

Lectures on Random Matrices (Spring 2025)

Lecture 1: Moments of random variables and random matrices

Leonid Petrov

Monday, January 13, 2025*

Contents

1 Why study random matrices?	2
2 Recall Central Limit Theorem	3
2.1 Central Limit Theorem and examples	3
2.2 Moments of the normal distribution	5
2.3 Moments of sums of iid random variables	5
2.3.1 Computation of moments	6
2.3.2 n -dependent factor	6
2.3.3 Combinatorial factor	7
2.3.4 Putting it all together	7
2.4 Convergence in distribution	7
3 Random matrices and semicircle law	7
3.1 Where can randomness in a matrix come from?	7
3.2 Real Wigner matrices	8
3.3 Empirical spectral distribution	9
3.4 Expected moments of traces of random matrices	10
3.5 Immediate next steps	12
A Problems (due 2025-02-13)	12
A.1 Normal approximation	12
A.2 Convergence in distribution	12
A.3 Moments of sum justification	12
A.4 Distribution not determined by moments	13
A.5 Uniqueness of the normal distribution	13
A.6 Quaternions	13
A.7 Ensemble $UD_\lambda U^\dagger$	13
A.8 Invariance of the GOE	13

*Course webpage • Live simulations • TeX Source • Updated at 07:18, Wednesday 12th February, 2025

A.9 Counting n -powers in the real Wigner matrix	13
A.10 Counting trees	13

1 Why study random matrices?

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.

2 Recall Central Limit Theorem

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 2.1. 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 $\{X_1, X_2, \dots, X_n\}$, the joint distribution factors as the product of the individual distributions:

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

Theorem 2.2 (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 $\sigma^2 = \text{Var}(X_i)$. Define the normalized sum*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu). \quad (2.1)$$

Then, as $n \rightarrow \infty$, 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

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq x) = \mathbb{P}(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt \quad \text{for all } x \in \mathbb{R}, \quad (2.2)$$

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

Remark 2.3. For a general random variable instead of $Z \sim \mathcal{N}(0, \sigma^2)$, the convergence in distribution (2.2) 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 2.4. 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:

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

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

$$\mathbb{P}(X_1 + \dots + X_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

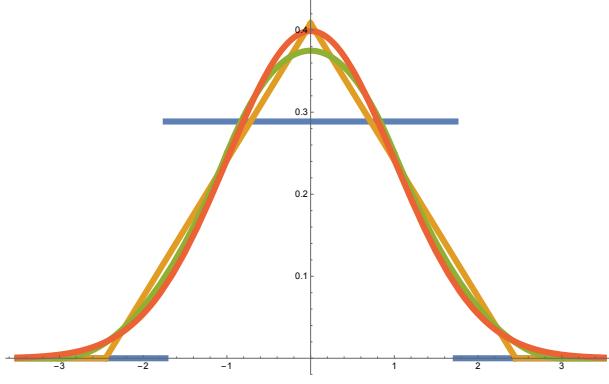


Figure 1: Densities of U_1 , $U_1 + U_2$, $U_1 + U_2 + U_3$ (where U_i are iid uniform on $[0, 1]$), and $\mathcal{N}(0, 1)$, normalized to have the same mean and variance.

Introduce the normalized quantity

$$z = \frac{k - np}{\sqrt{np(1 - p)}}, \quad (2.3)$$

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 $n \rightarrow \infty$, the following holds:

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

For large n , Stirling's formula gives

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

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

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

where $p \approx k/n$ thanks to the fact that z (2.3) 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 (2.2) in the classical CLT; but one can get the latter from the local CLT by integration.)

2.2 Moments of the normal distribution

Proposition 2.5. *The moments of a random variable $Z \sim \mathcal{N}(0, \sigma^2)$ are given by:*

$$\mathbb{E}[Z^k] = \begin{cases} 0, & \text{if } k \text{ is odd,} \\ \sigma^k (k-1)!! = \sigma^k \cdot (k-1)(k-3)\cdots 1, & \text{if } k \text{ is even.} \end{cases} \quad (2.4)$$

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. \square

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 2.6. 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.

