

$$\mathbb{E}\Big[\prod_{x \in X} f(x)\Big] = \det\Big[I + (\Delta_f - I)K\Big]$$

holds, where Δ_f is the multiplication operator defined by $(\Delta_f g)(x) = f(x)g(x)$. *Hint:* Expand the Fredholm determinant series and compare with the definition of the correlation functions.

5.4.3 Variance

Let I be a finite interval, and let $N(I)$ be the number of points of a determinantal point process in I with the kernel $K(x, y)$. Find $\mathrm{Var}(I)$ in terms of the kernel $K(x, y)$.

5.4.4 Formula for the Hermite polynomials

Show that the monic Hermite polynomials $p_j(x)$ are given by

$$p_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.$$

5.4.5 Generating function for the Hermite polynomials

Show that

$$\sum_{n=0}^{\infty} \frac{t^n}{n!} p_n(x) = e^{tx - t^2/2}.$$

5.4.6 Projection Property of the GUE Kernel

Show that the kernel

$$K_n(x, y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y),$$

(with the orthonormal functions ψ_j defined as in the lecture) acts as an orthogonal projection operator on $L^2(\mathbb{R})$. In other words, prove that for all $x, y \in \mathbb{R}$

$$\int_{-\infty}^{\infty} K_n(x, z)K_n(z, y)\, dz = K_n(x, y).$$

5.4.7 Recurrence Relation for the Hermite Polynomials

Show that the monic Hermite polynomials defined by

$$p_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}$$

satisfy the three-term recurrence relation

$$p_{n+1}(x) = x\, p_n(x) - n\, p_{n-1}(x),$$

with the convention $p_{-1}(x) = 0$.

5.4.8 Differential Equation for the Hermite Polynomials

Prove that the monic Hermite polynomials $p_n(x)$ satisfy the second-order differential equation

$$p_n''(x) - x\, p_n'(x) + n\, p_n(x) = 0.$$

5.4.9 Norm of the Hermite Polynomials

Show that

$$h_n = \int_{-\infty}^{\infty} p_n(x)^2\, e^{-x^2/2}\, dx = n!\sqrt{2\pi}.$$

5.4.10 Existence of Determinantal Point Processes with a Given Kernel

Let X be a locally compact Polish space equipped with a reference measure μ, and let $K(x, y)$ be the kernel of an integral operator K acting on $L^2(X, \mu)$. Suppose that:

1. K is Hermitian (i.e. $K(x, y) = \overline{K(y, x)}$),
2. K is locally trace class, and
3. $0 \leq K \leq I$ as an operator, that is, both the operator K and the operator $I - K$ are nonnegative definite. For K, this condition is

   $$\int_X \int_X f(x)\overline{K(x, y)} f(y)\, d\mu(x)\, d\mu(y) \geq 0$$

   for all $f \in L^2(X, \mu)$.

Under these conditions there exists a unique determinantal point process on X with correlation functions given by

$$\rho_n(x_1, \ldots, x_n) = \det\Big[K(x_i, x_j)\Big]_{i,j=1}^{n}.$$

Explain why the condition $0 \leq K \leq I$ is necessary. For the proof of the existence and uniqueness of the determinantal point process, see [Sos00].

Chapter 6

Double contour integral kernel. Steepest descent and semicircle law

6.1 Recap: Determinantal structure of the GUE

Last time, we proved the following result:

Theorem 6.1. *The GUE correlation functions are given by*

$$\rho_k(x_1, \ldots, x_k) = \det\Big[K_n(x_i, x_j)\Big]_{i,j=1}^{k},$$

with the correlation kernel

$$K_n(x, y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y).$$

Here

$$\psi_j(x) = \frac{1}{\sqrt{h_j}}\, p_j(x)\, e^{-x^2/4},$$

where $p_j(x)$ are the monic Hermite polynomials, and h_j are the normalization constants so that $\psi_j(x)$ are orthonormal in $L^2(\mathbb{R})$.

For this theorem, we need Cauchy–Binet summation formula and Andreief identity (which is essentially the same as Cauchy–Binet, but when summation is replaced by integration). Having these, we can write

$$\rho_k(x_1, \ldots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \ldots, x_n)\, dx_{k+1} \cdots dx_n$$

$$= \frac{1}{(n-k)!\,\widehat{Z}_{n,2}} \sum_{\substack{\sigma,\tau\in S_n \\ \sigma(k+1)=\tau(k+1),\ldots,\sigma(n)=\tau(n)}} \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)\prod_{i=1}^{k}\psi_{\sigma(i)-1}(x_i)\psi_{\tau(i)-1}(x_i)$$

$$= \mathrm{const}_n \sum_{I\subseteq[n],\,|I|=k}\ \sum_{\sigma',\tau'\in S(I)} \operatorname{sgn}(\sigma')\operatorname{sgn}(\tau')\prod_{i=1}^{k}\psi_{\sigma'(i)-1}(x_i)\psi_{\tau'(i)-1}(x_i)$$

$$= \mathrm{const}_n \sum_{I\subseteq[n],\,|I|=k} \det\left[\psi_{i_\alpha}(x_j)\right]_{\alpha,j=1}^{k}\det\left[\psi_{i_\alpha}(x_j)\right]_{\alpha,j=1}^{k},$$

where $I=\{i_1,\ldots,i_k\}$ is a subset of $[n]$ of size k, and $S(I)$ is the set of permutations of I. The last sum of products of two determinants is written by the Cauchy–Binet formula as

$$\mathrm{const}_n\cdot\det\left[\sum_{j=0}^{n-1}\psi_j(x_\alpha)\psi_j(x_\beta)\right]_{\alpha,\beta=1}^{k},$$

and finally the constant is equal to 1 by Andreief identity.

6.2 Double Contour Integral Representation for the GUE Kernel

6.2.1 One contour integral representation for Hermite polynomials

Recall that the GUE kernel is defined by

$$K_N(x,y)=\sum_{n=0}^{N-1}\psi_n(x)\psi_n(y),$$

with the orthonormal functions

$$\psi_n(x)=\frac{1}{\sqrt{h_n}}\,p_n(x)\,e^{-x^2/4},$$

where the (monic, probabilists') Hermite polynomials are given by

$$p_n(x)=(-1)^n e^{x^2/2}\frac{d^n}{dx^n}\,e^{-x^2/2}. \tag{6.1}$$

Note that the monic Hermite polynomials are uniquely defined by the orthogonality property. We are not proving (6.1) here, it is an exercise.

Lemma 6.2 (Generator function for Hermite polynomials)**.** *We have*

$$\exp\Big(xt - \frac{t^2}{2}\Big) = \sum_{n\geq 0} p_n(x)\frac{t^n}{n!}.$$

The series converges for all t since the left-hand side is an entire function of t.

Proof. Write the generating function as

$$\sum_{n\geq 0} p_n(x)\frac{t^n}{n!} = \sum_{n\geq 0} \frac{(-1)^n t^n}{n!}\, e^{x^2/2}\frac{d^n}{dx^n}\, e^{-x^2/2}.$$

Since the factor $e^{x^2/2}$ does not depend on n, we can factor it out:

$$\sum_{n\geq 0} p_n(x)\frac{t^n}{n!} = e^{x^2/2}\sum_{n\geq 0} \frac{(-t)^n}{n!}\frac{d^n}{dx^n}\, e^{-x^2/2}.$$

Now, recall Taylor's theorem: for any holomorphic function f we have

$$f(x-t) = \sum_{n\geq 0} \frac{(-t)^n}{n!} f^{(n)}(x).$$

Applying this with $f(x) = e^{-x^2/2}$, we deduce that

$$\sum_{n\geq 0} \frac{(-t)^n}{n!}\frac{d^n}{dx^n}\, e^{-x^2/2} = e^{-(x-t)^2/2}.$$

Thus, our generating function becomes

$$\sum_{n\geq 0} p_n(x)\frac{t^n}{n!} = e^{x^2/2}\, e^{-(x-t)^2/2},$$

as desired. □

By Cauchy's integral formula we can write using Lemma 6.2:

$$p_n(x) = \frac{n!}{2\pi i}\oint_C \frac{\exp\Big(xt - \frac{t^2}{2}\Big)}{t^{n+1}}\, dt, \tag{6.2}$$

where the contour C is a simple closed curve encircling the origin. Indeed, here we use the complex analysis property

$$\frac{1}{2\pi i}\oint_C \frac{1}{z^{k+1}}\, dz = \begin{cases} 1, & \text{if } k = 0,\\ 0, & \text{if } k \neq 0,\end{cases}$$

so (6.2) is simply a complex analysis version of the operation of extracting the coefficient of t^n in the Taylor expansion.

Therefore,

$$\psi_n(x) = \frac{1}{\sqrt{h_n}}\, p_n(x)\, e^{-x^2/4} = \frac{e^{-x^2/4}}{\sqrt{h_n}}\, \frac{n!}{2\pi i} \oint_C \frac{\exp\Big(xt - \frac{t^2}{2}\Big)}{t^{n+1}}\, dt.$$

6.2.2 Another contour integral representation for Hermite polynomials

We start with the Fourier transform identity

$$\int_{-\infty}^{\infty} \exp\Big(-\frac{t^2}{2} + i\,t\,x\Big)\, dt = \sqrt{2\pi}\, e^{-x^2/2}.$$

Differentiating both sides n times with respect to x yields

$$\frac{d^n}{dx^n}\Big(e^{-x^2/2}\Big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (i\,t)^n\, e^{-t^2/2 + i\,t\,x}\, dt.$$

Recalling the definition

$$p_n(x) = (-1)^n\, e^{x^2/2}\, \frac{d^n}{dx^n}\Big(e^{-x^2/2}\Big),$$

we obtain

$$p_n(x) = \frac{(-1)^n\, e^{x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (i\,t)^n\, e^{-t^2/2 + i\,t\,x}\, dt.$$

Next, perform the change of variable

$$s = i\,t, \quad \text{so that} \quad t = -i\,s, \quad dt = -i\,ds.$$

Under this substitution the factors transform as follows:

$$(i\,t)^n = s^n,$$

and the exponent becomes

$$-\frac{t^2}{2} + i\,t\,x = -\frac{(-i\,s)^2}{2} + i\,(-i\,s)\,x = \frac{s^2}{2} + s\,x.$$

Thus, the integral transforms into

$$\int_{-\infty}^{\infty} (i\,t)^n\, e^{-t^2/2 + i\,t\,x}\, dt = -i \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2 + s\,x}\, ds.$$

Substituting back we have

$$p_n(x) = \frac{(-1)^n\, e^{x^2/2}}{\sqrt{2\pi}}\, (-i) \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2+s\,x}\, ds.$$

That is,

$$p_n(x) = \frac{i\,(-1)^{n+1}\, e^{x^2/2}}{\sqrt{2\pi}} \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2+s\,x}\, ds.$$

Finally, change the sign of s, and we get:

$$p_n(x) = \frac{i\, e^{x^2/2}}{\sqrt{2\pi}} \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2-s\,x}\, ds.$$

Therefore,

$$\psi_n(x) = \frac{i\, e^{x^2/4}}{\sqrt{2\pi\, h_n}} \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2-s\,x}\, ds.$$

6.2.3 Normalization of Hermite polynomials

Lemma 6.3. *We have*

$$h_n = \int_{-\infty}^{\infty} p_n(x)^2\, e^{-x^2/2}\, dx = n!\sqrt{2\pi}.$$

Proof. Multiply the generating function

$$\exp\Big(xt - \frac{t^2}{2}\Big) = \sum_{n\geq 0} p_n(x)\frac{t^n}{n!}$$

with a second copy (with parameter s):

$$\exp\Big(xs - \frac{s^2}{2}\Big) = \sum_{m\geq 0} p_m(x)\frac{s^m}{m!}.$$

Then,

$$\exp\Big(xt - \frac{t^2}{2}\Big)\exp\Big(xs - \frac{s^2}{2}\Big) = \sum_{n,m\geq 0} p_n(x)p_m(x)\frac{t^n s^m}{n!m!}.$$

Integrate both sides against $e^{-x^2/2}\, dx$. Using the orthogonality

$$\int_{-\infty}^{\infty} p_n(x)p_m(x)e^{-x^2/2}dx = h_n\delta_{nm},$$

the right-hand side becomes

$$\sum_{n\geq 0} \frac{h_n}{(n!)^2}(ts)^n.$$

On the left-hand side, we have

$$\int_{-\infty}^{\infty} e^{-x^2/2} \exp\Big(x(t+s) - \frac{t^2+s^2}{2}\Big)dx.$$

Completing the square in x or recalling the standard Gaussian integral yields

$$\sqrt{2\pi}\exp\Big(\frac{(t+s)^2-(t^2+s^2)}{2}\Big) = \sqrt{2\pi}\exp(ts).$$

Thus, we obtain

$$\sqrt{2\pi}\exp(ts) = \sum_{n\geq 0} \frac{h_n}{(n!)^2}(ts)^n.$$

Expanding the left side as

$$\sqrt{2\pi}\sum_{n\geq 0}\frac{(ts)^n}{n!},$$

and comparing coefficients, we conclude that

$$\frac{h_n}{(n!)^2} = \frac{\sqrt{2\pi}}{n!} \quad\Longrightarrow\quad h_n = n!\sqrt{2\pi}.$$

This completes the proof. □

6.2.4 Double contour integral representation for the GUE kernel

We can sum up the kernel (essentially, this is another proof of the Christoffel–Darboux formula):

$$\begin{aligned} K_n(x,y) &= \sum_{k=0}^{n-1}\psi_k(x)\psi_k(y) \\ &= \frac{e^{\frac{x^2-y^2}{4}}}{(2\pi)^2}\oint_C dt\int_{-i\infty}^{i\infty} ds\, \exp\Big\{-\frac{t^2}{2}+xt+\frac{s^2}{2}-ys\Big\}\underbrace{\sum_{k=0}^{n-1}s^k t^{-k-1}}_{\frac{1-(s/t)^n}{t-s}}. \end{aligned} \tag{6.3}$$

Here we used the two contour integral representations for Hermite polynomials, and the explicit norm (Lemma 6.3). At this point, the t contour is a small circle around 0, and the s contour is a vertical line in the complex plane. Their mutual position can be arbitrary at this point — the s contour goes along the imaginary line. Indeed, the fraction $\frac{1-(s/t)^n}{t-s}$ does not have a singularity at $s = t$ due to the cancellation.

Let us now move the s contour to be to the left of the t contour, as in Figure 6.1. On the new contours, we have $|s| > |t|$. Now we can add the summands $s^k t^{-k-1}$ for all $k \leq -1$ into the sum in (6.3). Indeed, for $|s| > |t|$, the series in k converges, while the summand $s^k t^{-k-1}$ has zero residue at 0 and thus adding the summands does not change the value of the integral.

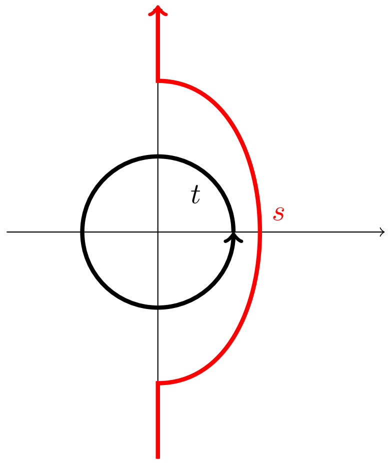

Figure 6.1: Integration contours for the GUE kernel (6.4).

With this extension of the sum, formula (6.3) becomes

$$K_n(x,y) = \frac{e^{(y^2-x^2)/4}}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \, \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n. \quad (6.4)$$

Remark 6.4. The s contour passes to the right of the t contour, but it might as well pass to the left of it. Indeed, one can deform the s contour to the left while picking the residue at $s = t$:

$$2\pi i \operatorname{Res}_{s=t} \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n = -e^{t(x-y)}.$$

This function is entire in t, and its integral over the t contour is zero. Therefore, there is no difference where the s contour passes with respect to the t

contour.

6.2.5 Conjugation of the kernel

The kernel $K_n(x,y)$ contains a factor $e^{\frac{y^2-x^2}{4}} = g(x)/g(y)$, where $g(\cdot)$ is a nonvanishing function. This factor can be safely removed, since in all determinants $\det[K_n(x_i,x_j)]_{i,j=1}^k$ representing the correlation functions, the conjugation factors $g(x_i)/g(x_j)$ do not affect the value of the determinant. Thus, we can and will deal with the correlation kernel

$$K_n(x,y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \, \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n, \qquad (6.5)$$

and will use the same notation for it. Throughout the asymptotic analysis in Section 6.4 below, other conjugation factors may appear, but we can similarly remove them.

6.2.6 Extensions

Many other versions of the GUE / unitary invariant ensembles admit determinantal structure:

1. The GUE corners process [JN06]

2. The Dyson Brownian motion (e.g., add a GUE to a diagonal matrix) [NF98]

3. GUE corners plus a fixed matrix [FF14]

4. Corners invariant ensembles with fixed eigenvalues $UDU^\dagger$, where D is a fixed diagonal matrix and U is Haar distributed on the unitary group [Met13]

We will discuss the corners process structure in the next Chapter 7.

6.3 Steepest descent — generalities for single integrals

6.3.1 Setup

In many problems arising in random matrix theory—as well as in asymptotic analysis more generally—it is necessary to evaluate integrals of the form

$$I(\Lambda) = \int_\gamma e^{\Lambda f(z)} \phi(z)\, dz, \tag{6.6}$$

where

- $\Lambda > 0$ is a large parameter,
- $f(z)$ and $\phi(z)$ are holomorphic functions in a neighborhood of the contour $\gamma \subset \mathbb{C}$,
- and the contour γ is chosen in such a way that the integral converges.

The *method of steepest descent* (also known as the *saddle point method*) provides a systematic procedure for obtaining the asymptotic behavior of $I(\Lambda)$ as $\Lambda \to +\infty$.

The key observation is that for large Λ, the exponential term $e^{\Lambda f(z)}$ is highly oscillatory or decaying, so that the main contributions to the integral come from small neighborhoods of points where the real part of $f(z)$ is maximal. Moreover, since we can deform the integration contour γ to pick points where $\operatorname{Re} f(z)$ is even bigger, it makes sense to find points *not only on the original contour* where $\operatorname{Re} f(z)$ is maximal. Such *critical* (or *saddle*) points are found from the equation with the complex derivative:

$$f'(z) = 0$$

Indeed, since $\operatorname{Re} f(z)$ is harmonic and $f(z)$ satisfies the Cauchy–Riemann equations, the condition $f'(z) = 0$ is equivalent to the condition that $\operatorname{Re} f(z)$ has zero gradient. Moreover, by harmonicity, all critical points of $\operatorname{Re} f(z)$ are saddle-like.

Once the saddle points are identified, one deforms the contour γ to Γ so that Γ passes through the saddle point(s) with the maximal value of $\operatorname{Re} f(z)$, and, moreover, such that on the rest of the new contour Γ the real part of $f(z)$ is strictly less than the value(s) at the saddle point(s). The decrease of $\operatorname{Re} f(z)$ along Γ may be ensured if one picks Γ to be *steepest descent* for $\operatorname{Re} f(z)$. By holomorphicity of $f(z)$, the steepest descent of Re is equivalent to the condition that the imaginary part of $f(z)$ is constant along Γ.

Remark 6.5. In practical applications, one does not need Γ to be fully steepest descent (it is usually hard to control). One can either choose Γ to be steepest descent in a neighborhood of the critical point and estimate the real part outside, or simply estimate the change of $\operatorname{Re} f(z)$ directly along a given contour.

Remark 6.6. The function $\phi(z)$ might not be holomorphic, and might have poles. The deformation of the contour from γ to Γ might pick residues at these poles. These residues can be harmless (easy to account for) or not (hard to account for; or affect the asymptotics of the integral), and one has to be careful with the contour deformation.

Despite the caveats in Remark 6.5 and **??**, in what follows in this section we will discuss the easiest case of steepest descent analysis. We also assume that there is only one saddle point z_0 to take care of.

6.3.2 Saddle points and steepest descent paths

Definition 6.7 (Saddle point)**.** A point $z_0 \in \mathbb{C}$ is called a *saddle point* of $f(z)$ if

$$f'(z_0) = 0.$$

We shall assume in what follows that at every saddle point under consideration the second derivative satisfies

$$f''(z_0) \neq 0.$$

Definition 6.8 (Steepest descent path)**.** Let z_0 be a saddle point of $f(z)$. A curve $\Gamma \subset \mathbb{C}$ passing through z_0 is called a *steepest descent path* for $f(z)$ if along Γ the imaginary part of $f(z)$ is constant (i.e., $\operatorname{Im}\big(f(z)\big) = \operatorname{Im}\big(f(z_0)\big)$ for all $z \in \Gamma$), which implies that the real part $\operatorname{Re}\big(f(z)\big)$ decreases away from z_0.

In a neighborhood of a saddle point z_0,

$$z = z_0 + w, \qquad f(z) = f(z_0) + \frac{1}{2} f''(z_0) w^2 + O(w^3).$$

If we denote

$$f''(z_0) = |f''(z_0)| e^{i\theta_0},$$

then writing $w = r\, e^{i\varphi}$, we obtain

$$f(z) = f(z_0) + \frac{1}{2} |f''(z_0)| r^2 e^{i(2\varphi+\theta_0)} + O(r^3).$$

For the imaginary part to remain constant in a neighborhood of z_0, and, moreover, for the phase of the quadratic term to be π modulo 2π, one must choose φ so that

$$2\varphi + \theta_0 = \pi \pmod{2\pi}. \tag{6.7}$$

We need the phase π so that the exponent is negative, for the integral to converge.

There are two directions satisfying (6.7) through z_0, and we use both of them for our contour Γ. Along these directions, one finds that

$$\mathrm{Re}\big(f(z)\big) = \mathrm{Re}\big(f(z_0)\big) - \frac{1}{2}|f''(z_0)|r^2 + O(r^3),$$

so that $\mathrm{Re}(f(z))$ is maximal at $z = z_0$ and decays quadratically as one moves away from z_0 along the steepest descent paths.

6.3.3 Local asymptotic evaluation near a saddle point

Assume now that the contour γ in (6.6) has been deformed so that it passes through a saddle point z_0 along a steepest descent path. In a small neighborhood of z_0, we write

$$z = z_0 + w/\sqrt{\Lambda},$$

so the local contribution of a neighborhood of z_0 to the integral is

$$I_{z_0}(\Lambda) = e^{\Lambda f(z_0)}\phi(z_0)\frac{1}{\sqrt{\Lambda}}\Big(1 + O\Big(\frac{1}{\Lambda^{\frac{1}{2}}}\Big)\Big)\int_{-\infty}^{\infty} e^{\frac{1}{2}f''(z_0)w^2}\,dw. \tag{6.8}$$

Here the integration is taken along the steepest descent direction, so that the quadratic term in the exponent is real and negative. (That is, by the choice (6.7), we have $\mathrm{Re}\big(f''(z_0)w^2\big) = -|f''(z_0)|r^2$.) Then the Gaussian integral evaluates to

$$\int_{-\infty}^{\infty} e^{-\frac{|f''(z_0)|}{2}w^2}\,dw = \sqrt{\frac{2\pi}{|f''(z_0)|}}.$$

Hence, we arrive at the following fundamental result.