2.3.1 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:

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

Taking the expectation and using linearity, we have:

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

The sum over all j_1, \dots, j_n with $j_1 + \dots + j_n = k$ 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 $\leq n$ parts. Since n is large, we simply sum over all partitions of k . For each partition λ of k (where $k = \lambda_1 + \lambda_2 + \dots + \lambda_n$ and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$), we must count the number of distinct multisets of indices (j_1, j_2, \dots, j_n) that yield the same collection $\{\lambda_1, \lambda_2, \dots\}$. Then,

$$\mathbb{E}[Y_{j_1} Y_{j_2} \dots Y_{j_n}] = m_{\lambda_1} m_{\lambda_2} \dots m_{\lambda_n},$$

where $m_j = \mathbb{E}[Y^j]$ (recall the identical distribution of Y_i). Note that $m_0 = 1$ and $m_1 = 0$. Let us illustrate this with an example.

Example 2.7. 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 $\mathbb{E}[Y_{j_1} Y_{j_2} Y_{j_3} Y_{j_4}] = m_4$) 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 $\mathbb{E}[Y_{j_1} Y_{j_2} Y_{j_3} Y_{j_4}] = m_2^2$) 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.

2.3.2 n -dependent factor

Consider first the n -dependent factor. In the case k is even and $\lambda = (2, 2, \dots, 2)$, 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 $n \rightarrow \infty$. When k is odd, the “best” we can do (without parts equal to 1) is going to be $\lambda = (3, 2, \dots, 2)$ with $(k-1)/2$ parts, so the power of n is $n^{(k-1)/2}$. This is also subleading in the limit $n \rightarrow \infty$.

2.3.3 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:

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

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.

2.3.4 Putting it all together

We have as $n \rightarrow \infty$:

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

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 (2.1) converge to the moments of the standard normal distribution.

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 A.5), so the CLT holds in this case, provided that the original iid random variables X_i have finite moments of all orders.

3 Random matrices and semicircle law

We now turn to random matrices.

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:

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

Here D_λ is a diagonal matrix with constant eigenvalues $\lambda = (\lambda_1, \dots, \lambda_n)$. 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 $\lambda_1 \geq \dots \geq \lambda_n$, since the unitary conjugation can permute the eigenvalues.

3.2 Real Wigner matrices

Definition 3.1 (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;

¹Real 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 $q = a + bi + cj + dk$ can be represented by the complex 2×2 matrix

$$q \longmapsto \begin{pmatrix} a + ib & c + id \\ -c + id & a - ib \end{pmatrix}.$$

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.

²The 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.

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 3.2 (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 3.3 (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.

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} :

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}, \quad (3.1)$$

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 (3.1) 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

$$\int_{\mathbb{R}} x \mu_n(dx) = \frac{1}{n} \sum_{i=1}^n \lambda_i = \frac{1}{n} \text{Tr}(W) = \frac{1}{n} \sum_{i=1}^n X_{ii}, \quad (3.2)$$

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

$$\int_{\mathbb{R}} x^2 \mu_n(dx) = \frac{1}{n} \sum_{i=1}^n \lambda_i^2 = \frac{1}{n} \text{Tr}(W^2) = \frac{1}{n} \sum_{i,j=1}^n X_{ij}^2. \quad (3.3)$$

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 3.4 (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 $n \rightarrow \infty$, the ESD of $W/(\sigma\sqrt{n})$ converges in distribution to the semicircular law:*

$$\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i/\sqrt{n}} \longrightarrow \mu_{\text{sc}}, \quad (3.4)$$

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

$$\mu_{\text{sc}}(dx) := \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{|x| \leq 2} dx. \quad (3.5)$$

Remark 3.5. The convergence in (3.4) 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

$$\int_{\mathbb{R}} f(x) \nu_n(dx) \longrightarrow \int_{\mathbb{R}} f(x) \mu_{\text{sc}}(dx), \quad n \rightarrow \infty, \quad (3.6)$$

where in (3.6) the convergence is in probability. Indeed, the left-hand side of (3.6) 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 (3.6) 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 $\int_{\mathbb{R}} f(x) \mu_{\text{sc}}(dx)$.

Remark 3.6. There exists a version of the limiting ESD for the Wishart matrices (Remark 3.3). In this case, the limiting distribution is the *Marchenko-Pastur law* [MP67].

3.4 Expected moments of traces of random matrices

The main computation in the proof of Theorem 3.4 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 3.7 (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

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

Our first goal is to study the asymptotic behavior of $\mathbb{E}[m_k^{(n)}]$ as $n \rightarrow \infty$ for each fixed $k \geq 1$, just like we did in (3.2)–(3.3) for $k = 1, 2$:

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

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 $n \rightarrow \infty$.

Lemma 3.8 (Convergence of Expected Moments). *For each fixed $k \geq 1$, we have*

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

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

$$\mathrm{Tr}(W^k) = \sum_{i_1, \dots, i_k=1}^n X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}. \quad (3.7)$$

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 $n \rightarrow \infty$.

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 $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$. 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 A.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 A.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 3.8. \square

3.5 Immediate next steps

The proof of Theorem 3.4 is continued in the next [Lecture 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.

A Problems (due 2025-02-13)

Each problem is a subsection (like Problem A.1), and may have several parts.

A.1 Normal approximation

1. In Figure 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:

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

A.2 Convergence in distribution

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

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

for all bounded continuous functions f . Show that convergence in distribution is equivalent to the condition outlined in (2.2):

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

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

A.3 Moments of sum justification

Justify the computations of the power of n in Section 2.3.2.

A.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.

A.5 Uniqueness of the normal distribution

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

A.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.

A.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 $\lambda = (\lambda_1, \dots, \lambda_N)$, $\lambda_1 \geq \dots \geq \lambda_N$. Let us fix λ to be, say, $\lambda = (1, 1, \dots, 1, 0, 0, \dots, 0)$, 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 $k = 1, 2, \dots, N$. For a fixed k , let $\lambda_1^{(k)} \geq \dots \geq \lambda_k^{(k)}$ be the eigenvalues of the top-left corner of size $k \times k$. Plot the two-dimensional array

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

A.8 Invariance of the GOE

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

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

for all orthogonal matrices O and Borel sets A .

A.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.

A.10 Counting trees

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

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

References

- [Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. ↑[5](#), [7](#)
- [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. ↑[2](#)
- [Dys62a] F.J. Dyson, *A Brownian motion model for the eigenvalues of a random matrix*, Jour. Math. Phys. **3** (1962), no. 6, 1191–1198. ↑[2](#)
- [Dys62b] F.J. Dyson, *Statistical Theory of the Energy Levels of Complex Systems. I-III*, Jour. Math. Phys. **3** (1962), no. 166, 140–175. ↑[2](#)
- [Fil10] Y. Filmus, *Two proofs of the central limit theorem*, 2010. <http://www.cs.toronto.edu/~yuvalf/CLT.pdf>. ↑[5](#)
- [Hur97] A. Hurwitz, *Über die Erzeugung der Invarianten durch Integration*, Nachr. Ges. Wiss. Göttingen (1897), 71–90. ↑[2](#)
- [MP67] V. Marchenko and L. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Math. USSR Sb. **1** (1967), 457–483. ↑[10](#)
- [Wig55] E. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics (1955), 548–564. ↑[2](#)

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

Lectures on Random Matrices (Spring 2025)

Lecture 2: Wigner semicircle law

Leonid Petrov

Wednesday, January 15, 2025*

Contents

1 Recap	2
2 Two computations	2
2.1 Moments of the semicircle law	2
2.2 Counting trees and Catalan numbers	3
3 Analysis steps in the proof	5
3.1 The semicircle distribution is determined by its moments	5
3.2 Convergence to the semicircle law	6
3.2.1 A concentration bound and the Borel–Cantelli lemma	7
3.2.2 Tightness of $\{\nu_n\}$ and subsequential limits	7
3.2.3 Characterizing the limit measure	7
4 Proof of Proposition 3.5: bounding the variance	8
5 Remark: Variants of the semicircle law	9
B Problems (due 2025-02-15)	11
B.1 Standard formula	11
B.2 Tree profiles	11
B.3 Ballot problem	11
B.4 Reflection principle	11
B.5 Bounding probability in the proof	11
B.6 Almost sure convergence and convergence in probability	12
B.7 Wigner’s semicircle law for complex Wigner matrices	12
B.8 Semicircle law without the moment condition	12

*Course webpage • Live simulations • TeX Source • Updated at 04:33, Saturday 15th February, 2025

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} :

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

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

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

Remark 1.1. 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 Two computations