Theorem 6.9 (Local asymptotics via steepest descent)**.** *Let z_0 be a saddle point of $f(z)$ with $f'(z_0) = 0$ and $f''(z_0) \neq 0$, and assume that $\phi(z)$ is holomorphic in a neighborhood of z_0. Then, as $\Lambda \to +\infty$, the contribution of a small neighborhood of z_0 to the integral (6.6) is given by*

$$I_{z_0}(\Lambda) \sim e^{\Lambda f(z_0)}\,\phi(z_0)\sqrt{\frac{2\pi}{\Lambda\,|f''(z_0)|}}, \qquad \Lambda \to +\infty. \tag{6.9}$$

Moreover, the behavior (6.9) captures the full asymptotic behavior of the integral (6.6) as long as on the new contour Γ, the real part of $f(z)$ is maximized at z_0 and is separated from $\operatorname{Re} f(z_0)$ everywhere else on Γ outside of a small neighborhood of z_0.

Under appropriate assumptions (typically, if f and ϕ are holomorphic on a neighborhood that can be reached by the deformed contour and if the contributions away from the saddle points are exponentially small), one may show that the error in approximating the full integral by the sum of the local contributions is itself exponentially small relative to the leading order terms. In many cases, the next-order corrections can be computed by carrying the expansion in (6.8) to higher order in w. (See, e.g., [Olv74] for a systematic treatment.)

6.4 Steepest descent for the GUE kernel

6.4.1 Scaling

Let us now consider the GUE kernel (6.5),

$$K_n(x,y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \, \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n,$$

where the integration contours are as in Figure 6.1.

We know from the Wigner semicircle law (established for real symmetric matrices with general iid entries in in Chapter 2, and for the GUE in Chapter 4) that the eigenvalues live on the scale $\sqrt{n}$. This means that to capture the local asymptotics, we need to scale

$$x = X\sqrt{n} + \frac{\Delta x}{\sqrt{n}}, \qquad y = Y\sqrt{n} + \frac{\Delta y}{\sqrt{n}}, \qquad \Delta x, \Delta y \in \mathbb{R}. \tag{6.10}$$

Moreover, if $X \neq Y$ (i.e., different global positions), one can check that the kernel vanishes. In other words, the local behaviors at different global positions are independent. See Problem 6.5.1. In what follows, we take $Y = X$.

Let us also make a change of the integration variables:

$$t = z\sqrt{n}, \qquad s = w\sqrt{n}.$$

The integration contours for z and w look the same as in Figure 6.1, up to a rescaling. However, as 0 and $t = s$ are the only singularities in the integrand, we can deform the z, w contours as we wish, while keeping $|z| < |w|$ and the general shape as in Figure 6.1.

We thus have:

$$K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n})$$
$$= \frac{\sqrt{n}}{(2\pi)^2} \oint_C dz \int_{-i\infty}^{i\infty} dw \, \frac{\exp\left\{n\left(\log w - \log z + \frac{w^2}{2} - \frac{z^2}{2} + X(z-w) + \frac{z\Delta x - w\Delta y}{n}\right)\right\}}{w-z}. \tag{6.11}$$

Remark 6.10. The logarithms in the exponent are harmless, since for the estimates we only need the real parts of the logarithms, and for the main contributions, we will have $z \approx w$, so any phases of the logarithms would cancel.

The asymptotic analysis of double contour integrals like (6.11) in the context of determinantal point processes was pioneered in [Oko02, Section 3].

6.4.2 Critical points

Let us define

$$S(z) := \frac{z^2}{2} + \log z - Xz.$$

Then the exponent contains $n\,(S(w) - S(z))$. According to the steepest descent ideology, we should deform the integration contours to pass through the critical point(s) z_{cr} of $S(z)$. Moreover, the new w contour should maximize the real part of $S(z)$ at z_{cr}, and the new z contour should minimize it. If $S''(z_{cr}) \neq 0$, it is possible to locally choose such contours, they will be perpendicular to each other at z_{cr}.

Thus, we need to find the critical points of $S(z)$. They are found from the quadratic equation:

$$S'(z) = z + \frac{1}{z} - X = 0, \qquad z_{cr} = \frac{X \pm \sqrt{X^2 - 4}}{2}. \tag{6.12}$$

Depending on whether $|X| < 2$, there are three cases. Unless $|X| = 2$ (when equation (6.12) has a single root), we have $S''(z_{cr}) \neq 0$.

In this lecture, we focus on the density function, which is obtained by taking the asymptotics of the kernel $K(x, x)$. In the next Chapter 7, we discuss limits of the correlation functions.

6.4.3 Imaginary critical points: $|X| < 2$, "bulk"

When $|X| < 2$, the critical points are complex conjugate. Denote them by z_{cr} and $\overline{z_{cr}}$. Since $S(z)$ has real coefficients, we have

$$\operatorname{Re} S(z_{cr}) = \operatorname{Re} S(\overline{z_{cr}}).$$

Thus, we need to consider the contribution from both points. The behavior of $\operatorname{Re} S(z)$ on the complex plane can be illustrated by a 3D plot or by a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ has constant sign. See Figure 6.2 for an illustration in the case $X = \frac{1}{2}$.

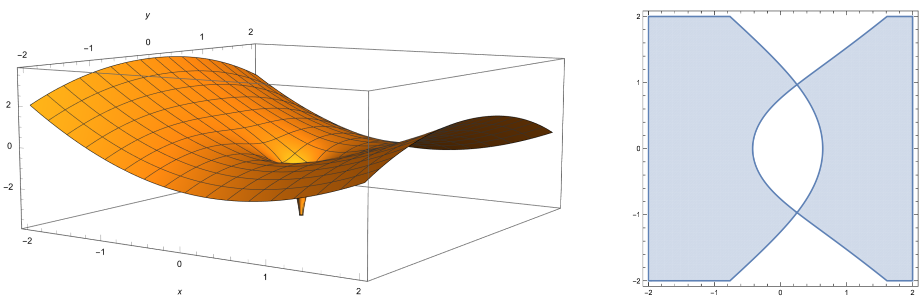

Figure 6.2: A 3D plot and a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ is positive (highlighted) or negative, in the case $X = \frac{1}{2}$. In this case, $z_{cr} \approx 0.25 + 0.96i$.

From the region plot, we see that the new z contour should pass through the shaded region $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr}) > 0$, and the new w contour should pass through the unshaded region $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr}) < 0$.

Deforming the contours from Figure 6.1 to the new contours is impossible without passing through the residue at $w = z$. Moreover, this residue appears only for certain values of z. Namely, let us first make the z contour to be the positively (counterclockwise) oriented unit circle. It passes through the critical points z_{cr} and $\overline{z_{cr}}$. Since the original w contour is to the right of the z contour, we only encounter the residue when z is in the right half of the arc.

Thus, we can write

$$\oiint_{\text{old contours}} = \oiint_{\text{new contours}} + \int_{\overline{z_{cr}}}^{z_{cr}} 2\pi i \operatorname{Res}_{w=z} dz, \tag{6.13}$$

where in the single integral, the z contour passes to the right of the origin, along the right half of the unit circle.

It remains to consider the two integrals in the right-hand side of (6.13). Recall that the correlation functions are defined relative to a reference measure, and the right object to scale is

$$K_n(x,y)dy = \frac{1}{\sqrt{n}} K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n})\, d\,(\Delta y)\,.$$

The extra factor $n^{-1/2}$ compensates the prefactor $\sqrt{n}$ in (6.11).

The single integral takes the form

$$\frac{-i}{2\pi}\int_{\overline{z_{cr}}}^{z_{cr}} dz = \frac{\sin(\arg z_{cr})}{\pi}. \tag{6.14}$$

The double integral in (6.13) has both contours in the "steepest descent" regime, which means that the main contribution is

$$\mathrm{const}\cdot \frac{e^{n(\operatorname{Re} S(z_{cr}) - \operatorname{Re} S(z_{cr}))}}{\sqrt{n}} \sim \frac{\mathrm{const}}{\sqrt{n}}.$$

At this rate, the double integral over the new contours *does not* contribute to the asymptotics of the correlation functions. Recall that the correlation functions are expressed as finite-dimensional determinants of the kernel $K_n(x,y)$, and the error $O(n^{-1/2})$ is negligible in the limit $n \to +\infty$. This is because the main term comes from the single integral, which does not vanish.

Note that

$$z_{cr} = \frac{X \pm \sqrt{X^2-4}}{2}, \qquad \sin(\arg z_{cr}) = \frac{\sqrt{4-X^2}}{2}.$$

This again establishes the *Wigner semicircle law* for the GUE kernel.

Remark 6.11. This is already the third proof — we worked with trees, the tridiagonal form, and now via steepest descent. The steepest descent method is the least general one, but it allows to access local correlations in the bulk and at the edge.

We will consider other regimes, $|X| > 2$ and $|X| = 2$, in the next Chapter 7.

6.5 Problems

6.5.1 Different global positions

Show that if in (6.10) we take $X \neq Y$, then $K_n(x,y)$ vanishes as $n \to +\infty$. Moreover, establish the rate of decay in n. Is it power-law or exponential?

6.5.2 Sine kernel

Compute the integral (6.14).

6.5.3 Discrete sine process

Define the discrete sine kernel on $\mathbb{Z}$ by

$$K_{\mathrm{dsine}}(x,y) \coloneqq \begin{cases} \dfrac{\sin \rho(x-y)}{\pi(x-y)}, & x \neq y, \\ \dfrac{\rho}{\pi}, & x = y, \end{cases}$$

where $\rho \in [0,1]$ is the density parameter.

Let $\rho = 1/2$. Compute (numerically) the asymptotics of the two events under the discrete sine process:

$$\mathbb{P}\Big(\underbrace{\circ\circ\ldots\circ}_{n \text{ times}}\underbrace{\bullet\bullet\ldots\bullet}_{n \text{ times}}\Big), \qquad \mathbb{P}\Big(\underbrace{\circ\bullet\circ\bullet\ldots\circ\bullet}_{2n \text{ points}}\Big),$$

If the sine process was of independent random points (with the same density $1/2$), both events would have the same probability 2^{-2n}. Which event is more favored by the sine process?

Chapter 7

Steepest descent and local statistics. Cutting corners

7.1 Steepest descent for the GUE kernel

7.1.1 Recap

We continue the asymptotic analysis of the GUE kernel.

The GUE correlation kernel is defined by

$$K_n(x,y) = \sum_{j=0}^{n-1} \psi_j(x)\psi_j(y),$$

where the functions

$$\psi_j(x) = \frac{1}{\sqrt{h_j}}\, p_j(x)\, e^{-x^2/4}$$

are built from the monic Hermite polynomials $p_j(x)$ with normalization constants h_j ensuring that the ψ_j's form an orthonormal system in $L^2(\mathbb{R})$.

Using the generating function

$$\exp\Big(xt - \frac{t^2}{2}\Big) = \sum_{n\ge 0} p_n(x)\frac{t^n}{n!},$$

one obtains by Cauchy's integral formula

$$p_n(x) = \frac{n!}{2\pi i}\oint_C \frac{\exp\Big(xt - \frac{t^2}{2}\Big)}{t^{n+1}}\, dt,$$

which leads to

$$\psi_n(x) = \frac{e^{-x^2/4}}{\sqrt{h_n}} \frac{n!}{2\pi i} \oint_C \frac{\exp\Big(xt - \frac{t^2}{2}\Big)}{t^{n+1}}\, dt.$$

Starting from the Fourier transform identity

$$\int_{-\infty}^{\infty} \exp\Big(-\frac{t^2}{2} + i\, t\, x\Big)\, dt = \sqrt{2\pi}\, e^{-x^2/2},$$

and differentiating with respect to x, then changing variables, one obtains

$$\psi_n(x) = \frac{i\, e^{x^2/4}}{\sqrt{2\pi\, h_n}} \int_{-i\infty}^{i\infty} s^n\, e^{s^2/2 - s\, x}\, ds.$$

By inserting the above representations for $\psi_n(x)$ into the kernel sum, one arrives at the double contour integral formula (after conjugation and the trick with removing $1/(s-t)$):

$$K_n(x,y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds\, \frac{\exp\Big\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\Big\}}{s-t} \left(\frac{s}{t}\right)^n.$$

The integration contour C is a small contour around 0, and s is passing to the right of C.

This representation is especially useful for performing asymptotic analysis (for example, via the steepest descent method) and for deriving results such as the semicircle law.

7.1.2 Scaling

Let us now consider the GUE kernel,

$$K_n(x,y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds\, \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n.$$

We know from the Wigner semicircle law (established for real symmetric matrices with general iid entries in in Chapter 2, and for the GUE in Chapter 4) that the eigenvalues live on the scare $\sqrt{n}$. This means that to capture the local asymptotics, we need to scale

$$x = X\sqrt{n} + \frac{\Delta x}{\sqrt{n}}, \qquad y = Y\sqrt{n} + \frac{\Delta y}{\sqrt{n}}, \qquad \Delta x, \Delta y \in \mathbb{R}. \tag{7.1}$$

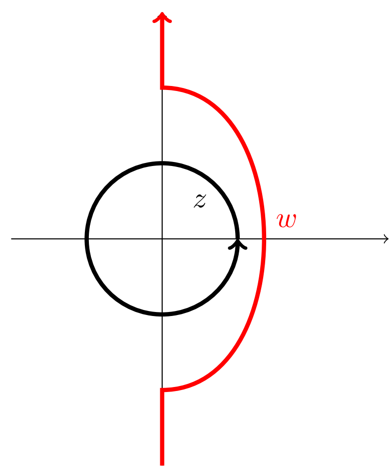

Figure 7.1: Integration contours for the GUE kernel.

Moreover, if $X \neq Y$ (i.e., different global positions), one can check that the kernel vanishes. In other words, the local behaviors at different global positions are independent. In what follows, we take $Y = X$.

Let us also make a change of the integration variables:

$$t = z\sqrt{n}, \qquad s = w\sqrt{n}.$$

The integration contours for z and w look the same as for t and s, up to a rescaling (Figure 7.1). However, as 0 and $t = s$ are the only singularities in the integrand, we can deform the z, w contours as we wish, while keeping $|z| < |w|$ and the general shape as in Figure 7.1.

We thus have:

$$\begin{aligned} &K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n}) \\ &= \frac{\sqrt{n}}{(2\pi)^2} \oint_C dz \int_{-i\infty}^{i\infty} dw \, \frac{\exp\left\{n\left(\log w - \log z + \frac{w^2}{2} - \frac{z^2}{2} + X(z-w) + \frac{z\Delta x - w\Delta y}{n}\right)\right\}}{w - z}. \end{aligned} \tag{7.2}$$

Remark 7.1. The logarithms in the exponent are harmless, since for the estimates we only need the real parts of the logarithms, and for the main contributions, we will have $z \approx w$, so any phases of the logarithms would cancel.

The asymptotic analysis of double contour integrals like (7.2) in the context of determinantal point processes was pioneered in [Oko02, Section 3].

7.1.3 Critical points

Let us define

$$S(z) := \frac{z^2}{2} + \log z - Xz.$$

Then the exponent contains $n\,(S(w) - S(z))$. According to the steepest descent ideology, we should deform the integration contours to pass through the critical point(s) z_{cr} of $S(z)$. Moreover, the new w contour should maximize the real part of $S(z)$ at z_{cr}, and the new z contour should minimize it. If $S''(z_{cr}) \neq 0$, it is possible to locally choose such contours, they will be perpendicular to each other at z_{cr}.

Thus, we need to find the critical points of $S(z)$. They are found from the quadratic equation:

$$S'(z) = z + \frac{1}{z} - X = 0, \qquad z_{cr} = \frac{X \pm \sqrt{X^2 - 4}}{2}. \tag{7.3}$$

Depending on whether $|X| < 2$, there are three cases. Unless $|X| = 2$ (when equation (7.3) has a single root), we have $S''(z_{cr}) \neq 0$. We will consider the three cases in Section 7.1.4 and **????** below.

7.1.4 Imaginary critical points: $|X| < 2$, "bulk"

When $|X| < 2$, the critical points are complex conjugate. Denote them by z_{cr} and $\overline{z_{cr}}$. Since $S(z)$ has real coefficients, we have

$$\operatorname{Re} S(z_{cr}) = \operatorname{Re} S(\overline{z_{cr}}).$$

Thus, we need to consider the contribution from both points. For simplicity of the computations, let us consider only the case $X = 0$. See Problem 7.5.1. We have

$$z_{cr} = i, \qquad S''(z_{cr}) = 2.$$

The behavior of $\operatorname{Re} S(z)$ on the complex plane can be illustrated by a 3D plot or by a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ has constant sign. See Figure 7.2 for an illustration in the case $X = \frac{1}{2}$. (We take $X \neq 0$ to break symmetry, for a better intuition.)

From the region plot, we see that the new z contour should pass through the shaded region $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr}) > 0$, and the new w contour should pass through the unshaded region $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr}) < 0$.

Deforming the contours from Figure 7.1 to the new contours is impossible without passing through the residue at $w = z$. Moreover, this residue

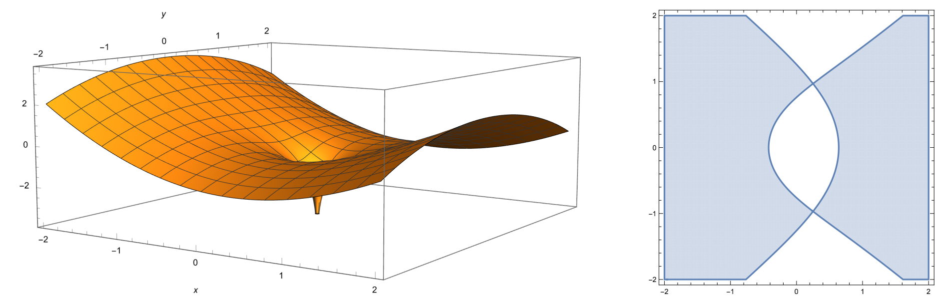

Figure 7.2: A 3D plot and a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ is positive (highlighted) or negative, in the case $X = \frac{1}{2}$. In this case, $z_{cr} \approx 0.25 + 0.96i$.

appears only for certain values of z. Namely, for $X = 0$, let us first make the z contour to be the positively (counterclockwise) oriented unit circle. It passes through the critical points $z_{cr} = i$ and $\overline{z_{cr}} = -i$. Since the original w contour is to the right of the z contour, we only encounter the residue when z is in the right half of the circle.

Thus, we can write

$$\oiint_{\text{old contours}} = \oiint_{\text{new contours}} + \int_{-i}^{i} 2\pi i \operatorname{Res}_{w=z} dz, \tag{7.4}$$

where in the single integral, the z contour passes to the right of the origin, along the right half of the unit circle.

It remains to consider the two integrals in the right-hand side of (7.4). Recall that the correlation functions are defined relative to a reference measure, and the right object to scale is

$$K_n(x,y)dy = \frac{1}{\sqrt{n}} d\left(\Delta y\right).$$

The extra factor $n^{-1/2}$ compensates the prefactor $\sqrt{n}$ in (7.2).

The single integral takes the form

$$\frac{-i}{2\pi} \int_{-i}^{i} e^{z(\Delta x - \Delta y)} dz = \frac{\sin\left(\Delta x - \Delta y\right)}{\pi(\Delta x - \Delta y)}, \qquad \Delta x, \Delta y \in \mathbb{R}. \tag{7.5}$$

Definition 7.2. The *sine kernel* is defined as

$$K_{\text{sine}}(x,y) := \begin{cases} \dfrac{\sin(x-y)}{\pi(x-y)}, & x \neq 0, \\ \dfrac{1}{\pi}, & x = 0. \end{cases}$$

(The value at $x = y$ is defined by continuity.)

This kernel is translation invariant, and is often defined with a single argument, as $K_{\text{sine}}(x - y)$.

The double integral has both contours in the "steepest descent" regime, which means that the main contribution is

$$\text{const} \cdot \frac{e^{n(\operatorname{Re} S(z_{cr}) - \operatorname{Re} S(z_{cr}))}}{\sqrt{n}} \sim \frac{\text{const}}{\sqrt{n}}.$$

At this rate, the double integral over the new contours *does not* contribute to the asymptotics of the correlation functions. Recall that the correlation functions are expressed as finite-dimensional determinants of the kernel $K_n(x, y)$, and the error $O(n^{-1/2})$ is negligible in the limit $n \to +\infty$. This is because the main term comes from the single integral, which does not vanish.

We have established the following result:

Proposition 7.3 (Bulk asymptotics at $X = 0$)**.** *The correlation kernel K_n of the GUE has the following asymptotics close to zero as $n \to +\infty$:*

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}} K_n\left(\frac{\Delta x}{\sqrt{n}}, \frac{\Delta y}{\sqrt{n}}\right) = K_{\text{sine}}(\Delta x, \Delta y), \qquad \Delta x, \Delta y \in \mathbb{R}.$$

Consequently, the eigenvalues of the GUE converge to the sine process determined by the sine kernel (Definition 7.2), in the sense of finite-dimensional distributions.

Remark 7.4. Beyond $X = 0$, the local correlations are essentially the same, up to rescaling of the real line by a constant factor (depending on the semicircle density). See Problem 7.5.1.

7.1.5 Real critical points: $|X| > 2$, "large deviations"

For $X^2 > 4$, both solutions (7.3) are real. Let us assume $X > 2$, the case $X < 2$ is similar. For $X > 2$, both solutions are positive. Label these solutions as

$$z_+ = \frac{X + \sqrt{X^2 - 4}}{2}, \qquad z_- = \frac{X - \sqrt{X^2 - 4}}{2}, \quad \text{so that} \quad z_+ z_- = 1.$$

A straightforward check reveals that $z_+ > 1$ and $z_- < 1$ (for $X > 2$). Note that $S''(z) = 1 - z^{-2}$, which is positive for $z_+ > 1$ and negative for $z_- <$

1. Thus, the critical points z_+ and z_- are a local minimum and a local maximum. A crucial observation is that

$$S(z_+) < S(z_-).$$

One can deform the z integration contour to pass through z_- and the w contour to pass through z_+. Then, on these contours, one can show that

$$\operatorname{Re} S(w) - \operatorname{Re} S(z) < 0.$$

According to the steepest descent ideology, we see that the main exponential behavior of the double contour integral is

$$\exp\left\{n\left(\operatorname{Re} S(z_+) - \operatorname{Re} S(z_-)\right)\right\} = O(e^{-\delta(X)n}), \qquad |X| > 2. \tag{7.6}$$

Here $\delta(X) > 0$ for $|X| > 2$, and $\delta(X) \to 0$ when $|X| \to 2$.

The outcome (7.6) reflects the fact that the Wigner semicircle law places all eigenvalues inside the interval $|X| \leq 2$. The probability to see even a single eigenvalue outside $[-2, 2]$ is exponentially small.

This exponential decay corresponds to a large deviation regime. Indeed, if at least one of the diagonal entries of the matrix is unusually large, this corresponds to the maximal eigenvalue to get outside the interval $[-2, 2]$. See also Problem 7.5.2.

7.1.6 Double critical point: $|X| = 2$, "edge"

Throughout the subsection, we assume that $X = 2$. The case $X = -2$ is symmetric.

When $X = 2$, the two solutions in (7.3) merge into a double critical point $z_{cr} = 1$. We have

$$S'(1) = 0, \qquad S''(1) = 0, \qquad S'''(1) = 2.$$

Thus, the usual quadratic approximation fails and one must expand to third order. Writing

$$z = 1 + u, \qquad w = 1 + v,$$

with u, v small, we have

$$S(1+u) = S(1) + \frac{S'''(1)}{6}\,u^3 + O(u^4) = S(1) + \frac{u^3}{3} + O(u^4),$$

and similarly for $S(1+v)$. Hence, the difference in the exponents becomes

$$S(1+v) - S(1+u) = \frac{v^3 - u^3}{3} + O(u^4 + v^4).$$

To capture the correct asymptotics, we rescale the local variables by setting

$$u = \frac{U}{n^{1/3}}, \qquad v = \frac{V}{n^{1/3}},$$

so that

$$n\Big[S(1+v) - S(1+u)\Big] = \frac{V^3 - U^3}{3} + O\Big(n^{-1/3}\Big).$$

Moreover, the correct edge scaling for the spatial variables is obtained by writing

$$x = 2\sqrt{n} + \frac{\xi}{n^{1/6}}, \qquad y = 2\sqrt{n} + \frac{\eta}{n^{1/6}}, \qquad \xi, \eta \in \mathbb{R}.$$

We have

$$n\left(S(w) - S(z)\right) = n^{1/3}(\xi - \eta) + \frac{V^3 - U^3}{3} + \xi U - \eta V + O\Big(n^{-1/3}\Big).$$

The terms $n^{1/3}(\xi - \eta)$ are harmless as they can be removed by conjugation.