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

2.1 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 B.1)

$$\int_0^{\pi/2} \sin^{2n} \theta \, d\theta = \frac{\pi}{2} \frac{(2n)!}{2^{2n}(n!)^2}, \quad (2.1)$$

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 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 2.1 (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 2.2 (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 2.3 (Catalan Numbers). The sequence of *Catalan numbers* $\{C_m\}_{m \geq 0}$ is defined recursively by:

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

Alternatively, they have the closed form¹

$$C_m = \frac{1}{m+1} \binom{2m}{m} = \binom{2m}{m} - \binom{2m}{m+1}. \quad (2.3)$$

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

Lemma 2.4. *Formulas (2.2) and (2.3) are equivalent.*

¹See Problem B.4 for a combinatorial proof of the second inequality.

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

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

Solving for $C(z)$, we get

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

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. \square

Remark 2.5. Catalan numbers enumerate many (too many!) combinatorial objects. For a comprehensive treatment, see [Sta15].

Proposition 2.6 (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 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. \square

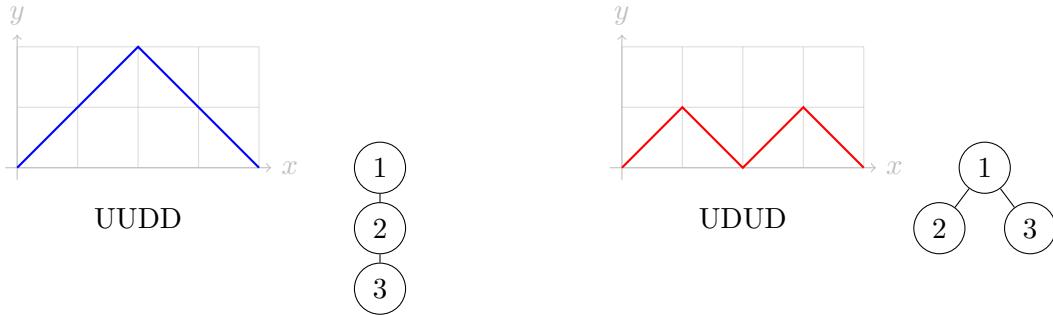


Figure 1: The two possible Dyck paths of semilength $m = 2$ and their corresponding rooted plane trees.

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 2.6, it suffices to consider only Dyck paths.

Proposition 2.7. *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 $j = 0, \dots, m$. 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. \square

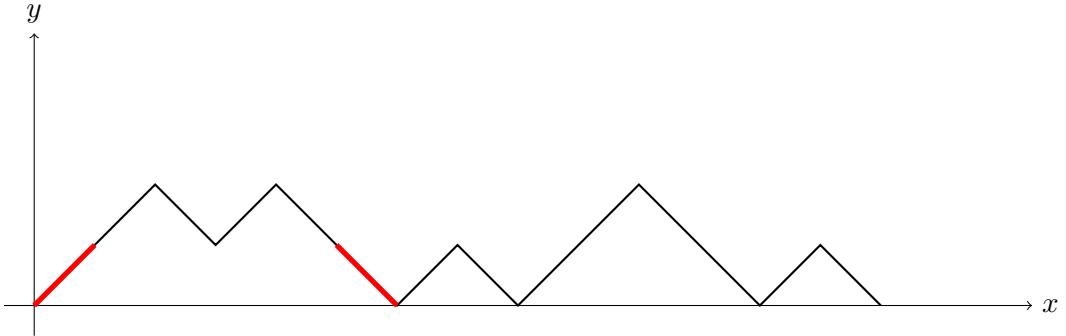


Figure 2: Illustration of a Dyck path decomposition for the proof of Proposition 2.7.

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 [Lecture 1](#).

Let us remember that so far, we showed that

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

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

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 3.1 (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*

$$\sum_{k=1}^{\infty} (m_{2k})^{-1/(2k)} = \infty, \quad (3.1)$$

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

Remark 3.2. 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 (3.1), as for any infinite series, is only determined by arbitrarily large moments, for the same reason.

Remark 3.3. 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 3.4. See also Problem A.4 from [Lecture 1](#) on an example of a distribution not determined by its moments.

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 [Lecture 1](#)). which deals with random variables

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

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):

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

where m_k are the moments of the semicircle distribution, and ν_n is the ESD corresponding to the scaling of the eigenvalues as λ_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 $\mathbb{E}(Y_{n,k}) \rightarrow m_k$. Let us assume the following:

Proposition 3.5 (Variance bound). *For each fixed integer $k \geq 1$ and large enough n , we have*

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

We will prove Proposition 3.5 in Section 4 below. Let us finish the proof of convergence to the semicircle law modulo Proposition 3.5.

3.2.1 A concentration bound and the Borel–Cantelli lemma

From Chebyshev's inequality,

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

where in the last step we used Proposition 3.5.

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 $\mathbb{E}[Y_{n,k}] \rightarrow m_k$, we conclude

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

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

Since $|Y_{n,k}| = |\int x^k \nu_n(dx)|$ stays almost surely bounded for each k , one readily checks (Problem B.5) that almost surely, for each fixed k ,

$$\nu_n(\{x : |x| > M\}) \leq \frac{C}{M^k}. \tag{3.2}$$

By choosing k large, we see that ν_n puts arbitrarily little mass outside any large interval $[-m, m]$. 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 .

3.2.3 Characterizing the limit measure

We claim that $\nu^* = \mu_{\text{sc}}$, 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 $\nu_{n_j} \rightarrow \nu^*$ weakly, we have

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

On the other hand, we have already shown

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

Thus

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

By Proposition 3.1, the measure ν^* is uniquely determined by its moments. Hence ν^* must coincide with μ_{sc} .

Remark 3.6. In Sections 3.2.2 and 3.2.3 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 $\nu_n \rightarrow \nu$ 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 3.5. (See also Problem B.6 for the weakly in probability convergence.)

4 Proof of Proposition 3.5: 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

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

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

$$\text{Var}(Y_{n,k}) = \frac{1}{n^{2+k}} \text{Var}\left(\sum_{I \in [n]^k} X_I\right) = \frac{1}{n^{2+k}} \sum_{I, J \in [n]^k} \text{Cov}(X_I, X_J).$$

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

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

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

is the product of the entries of our Wigner matrix corresponding to the directed ‘‘edges’’ $(i_1 \rightarrow i_2), (i_2 \rightarrow i_3), \dots, (i_k \rightarrow i_1)$. 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 $\text{Cov}(X_I, X_J) = 0$.
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 $\text{Cov}(X_I, X_J) = 0$ since $\mathbb{E}[X_{i_1 i_2}] = 0$. Indeed, for example if $X_{i_1 i_2}$ appears only in X_I , then

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

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. $q \leq k$. 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. $q \geq k + 1$. 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 3.5.

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

Also, see Problem [B.7](#) for the complex case of the Wigner semicircle law.

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

$$p_n(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_{n,\beta}} \exp\left(-\frac{\beta}{4} \sum_{i=1}^n \lambda_i^2\right) \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \quad (5.1)$$

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

Theorem 5.2 (Correlated entries [SSB05]). Let $W_n = \left(\frac{1}{\sqrt{n}} X_{pq} \right)_{1 \leq p,q \leq n}$ 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: $\sup_n \max_{p,q=1,\dots,n} \mathbb{E}[|X_{pq}|^k] < \infty$ for all $k \in \mathbb{N}$.
2. There exists an equivalence relation \sim_n on pairs of indices (p, q) in $\{1, \dots, n\}^2$ such that:
 - Entries $X_{p_1 q_1}, \dots, X_{p_j q_j}$ are independent when $(p_1, q_1), \dots, (p_j, q_j)$ belong to distinct equivalence classes.
 - The relation satisfies the following bounds:
 - (a) $\max_p \#\{(q, p', q') \in \{1, \dots, n\}^3 \mid (p, q) \sim_n (p', q')\} = o(n^2)$,
 - (b) $\max_{p,q,p'} \#\{q' \in \{1, \dots, n\} \mid (p, q) \sim_n (p', q')\} \leq B$ for some constant B ,
 - (c) $\#\{(p, q, p') \in \{1, \dots, n\}^3 \mid (p, q) \sim_n (q, p') \text{ and } p \neq p'\} = o(n^2)$.
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 $n \rightarrow \infty$, 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 5.3 ([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 5.4. 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 $\mathbb{E}[X_{ij}^2] = 1$.

²For $\beta = 1, 2, 4$, 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 (5.1) makes sense), and we can still treat all the β cases in a unified manner.