The region plot of $\operatorname{Re} S(z) - \operatorname{Re} S(1)$ (shown in Figure 7.3) makes sure that we can deform the z contour so that it passes through $z_{cr} = 1$ as the new U contour at the angles $\pm\frac{2\pi}{3}$ (where $\operatorname{Re} U^3 > 0$), we can deform the w contour so that it passes through $z_{cr} = 1$ as the new V contour at the angles $\pm\frac{\pi}{3}$ (where $\operatorname{Re} V^3 < 0$). This will ensure the convergence of the new double integral.

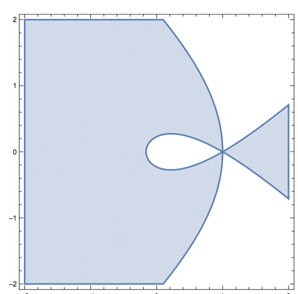

Figure 7.3: The plot of the region $\operatorname{Re} S(z) - \operatorname{Re} S(1) > 0$ for $X = 2$.

Thus, we have shown that under the rescaling, the GUE correlation kernel $K_n(x,y)\,dy$ converges to a new kernel.

Definition 7.5. Define the *Airy kernel* on $\mathbb{R}$ by

$$K_{\mathrm{Ai}}(\xi, \eta) = \frac{1}{(2\pi i)^2} \int_{e^{-\frac{\pi i}{3}}\infty}^{e^{\frac{\pi i}{3}}\infty} dV \int_{e^{-\frac{2\pi i}{3}}\infty}^{e^{2\frac{\pi i}{3}}\infty} dU\, \frac{\exp\Big\{\frac{V^3 - U^3}{3} + U\,\xi - V\,\eta\Big\}}{V - U}.$$

For another formula for the Airy kernel which does not involve integrals, see Problem 7.5.3.

Proposition 7.6. *We have*

$$\lim_{n\to\infty} \frac{1}{n^{1/6}} K_n\Big(2\sqrt{n} + \frac{\xi}{n^{1/6}},\, 2\sqrt{n} + \frac{\eta}{n^{1/6}}\Big) \to K_{\mathrm{Ai}}(\xi, \eta).$$

Consequently, the eigenvalue statistics at the edge of the spectrum converge to the Airy point process, in the sense of fine-dimensional distributions.

7.1.7 Airy kernel, Tracy–Widom distribution, and convergence of the maximal eigenvalue

Let us make a few remarks on the asymptotic results of Proposition 7.3 and **??**. First, a rigorous justification of convergence of contour integrals requires some estimates on the error terms in the steepest descent analysis, but these estimates are mild and not hard to obtain.

Second, the GUE has the maximal eigenvalue λ_{max}. It is reasonable to assume that the Airy process also (almost surely) admits a maximal point (usually denoted by $\mathfrak{a}_1$), and that λ_{max} converges to $\mathfrak{a}_1$ under appropriate rescaling:

$$\lim_{n\to\infty} n^{\frac{1}{6}} \left(\lambda_{max} - 2\sqrt{n}\right) = \mathfrak{a}_1. \tag{7.7}$$

This is indeed the case, but to show (7.7), one needs to show the convergence in distribution:

$$\lim_{n\to\infty} \mathbb{P}\Big(n^{1/6}(\lambda_{max} - 2\sqrt{n}) \le x\Big) \to \mathbb{P}(\mathfrak{a}_1 \le x). \tag{7.8}$$

Both events (7.8) are so-called *gap probabilities*, for example,

$$\mathrm{P}(\mathfrak{a}_1 \le x) = \mathrm{P}(\text{there are no eigenvalues in the interval } (x, \infty)),$$

which is expressed as the Fredholm determinant

$$\det\left(1 - K_{\mathrm{Ai}}\right)_{(x,\infty)} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \int_x^\infty dy_1 \int_x^\infty dy_2 \cdots \int_x^\infty dy_m \det_{i,j=1}^{m} K_{\mathrm{Ai}}(y_i, y_j). \tag{7.9}$$

Thus, to get (7.8)), one needs to show the convergence of sums like this for the GUE kernel to the corresponding sums for the Airy kernel. This is doable, but tedious.

Moreover, to get convergence in distribution of random variables, one would also have to argue either *tightness*, or independently show that (7.9) defines a cumulative probability distribution function in x:

$$F_2(x) = \det\left(1 - K_{\mathrm{Ai}}\right)_{(x,\infty)}. \tag{7.10}$$

The distribution (7.10) is known as the *GUE Tracy–Widom distribution*. The subscript 2 indicates that $\beta = 2$. There are distributions F_β for all beta, most notably, the GOE and GSE distributions. The classical distributions F_1, F_2, F_4 also appear as fluctuation distributions in interacting particle systems, while other beta values do not quite appear in the particle systems domain.

More details may be found in the original papers [TW93], [For93], [TW94].

7.1.8 Remark: what happens for general β?

- The determinantal structure exploited above is special to the $\beta = 2$ case. In contrast, for $\beta = 1$ (GOE) and $\beta = 4$ (GSE) the eigenvalue correlations are expressed in terms of *Pfaffians* rather than determinants. This happens before and after the scaling limit.

- Earlier attempts to extend the $\beta = 2$ techniques were determinantal. For example, one can replace the squared Vandermonde $\prod_{i<j}(x_i - x_j)^2$ with
$$\prod_{i<j}(x_i - x_j)(x_i^{\beta/2} - x_j^{\beta/2}).$$
This is known as the *Muttalib–Borodin* ensemble [FW17], and the kernel can be computed in a similar way using (bi)orthogonalization.

- Local eigenvalue statistics of general β-ensembles converge to the so-called *general β sine process* and *general β Airy process* in the bulk and at the edge, respectively. Detailed analyses of this convergence can be found in [RRV11], [VV09], [GS18], and the literature referenced in the recent work [GXZ24].

7.2 Cutting corners: setup

We begin a new topic, which will be the main focus for this and the next week.

In random matrix theory, one often studies the entire spectrum of an $n \times n$ matrix ensemble such as the Gaussian Unitary Ensemble (GUE), the

Gaussian Orthogonal Ensemble (GOE), or, more generally, β-ensembles. However, it is also natural to examine the spectra of *principal minors* of such matrices.

When we say "cutting corners," we typically refer to extracting a top-left $k \times k$ submatrix (or *corner*) out of an $n \times n$ random matrix H and then looking at the interplay among the eigenvalues of all corners $k = 1, \ldots, n$. This forms a *nested* family of spectra, often described by interlacing (or Gelfand–Tsetlin) patterns.

The *GUE corners process* is a classical example of this phenomenon. If H is an $n \times n$ GUE matrix, then the top-left $k \times k$ corners (for $1 \le k \le n$) have jointly distributed eigenvalues that exhibit a determinantal structure. We will employ the technique of *polynomial* (*characteristic function*) *equation* and then *loop equations* to study global limits (note that they are not suitable to get local limits like sine and Airy processes).

So far, we have the following access to eigenvalues and corners:

1. For $\beta = 1, 2, 4$, we have the actual matrices, and can cut the corners in the usual way.

2. For general β, we have the joint eigenvalue distribution with the interaction term $\prod_{i<j} |x_i - x_j|^\beta$, which is an interpolation.

3. For general β, we also have the Dumitriu–Edelman tridiagonal model [DE02].

Cutting corners from the tridiagonal matrix is not a good idea, for many reasons. The simplest might be that the $(n-1)\times(n-1)$ corner eigenvalues do not have the same distribution (up to changing n) as the general β ensemble eigenvalues. Maybe we might cut the lower right corners? Well, this is not a good idea either, because the total number of random variables (the "noise") in the tridiagonal matrix is $O(n)$, while the number of eigenvalues of all corners is $O(n^2)$.

7.3 Corners of Hermitian matrices

7.3.1 Principal corners

Let H be an $n \times n$ Hermitian matrix. For each $1 \le k \le n$, define the *top-left $k \times k$ corner* $H^{(k)}$ by

$$H^{(k)} \;=\; \big[H_{ij}\big]_{1 \le i,j \le k}.$$

Since H is Hermitian, each $H^{(k)}$ is also Hermitian. Let

$$\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \cdots \geq \lambda_k^{(k)}$$

denote the eigenvalues of $H^{(k)}$. Then the collection

$$\left\{\lambda_j^{(k)} : 1 \leq j \leq k \leq n\right\}$$

is called the *corners spectrum* (or *minor spectrum*) of H. When H is random, this triangular array of eigenvalues becomes a random point configuration in the two-dimensional set $\{1, \ldots, n\} \times \mathbb{R}$.

7.3.2 Interlacing

A fundamental feature of Hermitian matrices is that the eigenvalues of corners interlace with the eigenvalues of the full matrix:

Proposition 7.7. *If $\nu_1 \geq \cdots \geq \nu_n$ are the eigenvalues of H itself (i.e., the full $n \times n$ matrix), and $\mu_1 \geq \cdots \geq \mu_{n-1}$ are the eigenvalues of $H^{(n-1)}$, then we have:*

$$\nu_1 \geq \mu_1 \geq \nu_2 \geq \mu_2 \geq \ldots \geq \mu_{n-1} \geq \nu_n.$$

Proof. One can prove the statement using the Courant–Fischer (min–max) characterization of eigenvalues, often referred to as the variational principle. Recall that for an $n \times n$ Hermitian matrix H with ordered eigenvalues $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_n$, the j-th largest eigenvalue ν_j admits the variational characterization

$$\nu_j = \max_{\substack{V \subset \mathbb{F}^n \\ \dim(V)=j}} \min_{\substack{x \in V \\ x \neq 0}} \frac{x^* H\, x}{x^* x} = \min_{\substack{W \subset \mathbb{F}^n \\ \dim(W)=n-j+1}} \max_{\substack{x \in W \\ x \neq 0}} \frac{x^* H\, x}{x^* x},$$

where $\mathbb{F}$ is $\mathbb{R}$, $\mathbb{C}$, or the quaternions (depending on $\beta = 1, 2, 4$, respectively). We leave this as Problem 7.5.4. □

The same interlacing property holds for real symmetric matrices ($\beta = 1$), and in the case $\beta = 4$. Therefore, it is natural to require this property for all β-ensembles.

7.3.3 Orbital measure

It is natural to consider an extended setup, and take the matrix H to not just be GUE, but instead fix its eigenvalues. Let

$$H = U \Lambda U^{\dagger}, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where Λ is fixed and $U \in U(n)$ is Haar (uniformly) distributed. Denote the set of all such H by $\mathrm{Orbit}(\lambda)$, $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, $\lambda_1 \geq \cdots \geq \lambda_n$.

Then, if we understand the distribution structure of all corners of a random $H \in \mathrm{Orbit}(\lambda)$, we can then "average over" the GUE eigenvalue ensemble distribution of λ to get the GUE corners process.

Remark 7.8. The setting with orbits presents a bridge into "asymptotic representation theory". Namely, as $n \to \infty$, how does the corners distribution look like? We may ask for a characterization of *all the ways* how $\lambda^{(n)} = (\lambda_1^{(n)} \geq \ldots \lambda_n^{(n)})$ goes to infinity, in such a way that the corners spectrum converges on all levels $k = 1, \ldots, K$ for arbitrary K (independent of n). This problem was solved in [OV96]. More direct formulas for projections of orbital measures were obtained in [Ols13].

7.4 Polynomial equation and joint distribution

7.4.1 Derivation

Fix $\lambda = (\lambda_1 \geq \ldots \geq \lambda_n)$. Let $H \in \mathrm{Orbit}(\lambda)$ be a random matrix (in the case $\beta = 2$, but the proof works for $\beta = 1, 4$ as well). Let $\mu_1, \ldots, \mu_{n-1}$ be the eigenvalues of the $(n-1) \times (n-1)$ corner $H^{(n-1)}$.

Lemma 7.9. *The distribution of $\mu_1, \ldots, \mu_{n-1}$ is the same as the distribution of the roots of the polynomial equation*

$$\sum_{i=1}^{n} \frac{\xi_i}{z - \lambda_i} = 0, \tag{7.11}$$

where ξ_i are i.i.d. random variables with the distribution χ^2_β.

Proof. $\mu_1, \ldots, \mu_{n-1}$ are the roots of the following equation with the determinant of order $n+1$:

$$\det \begin{pmatrix} U \operatorname{diag}(\lambda) U^\dagger - z I_N & v^\top \\ v & 0 \end{pmatrix} = 0, \qquad v = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Indeed, expanding the determinant along the last row, we get the $(n-1)$th determinant, which corresponds to cutting the corner.

Next, multiply the determinant by $\begin{pmatrix} U^\dagger & 0 \\ 0 & 1 \end{pmatrix}$ on the left and $\begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix}$ on the right:

$$\det \begin{pmatrix} \mathrm{diag}(\lambda) - zI_N & u^\dagger \\ u & 0 \end{pmatrix} = 0,$$

where $u^\dagger = U^\dagger v^\top$ is the last row of $U^\dagger$. The determinant now can be expressed as

$$\det = -\prod_{i=1}^{n} (\lambda_i - z) \sum_{i=1}^{n} \frac{|u_i|^2}{\lambda_i - z}.$$

Since u is a row of a Haar unitary matrix, it is distributed uniformly on the unit sphere in $\mathbb{C}^n$. However, we can identify it with a normalized vector from a rotationally invariant measure on $\mathbb{C}^n$, the best of which is Gaussian. This completes the proof. $\square$

Remark 7.10. Lemma 7.9 provides another proof of the eigenvalue interlacing property. Indeed, assume that all ξ_i are rational. Then equation (7.11) is essentially $P'(z) = 0$, where $P(z)$ is a product of powers of the $(z - \lambda_i)$'s (the powers depend on the ξ_i's). As the roots of the derivative of a polynomial interlace with the roots of the polynomial, we get the interlacing property.

7.4.2 Inductive nature of the transition

Note that when we fix $\lambda = (\lambda_1 \geq \ldots \geq \lambda_n)$ and get random $\mu = (\mu_1 \geq \ldots \geq \mu_{n-1})$ by solving (7.11), we can then fix μ and get random $\nu = (\nu_1 \geq \ldots \geq \nu_{n-2})$, and so on. Here, ν corresponds to the $(n-2) \times (n-2)$ corner of H. Indeed, we can condition on μ, and conjugate H again by a unitary matrix of the form $U = \begin{pmatrix} U' & 0 \\ 0 & 1 \end{pmatrix}$, where $U' \in U(n-1)$ is Haar distributed. Since $U \in U(n)$, this extra conjugation does not change the distribution of $H \in \mathrm{Orbit}(\lambda)$, but it allows us to treat the passage from μ to ν on the same grounds as the passage from λ to μ.

Remark 7.11. In more detail, since the homogeneous space $U(n)/U(n-1)$ can be identified with S^{2n-1}, the $(2n-1)$-dimensional real sphere, we can construct a Haar-distributed unitary matrix $U \in U(n)$ by first picking a Haar-distributed unitary matrix $U' \in U(n-1)$, and then picking a random point on the sphere S^{2n-1}. Restricting H to $\mathbb{C}^{n-1}$ fixes the last component on the sphere (up to a complex phase), but the eigenbasis of the restriction $H^{(n-1)}$ is still Haar distributed, but now in $U(n-1)$.

This implies that in order to understand the full corners process, it is enough to understand the transition from λ to μ, where λ is fixed, and μ is obtained by solving (7.11).

7.4.3 Case $\beta = \infty$

In the limit $\beta \to +\infty$, the χ^2_β distribution obeys the law of large numbers:

$$\frac{\chi^2_\beta}{\beta} \to 1, \qquad \beta \to +\infty.$$

Thus, the equation (7.11) becomes deterministic:

$$\sum_{i=1}^{n} \frac{1}{z - \lambda_i} = 0.$$

Denote

$$P(z) = \prod_{i=1}^{n} (z - \lambda_i). \tag{7.12}$$

Then

Proposition 7.12. *The passage from $\lambda = (\lambda_1 \geq \ldots \geq \lambda_n)$ to $\mu = (\mu_1 \geq \ldots \geq \mu_{n-1})$ in the limit as $\beta = \infty$ is deterministic, and it the same as the passage from the roots of the polynomial $P(z)$ (7.12) to the roots of its derivative $P'(z)$.*

7.5 Problems

7.5.1 General bulk case

Perform the asymptotic analysis of the correlation kernel as in Section 7.1.4, but in the general case $-2 < X < 2$.

7.5.2 Large deviations

Let W_n be an $n \times n$ Wigner real or Hermitian matrix with finite variance entries. Assume that the matrix is normalized so that the variance of each diagonal entry is 1.

Assumption [BBP05]**.** *If a Wigner matrix is normalized to have diagonal variance 1, then a rank 1 perturbation of magnitude $c > 0$ is sufficient to* *shoot the maximum eigenvalue outside the support of the Wigner semicircle law. (For a simulation of this phenomenon, see here.)*

Consider the following large deviation event. For a fixed $\eta > 0$, let

$$E_{n,\eta} := \Big\{ \exists\, i \in \{1, \ldots, n\} \text{ such that } W_{ii} \geq \eta \Big\}.$$

Under the above assumption, if for some i the diagonal entry W_{ii} is unusually large, it will push the maximal eigenvalue of W_n outside the bulk.

1. Assuming that the entries are Gaussian, *lower bound* the probability of the event $E_{n,\eta}$ for large n.

2. Assuming another tail behavior of the diagonal entries (exponential or power-law tails), use the limit theorems for maxima of independent random variables to generalize the *lower bound* of $\mathbb{P}(E_{n,\eta})$.

7.5.3 Airy kernel

Define the Airy function by

$$Ai(\xi) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iU^3/3 + i\xi U} dU = \frac{1}{\pi} \int_0^{\infty} \cos\left(\frac{U^3}{3} + \xi U \right) dU.$$

This integral converges, but only conditionally. To improve convergence, one should instead integrate along a complex contour, from $e^{\frac{5\pi i}{6}}\infty$ to 0 to $e^{\frac{\pi i}{6}}\infty$.

Show that

$$K_{\mathrm{Ai}}(\xi, \eta) = \frac{Ai(\xi)\, Ai'(\eta) - Ai(\eta)\, Ai'(\xi)}{\xi - \eta}.$$

Note that this expression is parallel to the sine kernel,

$$\frac{\sin(x-y)}{\pi(x-y)} = \frac{\sin x \cos y - \cos x \sin y}{\pi(x-y)}, \qquad \cos x = (\sin x)'.$$

These correlation kernels are called *integrable* [IIKS90].

Hint for the problem: observe that

$$\exp\{-izx + iwy\} = \frac{i}{x-y} \left(\frac{\partial}{\partial z} + \frac{\partial}{\partial w} \right) \exp\{-izx + iwy\},$$

and use integration by parts in $K_{\mathrm{Ai}}(\xi, \eta)$ from Definition 7.5.

7.5.4 Interlacing proof

Finish the proof of Proposition 7.7.

Chapter 8

Cutting corners and loop equations

8.1 Cutting corners: polynomial equation and distribution

8.1.1 Recap: polynomial equation

Recall the polynomial equation we proved in the last Chapter 7. Fix $\lambda = (\lambda_1 \geq \ldots \geq \lambda_n)$. Let $H \in \mathrm{Orbit}(\lambda)$ be a random Hermitian matrix defined as

$$H = U \,\mathrm{diag}(\lambda_1, \ldots, \lambda_n) U^\dagger,$$

where U is Haar-distributed unitary matrix from $U(n)$. This is the case $\beta = 2$, but the statement holds for the cases $\beta = 1, 4$ with appropriate modifications. Let $\mu_1, \ldots, \mu_{n-1}$ be the eigenvalues of the $(n-1) \times (n-1)$ corner $H^{(n-1)}$.

Lemma 8.1. *The distribution of $\mu_1, \ldots, \mu_{n-1}$ is the same as the distribution of the roots of the polynomial equation*

$$\sum_{i=1}^{n} \frac{\xi_i}{z - \lambda_i} = 0, \tag{8.1}$$

where ξ_i are i.i.d. random variables with the distribution χ^2_β.

Recall also that this passage from λ to μ works inductively, and the distribution of the next level eigenvalues $\nu = (\nu_1 \geq \ldots \geq \nu_{n-2})$ is given by the same polynomial equation, but with λ replaced by μ. In this way, we can define a *Markov map* from λ to μ, which is then iterated to construct the full array of eigenvalues of the corners of H.

For $\beta = \infty$, this map is deterministic, and is equivalent to successive differentiating the characteristic polynomial of H.

8.1.2 Extension to general β

We extend the polynomial equation to general β, by *declaring* (defining) that the general β corners distribution is powered by the passage from $\lambda = (\lambda_1 \geq \ldots \geq \lambda_n)$ to $\mu = (\mu_1 \geq \ldots \geq \mu_{n-1})$, where μ solves (8.1) with ξ_i i.i.d. χ^2_β. In this way, μ interlaces with λ. For $\beta = 1, 2, 4$, this definition reduces to the one with invariant ensembles with fixed eigenvalues λ.

8.1.3 Distribution of the eigenvalues of the corners

Let μ be obtained from λ by the general β corners operation.

Theorem 8.2. *The density of μ with respect to the Lebesgue measure is given by*

$$\frac{\Gamma(N\beta/2)}{\Gamma(\beta/2)^n} \prod_{1\leq i<j\leq n-1} (\mu_i - \mu_j) \prod_{i=1}^{n-1} \prod_{j=1}^{n} |\mu_i - \lambda_j|^{\beta/2-1} \prod_{1\leq i<j\leq n} (\lambda_i - \lambda_j)^{1-\beta}.$$

Proof. Let $\varphi_i = \xi_i / \sum_{j=1}^n \xi_j$. It is well-known[1] the joint density of $(\varphi_1, \ldots, \varphi_n)$ is the (symmetric) Dirichlet density

$$\frac{\Gamma(N\beta/2)}{\Gamma(\beta/2)^n} w_1^{\beta/2-1} \ldots w_n^{\beta/2-1} dw_1 \ldots dw_{n-1}$$

(note that the density is $(n-1)$-dimensional).

We need to compute the Jacobian of the transformation from φ to μ, if we write

$$\sum_{i=1}^n \frac{\varphi_i}{z - \lambda_i} = \frac{\prod_{i=1}^{n-1}(z - \mu_i)}{\prod_{i=1}^n (z - \lambda_i)},$$

and compute (as a decomposition into partial fractions):

$$\varphi_a = \frac{\prod_{i=1}^{n-1}(\lambda_a - \mu_i)}{\prod_{i\neq a}(\lambda_a - \lambda_i)}.$$

[1] See Problem 8.4.3.