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

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 5.5 (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

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

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

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

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

B Problems (due 2025-02-15)

B.1 Standard formula

Prove formula (2.1):

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

B.2 Tree profiles

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

B.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?

B.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.

B.5 Bounding probability in the proof

Show inequality (3.2).

B.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.

B.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.

B.8 Semicircle law without the moment condition

Prove Theorem 5.3.

References

- [Akh65] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*, Translated by N. Kemmer, Hafner Publishing Co., New York, 1965. ↑6
- [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]. ↑10
- [Bil95] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley Ser. Probab. Stat., Wiley, 1995. ↑6, 7
- [For10] Peter J. Forrester, *Log-gases and random matrices*, Princeton University Press, 2010. ↑9
- [Joh98] K. Johansson, *On fluctuations of eigenvalues of random hermitian matrices*, Duke Math. J. **91** (1998), no. 1, 151–204. ↑9
- [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. ↑11
- [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. ↑10
- [ST43] J. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society (AMS), 1943. ↑6
- [Sta15] R. Stanley, *Catalan numbers*, Cambridge University Press, 2015. ↑4

L. PETROV, UNIVERSITY OF VIRGINIA, DEPARTMENT OF MATHEMATICS, 141 CABELL DRIVE, KERCHOFF HALL, P.O. Box 400137, CHARLOTTESVILLE, VA 22904, USA
E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)

Lecture 3: Gaussian and tridiagonal matrices

Leonid Petrov

Wednesday, January 22, 2025*

Contents

1 Recap	2
2 Gaussian ensembles	2
2.1 Definitions	2
2.2 Joint eigenvalue distribution for GOE	3
2.3 Step A. Joint density of matrix entries	4
2.4 Step B. Spectral decomposition	4
2.5 Step C. Jacobian	5
2.6 Step D. Final Form of the density	6
3 Other classical ensembles with explicit eigenvalue densities	7
3.1 Wishart (Laguerre) ensemble	7
3.1.1 Definition via SVD	7
3.1.2 Joint density of eigenvalues	7
3.2 Jacobi (MANOVA/CCA) ensemble	8
3.2.1 Setup	8
3.2.2 Jacobi ensemble	9
3.3 General Pattern and β -Ensembles	9
4 Tridiagonal form for real symmetric matrices	10
5 Tridiagonalization of random matrices	11
5.1 Dumitriu–Edelman tridiagonal model for GOE	12
5.2 Generalization to β -ensembles	13
C Problems (due 2025-02-22)	13
C.1 Invariance of GOE and GUE	13
C.2 Preimage size for spectral decomposition	13
C.3 Distinct eigenvalues	14

*Course webpage • Live simulations • TeX Source • Updated at 12:58, Wednesday 12th February, 2025

C.4	Testing distinctness of eigenvalues via rank-1 perturbations	14
C.5	Jacobian for GUE	14
C.6	Normalization for GOE	14
C.7	Wishart eigenvalue density	14
C.8	Householder reflection properties	15
C.9	Distribution of the Householder vector in random tridiagonalization	15
C.10	Householder reflection for GUE	15
C.11	Jacobi ensemble is related to two Wisharts	16

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 $n \rightarrow \infty$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i/\sqrt{n}} = \mu_{\text{sc}}, \quad (1.1)$$

where

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

The convergence in (1.1) is weakly almost sure. The way we got the result is by expanding $\mathbb{E} \text{Tr}(W^k)$ and counting trees, plus analytic lemmas which ensure that the convergence of expected powers of traces is enough to conclude the convergence (1.1) 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.

2 Gaussian ensembles

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} , $1 \leq i, j \leq n$, such that Y_{ij} are mean zero, variance 1. Then for $1 \leq i < j \leq n$, we have for the matrix $W = (X_{ij})$:

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

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*

$$Z = \frac{1}{\sqrt{2}}(Z^R + iZ^I), \quad \mathbb{E}(Z) = 0, \quad \text{Var}_{\mathbb{C}}(Z) := \mathbb{E}(|Z|^2) = \frac{\mathbb{E}(|Z^R|^2) + \mathbb{E}(|Z^I|^2)}{2} = 1,$$

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} , $1 \leq i, j \leq n$ distributed as Z , then the random matrix¹

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

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

$$\text{Var}_{\mathbb{C}}(X_{ii}) = 1, \quad \text{Var}_{\mathbb{C}}(X_{ij}) = \frac{1}{4}[\mathbb{E}(Z_{ij}^R + Z_{ji}^R)^2 + \mathbb{E}(Z_{ij}^I + Z_{ji}^I)^2] = 1.$$

Both GOE and GUE have real eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$. We are going to describe the joint distribution of these eigenvalues. Despite the fact that 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 C.1 for invariance of GOE/GUE under orthogonal/unitary conjugation (this is where the names “orthogonal” and “unitary” come from).

Remark 2.1. 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.

2.2 Joint eigenvalue distribution for GOE

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

Theorem 2.2 (GOE Joint Eigenvalue Density). *Let W be an $n \times n$ real symmetric matrix with the GOE distribution (Section 2.1). Then its ordered real eigenvalues $\lambda_1 \leq \dots \leq \lambda_n$ of $W/\sqrt{2}$ have a joint probability density function on \mathbb{R}^n given by:*

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

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

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

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

Remark 2.3. 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 Sections 2.3 to 2.6 below.

2.3 Step A. Joint density of matrix entries

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

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

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

$$f(x_{12}, x_{13}, \dots, x_{22}, x_{33}, \dots) \propto \exp\left(-\sum_{i < j} x_{ij}^2 - \frac{1}{2} \sum_{i=1}^n x_{ii}^2\right) = \exp\left(-\frac{1}{2} \sum_{i,j=1}^n x_{ij}^2\right), \quad (2.1)$$

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

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

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

$$f(W) dW = \pi^{-\frac{n(n-1)}{4}} (2\pi)^{-\frac{n}{4}} \exp\left(-\frac{1}{2} \text{Tr}(W^2)\right) dW,$$

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

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

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

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 C.2 and C.3.

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

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,

$$dW = \left| \det\left(\frac{\partial W}{\partial(\Lambda, Q)}\right) \right| d\Lambda dQ,$$

where dQ is the Haar measure² 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” $\lambda_1 \leq \dots \leq \lambda_n$ if we want an ordering, this introduces the simple factor $n!$ in the final density.

Lemma 2.4 (Jacobian for Spectral Decomposition). *For real symmetric $W = Q\Lambda Q^\top$, one has*

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

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

Remark 2.5. Equivalently, one often writes

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

is the *Vandermonde determinant*.

We prove Lemma 2.4 in the rest of this subsection.

Consider small perturbations of Λ and Q . Write

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

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

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

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.

²Recall 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 $d\lambda_1 \dots d\lambda_n$ 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 $1 \leq i < j \leq n$. 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 2.4. See also Problem C.5 for the GUE Jacobian.

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 $\{\lambda_1, \dots, \lambda_n\}$ is, up to normalization depending only on n , equal to

$$\prod_{i < j} |\lambda_i - \lambda_j| \exp\left(-\frac{1}{2} \sum_{k=1}^n \lambda_k^2\right). \tag{2.2}$$

We leave the computation of the normalization constant in Theorem 2.2 as Problem C.6.

Remark 2.6. We emphasize that in the GOE case, the normalization $W/\sqrt{2}$ for (2.2) is so that the variance is 1 on the diagonal and $\frac{1}{2}$ off the diagonal.

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).

3.1 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.

3.1.1 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 $n \leq m$. 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), $s_1, \dots, s_n \geq 0$ 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

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

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

3.1.2 Joint density of eigenvalues

Theorem 3.1 (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*

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

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 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 $\exp(-\frac{1}{2} \sum s_i^2)$ when reshaped to $\exp(-\frac{1}{2} \sum \lambda_i)$. \square

Remark 3.2. 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*.

3.2 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.

3.2.1 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 3.3 (Projectors and canonical correlations). Denote by

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

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 $\text{rank}(P_X P_Y) \leq \min(n, k)$, 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 3.4. For $n = k = 1$, we have

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

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 .

3.2.2 Jacobi ensemble

Theorem 3.5 (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 $\lambda_1, \dots, \lambda_n$ of the matrix $P_X P_Y$ lie in the interval $[0, 1]$ and have the joint density function of the form

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

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 3.6. 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 $n \leq t$ 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 General Pattern and β -Ensembles

We have now seen three classical examples:

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

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

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

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 3.7. 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 3.8. 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.