Therefore,

$$\frac{\partial \varphi_a}{\partial \mu_b} = \frac{\prod_{i=1}^{n-1}(\lambda_a - \mu_i)}{\prod_{i\neq a}(\lambda_a - \lambda_i)} \frac{1}{\mu_b - \lambda_a}, \qquad a = 1, \ldots, n, \quad b = 1, \ldots, n-1. \tag{8.2}$$

The Jacobian is essentially the determinant of the matrix $1/(\mu_b - \lambda_a)$, which is the Cauchy determinant (Problems 8.4.1 and 8.4.2). The final density is obtained from the symmetric Dirichlet density, but we plug in $w = \varphi$, and also multiply by the inverse of the Jacobian determinant (8.2). After the necessary simplifications, this completes the proof. ☐

Corollary 8.3 (Joint density of the corners)**.** *The eigenvalues $\lambda^(k)_j$, $1 \le j \le k \le n$, of a random matrix from* Orbit(λ) *form an interlacing array, with the joint density*

$$\propto \prod_{k=1}^{n} \prod_{1\le i<j\le k} \left(\lambda_j^{(k)} - \lambda_i^{(k)}\right)^{2-\beta} \prod_{a=1}^{k+1}\prod_{b=1}^{k} \left|\lambda_a^{(k+1)} - \lambda_b^{(k)}\right|^{\beta/2-1}.$$

For $\beta = 2$, all factors disappear, and we get the uniform distribution on the interlacing array. This is the *uniform Gibbs property* which is important for other models, including discrete ensembles.

8.2 Loop equations

Let us write down the *loop equations* for the passage from the eigenvalues λ to the eigenvalues μ. These loop equations are due to [GH24] by a limit from a discrete system (related to Jack symmetric polynomials). Note that despite the name, these are not **equations**, but rather a statement that some expectations are holomorphic. We stick to the random matrix setting, and present a formulation and a proof given by [Gor25].

8.2.1 Formulation

Theorem 8.4. *We fix $n = 1, 2, \ldots$ and $n+1$ real numbers $\lambda_1 \ge \cdots \ge \lambda_{n+1}$. For $\beta > 0$, consider $n+1$ i.i.d. χ^2_β random variables ξ_i and set*

$$w_i = \frac{\xi_i}{\sum_{j=1}^{n+1}\xi_j}, \qquad 1 \le i \le n+1.$$

We define n random points $\{\mu_1, \ldots, \mu_n\}$ as n solutions to the equation

$$\sum_{i=1}^{n+1} \frac{w_i}{z - \lambda_i} = 0. \tag{8.3}$$

Take any polynomial $W(z)$ *and consider the complex function:*

$$f_W(z) = \mathbb{E}\left[\prod_{j=1}^{n} \exp\bigl(W(\mu_j)\bigr) \frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)} \left(W'(z) + \sum_{i=1}^{n+1} \frac{\beta/2-1}{z-\lambda_i} + \sum_{j=1}^{n} \frac{1}{z-\mu_j}\right)\right]. \tag{8.4}$$

Then $f_W(z)$ *is an* entire function *of* z*, in the following sense:*

- *For* $z \in \mathbb{C} \setminus [\lambda_{n+1}, \lambda_1]$*, the expectation in* (8.4) *defines a holomorphic function of* z*.*
- *This function has an analytic continuation to* $\mathbb{C}$*, which has no singularities.*

Remark 8.5. Note that for z in $[\lambda_{n+1}, \lambda_1]$, the integral determining (8.4) might be divergent, and, therefore, analytic continuation is the proper way to define $f_W(z)$, $z \in [\lambda_{n+1}, \lambda_1]$.

Corollary 8.6. *We have*

$$f_0(z) = \frac{(n+1)\beta}{2} - 1.$$

Here f_0 *means* f_W *with* $W \equiv 0$*.*

Proof. This is obtained by sending $z \to \infty$ in (8.4). □

8.2.2 Proof of Theorem 8.4 for $\beta > 2$

Theorem 8.4 remains valid for $\beta > 0$, but we only prove it for $\beta > 2$ here. We also assume that $\lambda_1 > \ldots > \lambda_n$.

We begin by observing that for $z \in \mathbb{C} \setminus [\lambda_{n+1}, \lambda_1]$, the expectation in (8.4) is well-defined and holomorphic in z. This follows since for such z, the denominators $z-\lambda_i$ and $z-\mu_j$ are bounded away from zero with probability 1. The key challenge is to show that $f_W(z)$ can be analytically continued to an entire function. Potential singularities of $f_W(z)$ are inside the intervals $(\lambda_{i+1}, \lambda_1)$. We will show that these singularities do not actually occur.

Consider a specific interval (λ_2, λ_1). We need to show that $f_W(z)$ has no singularities in this interval. From Theorem 8.2, the probability distribution of $\mu = (\mu_1, \ldots, \mu_n)$ has density proportional to:

$$\prod_{1 \le i < j \le n} (\mu_i - \mu_j) \prod_{i=1}^{n} \prod_{j=1}^{n+1} |\mu_i - \lambda_j|^{\beta/2-1}.$$

Let us analyze the function in (8.4). For $z \in (\lambda_2, \lambda_1)$, we need to demonstrate that the expectation

$$\mathbb{E}\left[\prod_{j=1}^{n} \exp\bigl(W(\mu_j)\bigr) \frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)} \left(W'(z) + \sum_{i=1}^{n+1} \frac{\beta/2-1}{z-\lambda_i} + \sum_{j=1}^{n} \frac{1}{z-\mu_j}\right)\right]$$

is holomorphic. This expectation is an $(n-1)$-fold integral over $\mu_1, \dots, \mu_n$. For $z \in (\lambda_2, \lambda_1)$, we will show that the one-dimensional integral over μ_1 is already holomorphic, and the remaining integrals are over domains which do not encounter singularities in z. We need to consider the integral

$$\int\limits_{\lambda_2}^{\lambda_1} \prod_{1\le i<j\le n} (\mu_i - \mu_j) \prod_{j=1}^{n} \prod_{i=1}^{n+1} (\mu_j - \lambda_i)^{\beta/2-1} \prod_{j=1}^{n} e^{W(\mu_j)} \frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)} \times \left(W'(z) + \sum_{i=1}^{n+1} \frac{\beta/2-1}{z-\lambda_i} + \sum_{j=1}^{n} \frac{1}{z-\mu_j}\right) d\mu_2. \tag{8.5}$$

Note that (here we are using the fact that $\beta > 2$)

$$\begin{aligned} 0 &= \int_{\lambda_2}^{\lambda_1} d\mu_1 \frac{\partial}{\partial \mu_1} \left(\underbrace{\prod_{1\le i<j\le n} (\mu_i - \mu_j) \prod_{j=1}^{n} \prod_{i=1}^{n+1} (\mu_j - \lambda_i)^{\beta/2-1} \prod_{j=1}^{n} e^{W(\mu_j)} \frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)}}_{(*)} \right) \\ &= \int_{\lambda_2}^{\lambda_1} d\mu_1 \, (*) \cdot \left[\sum_{j=2}^{n} \frac{1}{\mu_1 - \mu_j} + \sum_{i=1}^{n+1} \frac{\beta/2-1}{\mu_1 - \lambda_i} + W'(\mu_1) + \frac{1}{z-\mu_1} \right]. \end{aligned}$$

Subtracting this expression from our original integral (8.5) and noting that

$$\left(W'(z) + \sum_{i=1}^{n+1} \frac{\beta/2-1}{z-\lambda_i} + \sum_{j=1}^{n} \frac{1}{z-\mu_j}\right) - \left(\sum_{j=2}^{n} \frac{1}{\mu_1-\mu_j} + \sum_{i=1}^{n+1} \frac{\beta/2-1}{\mu_1-\lambda_i} + W'(\mu_1) + \frac{1}{z-\mu_1}\right)$$

has zero at $z = \mu_1$, we conclude that our integral has no singularity at μ_1, and therefore no singularities in the $[\lambda_2, \lambda_1]$ interval. This completes the proof of Theorem 8.4 for $\beta > 2$.

8.3 Applications of loop equations

The loop equations provide a powerful tool for analyzing the spectral properties of random matrices through their eigenvalue distributions. Let us derive an equation for the Stieltjes transform of the empirical measures.

8.3.1 Stieltjes transform equations

Starting from Theorem 8.4 with $W = 0$, we have:

$$\mathbb{E}\left[\frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)}\left(\sum_{i=1}^{n+1}\frac{\beta/2-1}{z-\lambda_i}+\sum_{j=1}^{n}\frac{1}{z-\mu_j}\right)\right]=\frac{(n+1)\beta}{2}-1. \tag{8.6}$$

Let us introduce the empirical Stieltjes transforms:

$$G_\lambda(z)=\frac{1}{n}\sum_{i=1}^{n+1}\frac{1}{z-\lambda_i},$$
$$G_\mu(z)=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{z-\mu_j}.$$

We also define the "logarithmic potentials" (indefinite integrals of the Stieltjes transforms):

$$\int G_\lambda(z)dz=\frac{1}{n}\sum_{i=1}^{n+1}\ln(z-\lambda_i),$$
$$\int G_\mu(z)dz=\frac{1}{n}\sum_{j=1}^{n}\ln(z-\mu_j).$$

We understand the integrals up to the same integration constant (and branch), so the exponent of the difference yields the original product:

$$\frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)}=\exp\left(n\left(\int G_\lambda(z)-\int G_\mu(z)\right)\right)$$

We can rewrite equation (8.6) as:

$$\mathbb{E}\left[\exp\left(n\left(\int G_\lambda(z)\,dz-\int G_\mu(z)\,dz\right)\right)\left(\left(\frac{\beta}{2}-1\right)G_\lambda(z)+G_\mu(z)\right)\right]=\frac{\beta}{2}+\frac{1}{n}\left(\frac{\beta}{2}-1\right). \tag{8.7}$$

8.3.2 Asymptotic behavior

Equation (8.7) can be reinterpreted in terms of a time evolution of eigenvalue distributions. This perspective offers significant insights into the asymptotic behavior of the corners process.

If we think of λ as configuration at time $t = 1$ and μ as configuration at time $t = 1 - \frac{1}{n}$, then denoting the general time parameter as t and setting $G_\lambda = G_1$, $G_\mu = G_{1-\frac{1}{n}}$, we obtain a continuous time evolution of Stieltjes transforms. (And similarly for all t, of course.)

As $n \to \infty$, equation (8.7) transforms into:

$$\frac{\beta}{2} \exp\left(\frac{\partial}{\partial t} \int G_t(z)\, dz\right) \cdot G_t(z) = \frac{\beta}{2}.$$

This implies

$$\frac{\partial}{\partial t} \int G_t(z)\, dz + \ln G_t(z) = 0.$$

Taking the derivative with respect to z, we get:

$$\frac{\partial}{\partial t} G_t(z) + \frac{1}{G_t(z)} \frac{\partial}{\partial z} G_t(z) = 0. \tag{8.8}$$

This is the inviscid Burgers equation, a fundamental nonlinear PDE in fluid dynamics — but with complex z. The complex Burgers equation has appeared in descriptions of limit shapes of models in statistical mechanics, such as lozenge tilings [KO07].

Remark 8.7. We see that the Burgers equation (8.8) does not depend on β, which is expected. Indeed, for example, GβE eigenvalues have the same Wigner semicircle law as $\beta = 2$, up to an overall rescaling.

8.3.3 Example: GβE and the semicircle law

The Stieltjes transform of the semicircular law is given by:

$$G(z) = \int_{-2}^{2} \frac{1}{z-x} \frac{\sqrt{4-x^2}}{2\pi} dx = \frac{1}{2}\left(z - \sqrt{z^2-4}\right).$$

We take this as the function $G_t(z)$ for $t = 1$. Then, for each $0 \le t \le 1$, the GβE solution should be

$$\frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \frac{1}{z - \lambda_i^{(\lfloor nt \rfloor)}} \to t\, G^{(\sqrt{t})}(z),$$

where
$$G^{(c)}(z) := \frac{z - \sqrt{z^2 - 4c^2}}{2c^2},$$
is the Stieltjes transform of the semicircular law on $[-2c, 2c]$.

Lemma 8.8. *The function $G_t(z) := t\,G^{(\sqrt{t})}(z)$ satisfies the Burgers equation* (8.8).

Proof. Straightforward verification. □

8.4 Problems

8.4.1 Cauchy determinant

Prove the Cauchy determinant formula:
$$\det\left(\frac{1}{x_i - y_j}\right)_{1\le i,j\le n} = \frac{\prod_{i<j}(x_i - x_j)(y_i - y_j)}{\prod_{i,j}(x_i - y_j)}.$$

8.4.2 Jacobian from $n-1$ to n dependent variables

Explain how the factor $\prod_{i=1}^{n-1}\prod_{j=1}^{n}|\mu_i - \lambda_j|$ appears from the Jacobian of the transformation from φ to μ (8.2), even though $\partial\varphi_a/\partial\mu_b$ is defined for $a = 1,\dots,n$, $b = 1,\dots,n-1$, but the φ_i's are not independent.

8.4.3 Dirichlet density

Find in the literature or prove on your own the first statement in the proof of Theorem 8.2 about the symmetric Dirichlet density arising from normalizing the ξ_i's to φ_i's.

8.4.4 General beta Gaussian density and cutting corners

Show that if $\lambda_1,\dots,\lambda_{n+1}$ have the Gaussian beta density of order $n+1$,
$$\propto \prod_{1\le i<j\le n+1}(\lambda_i - \lambda_j)^{\beta}\prod_{i=1}^{n+1} e^{-\beta\lambda_i^2/2},$$
and $\mu_1,\dots,\mu_n$ are obtained from $\lambda_1,\dots,\lambda_{n+1}$ by cutting the corner (so have the conditional density as in Theorem 8.2), then $\mu_1,\dots,\mu_n$ have the Gaussian beta density of order n.

8.4.5 General β Corners Process Simulation

This problem explores computational aspects of the general β corners process.

(a) Write code for generating a sample from the distribution of $\mu = (\mu_1, \ldots, \mu_{n-1})$ given $\lambda = (\lambda_1, \ldots, \lambda_n)$ for arbitrary $\beta > 0$, using the polynomial equation characterization.

(b) Let $\lambda = (n, n-1, \ldots, 2, 1)$. For $n = 7$, compute (numerically) the expected values $\mathbb{E}[\mu_i]$ for each i, when $\beta = 1, 2, 4$, and 10. Describe the behavior as β increases.

Chapter 9

Loop equations and asymptotics to Gaussian Free Field

9.1 Recap

9.1.1 (Dynamical) loop equations

Theorem 9.1. *We fix $n = 1, 2, \dots$ and $n+1$ real numbers $\lambda_1 \geq \dots \geq \lambda_{n+1}$. For $\beta > 0$, consider $n+1$ i.i.d. χ^2_β random variables ξ_i and set*

$$w_i = \frac{\xi_i}{\sum_{j=1}^{n+1} \xi_j}, \qquad 1 \leq i \leq n+1.$$

We define n random points $\{\mu_1, \dots, \mu_n\}$ as n solutions to the equation

$$\sum_{i=1}^{n+1} \frac{w_i}{z - \lambda_i} = 0. \tag{9.1}$$

Take any polynomial *$W(z)$ and consider the complex function:*

$$f_W(z) = \mathbb{E}\left[\prod_{j=1}^{n} \exp\big(W(\mu_j)\big) \frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)} \left(W'(z) + \sum_{i=1}^{n+1} \frac{\beta/2 - 1}{z - \lambda_i} + \sum_{j=1}^{n} \frac{1}{z - \mu_j}\right)\right]. \tag{9.2}$$

Then $f_W(z)$ is an entire function *of z, in the following sense:*

- *For $z \in \mathbb{C} \setminus [\lambda_{n+1}, \lambda_1]$, the expectation in (9.2) defines a holomorphic function of z.*

- *This function has an analytic continuation to $\mathbb{C}$, which has no singularities.*

We proved this statement for $\beta > 2$, but it is valid for all $\beta > 0$.

9.1.2 Loop equations for $W = 0$

When $W = 0$, the loop equation (9.2) becomes

$$f_0(z) = \frac{(n+1)\beta}{2} - 1,$$

so

$$\mathbb{E}\left[\frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)}\left(\sum_{i=1}^{n+1}\frac{\beta/2-1}{z-\lambda_i} + \sum_{j=1}^{n}\frac{1}{z-\mu_j}\right)\right] = \frac{(n+1)\beta}{2} - 1.$$

Recall that we defined

$$G_\lambda(z) = \frac{1}{n}\sum_{i=1}^{n+1}\frac{1}{z-\lambda_i}, \qquad G_\mu(z) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{z-\mu_j}.$$

We also define the "logarithmic potentials" (indefinite integrals of the Stieltjes transforms):

$$\int G_\lambda(z)dz = \frac{1}{n}\sum_{i=1}^{n+1}\ln(z-\lambda_i), \qquad \int G_\mu(z)dz = \frac{1}{n}\sum_{j=1}^{n}\ln(z-\mu_j).$$

We understand the integrals up to the same integration constant (and branch), so the exponent of the difference yields the original product:

$$\frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{\prod_{j=1}^{n}(z-\mu_j)} = \exp\left(n\left(\int G_\lambda(z) - \int G_\mu(z)\right)\right)$$

We can rewrite the loop equation as:

$$\mathbb{E}\left[\exp\left(n\left(\int G_\lambda(z)\,dz - \int G_\mu(z)\,dz\right)\right)\left(\left(\frac{\beta}{2}-1\right)G_\lambda(z) + G_\mu(z)\right)\right] = \frac{\beta}{2} + \frac{1}{n}\left(\frac{\beta}{2}-1\right). \tag{9.3}$$

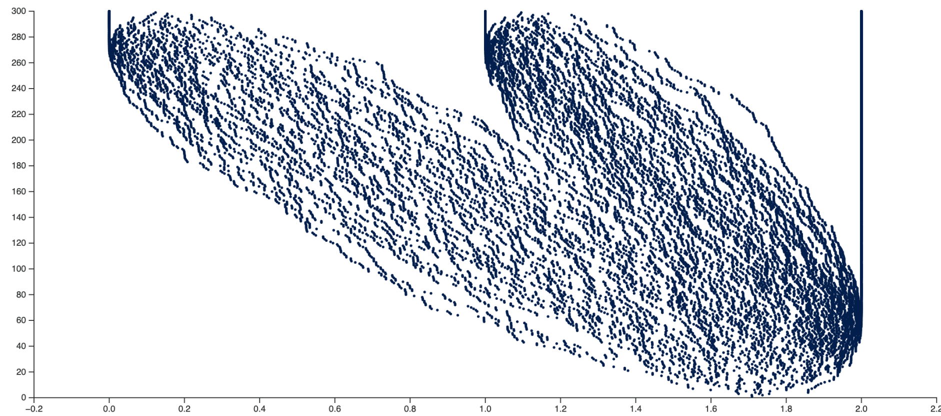

Figure 9.1: Corners process for $n = 300$, $\beta = 1$, with $n/10$ points at 0, $n/10$ points at 1, and $8n/10$ points at 2 on the top level.

9.1.3 The full corners process

Assume n is going to infinity, and we fix a sequence of top-level eigenvalues $\lambda_j^{(n)}$, $1 \le j \le n$, growing in some way. This sequence can be random (like GβE rescaled to have eigenvalues in a bounded interval) or deterministic (for example, $\lambda^{(n)}$ has $n/10$ points at 0, $n/10$ points at 1, and $8n/10$ points at 2, see Figure 9.1).

Denote the eigenvalues of the $k \times k$ beta corner (that is, obtained by successively solving the polynomial equation (9.1) $n - k$ times) by $\lambda_j^{(k)}$, $1 \le j \le k$. As $n \to \infty$, we postulate that

> The empirical distribution of $\lambda_j^{(k)}$ converges to some deterministic probability measure $\mathfrak{m}_t$, where $k/n \to t \in [0, 1]$. Consequently, the Stieltjes transform $G_{\lambda^{(k)}}(z)$ converges to $G_t(z)$, for z in a complex domain outside of the support of $\mathfrak{m}_t$.

Note that we do not assume the scaling of the $\lambda_j^{(k)}$'s, for convenience.

Denote by $G_t(z) = \displaystyle\int_{\mathbb{R}} \frac{\mathfrak{m}_t(dx)}{z - x}$ the Stieltjes transform of the measure $\mathfrak{m}_t$.

Proposition 9.2. *The functions $G_t(z)$ satisfy the complex Burgers equation*

$$\frac{\partial}{\partial t} G_t(z) + \frac{1}{G_t(z)} \frac{\partial}{\partial z} G_t(z) = 0.$$

Proof. We have in (9.3), if λ and μ live on levels t and $t-\frac{1}{n}$, respectively:

$$G_\lambda(z)-G_\mu(z) \approx \frac{1}{n}\frac{\partial}{\partial t}G_t(z), \qquad \left(\frac{\beta}{2}-1\right)G_\lambda(z)+G_\mu(z) \approx \frac{\beta}{2}G_t(z)-\frac{1}{n}\frac{\partial}{\partial t}G_t(z) \approx \frac{\beta}{2}G_t(z).$$

Due to the concentration assumption, we can ignore the expectation. Then, taking the logarithm of (9.3), and differentiating with respect to z, we get the Burgers equation. □

9.1.4 Example: GβE and the semicircle law

The Stieltjes transform of the semicircular law is given by:

$$G(z) = \int_{-2}^{2} \frac{1}{z-x}\frac{\sqrt{4-x^2}}{2\pi}dx = \frac{1}{2}\left(z-\sqrt{z^2-4}\right).$$

We take this as the function $G_t(z)$ for $t=1$. Then, for each $0 \le t \le 1$, the GβE solution should be

$$\frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor}\frac{1}{z-\lambda_i^{(\lfloor nt \rfloor)}} \to t\,G^{(\sqrt{t})}(z),$$

where

$$G^{(c)}(z) := \frac{z-\sqrt{z^2-4c^2}}{2c^2},$$

is the Stieltjes transform of the semicircular law on $[-2c, 2c]$.

Lemma 9.3. *The function $G_t(z) := t\,G^{(\sqrt{t})}(z)$ satisfies the Burgers equation.*

Proof. Straightforward verification. □

9.2 Gaussian Free Field

The *Gaussian Free Field* (GFF) is a fundamental object in probability theory and mathematical physics. Roughly speaking, it can be viewed as a multi-dimensional analog of Brownian motion: instead of one-dimensional "time," the underlying parameter space is a multi-dimensional domain (often two-dimensional). In one dimension, the GFF reduces to an ordinary Brownian bridge (or motion). In higher dimensions, it becomes a random generalized function (a "distribution") whose covariance structure is governed by the appropriate Green's function of the Laplacian. Below we provide an introduction, starting from finite-dimensional Gaussian vectors and culminating in the GFF as a random distribution.

9.2.1 Gaussian correlated vectors and random fields

Recall that an n-dimensional real-valued random vector $X = (X_1, \ldots, X_n)$ is called *Gaussian* if every linear combination

$$\alpha_1 X_1 + \cdots + \alpha_n X_n$$

of its components is a univariate Gaussian random variable. The law of such a vector is completely determined by its mean vector $m \in \mathbb{R}^n$ and its covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$. The density function, for invertible Σ, is

$$f_X(x) \;=\; \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\Big(-\tfrac{1}{2}\,(x-m)^\top \Sigma^{-1}(x-m)\Big).$$

For simplicity, we will assume that $m = 0$ (the centered case).

9.2.2 Gaussian fields as random generalized functions

A natural extension from finite-dimensional Gaussian vectors to infinite-dimensional settings leads us to Gaussian fields. Informally, a Gaussian field is a collection of Gaussian random variables indexed by points in some space.

For a domain $D \subset \mathbb{R}^d$, we might wish to define a random function $\Phi : D \to \mathbb{R}$ such that for any finite collection of points $x_1, \ldots, x_n \in q$, the vector $(\Phi(x_1), \ldots, \Phi(x_n))$ is a Gaussian vector. However, such a random function may not exist as a proper function in the usual sense. The reason is that we would like to consider analogues of linear combinations of the form

$$\Phi(f) = \int_D \Phi(x) f(x)\, dx, \tag{9.4}$$

For example, if we wish the vector $(\Phi(x_1), \ldots, \Phi(x_n))$ to have independent components, we would need to assign a value to each point in D. This means that the hypothetical function Φ would be too irregular, and even non-measurable, and the integral (9.4) would not be well-defined.

Instead, for the field with independent values at all points, we would like $\Phi(f)$ to be normal with mean zero and variance (paralleling the finite-dimensional story)

$$\mathrm{Var}\,(\Phi(f)) = \|f\|^2_{L^2(D)} \;=\; \int_D f(x)^2\,dx.$$

So, Gaussian fields (in particular, our topic, the *Gaussian Free Field*)a are defined as random distributions, not as functions. That is, rather than assigning a value to each point, we assign a random value to each test function f in some appropriate space via (9.4).

The covariance structure of the mean zero Gaussian random variables $\Phi(f_1), \dots, \Phi(f_n)$ is given by a certain bilinear form determined by the domain D.

9.2.3 Concrete treatment via orthogonal functions

Let us now construct the Gaussian Free Field more concretely. Consider a bounded domain $D \subset \mathbb{R}^d$ with smooth boundary. Let $\{f_n\}_{n=1}^{\infty}$ be an orthonormal basis of $L^2(D)$ consisting of eigenfunctions of the Laplacian with Dirichlet boundary conditions:

$$\begin{cases} -\Delta f_n = \lambda_n f_n & \text{in } D, \\ f_n = 0 & \text{on } \partial D, \end{cases} \tag{9.5}$$

where $0 < \lambda_1 \leq \lambda_2 \leq \dots$ are the corresponding eigenvalues.

We can now define the Gaussian Free Field on D as:

$$\Phi = \sum_{n=1}^{\infty} \frac{\alpha_n}{\sqrt{\lambda_n}} f_n, \tag{9.6}$$

where $\{\alpha_n\}_{n=1}^{\infty}$ are independent standard Gaussian random variables. This series does not converge pointwise, but it does converge in the space of distributions almost surely.

For any test function $g \in C_0^{\infty}(D)$, we have:

$$\Phi(g) = \int_D \Phi(x) g(x)\,dx = \sum_{n=1}^{\infty} \frac{\alpha_n}{\sqrt{\lambda_n}} \int_D f_n(x) g(x)\,dx, \tag{9.7}$$

which is a well-defined Gaussian random variable.

9.2.4 Connection to Brownian bridge

The Gaussian Free Field in one dimension is closely related to the Brownian bridge. Consider the interval $[0, 1]$ with the Dirichlet Laplacian. The eigenfunctions are $f_n(x) = \sqrt{2}\sin(n\pi x)$ with eigenvalues $\lambda_n = n^2\pi^2$. The Gaussian Free Field on $[0, 1]$ can be expressed as:

$$\Phi(x) = \sqrt{2}\sum_{n=1}^{\infty} \frac{\alpha_n}{n\pi} \sin(n\pi x). \tag{9.8}$$

This series representation converges to a continuous function, which is precisely the Brownian bridge on $[0, 1]$. The Brownian bridge is a Gaussian process B_t with mean zero and covariance function:

$$\mathbb{E}[B_s B_t] = \min(s, t) - st. \tag{9.9}$$

The key difference between the one-dimensional and higher-dimensional cases is that in one dimension, the Gaussian Free Field is a continuous function, whereas in dimensions two and higher, it is a genuine distribution (not a function). This reflects the fact that Brownian motion is a continuous path in one dimension but becomes increasingly irregular in higher dimensions.

9.2.5 Covariance structure and Green's function

The covariance structure of the Gaussian Free Field is intimately connected to the Green's function of the Laplacian. For test functions $f, g \in C_0^\infty(D)$, we have:

$$\begin{aligned} \mathbb{E}[\Phi(f)\Phi(g)] &= \mathbb{E}\left[\sum_{n,m=1}^{\infty} \frac{\alpha_n\alpha_m}{\sqrt{\lambda_n\lambda_m}} \int_D f_n(x)f(x)\,dx \int_D f_m(y)g(y)\,dy\right] &(9.10)\\ &= \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \int_D f_n(x)f(x)\,dx \int_D f_n(y)g(y)\,dy. &(9.11) \end{aligned}$$

Define the Green's function $G_D(x, y)$ for the Dirichlet Laplacian on D as the solution to:

$$\begin{cases} -\Delta_x G_D(x, y) = \delta(x - y) & \text{for } x, y \in D, \\ G_D(x, y) = 0 & \text{for } x \in \partial D \text{ or } y \in \partial D. \end{cases} \tag{9.12}$$

The Green's function has the eigenfunction expansion:

$$G_D(x,y) = \sum_{n=1}^{\infty} \frac{f_n(x) f_n(y)}{\lambda_n}. \tag{9.13}$$

Using this, we can rewrite the covariance as:

$$\mathbb{E}[\Phi(f)\Phi(g)] = \int_D \int_D G_D(x,y) f(x) g(y) \, dx \, dy. \tag{9.14}$$

This relationship between the covariance of the GFF and the Green's function is fundamental. It shows that the GFF can be viewed as a random solution to the equation $-\Delta\Phi = W$, where W is white noise. Here the white noise is the Gaussian field with covariance $\delta(x-y)$ — the object which is the correct way of constructing a Gaussian field with i.i.d. values at all points.

9.2.6 The GFF on the upper half-plane

In the complex upper half-plane $\{\operatorname{Im} z > 0\}$ with $\mathbb{R}$ as the boundary, the Green function has the form

$$G(z,w) = -\frac{1}{\pi} \ln |z-w| + \frac{1}{\pi} \ln |z - \overline{w}|.$$

The covariance is

$$\mathbb{E}\left[\Phi(f)\,\Phi(g)\right] = \int \int |dz|^2 \, |dw|^2 f(z) g(w) G(z,w).$$

9.3 Fluctuations

9.3.1 Height function and related definitions

Let us define the *height function* using the corners process $\{\lambda_j^{(k)} : 1 \le j \le k \le n\}$:

$$h(t,x) := \#\{\text{eigenvalues } \lambda_j^{(\lfloor nt \rfloor)} \text{ which are} \le x\}.$$

Recall that in our regime, we do not scale x. Throughout the following, we will interchangeably use the parameters n and $\varepsilon := 1/n$.

Our goal is to understand the asymptotic behavior of the centered height function

$$h(t,x) - \mathbb{E}[h(t,x)],$$

defined inside the region of the (t, x) plane. Note that in contrast with the usual Central Limit Theorem, **the fluctuations are not scaled by $\varepsilon^{1/2}$, but rather converge to a certain object without any scaling**. Note that the law of large numbers is going to be

$$\varepsilon h(t,x) \to \mathfrak{h}(t,x),$$

where $\mathfrak{h}(t,x)$ is the limiting height function (for a fixed t, this is the cumulative distribution function of the measure $\mathfrak{m}_t$). We will see that these unscaled fluctuations are converging to a Gaussian Free Field. Thus, the unscaled fluctuations are "just barely" going to infinity, while retaining nontrivial and bounded correlations.

9.3.2 Main results on Gaussian fluctuations

Recall that our main assumption is that the distribution at the top row converges (with a good control) to a deterministic measure $\mathfrak{m}_1$:

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i^{(n)}} \to \mathfrak{m}_1.$$

For example, in Figure 9.1, the measure $\mathfrak{m}_1$ has three atoms.

Denote the centered Stieltjes transforms by

$$\tilde{G}_\lambda(z) \coloneqq G_\lambda(z) - \mathbb{E}[G_\lambda(z)].$$

Theorem 9.4. *Fix an integer $k \geq 1$ and pick k pairs $\big(t_i, u_i\big)$, $1 \leq i \leq k$. Consider the random variables*

$$\varepsilon^{-1}\tilde{G}_{\lambda([nt_i])}\big(z(t_i,u_i)\big), \quad 1 \leq i \leq k,$$

where $z(\cdot,\cdot)$ is a conformal structure on the liquid region in the corners process.[1]

Then, as $\varepsilon \to 0$, these k random variables converge (in the sense of moments, uniformly over (t_i, u_i) in compact sets) to a k-dimensional Gaussian vector of mean zero. Their limiting covariances are

$$\lim_{\varepsilon\to 0} \varepsilon^{-2}\mathbb{E}\Big[\tilde{G}_{\varepsilon^{-1}t_i}\big(z(t_i,u_i)\big)\,\tilde{G}_{\varepsilon^{-1}t_j}\big(z(t_j,u_j)\big)\Big] \;=\; \frac{1}{\partial_{u_i} z(t_i,u_i)\,\partial_{u_j} z(t_j,u_j)}\,\partial_{u_i}\partial_{u_j}\,\ln\Bigg[\frac{u_i \;-\; u_j}{z\big(\tau,u_i\big) \;-\; z\big(\tau,u_j\big)}\Bigg],$$

where $\tau = \min(t_i, t_j)$.

[1] It exists, and can be characterized rather explicitly, but we will not go into details here.

Corollary 9.5. *Again assuming $b(z) = z$ for all z, fix an integer $k > 0$ and real parameters*

$$0 < t_1 \ \leq \ t_2 \ \leq \ \cdots \ \leq \ t_k < T,$$

along with real-analytic functions $f_1(x), \dots, f_k(x)$ in a neighborhood of the real axis. Define the random vector

$$\left(\sqrt{\pi} \int_{l(t_i)}^{r(t_i)} f_i(x) \left[h\big(t_i,\, \varepsilon^{-1}x\big) - \mathbb{E}\big(h(t_i, \varepsilon^{-1}x)\big) \right] dx \right)_{i=1}^{k},$$

where $[l(t_i), r(t_i)]$ contains the support of the t_i-th slice of the corners process.

As $\varepsilon \to 0$, this random k-vector converges (in the sense of moments) to a centered Gaussian vector, whose covariance is

$$-\frac{1}{4\pi} \oint_{C_i} \oint_{C_j} \partial_{w_i} \partial_{w_j} \left[\log\big(w_i - w_j\big) \right] F_i\big(w_i\big)\, F_j\big(w_j\big)\, dw_i\, dw_j,$$

where C_i and C_j are positively oriented contours enclosing the real interval $[\, l(t_i),\, r(t_i)\,]$ and $[\, l(t_j),\, r(t_j)\,]$, respectively, inside their regions of analyticity, and $F_i(x)$ is such that $f_i(x) = \partial_x[F_i(x)]$.

9.3.3 Deformed ensemble

The rest of this section illustrates the beginning of the argument in [GH24], but in our random matrix setting. In the interest of time, we are following the main steps in a non-rigorous manner, (in particular, following [GH24, Section 4.2]), and do not present a complete proof. The goal here is to illustrate the main idea how the loop equation can be useful for analyzing asymptotics.

This theorem is an asymptotic expansion of the Stieltjes transform of the one-step transition from λ to μ. We assume that the support of λ is in $[l, r]$. Denote

$$\Pi_\lambda(z) \coloneqq \prod_{i=1}^{n+1} (z - \lambda_i), \qquad \Pi_\mu(z) \coloneqq \prod_{j=1}^{n} (z - \mu_j).$$

Also assume that $W(z)$ is fixed and nice, and that μ_j are distributed according to a modified density, which includes $W(z)$:

$$\frac{1}{Z} \prod_{1 \leq i < j \leq n} (\mu_i - \mu_j) \prod_{i=1}^{n} \prod_{j=1}^{n+1} |\mu_i - \lambda_j|^{\beta/2-1} \prod_{1 \leq i < j \leq n+1} (\lambda_i - \lambda_j)^{1-\beta} \prod_{j=1}^{n} e^{W(\mu_j)}.$$

From now on, all expectations will be over the W-modified density.

We aim to analyze the quantity

$$\mathcal{A}(z) := \mathbb{E}\left[\frac{\Pi_\lambda(z)}{z\,\Pi_\mu(z)}\right],$$

which enters the loop equation. Moreover, the loop equation states the holomorphicity of

$$\mathcal{C}(z) = \mathcal{A}(z)\left[zW'(z)+\frac{\beta}{2}\frac{1}{n}\sum_{i=1}^{n+1}\frac{z}{z-\lambda_i}\right]+\mathbb{E}\left[\frac{\Pi_\lambda(z)}{\Pi_\mu(z)}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{z-\mu_j}-\frac{1}{n}\sum_{i=1}^{n+1}\frac{1}{z-\lambda_i}\right)\right].$$

The first summand is the leading term, and the second summand will be negligible. Indeed, it contains the difference of $G_\mu(z)$ and $G_\lambda(z)$, and these Stieltjes transforms are close to each other, so the difference is $O(\varepsilon)$.

9.3.4 Wiener-Hopf like factorization

Denote

$$\mathcal{B}(z) = zW'(z) + \frac{z\beta}{2}G_\lambda(z).$$

Decompose $\mathcal{B}(z)$ using the Cauchy residue formula:

$$\ln\mathcal{B}(z) = \frac{1}{2\pi i}\oint_{\omega_+}\frac{\ln\mathcal{B}(w)}{w-z}dw - \frac{1}{2\pi i}\oint_{\omega_-}\frac{\ln\mathcal{B}(w)}{w-z}dw,$$

where ω_+ is positively oriented and encloses $[l,r]$ and z, while ω_- is also positively oriented and encloses $[l,r]$ but not z. Then define

$$h_+(u) := \frac{1}{2\pi i}\oint_{\omega_+}\frac{\ln\mathcal{B}(w)}{w-u}dw, \qquad h_-(u) := \frac{1}{2\pi i}\oint_{\omega_-}\frac{\ln\mathcal{B}(w)}{w-u}dw.$$

Thus, we get the Wiener-Hopf like factorization

$$\mathcal{B}(z) = e^{h_+(z)}e^{-h_-(z)},$$

where h_+ is holomorphic in a neighborhood of $[l,r]$, and h_- is holomorphic in a neighborhood of ∞, with behavior $O(1/u)$ at infinity. The factorization is valid in an annulus between the two contours ω_+ and ω_-.

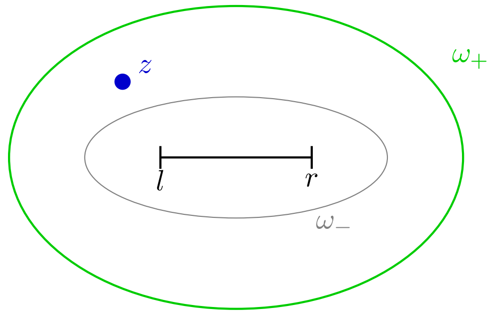

Figure 9.2: Positively oriented contours ω_+ and ω_- in the complex plane.

9.3.5 First order asymptotics of $\mathcal{A}(z)$

The next step is to understand the asymptotics of $\mathcal{A}(z)$. Recall that

$$\mathcal{A}(z) = \mathbb{E}\left[\frac{\Pi_\lambda(z)}{z\Pi_\mu(z)}\right]. \tag{9.15}$$

From the loop equation, we know that $\mathcal{C}(z)$ is entire, and the leading term involves $\mathcal{A}(z)\mathcal{B}(z)$. That is,

$$\mathcal{A}(z)\mathcal{B}(z) = \text{entire function} + O(\varepsilon). \tag{9.16}$$

Using the Wiener-Hopf factorization of $\mathcal{B}(z)$, let us multiply (9.16) by $e^{-h_+(z)}$. The entire function remains entire in a complex neighborhood of $[l, r]$. Therefore, we can integrate over ω_-, and get

$$\begin{aligned} 0 = \frac{1}{2\pi i}\oint_{\omega_-}\frac{\mathcal{C}(w)e^{-h_+(w)}dw}{w-z} &= \frac{1}{2\pi i}\oint_{\omega_-}\frac{\mathcal{A}(w)e^{-h_-(w)}dw}{w-z} + O(\varepsilon) \\ &= -\mathcal{A}(z)e^{-h_-(z)} + \frac{1}{2\pi i}\oint_{\omega_+}\frac{\mathcal{A}(w)e^{-h_-(w)}}{w-z}dw + O(\varepsilon). \end{aligned}$$

In the last equality, we took a residue at $w = z$, and replaced the integral by an integral over ω_+.

The integrand has no singularities outside ω_+, and thus is just the residue at infinity. Using the fact that $e^{-h_-(u)} = e^{1+O(1/u)} = 1+O(1/u)$, $u \to \infty$ and the fact that the expectation $\mathcal{A}(u)$ is balanced in u (hence it is $1+O(1/u)$), we see that the residue at infinity is simply equal to 1. Therefore,

$$0 = -\mathcal{A}(z)e^{-h_-(z)} + 1 + O(\varepsilon), \qquad \ln\mathcal{A}(z) = h_-(z) + O(\varepsilon).$$

We emphasize that this equation stays valid for all functions $W(z)$.

9.3.6 Outlook of further steps

Let us rewrite the last equation explicitly, inserting W into the expectation, and taking $\mathbb{E}_0$ to be the undeformed expectation over the GβE corner:

$$\mathbb{E}_0\left[\frac{\prod_{i=1}^{n+1}(z-\lambda_i)}{z\prod_{j=1}^{n} e^{-W(\mu_j)}(z-\mu_j)}\right] = \frac{1}{2\pi i}\oint_{\omega_-} \frac{\ln\left(W'(w)+\frac{\beta}{2}G_\lambda(w)\right)}{w-z}\,dw + O(\varepsilon). \tag{9.17}$$

There are the following extra steps required to complete the proofs of the main results:

- Continue the expansion (9.17) to higher orders of ε.
- Extract probabilistic information from the formula in the left-hand side of (9.17).
- Carefully execute the analysis, including all the required estimates, to get the asymptotic behavior of the Stieltjes transforms.
- From the Stieltjes transforms, extract the asymptotic behavior of the height function.

We do not perform this analysis here, but direct the reader to [GH24] for the full details, in a setting of lozenge tilings with q-Racah weights.

9.4 Problems

9.4.1 Brownian bridge

Derive the covariance structure of the Brownian bridge (9.9) from the series representation (9.8).

Chapter 10

Dyson Brownian Motion

10.1 Motivations

10.1.1 Why introduce time?

Our previous lectures dealt with static matrix ensembles (e.g., GUE, GOE, and so on). However, there are both *physical* and *mathematical* reasons to study a dynamical model for random matrices. For instance:

1. In physics, one often interprets random matrices as Hamiltonians of quantum systems. It is natural to let these Hamiltonians vary in time and to describe how spectra evolve.

2. Such time-dependent models are vital for studying *universality results* in random matrix theory. Rigorous proofs of local eigenvalue correlations often involve coupling or evolving an ensemble toward (or away from) a known reference ensemble.

3. Dynamical extensions yield intriguing connections to 2D statistical mechanics, representation theory, and Markov chain interpretations such as *nonintersecting path ensembles*.

10.1.2 Simple example: 1×1 case

When $N = 1$, an $N \times N$ Hermitian matrix is just a single real entry. Thus GUE/GOE/GSE distributions each reduce to a real Gaussian variable with mean 0 and variance 1. If we allow *time*, the natural time evolution is standard *Brownian motion* $B(t)$ on $\mathbb{R}$.

Recall that a standard one–dimensional Brownian motion $B(t)$ is a continuous stochastic process with the following key properties:

1. **Continuity:** $t \mapsto B(t)$ is almost surely continuous.
2. **Independent increments:** For any $0 \le s < t$, the increment $B(t) - B(s)$ is independent of the past $\{B(u) : 0 \le u \le s\}$.
3. **Gaussian increments:** $B(t) - B(s)$ is normally distributed with mean 0 and variance $t - s$; that is,

$$B(t) - B(s) \sim \mathcal{N}(0,\, t-s).$$

Thus, if the process starts at $B(0) = a$, then for any fixed $t > 0$,

$$B(t) \sim \mathcal{N}(a,\, t).$$

Our goal is to generalize this to the case of *matrix-valued* Brownian motion and, ultimately, to see how the *eigenvalues* of such a matrix evolve.

10.2 Matrix Brownian motion and its eigenvalues

10.2.1 Definition

Let $X(t)$ be an $N \times N$ matrix whose entries are i.i.d. real/complex Brownian motions (depending on $\beta = 1, 2$). For instance:

- If $\beta = 1$: $X(t)$ has entries that are i.i.d. real Brownian motions.
- If $\beta = 2$: $X(t)$ has entries that are i.i.d. complex Brownian motions (independent real and imaginary parts).

Since $X(t)$ may not be Hermitian, define

$$\mathcal{M}(t) \;=\; \frac{1}{\sqrt{2}}\big(X(t) + X^{\dagger}(t)\big).$$

Here $X^{\dagger}(t)$ is the conjugate transpose. Then $\mathcal{M}(t)$ is an *Hermitian* matrix (or real symmetric for $\beta = 1$).

Lemma 10.1. *If $\mathcal{M}(0) = A$ is a fixed deterministic matrix, then $\mathcal{M}(t)$ at time t is distributed as*

$$A \;+\; \sqrt{t}\, G_{\beta},$$

where G_{β} is a random Hermitian matrix from the Gaussian ensemble with $\beta = 1$ or 2.

Sketch of proof. Straightforward observation. □

For the one-dimensional case, notice that $a + \sqrt{t}Z$, where $Z \sim \mathcal{N}(0, 1)$, is a Gaussian random variable with mean a and variance t, and every such Gaussian variable can be represented in this form.

10.2.2 Eigenvalues as Markov process

We now focus on $\lambda_i(t)$, the (ordered) eigenvalues of $\mathcal{M}(t)$. Denote

$$\lambda(t) = \big(\lambda_1(t) \geq \cdots \geq \lambda_N(t)\big).$$

Theorem 10.2. *As t varies, the process $\lambda(t)$ is a continuous-time Markov process in $\mathbb{R}^N$.*

Sketch of proof. Assume $\beta = 2$, the case $\beta = 1$ is similar. We need to show that $\lambda(t)$ depends on its future and past only through its instantaneous value. Using the independent increment property of $X(t)$, consider times $0 < u < t$. We have

$$\mathcal{M}(t) \;=\; \mathcal{M}(u) \;+\; \big(\mathcal{M}(t) - \mathcal{M}(u)\big).$$

Since $\mathcal{M}(u)$ diagonalizes to $\operatorname{diag}\big(\lambda_1(u), \ldots, \lambda_N(u)\big)$ by some unitary U_u, we can write

$$U_u^\dagger \, \mathcal{M}(t) \, U_u \;=\; \operatorname{diag}\big(\lambda_1(u), \ldots, \lambda_N(u)\big) \;+\; U_u^\dagger \big(\mathcal{M}(t) - \mathcal{M}(u)\big) \, U_u.$$

The second term again has i.i.d. random entries (due to unitary invariance of GUE), independent of $\mathcal{M}(s)$ for $s \leq u$. Therefore, conditioned on $\mathcal{M}(s)$, $s \leq u$, the dependence only comes through $\lambda(u)$, and the eigenvalues $\lambda_i(s)$ for $s \geq u$ follow the same dynamics. This proves the Markov property. $\square$

10.3 Dyson Brownian Motion

We now describe the stochastic differential equation (SDE) for $\lambda(t)$ explicitly, following the classical result due to Dyson [Dys62a]. Let us first briefly discuss what is an SDE.

10.3.1 Stochastic differential equations - an informal introduction

In order to describe the eigenvalues of a time-dependent Hermitian matrix, we rely on *stochastic differential equations* (SDEs). These are differential equations where one or more of the terms involve *random noise*. For simplicity, we start with the one-dimensional setup and later extend it to systems of equations such as those arising in Dyson Brownian Motion.

In an ordinary differential equation (ODE), a function $x(t)$ evolves according to a deterministic rule of the form

$$\frac{dx(t)}{dt} \;=\; b\big(x(t)\big),$$

where $b(\,\cdot\,)$ is a deterministic function called the *drift*. If one imposes an initial condition $x(0) = x_0$, then classical theorems guarantee that, under mild regularity assumptions, a unique solution exists for all $t \geq 0$.

An SDE generalizes this setup by adding a *stochastic (or noise) term* to the right-hand side. Concretely, suppose $W(t)$ is a standard one-dimensional Brownian motion. Then the simplest SDE has the form

$$dx(t) \;=\; \sigma\, dW(t),$$

where σ is a nonnegative constant. This equation may be formally interpreted as

$$\frac{dx(t)}{dt} \;=\; \sigma\,\frac{dW(t)}{dt},$$

but it should be emphasized that $\frac{dW}{dt}$ does not exist in the usual sense of classical calculus (Brownian motion is nowhere differentiable almost surely). Instead, one interprets the equation via the *Itô integral*

$$x(t) \;=\; x(0) \;+\; \int_0^t \sigma\, dW(s).$$

This integral is defined carefully through a limit of sums involving the increments $W(t_{k+1}) - W(t_k)$, yielding an *almost sure* continuous stochastic process $t \mapsto x(t)$.

More generally, one allows both *drift* and *diffusion* terms:

$$dx(t) \;=\; b\big(x(t)\big)\, dt \;+\; \sigma\big(x(t)\big)\, dW(t). \tag{10.1}$$

Here,

- $b(\cdot)$ is the *drift coefficient*, capturing deterministic motion;
- $\sigma(\cdot)$ is the *diffusion coefficient*, encoding how strongly the process is randomized by Brownian motion.

Under suitable Lipschitz and growth conditions on b and σ, one can show *existence and pathwise uniqueness* of strong solutions to (10.1). Concretely, this means there is almost surely a unique process $x(t)$ satisfying (10.1) for each realization of the Brownian motion $W(t)$. One constructs such a solution, for example, by an iterative limit of approximations. The simplest discrete-time approximation, analogous to Euler's method for ordinary differential equations. Over a small time step Δt, one approximates

$$x_{n+1} \;=\; x_n \;+\; b(x_n)\,\Delta t \;+\; \sigma(x_n)\,\big(W(t_{n+1}) - W(t_n)\big).$$

This scheme converges to the true solution pathwise under standard Lipschitz conditions on b and σ.

A major utility of SDEs is in performing *Itô calculus*. Suppose $x(t)$ solves the SDE (10.1) and let $f\colon \mathbb{R} \to \mathbb{R}$ be a sufficiently smooth function. One might try to apply the usual chain rule to $f(x(t))$, but must account for the extra "noise" term. The correct extension is the *Itô formula*:

$$df\big(x(t)\big) \;=\; \frac{\partial f}{\partial x}\big(x(t)\big)\,dx(t) \;+\; \frac{1}{2}\,\frac{\partial^2 f}{\partial x^2}\big(x(t)\big)\,\big(dW(t)\big)^2,$$

where $(dW(t))^2$ is interpreted as dt in a formal sense. Substituting (10.1) yields:

$$df\big(x(t)\big) \;=\; b\big(x(t)\big)\,\frac{\partial f}{\partial x}\big(x(t)\big)\,dt + \sigma\big(x(t)\big)\,\frac{\partial f}{\partial x}\big(x(t)\big)\,dW(t) + \frac{1}{2}\,\sigma^2\big(x(t)\big)\,\frac{\partial^2 f}{\partial x^2}\big(x(t)\big)\,dt.$$

This identity is an indispensable tool for analyzing stochastic processes, both in theoretical and applied contexts.

To handle matrix-valued processes, one must consider multi-dimensional (or matrix-dimensional) analogs of (10.1). For instance, if $X(t) \in \mathbb{R}^n$ is an n-dimensional stochastic process, the SDE becomes

$$dX(t) \;=\; b\big(X(t)\big)\,dt \;+\; \sigma\big(X(t)\big)\,dW(t),$$

where $b(\cdot)\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma(\cdot)\colon \mathbb{R}^n \to \mathbb{R}^{n\times n}$. Here $W(t)$ is an n-dimensional Brownian motion, and the product $\sigma\big(X(t)\big)\,dW(t)$ is understood as a matrix-vector multiplication in each small time increment. Existence, uniqueness, and Itô's formula all generalize naturally under suitable regularity assumptions.

Summary Although SDEs can be introduced rigorously via measure-theoretic tools, the above *informal* derivation and discussion provide a workable framework for many typical computations. The key points are:

- Brownian motion's roughness prevents classical differential calculus, so new techniques (Itô integrals) are needed.

- The Itô formula extends the classical chain rule by adding a second-order correction term.
- Existence and uniqueness theorems ensure that SDEs define well-posed dynamical systems in a stochastic setting.
- Extending to matrix-valued (or multi-dimensional) settings is conceptually straightforward but requires careful linear algebraic bookkeeping and additional regularity arguments.

Equipped with these ideas, we can rigorously address how the eigenvalues of a random matrix evolve over continuous time, culminating in the Dyson Brownian Motion description of Hermitian ensembles.

10.3.2 Heuristic derivation of the SDE for the Dyson Brownian Motion

Let $\mathcal{M}(t)$ be an $n \times n$ Hermitian matrix evolving as $\mathcal{M}(0) = A$ plus i.i.d. Gaussian increments in time. Denote its ordered eigenvalues at time t by

$$\lambda_1(t) \;\geq\; \dots \;\geq\; \lambda_n(t).$$

We aim to find an SDE for $\lambda_i(t)$.

For a small increment Δt, we have

$$\mathcal{M}(t + \Delta t) \;=\; \mathcal{M}(t) \;+\; \Delta\mathcal{M},$$

where the entries of $\Delta\mathcal{M}$ are (approximately) independent $\mathcal{N}(0, \Delta t)$ random variables (real or complex). Suppose we diagonalize $\mathcal{M}(t) = U\, \mathrm{diag}(\lambda_1(t), \dots, \lambda_n(t))\, U^\dagger$.

Sketch of the computation. Search for the i-th eigenvalue of the form

$$\lambda \;=\; \lambda_i(T) \;+\; \Delta\lambda \quad \big[\text{expect } \Delta\lambda \approx O(\sqrt{\Delta t})\big].$$

We want to solve

$$\det\begin{pmatrix} \lambda_1(T) - \lambda_i(T) + B_{11}(\Delta t) - \Delta\lambda & \cdots & \frac{1}{\sqrt{2}}\, B_{i1}(\Delta t) \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}}\, B_{1i}(\Delta t) & \cdots & \lambda_n(T) - \lambda_i(T) + B_{nn}(\Delta t) - \Delta\lambda \end{pmatrix} \;=\; 0.$$

In this matrix only $n-1$ diagonal elements — excluding the (i,i) entry — are bounded away from zero; the remaining (i,i)-th off-diagonal element is small. We have

$$\det = \prod_{m=1}^{n}\Big[\lambda_m(T)-\lambda_i(T)+B_{mm}(\Delta t)-\Delta\lambda\Big] - \sum_{j\neq i}\Bigg(\prod_{\substack{m\neq j\\ m\neq i}}\big[\lambda_m(T)-\lambda_i(T)+B_{mm}(\Delta t)-\Delta\lambda\big]\Bigg)\frac{1}{2}\,B_{ji}^2(\Delta t) + o$$

Here, the first product (diagonal part) involves all n diagonal-like terms, and the sum over $j \neq i$ ($n-1$ diagonal elements) accounts for corrections from the off-diagonal blocks. Higher-order terms are $o(\Delta t)$.

Divide by $\prod_{m\neq i}\big[\lambda_m(T)-\lambda_i(T)+B_m(\Delta t)-\Delta\lambda\big]$ to obtain

$$o(\Delta t) \;=\; -\,\Delta\lambda \;+\; B_{ii}(\Delta t) \;-\; \sum_{j\neq i} \frac{\frac{1}{2}\,B_{ji}^2(\Delta t)}{\lambda_j(T)-\lambda_i(T)+B_j(\Delta t)-\Delta\lambda}.$$

Hence, to leading order in small Δt, we can ignore $\Delta\lambda$ in the denominator, replace $B_{ji}^2(\Delta t)$ by Δt as its expectation,[1] ignore the random correction (as in Itô calculus), and obtain the desired SDE. We do not go into further details here, but the details are abundant in the literature, including the original work of Dyson [Dys62a]. □

Definition 10.3 (Dyson Brownian Motion)**.** Fix $\beta > 0$ and initial data $\big(\lambda_1(0) \geq \cdots \geq \lambda_n(0)\big)$. The *Dyson Brownian Motion* is the unique strong solution to the system of SDEs

$$d\lambda_i(t) \;=\; \frac{\beta}{2}\sum_{j\neq i}\frac{dt}{\lambda_i(t)-\lambda_j(t)} \;+\; dW_i(t), \quad i=1,\dots,n, \tag{10.2}$$

with the $W_i(t)$ being independent real standard Brownian motions. For $\beta = 1, 2, 4$, this coincides with the eigenvalue process of matrix Brownian motion (GOE, GUE, GSE).

Remark 10.4. Equation (10.2) succinctly captures the key idea that the eigenvalues repel each other. Note the singular drift term $\frac{1}{\lambda_i-\lambda_j}$ which pushes λ_i away from collisions with λ_j. This repulsion is so strong (for all $\beta > 0$) that eigenvalues will not cross (and thus remain ordered) with probability one.

[1] For other β, this will be $\beta\Delta t$, due to the dimensionality of the Brownian motion on the full rank matrix.

10.4 Mapping the GβE densities with the Dyson Brownian Motion

If the Dyson Brownian motion starts from zero[2] $\lambda_1(0) = \cdots = \lambda_N(0) = 0$, we expect that at time t, the density of eigenvalues is GβE,

$$\propto \prod_{i<j} |\lambda_i - \lambda_j|^\beta \exp\left\{-\frac{1}{2t}\sum_i \lambda_i^2\right\}.$$

This is evident for $\beta = 1, 2, 4$, when we have a matrix model, but not so much for other β. For other β, we would like to

- Make sense of the SDE and its solutions. We skip this part in the course.
- Make a computation checking that the above density is preserved under the SDE (10.2).

For example, in the $N = 1$ case, $d\lambda = dW(t)$ is a Markov process and one wants to show that

$$p(t,\lambda) \; = \; \frac{1}{\sqrt{2\pi\, t}} \exp\Big(-\frac{\lambda^2}{2\, t}\Big)$$

is preserved in the evolution. To verify this, one computes the generator of the semigroup, which for Brownian motion is

$$\frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}.$$

One then checks that

$$\frac{\partial}{\partial t}\, p(t,\lambda) \; = \; \frac{1}{2}\,\frac{\partial^2}{\partial\lambda^2}\, p(t,\lambda).$$

This is a direct computation.

For larger N, one needs to write down the corresponding generator and check that the same type of equation is satisfied. See Problem 10.7.3.

[2] And then the particles immediately repel each other and stay ordered for the whole time.

10.5 Determinantal structure for $\beta = 2$

To understand the determinantal structure of the Dyson Brownian Motion, we first need the explicit transition probabilities:

Theorem 10.5 ($\beta = 2$ Dyson Brownian Motion transition probabilities)**.** *For $\beta = 2$, let $\lambda(t) = (\lambda_1(t) \geq \cdots \geq \lambda_N(t))$ follow Dyson Brownian Motion starting at $\lambda(0) = \mathbf{a} = (a_1 \geq \cdots \geq a_N)$. Then for each fixed time $t > 0$,*

$$P\big(\lambda(t) = \mathbf{x} \mid \lambda(0) = \mathbf{a}\big) \;=\; N!\,\Big(\frac{1}{\sqrt{2\pi t}}\Big)^N \prod_{1\leq i<j\leq N} \frac{x_i - x_j}{a_i - a_j}\;\det\Big[\exp\Big(-\frac{(x_i-a_j)^2}{2t}\Big)\Big]_{i,j=1}^N,$$

where $x_1 \geq \cdots \geq x_N$.

The proof of this theorem is given in the next Chapter 11, based on the Harish–Chandra–Itzykson–Zuber formula that we outline next.

10.6 Harish-Chandra–Itzykson–Zuber (HCIZ) integral

In this section, we give a self-contained derivation of the Harish–Chandra–Itzykson–Zuber (HCIZ) integral from first principles, in a form commonly used in Random Matrix Theory and particularly in the derivation of Dyson Brownian Motion transition densities.

10.6.1 Statement of the HCIZ formula

Let A and B be two $N \times N$ Hermitian matrices with (real) eigenvalues

$$\mathrm{Spec}(A) = (a_1, \ldots, a_N), \quad \mathrm{Spec}(B) = (b_1, \ldots, b_N).$$

We want to compute the integral

$$\mathcal{I}(A,B) \;:=\; \int_{U(N)} \exp\big(\mathrm{Tr}(A\,U\,B\,U^\dagger)\big)\,dU,$$

where $U(N)$ is the group of $N \times N$ unitary matrices equipped with its normalized Haar measure dU. The Harish–Chandra–Itzykson–Zuber formula states that

$$\int_{U(N)} e^{\,\mathrm{Tr}(A\,U\,B\,U^\dagger)}\,dU \;=\; \Big(\prod_{k=1}^{N-1} k!\Big)\frac{\det\big[e^{\,a_i b_j}\big]_{i,j=1}^N}{\prod_{1\leq i<j\leq N}(a_j - a_i)\,\prod_{1\leq i<j\leq N}(b_j - b_i)},$$

up to conventions for the normalization of the Haar measure. Many references fix the normalization constant as above.

10.6.2 Reduction to the diagonal case

The integrand $\exp(\mathrm{Tr}(A\,U\,B\,U^\dagger))$ depends on U only via conjugation. Exploiting the Haar measure's bi-invariance:

1. Diagonalize $A = V_A \,\mathrm{diag}(a_1, \dots, a_N)\, V_A^\dagger$.
2. Diagonalize $B = V_B \,\mathrm{diag}(b_1, \dots, b_N)\, V_B^\dagger$.
3. Notice

$$\mathrm{Tr}(A\,U\,B\,U^\dagger) \;=\; \mathrm{Tr}\Big(\mathrm{diag}(a)\,\big(V_A^\dagger U V_B\big)\,\mathrm{diag}(b)\,\big(V_B^\dagger U^\dagger V_A\big)\Big).$$

Setting $W = V_A^\dagger\, U\, V_B$ preserves the Haar measure. Thus

$$\int_{U(N)} e^{\,\mathrm{Tr}(A\,U\,B\,U^\dagger)}\,dU \;=\; \int_{U(N)} e^{\,\mathrm{Tr}(\mathrm{diag}(a)\,W\,\mathrm{diag}(b)\,W^\dagger)}\,dW.$$

Therefore, we may assume $A = \mathrm{diag}(a)$ and $B = \mathrm{diag}(b)$. In that case,

$$\mathrm{Tr}\big(A\,U\,B\,U^\dagger\big) \;=\; \sum_{i,j=1}^{N} a_i\,b_j\,|U_{ij}|^2.$$

Hence

$$\int_{U(N)} \exp\Big(\mathrm{Tr}(A\,U\,B\,U^\dagger)\Big)\,dU \;=\; \int_{U(N)} \exp\Big(\sum_{i,j=1}^{N} a_i\,b_j\,|U_{ij}|^2\Big)\,dU. \tag{10.3}$$

10.6.3 Symmetry

Let $f(A,B)$ denote the right-hand side of (10.3). We have established that $f(A,B)$ must be:

1. Symmetric in the eigenvalues $\{a_1, \dots, a_N\}$ of A
2. Symmetric in the eigenvalues $\{b_1, \dots, b_N\}$ of B
3. Analytic in all variables when the eigenvalues are distinct

When some eigenvalues coincide, the function must behave appropriately. Specifically:

Lemma 10.6. *If $a_i = a_j$ for some $i \neq j$, then $f(A,B)$ must be invariant under permuting the corresponding b_i and b_j.*

Proof. When eigenvalues coincide, the corresponding eigenvectors can be chosen arbitrarily within the degenerate subspace. This means that when $a_i = a_j$, we can apply a unitary transformation that effectively swaps the roles of b_i and b_j without changing the integral. □

Remark on rigor. To make these symmetry arguments fully rigorous, one notes that $f(A, B)$ can be extended to an analytic function of the eigenvalues (even when they are treated as complex variables close to the real axis). Moreover, if some $a_i = a_j$, the existence of a unitary acting within the degenerate subspace justifies the required symmetry in (b_i, b_j). One also checks that $f(A, B)$ remains finite in the limit $(a_j - a_i) \to 0$ or $(b_j - b_i) \to 0$, enforcing vanishing at a rate that compensates for the factor in the denominator.

This constraint, combined with analyticity, forces $f(A, B)$ to vanish as $(a_j - a_i) \to 0$ or $(b_j - b_i) \to 0$ at a rate that exactly cancels the denominator's singularity. The form of the answer must therefore be:

$$f(A, B) = \frac{g(A, B)}{\prod_{1 \le i < j \le N}(a_j - a_i) \prod_{1 \le i < j \le N}(b_j - b_i)},$$

where $g(A, B)$ is analytic and antisymmetric in the $\{a_i\}$ and in the $\{b_i\}$ variables.

10.6.4 Conclusion of the argument

By the fundamental theorem of antisymmetric polynomials, $g(A, B)$ must be expressible as a product of the Vandermonde determinants and a symmetric function. Moreover, by examining the behavior under the scaling $A \mapsto tA$ and $B \mapsto B/t$, one shows that the only function with the correct analytic properties and scaling behavior is

$$g(A, B) = C_N \cdot \det[e^{\,a_i b_j}]_{i,j=1}^{N},$$

where C_N is a constant depending only on N. One can alternatively pin this down by checking that $f(A, B)$ satisfies a certain heat equation in A (or B), and thus matches the known solution $\det[e^{\,a_i b_j}]$ up to a constant.

Therefore, we have established that

$$\int_{U(N)} e^{\,\mathrm{Tr}(A\,U\,B\,U^\dagger)}\,dU \;=\; \Phi_N\,\frac{\det[e^{\,a_i b_j}]_{i,j=1}^{N}}{\prod_{1 \le i < j \le N}(a_j - a_i) \prod_{1 \le i < j \le N}(b_j - b_i)},$$

where $\Phi_N = C_N$ is a normalization constant independent of the eigenvalues. Through a separate calculation (see Problem 10.7.4), often involving either a small-time heat-kernel expansion or a rank-one reduction, one can determine that

$$\Phi_N = \prod_{k=1}^{N-1} k!. \tag{10.4}$$

10.7 Problems

10.7.1 Collisions

Show that two independent standard 1D Brownian motions, started at $a_1 \neq a_2$, almost surely intersect.

10.7.2 Estimate on the modulus of continuity

Let $B(t)$ be a standard 1D Brownian motion with $B(0) = 0$, defined as a process with independent increments and $B(t) - B(s) \sim \mathcal{N}(0, t-s)$, without any continuity assumptions.

Show that

$$\mathbb{E}\,|B(t) - B(s)|^2 \;\leq\; |t-s|$$

implies that that one can take an almost surely continuous modification of the function $t \mapsto B(t)$.

10.7.3 Generator for Dyson Brownian Motion

Consider the Dyson Brownian Motion(Definition 10.3) for general $\beta > 0$. The invariant measure for this process when started from zero is expected to be the distribution with density proportional to:

$$p_\beta(\lambda_1, \dots, \lambda_N) \propto \prod_{i<j} |\lambda_i - \lambda_j|^\beta \exp\left\{ -\frac{1}{2} \sum_{i=1}^N \lambda_i^2 \right\}.$$

Prove that this density is invariant under the Dyson SDE (10.2) by showing

$$\mathcal{L} p_\beta = 0,$$

where $\mathcal{L}$ is the infinitesimal generator of the process. Specifically, compute:

$$\mathcal{L}\rho = \frac{1}{2} \sum_{i=1}^N \frac{\partial^2}{\partial \lambda_i^2} \rho - \frac{\beta}{2} \sum_{i=1}^N \sum_{j \neq i} \frac{\partial}{\partial \lambda_i} \left(\frac{1}{\lambda_i - \lambda_j} \rho \right),$$

and verify that it indeed annihilates p_β.

10.7.4 Constant in the HCIZ formula

Show that in the Harish–Chandra–Itzykson–Zuber formula, the constant Φ_N is given by

$$\Phi_N = \prod_{k=1}^{N-1} k!,$$

by *directly* evaluating the left-hand side for the special case

$$A = \mathrm{diag}(x, 0, \ldots, 0), \quad B = \mathrm{diag}(y, 0, \ldots, 0).$$

In this rank-one case, note that

$$\mathrm{Tr}(A\,U\,B\,U^\dagger) \;=\; x\,y\,\bigl|U_{11}\bigr|^2.$$

You can then reduce the integral to one over the distribution of the first column of U, which is a vector uniformly distributed on the complex unit sphere $\mathbb{C}^N$ (under the normalized Haar measure). Use the known Jacobian for this parametrization to perform the integral and match it with the right-hand side evaluated at $(a_1, b_1) = (x, y)$ and $(a_2 = \cdots = a_N = b_2 = \cdots = b_N = 0)$.

Chapter 11

Asymptotics of Dyson Brownian Motion with an outlier

11.1 Recap

11.1.1 Dyson Brownian Motion (DBM)

We introduced a time-dependent model of random matrices by letting an $N \times N$ Hermitian matrix $\mathcal{M}(t)$ evolve in time so that each off-diagonal entry follows independent Brownian increments (real or complex depending on the symmetry class). Setting

$$\mathcal{M}(t) \;=\; \frac{1}{\sqrt{2}}\big(X(t) + X^{\dagger}(t)\big),$$

where $X(t)$ is an $N \times N$ matrix of i.i.d. Brownian motions, produces a self-adjoint matrix with a stochastically evolving spectrum. This model is full-rank matrix Brownian motion, and works well for $\beta = 1, 2, 4$. For other β, we need an SDE to describe the evolution of the eigenvalues (particles).

11.1.2 Eigenvalue SDE

Denote by $\lambda_1(t) \geq \cdots \geq \lambda_N(t)$ the ordered eigenvalues of $\mathcal{M}(t)$. Dyson showed that these eigenvalues form a continuous-time Markov process satisfying the SDE

$$d\lambda_i(t) \;=\; \frac{\beta}{2} \sum_{j \neq i} \frac{dt}{\lambda_i(t) - \lambda_j(t)} \;+\; dW_i(t), \quad i = 1, \ldots, N,$$

where $\beta > 0$ and $W_i(t)$ are independent standard real Brownian motions. For classical random matrix ensembles ($\beta = 1, 2, 4$), this SDE describes how the eigenvalues evolve under real symmetric (GOE), Hermitian (GUE), or quaternionic (GSE) Brownian motion — in the last Chapter 10 we discussed the cases $\beta = 1, 2$ in detail. A key feature is the *repulsion* term $\frac{1}{\lambda_i - \lambda_j}$, which prevents collisions (and ensures the ordering remains intact).

11.1.3 Preservation of GβE density

A fundamental result is that starting from all eigenvalues at 0, the distribution of $\lambda(t)$ at time t has the joint density proportional to

$$\prod_{i<j} |\lambda_i - \lambda_j|^\beta \ \exp\Big\{-\tfrac{1}{2t} \sum_i \lambda_i^2\Big\},$$

matching the Gaussian β-Ensemble (GβE) law. Hence DBM provides a dynamical realization of GβE. Invariance can be checked by verifying that this density is annihilated by the generator of the SDE.

11.1.4 Harish–Chandra–Itzykson–Zuber (HCIZ) integral

The HCIZ integral is a key tool for computing matrix integrals involving traces. For two Hermitian matrices A and B with eigenvalues $(a_1, \ldots, a_N)$ and $(b_1, \ldots, b_N)$, it states (in one common normalization):

$$\int_{U(N)} \exp\big(\mathrm{Tr}(A\,U\,B\,U^\dagger)\big)\, dU \;=\; \prod_{k=1}^{N-1} k!\, \frac{\det\big[e^{\,a_i b_j}\big]_{i,j=1}^N}{\prod_{1\le i<j\le N}(a_j - a_i)\, \prod_{1\le i<j\le N}(b_j - b_i)}\,.$$

This formula is instrumental in deriving transition densities for $\beta = 2$ Dyson Brownian Motion.

11.2 Optional: proof of HCIZ integral via representation theory

In this section, we outline a standard argument (adapted from the theory of symmetric functions and representation theory of the unitary group) that leads to a proof of the Harish–Chandra–Itzykson–Zuber formula. It is often referred to as the "orbital integral" or "character expansion" approach.

Step 1. Setting up the integral and Schur expansions. Let A and B be two $N \times N$ diagonalizable matrices, with eigenvalues $a_1, \ldots, a_N$

and $\lambda_1, \ldots, \lambda_N$ respectively. Denote by $D_a = \mathrm{diag}(a_1, \ldots, a_N)$ and $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. We want to evaluate the integral

$$I \;=\; \int_{U(N)} \exp\bigl(\mathrm{Tr}(D_a\, U\, D_\lambda\, U^\dagger)\bigr)\, dU$$

over the Haar measure on $U(N)$.

Since $\mathrm{Tr}(B) = p_1(B)$ in the language of power sums (where $p_1(x_1, x_2, \ldots) = x_1 + x_2 + \ldots$), we have

$$\exp\bigl(\mathrm{Tr}(B)\bigr) \;=\; \exp\bigl(p_1(B)\bigr).$$

One can use a known expansion [Mac95]

$$e^{p_1(B)} \;=\; \sum_{m=0}^{\infty} \frac{p_1^m(B)}{m!} \;=\; \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\mu:|\mu|=m} \dim(\mu)\, s_\mu(B),$$

where the sum is over all partitions μ of size m, and $s_\mu(\cdot)$ is the Schur polynomial (or Schur function) indexed by μ. The coefficient $\dim(\mu)$ is the dimension of the corresponding representation of S_m.

We set $B = D_a\, U\, D_\lambda\, U^\dagger$ and write

$$I \;=\; \int_{U(N)} \exp\Bigl(\mathrm{Tr}(D_a\, U\, D_\lambda\, U^\dagger)\Bigr)\, dU \;=\; \int_{U(N)} \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\mu:|\mu|=m} \dim(\mu)\; s_\mu\bigl(D_a\, U\, D_\lambda\, U^\dagger\bigr)\, dU.$$

One can exchange the integral and the sum (the series converges absolutely for all matrix arguments), giving

$$I \;=\; \sum_{m=0}^{\infty} \sum_{\mu:|\mu|=m} \frac{\dim(\mu)}{m!} \int_{U(N)} s_\mu(D_a\, U\, D_\lambda\, U^\dagger)\, dU. \tag{11.1}$$

Step 2. Orthogonality of characters and the Unitary group. The Schur functions $s_\mu(\cdot)$ can be seen as irreducible characters of the unitary group $U(N)$ (up to a normalization factor) when restricted to N-tuples of eigenvalues.[1]

[1] s_μ for $\ell(\mu) \le N$ can be viewed as the character of the corresponding polynomial representation of $GL(N, \mathbb{C})$, then restricted to $U(N)$. If $\ell(\mu) > N$, the function s_μ vanishes on $U(N)$. Thus, we need to impose the condition $|a_i| = |\lambda_i| = 1$ (so that $D_a, D_\lambda \in U(N)$) to ensure immediate applicability of representation theory of $U(N)$, then extend to general $\{a_i\}$ and $\{\lambda_i\}$ by analytic continuation.

Proposition 11.1 (Functional equation for characters of compact groups)**.** *Let G be a compact group with normalized Haar measure dh, and let χ be an irreducible character of a finite-dimensional representation of G. Then for any elements $g_1, g_2 \in G$, the following relation holds:*

$$\int_G \chi(g_1 h g_2 h^{-1}) dh = \frac{\chi(g_1)\chi(g_2)}{\dim V}, \tag{11.2}$$

where $\dim V = \chi(e)$ is the dimension of the representation space.

Remark 11.2. A similar relation holds for characters of finite groups.

By Proposition 11.1, the integral over $U(N)$ in (11.1) can be evaluated as

$$\int_{U(N)} s_\mu(D_a\, U\, D_\lambda\, U^\dagger)\, dU \;=\; \frac{1}{\mathrm{Dim}_N(\mu)}\, s_\mu(a)\, s_\mu(\lambda),$$

where $\mathrm{Dim}_N(\mu)$ is the dimension of the corresponding irreducible representation of $U(N)$. Substituting back into (11.1) yields

$$I \;=\; \sum_{m=0}^{\infty} \sum_{\mu:|\mu|=m,\, \ell(\mu)\le N} \frac{\dim(\mu)}{m!}\, \frac{1}{\mathrm{Dim}_N(\mu)}\, s_\mu(a)\, s_\mu(\lambda),$$

where $\ell(\mu) \le N$ is needed for $s_\mu(\cdot)$ not to vanish on $U(N)$.

Step 3. Hook-length formulas and the final determinant. Next, one applies the hook-length formula and the hook-content formula to dimensions:

$$\dim \mu = \frac{|\mu|!}{\prod_{\square\in\mu} h(\square)}, \qquad \mathrm{Dim}_N(\mu) = \frac{\prod_{\square\in\mu}(N + c(\square))}{\prod_{\square\in\mu} h(\square)},$$

We have

$$\prod_{\square\in\mu}(N + c(\square)) = \prod_{i=1}^{N} \frac{(\mu_i + N - i)!}{(N-i)!},$$

so identifying $m_i = \mu_i + N - i$ gives

$$I = 0!1!\cdots(N-1)! \sum_{m_1>\ldots>m_N\ge 0} \frac{s_\mu(a)s_\mu(\lambda)}{m_1!\cdots m_N!},$$

which yields the HCIZ formula by the Cauchy-Binet summation.

11.3 Determinantal structure for $\beta = 2$

11.3.1 Transition density

Theorem 11.3 ($\beta = 2$ Dyson Brownian Motion Transition Probabilities)**.** *For $\beta = 2$, let $\lambda(t) = (\lambda_1(t) \geq \cdots \geq \lambda_N(t))$ follow Dyson Brownian Motion starting at $\lambda(0) = \mathbf{a} = (a_1 \geq \cdots \geq a_N)$. Then for each fixed time $t > 0$,*

$$P\big(\lambda(t) = \mathbf{x} \mid \lambda(0) = \mathbf{a}\big) \;=\; N!\,\big(\frac{1}{\sqrt{2\pi t}}\big)^N \prod_{1 \leq i < j \leq N} \frac{x_i - x_j}{a_i - a_j}\; \det\Big[\exp\Big(-\frac{(x_i - a_j)^2}{2t}\Big)\Big]_{i,j=1}^N,$$

where $x_1 \geq \cdots \geq x_N$.

Proof. Consider an $N \times N$ Hermitian matrix process $X(t)$ whose entries perform independent complex Brownian motions (so that $X(t)$ is distributed as $A + \sqrt{t}\,\mathrm{GUE}$ at each fixed time, with $A = \mathrm{diag}(a_1, \ldots, a_N)$). Its eigenvalues $\lambda_1(t) \geq \cdots \geq \lambda_N(t)$ evolve exactly according to the $\beta = 2$ Dyson Brownian Motion.

The density of X at time t, viewed as a random matrix, is proportional to

$$\exp\Big(-\tfrac{1}{2t}\,\mathrm{Tr}\big(X - A\big)^2\Big).$$

If we replace A by $U\,A\,U^\dagger$ for any fixed unitary U, the law of X remains the same (this follows from the unitary invariance of the GUE). Thus the distribution of the eigenvalues of X is unchanged by such conjugation.

One writes

$$\int_{U(N)} \exp\Big(-\tfrac{1}{2t}\,\mathrm{Tr}\big(X - U\,A\,U^\dagger\big)^2\Big)\,dU \;=\; (\text{const.}) \times [\text{HCIZ integral in the variables } (X, A)],$$

which by the Harish–Chandra–Itzykson–Zuber formula leads to a product of determinants and a factor that is precisely

$$\exp\Big(-\tfrac{1}{2t}\sum_{i=1}^N x_i^2 - \tfrac{1}{2t}\sum_{i=1}^N a_i^2\Big)\,\frac{\det\Big[\exp\big(\frac{x_i\,a_j}{t}\big)\Big]}{\prod_{i<j}(x_i - x_j)\,(a_i - a_j)},$$

where $x_1, \ldots, x_N$ are the eigenvalues of X.

To convert this matrix distribution into a distribution on eigenvalues alone, we multiply by the usual Vandermonde Jacobian $\prod_{i<j}(x_i - x_j)^2$ (which comes from integrating out the unitary degrees of freedom). This produces exactly

$$N!\,\big(\tfrac{1}{\sqrt{2\pi t}}\big)^N \prod_{i<j} \frac{x_i - x_j}{a_i - a_j}\; \det\Big[\exp\Big(-\tfrac{(x_i - a_j)^2}{2t}\Big)\Big].$$

Hence we obtain the stated transition probability for the Dyson Brownian Motion at $\beta = 2$. □

Remark 11.4. The factor $N!\,(\frac{1}{\sqrt{2\pi t}})^N$ arises naturally from normalizing the Gaussian increments and accounts for the ordering $\lambda_1 \geq \cdots \geq \lambda_N$. The determinant and product factors encode the eigenvalue "repulsion" characteristic of $\beta = 2$ random matrices.

11.3.2 Determinantal correlations

Theorem 11.5 (Determinantal structure for $\beta = 2$ DBM)**.** *Let $\{x_1(t), \ldots, x_n(t)\}$ be the eigenvalues at time $t > 0$ of the $\beta = 2$ Dyson Brownian Motion started at initial locations $(a_1, \ldots, a_n)$ at time 0. Equivalently, these $x_i(t)$ are the eigenvalues of*

$$A + \sqrt{t}\,G,$$

where $A = \operatorname{diag}(a_1, \ldots, a_n)$ and G is a random Hermitian matrix from the GUE. Then the (random) point configuration $\{x_i(t)\}$ forms a determinantal point process with correlation kernel

$$K_t(x, y) = \frac{1}{(2\pi)^2\, t} \int\int \exp\Bigl(\frac{w^2 - 2\,y\,w}{2\,t}\Bigr) \Big/ \exp\Bigl(\frac{z^2 - 2\,x\,z}{2\,t}\Bigr) \prod_{i=1}^{n} \frac{w - a_i}{z - a_i}\, \frac{dw\,dz}{w - z}.$$

Here z goes around all the points $a_1, \ldots, a_n$ in the positive direction, and the w contour passes from $-i\infty$ to $i\infty$, to the right of the z contour.

- If $a_1 = \cdots = a_n = 0$ and $t = 1$, this kernel reduces to the familiar correlation kernel of the GUE (see Chapter 6).
- One can use this formula to study the Baik–Ben Arous–Péché (BBP) [BBP05] phase transition for $\beta = 2$, which deals with finite rank perturbations of the GUE random matrix ensemble. Indeed, rank r perturbation corresponds to taking $a_1, \ldots, a_r \neq 0$, and $a_{r+1} = \cdots = a_n = 0$.

11.3.3 On the proof of determinantal structure

The idea of the proof of Theorem 11.5 is to represent the measure (the transition density) as a product of determinants. In general, if a measure is given as a product of determinants, there is a well-studied method (biorthogonal ensembles and, more generally, the Eynard–Mehta theorem) to compute the determinantal correlation kernel. We refer to [BR05], [Bor11] for a detailed exposition in the discrete case (which is arguably more transparent). The first step for the Dyson Brownian Motion is as follows.

Lemma 11.6 (Density representation)**.** *Let $P_t(x \to y)$ be the transition probability kernel of standard Brownian motion,*

$$P_t(x \to y) \;=\; \frac{1}{\sqrt{2\pi\, t}}\, \exp\Big(-\tfrac{(x-y)^2}{2\,t}\Big).$$

Then the density of the eigenvalues $(x_1, \ldots, x_N)$ of DBM started at $(a_1, \ldots, a_N)$ at time 0 admits the representation

$$\lim_{s\to\infty} \left(\frac{1}{Z}\right) \det\Big[P_t\big(a_i \to x_j\big)\Big]_{i,j=1}^{N} \det\Big[P_s\big(x_i \to k-1\big)\Big]_{i,k=1}^{N}. \tag{11.3}$$

Remark 11.7. This representation (11.3) is related to an alternative description of the $\beta = 2$ Dyson Brownian Motion as an ensemble of noncolliding Brownian motions (that is, independent Brownian motions, conditioned to never collide).

Proof of Lemma 11.6. The first determinant (as $s \to \infty$) matches the determinant we have in Theorem 11.3. It remains to analyze the second determinant

$$\det\Big[P_s\big(x_j \to k-1\big)\Big]_{j,k=1}^{N} \;=\; \det\Big[\tfrac{1}{\sqrt{2\pi\, s}}\, \exp\Big(-\tfrac{\big((k-1)-x_j\big)^2}{2\,s}\Big)\Big]_{j,k=1}^{N}.$$

We may ignore the factor $\frac{1}{\sqrt{2\pi\, s}}$ in each entry since it does not depend on x_j. Inside the exponential,

$$-\frac{\big((k-1)-x_j\big)^2}{2\,s} \;=\; -\frac{x_j^2}{2\,s} \;+\; \frac{x_j\,(k-1)}{s} \;-\; \frac{(k-1)^2}{2\,s}.$$

Thus, up to the factor $\exp\big(-\frac{(k-1)^2}{2\,s}\big)$ (which depends only on k and hence is independent of each x_j), we can factor out $\exp\big(-\frac{x_j^2}{2\,s}\big)$ from row j. Consequently, the nontrivial part of the determinant becomes

$$\det\Big[e^{\frac{x_j\,(k-1)}{s}}\Big]_{j,k=1}^{N}.$$

Recognize this as a Vandermonde-type determinant in the variables $e^{x_j/s}$. Indeed,

$$\det\Big[e^{\frac{x_j\,(k-1)}{s}}\Big]_{j,k=1}^{N} \;=\; \prod_{1\le i<j\le N} \Big(e^{\frac{x_i}{s}} - e^{\frac{x_j}{s}}\Big).$$

As $s \to \infty$, we expand $e^{\frac{x_i}{s}} = 1 + \frac{x_i}{s} + O\left(\frac{1}{s^2}\right)$, so each difference $\left(e^{\frac{x_i}{s}} - e^{\frac{x_j}{s}}\right) \sim \frac{x_i - x_j}{s}$. Hence,

$$\prod_{1 \le i < j \le N} \left(e^{\frac{x_i}{s}} - e^{\frac{x_j}{s}}\right) \sim \frac{1}{s^{\frac{N(N-1)}{2}}} \prod_{1 \le i < j \le N} (x_i - x_j).$$

Combining all these factors and matching with the first determinant (as $s \to \infty$) verifies the claimed product form, up to overall constants that do not depend on the variables x_j. This completes the proof. □

Then, the product of determinants idea (biorthogonal ensembles) applies to the density (11.3) before the limit $s \to \infty$, and simplifies after taking the limit. We omit the details here, see Problem 11.5.1.

11.4 Asymptotic analysis: signal plus noise

11.4.1 Setup

In this section, we provide a detailed derivation of how the rank-1 spike

$$A + \sqrt{G}, \qquad A = \operatorname{diag}(a, 0, \ldots, 0) \quad \text{with } a \in \mathbb{R},$$

affects the large-n and large-time behavior of the Dyson Brownian Motion at $\beta = 2$. See the simulation at https://lpetrov.cc/simulations/2025-01-28-bbp-transition/.[2]

We set $a_1 = a\sqrt{n}$ and $a_2 = a_3 = \cdots = a_n = 0$, which simplifies the product:

$$\prod_{i=1}^{n} (w - a_i) = (w - a\sqrt{n})\, w^{n-1}, \qquad \prod_{i=1}^{n} (z - a_i) = (z - a\sqrt{n})\, z^{n-1}.$$

Let us also take $t = 1$ for simplicity, so that the limit shape (at least in the case $a = 0$, but also in general) is supported by $[-2\sqrt{n}, 2\sqrt{n}]$. Let us also make the change of the integration variables $w \to w\sqrt{n}$, $z \to z\sqrt{n}$.

Hence, the correlation kernel becomes

$$K_t(x, y) = \frac{\sqrt{n}}{(2\pi)^2} \int\int \exp\left(\frac{nw^2 - 2\,y\,w\sqrt{n}}{2}\right) \Big/ \exp\left(\frac{nz^2 - 2\,x\,z\sqrt{n}}{2}\right) \frac{w - a}{z - a} \left(\frac{w}{z}\right)^{n-1} \frac{dw\,dz}{w - z}. \tag{11.4}$$

Here:

[2] Note that the simulation has $\beta = 1$ (real matrices), so the edge is at $\sqrt{2}$, and the critical value of the spike is at $1/\sqrt{2}$.

- The z-contour is a small positively oriented loop around $z = a$, and also around $z = 0$, so that it encircles these two singularities but excludes w.
- The w-contour is a vertical line (or an equivalent contour from $-i\infty$ to $i\infty$) passing to the right of all singularities (i.e. to the right of z).

Note that to capture the edge behavior, we need to set $x = y = 2$ plus lower order terms. Let us make this substitution $x = 2\sqrt{n} + x'$, $y = 2\sqrt{n} + y'$, and the scale of x', y' will be determined later (but for now we assume that they are $o(\sqrt{n})$).

11.4.2 Outline of the steepest descent approach

We aim to understand the behavior of (11.4) in the regime $n \to \infty$, especially near the largest eigenvalue $\lambda_1(t)$. Recall from standard GUE (i.e. $a = 0$) that the top of the spectrum is about $2\sqrt{n}$. The presence of the rank-1 spike a can drastically modify the top eigenvalue if a is large enough to produce an "outlier." Our goal is to detect precisely how this occurs by analyzing the double contour integral via steepest descent.

For large n, the integral localizes around these double critical point. Any crossing from z- to w-contour may pick up residues, which account for separate contributions (leading, for instance, to the Airy kernel in the unperturbed GUE). We track how the spike a changes these deformations.

11.4.3 Asymptotics

Set

$$S(w; y') \;=\; \tfrac{w^2}{2} - 2\,w - y'w/\sqrt{n} \;+\; \frac{n-1}{n}\ln(w).$$

Then the integrand in (11.4) is

$$\frac{\exp\left\{n\big[S(w; y') - S(z; x')\big]\right\}}{w - z}\,\frac{w - a}{z - a}.$$

To capture the Airy behavior, we can ignore y', and find the double critical point of $S(w; 0)$. It is equal to $w_c = 1$, and we would like to bring the z and w contours to intersect at $w_c = 1$. Note however that the old z contour must encircle $z = a$ and $z = 0$, and $z = a$ is a pole of the integrand. The w contour must always be to the right of the z contour.

We see that there are three regimes, which we consider in the next three subsections.

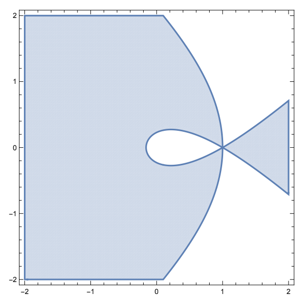

Figure 11.1: The plot of the region $\operatorname{Re} S(z) - \operatorname{Re} S(1) > 0$ at the edge, in the neighborhood of the double critical point $w_c = 1$. The new w contour should pass through the shaded region, and the new z must stay in the non-shaded region.

11.4.4 Airy kernel

If $a < 1$, we can deform the z contour to encircle $z = 0$ and $z = a$, and the w contour to pass through $w = 1$. This will lead to the Airy kernel, and the derivation is the same as in Chapter 7. We obtain[3]

$$z = 1 + \frac{Z}{n^{1/3}}, \quad w = 1 + \frac{W}{n^{1/3}}, \qquad x' = \frac{\xi}{n^{1/6}}, \quad y' = \frac{\eta}{n^{1/6}}, \qquad \frac{1}{n^{1/6}} K_n \to K_{\mathrm{Airy}}(\xi, \eta).$$

Here

$$K_{\mathrm{Airy}}(\xi, \eta) = \frac{1}{(2\pi i)^2} \iint \frac{\exp\left\{\frac{W^3}{3} - \xi W - \frac{Z^3}{3} + \eta Z\right\}}{W - Z} \, dW \, dZ.$$

Indeed, the only one new thing that happens here is that $a < 1$, and so

$$\frac{w-a}{z-a} = \frac{1 - a + W/n^{1/3}}{1 - a + Z/n^{1/3}} = 1 + O(n^{-1/3}), \tag{11.5}$$

so this term does not contribute to the asymptotics of the kernel.

11.4.5 BBP transition and the deformed Airy kernel

If $a = 1$, the behavior is going to be critical — we still will be able to get the same scaling, but the limiting kernel will be different. Moreover, looking at

[3]Here and below, we understand the convergence of the kernels is up to a gauge transformation of the form $K(x, y) \mapsto \frac{f(x)}{f(y)} K(x, y)$.

(11.5), we see that we need to critically rescale a, so

$$a = 1 + An^{-1/3}, \qquad \frac{w-a}{z-a} = \frac{W-A}{Z-A} + O(n^{-4/3}), \qquad \frac{1}{n^{1/6}} K_n \to \tilde{K}_{\mathrm{Airy}}(\xi,\eta),$$

where

$$\tilde{K}_{\mathrm{Airy}}(\xi,\eta) = \frac{1}{(2\pi i)^2} \iint \frac{\exp\left\{\frac{W^3}{3} - \xi W - \frac{Z^3}{3} + \eta Z\right\}}{W-Z} \frac{W-A}{Z-A}\, dW\, dZ.$$

This kernel is the BBP transition kernel, first obtained in the seminal paper by Baik–Ben Arous–Péché [BBP05]. The spiked top eigenvalue distribution (and the Tracy–Widom distribution) are widely used in statistics of high-dimensional, highly correlated data.

11.4.6 Gaussian regime

Finally, for $a > 1$, we cannot deform the integration contours so that they pass through the double critical point $w_c = 1$. Instead, we can make the contours pass through the point a itself, and scale the integration variables w, z around a on the scale $n^{-1/2}$ and not $n^{-1/3}$.

Moreover, we need to make x, y to scale around a different location instead of $2\sqrt{n}$. We can find this location by first considering $x = c\sqrt{n}$ and expanding as $n \to \infty$:

$$\begin{aligned} &n\left(\frac{w^2}{2} - yw/\sqrt{n} + \log w\right)\Big|_{w=a+W/\sqrt{n},\ y=c\sqrt{n}+\eta} \\ &= n\left(\frac{a^2}{2} - ac + \log(a)\right) + \sqrt{n}\left(-a\eta + aW + \frac{W}{a} - cW\right) - \frac{W^2}{2a^2} + \frac{W^2}{2} - \eta W. \end{aligned}$$

The term by n is the same in $S(w)$ and $S(z)$ and thus cancels out. The term by $\sqrt{n}$ depends on W and cannot be simply removed by a gauge transformation, so we need to match c. We have

$$c = a + \frac{1}{a}.$$

Remark 11.8. You can go to `https://lpetrov.cc/simulations/2025-01-28-bbp-transition/` and set the parameter θ (which is the same as a) to an integer, make N large, and check that the location of the top or bottom eigenvalue becomes exactly $a + 1/a$. (Despite the fact that the simulation at the link is for $\beta = 1$.)

Setting $c = a + 1/a$, we have

$$nS \sim -\frac{W^2}{2a^2} + \frac{W^2}{2} - \eta W,$$

and thus the distribution of the top eigenvalue is given by a Fredholm determinant with the kernel

$$K_G(\xi, \eta) = \frac{1}{(2\pi)^2} \int\int \exp\left\{\frac{a^{-2}-1}{2}(Z^2 - W^2) - \eta W + \xi Z\right\} \cdot \frac{W}{Z} \cdot \frac{dW\, dZ}{W - Z}$$

Note that the factor $\sqrt{n}$ in front of K_t is precisely removed by the scaling of w, z, and there is no additional scaling coming from the map $(x, y) \mapsto (\xi, \eta)$. The contribution from $(w - a)/(z - a)$ becomes W/Z.

The integration contours in K_G are such that $Re(W^2) > 0$ and $Re(Z^2) < 0$ on them, and this can be achieved by the contour deformation. Indeed, in the new variables, the behavior at $W = Z = 0$ is quadratic, so the Z contour must pass on the left, and the W contour must be on the right. One can check that this contour deformation is possible.

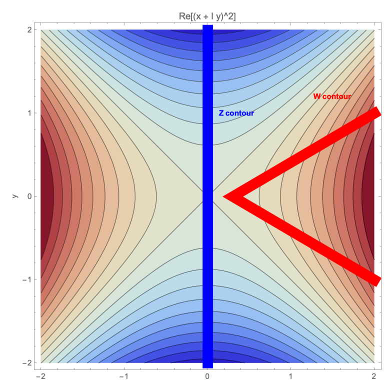

Figure 11.2: The contour plot of $\Re(Z^2)$ around zero. Blue shades correspond to negative values, and yellow to positive. The Z contour must pass through the blue region and becomes vertical, and the W contour must stay in the yellow region, and becomes a union of two half-lines, which are at the angle $< \frac{\pi}{4}$ from the real line.

11.4.7 Matching Fredholm determinant to the Gaussian distribution

Let us renormalize the integration variables to remove the factor $a^{-2} - 1$ in front of the squares, and match $\det(1 - K_G)_{\geq x}$ to the Gaussian distribution (see also Problem 11.5.5 for another way to match). We will work with

$$K_G(\xi,\eta) = \frac{1}{(2\pi)^2}\int\int \exp\left\{\frac{1}{2}(Z^2-W^2)-\eta W+\xi Z\right\}\cdot\frac{W}{Z}\cdot\frac{dW\,dZ}{W-Z}$$

The discussion below is informal, but can be easily made rigorous.

Step 1. Partial fractions and decomposition. Observe that

$$\frac{W}{Z(W-Z)} \;=\; \frac{1}{Z} \;+\; \frac{1}{W-Z}.$$

Thus we can write

$$K_G(\xi,\eta) = K^{(1)}(\xi,\eta) + K^{(2)}(\xi,\eta),$$

where

$$K^{(1)}(\xi,\eta) = \frac{1}{(2\pi)^2}\int\int \exp\Bigl(\tfrac{1}{2}(Z^2-W^2)+\xi Z-\eta W\Bigr)\,\frac{1}{Z}\,dW\,dZ,$$
$$K^{(2)}(\xi,\eta) = \frac{1}{(2\pi)^2}\int\int \exp\Bigl(\tfrac{1}{2}(Z^2-W^2)+\xi Z-\eta W\Bigr)\,\frac{1}{W-Z}\,dW\,dZ.$$

The term $K^{(1)}$ has a factor $\frac{1}{Z}$ independent of $W-Z$, while $K^{(2)}$ contains the remaining part $\frac{1}{W-Z}$.

Step 2. Analysis of $K^{(1)}$. Focus on

$$K^{(1)}(\xi,\eta) = \frac{1}{(2\pi)^2}\Bigl(\int e^{\frac{1}{2}Z^2+\xi Z}\,\frac{dZ}{Z}\Bigr)\Bigl(\int e^{-\frac{1}{2}W^2-\eta W}\,dW\Bigr).$$

The operator $K^{(1)}$ is a rank-1 operator in the variables ξ,η:

$$K^{(1)}(\xi,\eta) = u(\xi)\,v(\eta)$$

for some functions $u(\cdot)$ and $v(\cdot)$ of one variable each. Hence $K^{(1)}$ has at most one nonzero eigenvalue (its trace).

Step 3. Representation of $K^{(2)}$ and the key identity. For $K^{(2)}$, we use

$$\frac{1}{W-Z} = \int_0^\infty e^{-t(W-Z)}\,dt$$

(again justified by the choice of integration contours). Then

$$K^{(2)}(\xi,\eta) = \int_0^\infty \Bigl[\frac{1}{2\pi i}\int e^{\frac{1}{2}Z^2+(\xi+t)Z}\,dZ\Bigr]\Bigl[\frac{1}{2\pi i}\int e^{-\frac{1}{2}W^2-(\eta+t)W}\,dW\Bigr]dt.$$

Denote

$$A(\xi,t)=\frac{1}{2\pi i}\int e^{\frac{1}{2}Z^2+(\xi+t)Z}\,dZ,\quad B(t,\eta)=\frac{1}{2\pi i}\int e^{-\frac{1}{2}W^2-(\eta+t)W}\,dW.$$

Hence $K^{(2)}(\xi,\eta)=\int_0^\infty A(\xi,t)\,B(t,\eta)\,dt$. In operator form on suitable spaces, this reads $K^{(2)}=A\,B$, and one checks $B\,A=I$ (*identity on the t-variable space*), so $A\,B$ and $B\,A$ share the same nonzero spectrum. Indeed,

$$BA(s,t)=\frac{1}{(2\pi i)^2}\int_{\mathbb{R}}d\xi\int dW\,dZ e^{-\frac{1}{2}W^2+\frac{1}{2}Z^2}e^{-(s+\xi)W+(\xi+t)Z}.$$

Integrating over ξ in the Fourier sense yields the delta:

$$\int_{\mathbb{R}}d\xi e^{\xi(Z-W)}=2\pi\delta(Z-W).$$

Integrating in W is again an integral of $e^{(t-s)W}$, and thus, the second 2π disappears, and we arrive at $BA(s,t)=\delta(s-t)$, which is the kernel of the identity operator.

We conclude that AB is a projection, since $(AB)^2=ABAB=AB$.

For the rest of the analysis, continue to Problem 11.5.4.

11.5 Problems

11.5.1 Biorthogonal ensembles

Derive Theorem 11.5 from Lemma 11.6 using the orthogonalization process similar to Chapter 5, and then taking the limit as $s\to\infty$.

11.5.2 Scaling of the kernel

Let $a_i=0$ in Theorem 11.5. Find α such that $t^\alpha K_t(x/\sqrt{t},y/\sqrt{t})$ is independent of t. Can you explain this value of α?

11.5.3 Gaussian regime and integration contours

Check that the contour deformation from (z,w) to (Z,W) passing through a described in Section 11.4.6 is valid.

11.5.4 Gaussian kernel

Finish the proof of the Fredholm determinant representation of the Gaussian cumulative distribution function by manipulation with Fredholm determinants, which was started in Section 11.4.7.

11.5.5 GUE kernel

Consider the following generalization of the kernel K_G from Section 11.4.6:

$$K_G^m(\xi,\eta) = \frac{1}{(2\pi)^2}\int\int \exp\left\{\frac{1}{2}(Z^2 - W^2) - \eta W + \xi Z\right\}\cdot\left(\frac{W}{Z}\right)^m \frac{dW\,dZ}{W-Z},$$

where $m \geq 1$ is an integer and the contours are as in Figure 11.2. Show that the Fredholm determinant $\det\left(1 - K_G^m\right)_{L^2(x,+\infty)}$ is the cumulative distribution function of the largest eigenvalue of the $m \times m$ GUE matrix, that is, $\mathbb{P}(\lambda_{\max}^{(m\times m)} \leq x)$.

Chapter 12

Title TBD

12.1 Recap

In our last lecture, we explored the asymptotics of Dyson Brownian Motion with an outlier. We specifically focused on the phase transition that occurs when a rank-1 perturbation is applied to a random matrix ensemble.

12.1.1 Dyson Brownian Motion with Determinantal Structure

We established that for $\beta = 2$, the eigenvalues of the time-evolved process form a determinantal point process. The transition probability from an initial configuration $\mathbf{a} = (a_1 \geq \cdots \geq a_N)$ to a configuration $\mathbf{x} = (x_1 \geq \cdots \geq x_N)$ at time t is given by:

$$P(\lambda(t) = \mathbf{x} \mid \lambda(0) = \mathbf{a}) = N!\Big(\frac{1}{\sqrt{2\pi t}}\Big)^N \prod_{1\leq i<j\leq N} \frac{x_i - x_j}{a_i - a_j} \det\left[\exp\left(-\frac{(x_i-a_j)^2}{2t}\right)\right]_{i,j=1}^N \tag{12.1}$$

This determinantal structure enabled us to derive the correlation kernel:

$$K_t(x,y) = \frac{1}{(2\pi)^2 t}\int\int \exp\left(\frac{w^2-2yw}{2t}\right)\Big/ \exp\left(\frac{z^2-2xz}{2t}\right)\prod_{i=1}^n \frac{w-a_i}{z-a_i}\frac{dw\,dz}{w-z} \tag{12.2}$$

where the contours of integration are specified to maintain analytical properties.

12.1.2 The BBP Phase Transition

The central focus was the Baik-Ben Arous-Péché (BBP) phase transition that occurs with finite-rank perturbations of GUE matrices. For the rank-1 case, we analyzed:

$$A + \sqrt{t}G, \quad \text{where } A = \mathrm{diag}(a\sqrt{n}, 0, \ldots, 0) \tag{12.3}$$

Through asymptotic analysis using steepest descent methods, we identified three distinct regimes:

1. **Airy regime** ($a < 1$): The largest eigenvalue follows the Tracy-Widom GUE distribution, just as in the unperturbed case. The spike is too weak to escape the bulk.

2. **Critical regime** ($a = 1$): A transitional behavior occurs when $a = 1 + An^{-1/3}$, leading to a deformed Airy kernel:

$$\tilde{K}_{\mathrm{Airy}}(\xi, \eta) = \frac{1}{(2\pi i)^2} \iint \frac{\exp\left\{\frac{W^3}{3} - \xi W - \frac{Z^3}{3} + \eta Z\right\}}{W - Z} \frac{W - A}{Z - A} dW\, dZ \tag{12.4}$$

3. **Gaussian regime** ($a > 1$): The largest eigenvalue separates from the bulk, becoming an "outlier" centered at $a + 1/a$. Its fluctuations follow a Gaussian distribution rather than the Tracy-Widom law.

12.2 A window into universality: Airy line ensemble

The edge scaling limit of Dyson Brownian Motion is a universal object for $\beta = 2$ models and determinantal structures (and far beyond); which includes the GUE Tracy-Widom distribution as a marginal. This is also intimately related to the KPZ universality class, which we will touch upon in the next few lectures. GUE formulas provide us with a powerful lens through which to examine these universality phenomena. In this section, we discuss the limiting behavior of Dyson Brownian Motion near the spectral edge, highlighting two of its fundamental properties: Brownian Gibbs property and characterization.

Theorem 12.1 (Edge scaling limit to Airy line ensemble)**.** *Consider an $N \times N$ GUE (Gaussian Unitary Ensemble) Dyson Brownian motion, i.e.,* *the stochastic process of eigenvalues $(\lambda_1(t) \geq \cdots \geq \lambda_N(t))_{t\in\mathbb{R}}$ evolving under Dyson's eigenvalue dynamics. After centering at the spectral edge parallel to the vector $\mathbf{v}_t$ and applying the Airy scaling (tangent axis scaled by $N^{-1/3}$ and fluctuations scaled by $N^{-1/6}$), the top k eigenvalue trajectories converge as $N \to \infty$ to the **Airy line ensemble**. In particular, for each fixed $k \geq 1$ the rescaled process*

$$(N^{1/6}[\lambda_i(\langle N^{-1/3}, N^{-1/6}\rangle \cdot \mathbf{v}) - c_{N,t}])_{1\leq i\leq k}$$

*converges in distribution (uniformly on compact t-intervals) to $(\mathcal{P}_i(t))_{1\leq i\leq k}$, where $\{\mathcal{P}_i(t)\}_{i\geq 1}$ is the parabolic Airy line ensemble. Consequently, the top curve $\mathcal{L}_1(t)$ is the **Airy$_2$ process**, which is the t-stationary scaling limit of the largest eigenvalue process.*

Let us define $\mathcal{L}_i(t) = \mathcal{P}_i(t) + t^2$, and call $\mathcal{L}$ the Airy Line Ensemble (without the word "parabolic"). One can (correctly) think that the parabola comes from the scaling window, which is of different proportions in the horizontal and vertical directions.

Theorem 12.2 (Airy line ensemble is Brownian Gibbsian [CH16])**.** *The parabolic Airy line ensemble $\{\mathcal{P}_i(t)\}_{i\geq 1}$ satisfies the **Brownian Gibbs property**. Namely, for any fixed index $k \geq 1$ and any finite time interval $[a, b]$, conditioning on the outside portions of the ensemble (i.e., $\{\mathcal{P}_j(t) : t \notin [a, b]\}$ for all j, and $\{\mathcal{P}_j(t) : j \neq k\}$ for $t \in [a, b]$), the conditional law of the kth curve on $[a, b]$ is that of a **Brownian bridge** from $(a, \mathcal{P}_k(a))$ to $(b, \mathcal{P}_k(b))$ **conditioned** to stay above the $(k+1)$th curve and below the $(k-1)$th curve on $[a, b]$. In particular, the Airy line ensemble is invariant under this resampling of a single curve by a conditioned Brownian bridge.*

At the same time, the Airy line ensemble $\mathcal{L}$ is time-stationary.

Theorem 12.3 (Unique characterization of ALE [AH23])**.** *The parabolic Airy line ensemble is the **unique** Brownian Gibbs line ensemble satisfying a natural parabolic curvature condition on the top curve. More precisely, let $\boldsymbol{\mathcal{P}} = (\mathcal{P}_1, \mathcal{P}_2, \ldots)$ be any line ensemble that satisfies the Brownian Gibbs property. Suppose in addition that the top line $\mathcal{P}_1(t)$ **approaches a parabola** of curvature $1/\sqrt{2}$ at infinity. Then $\boldsymbol{\mathcal{L}}$ must coincide (in law) with the **parabolic Airy line ensemble**, up to an overall affine shift of the entire ensemble.*

12.3 Problems

Chapter 13

Title TBD

13.1 Problems

Chapter 14

Title TBD

14.1 Problems

Chapter 15

Random matrices and topology

15.1 Problems

Bibliography

[AH23] A. Aggarwal and J. Huang, *Strong Characterization for the Airy Line Ensemble*, arXiv preprint (2023). arXiv:2308.11908. ↑175

[Akh65] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*, Translated by N. Kemmer, Hafner Publishing Co., New York, 1965. ↑30

[BBP05] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643–1697. arXiv:math/0403022 [math.PR]. ↑121, 163, 168

[BGK16] F. Benaych-Georges and A. Knowles, *Lectures on the local semicircle law for Wigner matrices*, Advanced Topics in Random Matrices, 2016. arXiv:1601.04055 [math.PR]. ↑36

[Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. ↑14, 16, 31, 32

[Bor11] A. Borodin, *Determinantal point processes*, Oxford handbook of random matrix theory, 2011. arXiv:0911.1153 [math.PR]. ↑72, 78, 163

[BR05] A. Borodin and E.M. Rains, *Eynard–Mehta theorem, Schur process, and their Pfaffian analogs*, J. Stat. Phys **121** (2005), no. 3, 291–317. arXiv:math-ph/0409059. ↑163

[CH16] I. Corwin and A. Hammond, *KPZ line ensemble*, Probability Theory and Related Fields **166** (2016), no. 1-2, 67–185. arXiv:1312.2600 [math.PR]. ↑175

[DE02] I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, Journal of Mathematical Physics **43** (2002), no. 11, 5830–5847. arXiv:math-ph/0206043. ↑49, 53, 61, 75, 117

[Dix05] A. L. Dixon, *Generalization of Legendre's Formula* $KE' - (K - E)K' = \frac{1}{2}\pi$, Proc. Lond. Math. Soc. **2** (1905), no. 1, 206–224. ↑9

[Dys62a] F.J. Dyson, *A Brownian motion model for the eigenvalues of a random matrix*, Jour. Math. Phys. **3** (1962), no. 6, 1191–1198. ↑9, 147, 151

[Dys62b] F.J. Dyson, *Statistical Theory of the Energy Levels of Complex Systems. I-III*, Jour. Math. Phys. **3** (1962), no. 166, 140–175. ↑9

[FF14] P. Ferrari and R. Frings, *Perturbed GUE minor process and Warren's process with drifts*, J. Stat. Phys **154** (2014), no. 1-2, 356–377. arXiv:1212.5534 [math-ph]. ↑98

[Fil10] Y. Filmus, *Two proofs of the central limit theorem*, 2010. http://www.cs.toronto.edu/~yuvalf/CLT.pdf. ↑14

[For10] Peter J. Forrester, *Log-gases and random matrices*, Princeton University Press, 2010. ↑35

[For93] P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nuclear Physics B **402** (1993), 709–728. ↑116

[FW17] Peter J. Forrester and Dong Wang, *Muttalib–borodin ensembles in random matrix theory — realisations and correlation functions*, Electron. J. Probab. **22** (2017), 43 pp. arXiv:1502.07147 [math-ph]. ↑116

[GH24] V. Gorin and J. Huang, *Dynamical loop equation*, Ann. Probab. **52** (2024), no. 5, 1758–1863. arXiv:2205.15785 [math.PR]. ↑125, 141, 144

[Gor21] V. Gorin, *Lectures on random lozenge tilings*, Cambridge Studies in Advanced Mathematics. Cambridge University Press (2021). ↑49

[Gor25] V. Gorin, *Private communication*, 2025. ↑125

[GS18] V. Gorin and M. Shkolnikov, *Stochastic Airy semigroup through tridiagonal matrices*, Ann. Probab. **46** (2018), no. 4, 2287–2344. arXiv:1601.06800 [math.PR]. ↑116

[GXZ24] V. Gorin, J. Xu, and L. Zhang, *Airy*$_\beta$ *line ensemble and its Laplace transform*, arXiv preprint (2024). arXiv:2411.10829 [math.PR]. ↑116

[HKPV06] J.B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Determinantal processes and independence*, Probability Surveys **3** (2006), 206–229. arXiv:math/0503110 [math.PR]. ↑78

[Hur97] A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachr. Ges. Wiss. Göttingen (1897), 71–90. ↑9

[IIKS90] A.R. Its, A.G. Izergin, V.E. Korepin, and N.A. Slavnov, *Differential equations for quantum correlation functions*, Int. J. Mod. Phys. B **4** (1990), no. 5, 1003–1037. ↑122

[JN06] K. Johansson and E. Nordenstam, *Eigenvalues of GUE minors*, Electron. J. Probab. **11** (2006), no. 50, 1342–1371. arXiv:math/0606760 [math.PR]. ↑98

[Joh98] K. Johansson, *On fluctuations of eigenvalues of random hermitian matrices*, Duke Math. J. **91** (1998), no. 1, 151–204. ↑35

[KO07] R. Kenyon and A. Okounkov, *Limit shapes and the complex Burgers equation*, Acta Math. **199** (2007), no. 2, 263–302. arXiv:math-ph/0507007. ↑129

[KS96] R. Koekoek and R.F. Swarttouw, *The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue*, Technical report, Delft University of Technology and Free University of Amsterdam (1996). arXiv:math/9602214 [math.CA], report no. OP-SF 20 Feb 1996. Updated version available at https://fa.ewi.tudelft.nl/~koekoek/documents/as98.pdf. ↑49

[KT12] A. Kulesza and B. Taskar, *Determinantal Point Processes for Machine Learning*, Foundations and Trends in Machine Learning **5** (2012), no. 2–3, 123–286. arXiv:1207.6083 [stat.ML]. ↑78

[Mac95] I.G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed., Oxford University Press, 1995. ↑160

[Met13] A. Metcalfe, *Universality properties of Gelfand-Tsetlin patterns*, Probab. Theory Relat. Fields **155** (2013), no. 1-2, 303–346. arXiv:1105.1272 [math.PR]. ↑98

[MP67] V. Marchenko and L. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb. **1** (1967), 457–483. ↑20

[NF98] T. Nagao and P.J. Forrester, *Multilevel dynamical correlation functions for Dyson's Brownian motion model of random matrices*, Physics Letters A **247** (1998), no. 1-2, 42–46. ↑98

[Oko02] A. Okounkov, *Symmetric functions and random partitions*, Symmetric functions 2001: Surveys of developments and perspectives, 2002. arXiv:math/0309074 [math.CO]. ↑103, 109

[Ols13] G. Olshanski, *Projections of orbital measures, gelfand-tsetlin polytopes, and splines*, Journal of Lie Theory **23** (2013), no. 4, 1011–1022. arXiv:1302.7116 [math.RT]. ↑119

[Olv74] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, 1974. ↑102

[OV96] G. Olshanski and A. Vershik, *Ergodic unitarily invariant measures on the space of infinite Hermitian matrices*, Contemporary Mathematical Physics. F.A.Berezin memorial volume. American Mathematical Society Translations, (Advances in the Mathematical Sciences — 31), 1996, pp. 137–175. arXiv:math/9601215v1 [math.RT]. ↑119

[RRV11] J. Ramirez, B. Rider, and B. Virag, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, Jour. AMS **24** (2011), no. 4, 919–944. arXiv:math/0607331 [math.PR]. ↑116

[SB95] J. W. Silverstein and Z. D. Bai, *On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices*, J. Multivar. Anal. **54** (1995), no. 2, 175–192. ↑37

[Sok20] A. D. Sokal, *The Euler and Springer numbers as moment sequences*, Expo. Math. **38** (2020), no. 1, 1–26. arXiv:1804.04498 [math.CO]. ↑70

[Sos00] A. Soshnikov, *Determinantal random point fields*, Russian Mathematical Surveys **55** (2000), no. 5, 923–975. arXiv:math/0002099 [math.PR]. ↑78, 90

[SSB05] J. Schenker and H. Schulz-Baldes, *Semicircle law and freeness for random matrices with symmetries or correlations*, Math. Res. Lett. **12** (2005). arXiv:math-ph/0505003. ↑35

[ST43] J. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society (AMS), 1943. ↑30

[Sta15] R. Stanley, *Catalan numbers*, Cambridge University Press, 2015. ↑28

[TW93] C. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Physics Letters B **305** (1993), no. 1, 115–118. arXiv:hep-th/9210074. ↑116

[TW94] C. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys. **159** (1994), no. 1, 151–174. arXiv:hep-th/9211141. ↑116

[VV09] B. Valkó and B. Virág, *Continuum limits of random matrices and the Brownian carousel*, Invent. Math. **177** (2009), no. 3, 463–508. arXiv:0712.2000 [math.PR]. ↑116

[Wig55] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics (1955), 548–564. ↑9

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com