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 4.1. *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 4.2 (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 4.1. 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:

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

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

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

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. \square

Remark 4.3. 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.

5 Tridiagonalization of random matrices

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

5.1 Dumitriu–Edelman tridiagonal model for GOE

Theorem 5.1. *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 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 $\{d_i, \alpha_j\}_{1 \leq i \leq n, 1 \leq j \leq n-1}$ are mutually independent, with

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

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

Remark 5.2 (Chi-square distributions). The *chi-square distribution* with ν degrees of freedom, denoted by χ_ν^2 , 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, \dots, Z_ν 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., $Q \sim \chi_\nu^2$. In the context of the Dumitriu–Edelman tridiagonal model (Theorem 5.1), the subdiagonal entries α_j are defined as $\alpha_j = \sqrt{\frac{\chi_{n-j}^2}{2}}$. 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 5.1. This construction is essentially a specialized version of the Householder reduction in Section 4, set up so that each step matches precisely the distributions $\alpha_j \sim \sqrt{\frac{\chi_{n-j}^2}{2}}$ 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 .

5.2 Generalization to β -ensembles

The tridiagonal GOE construction (Theorem 5.1) 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

$$\exp\left(-\sum_{i=1}^n V(\lambda_i)\right) \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta$$

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 5.3 (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 $\alpha_j = \sqrt{\frac{\chi_{(n-j)\beta}^2}{2}}$, where χ_ν^2 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.

C Problems (due 2025-02-22)

C.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.

C.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 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ is diagonal with ($\lambda_1 \geq \dots \geq \lambda_n$), then there are exactly 2^n different choices of Q that give the same matrix W .

C.3 Distinct eigenvalues

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

C.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.

C.5 Jacobian for GUE

Arguing similarly to Section 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.

C.6 Normalization for GOE

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

$$\begin{aligned} \int_{\lambda_1 < \dots < \lambda_n} \prod_{i < j} (\lambda_i - \lambda_j) \exp\left(-\frac{1}{2} \sum_{k=1}^n \lambda_k^2\right) d\lambda_1 \cdots d\lambda_n. \\ = \frac{1}{n!} \int_{\mathbb{R}^n} \prod_{i < j} |\lambda_i - \lambda_j| \exp\left(-\frac{1}{2} \sum_{k=1}^n \lambda_k^2\right) d\lambda_1 \cdots d\lambda_n. \end{aligned}$$

Hint: The following identity might be useful:

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

C.7 Wishart eigenvalue density

Prove Theorem 3.1 (in the real case $\beta = 1$) by using the singular value decomposition of X and the properties of the Wishart ensemble.

C.8 Householder reflection properties

Show that the Householder reflection $H = I - 2vv^\top/\|v\|^2$ 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 .

C.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

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

Prove that:

1. $\|v\|^2$ follows a χ_ν^2 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} .

C.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).

C.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 $n \leq k \leq m$. Consider the matrix

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

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 (n, k, m) .
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

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

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.

References

- [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. [↑10, 13](#)
- [Gor21] V. Gorin, *Lectures on random lozenge tilings*, Cambridge Studies in Advanced Mathematics. Cambridge University Press (2021). [↑10](#)
- [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>. [↑10](#)

L. PETROV, UNIVERSITY OF VIRGINIA, DEPARTMENT OF MATHEMATICS, 141 CABELL DRIVE, KERCHOFF HALL, P.O. Box 400137, CHARLOTTESVILLE, VA 22904, USA
E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)

Lecture 4: Semicircle law for $G\beta E$ via tridiagonalization.

Beginning determinantal processes

Leonid Petrov

Wednesday, January 29, 2025*

Contents

1 Recap	2
1.1 Gaussian ensembles	2
1.2 Tridiagonalization	2
2 Tridiagonal random matrices	3
2.1 Distribution of the tridiagonal form of the GOE	3
2.2 Dumitriu–Edelman $G\beta E$ tridiagonal random matrices	4
2.3 The case $\beta = 2$	4
3 Wigner semicircle law via tridiagonalization	6
3.1 Moments for tridiagonal matrices	6
3.2 Asymptotics of chi random variables	7
3.3 Completing the proof: global semicircle behavior	8
4 Wigner semicircle law via Stieltjes transform	8
4.1 Tridiagonal structure and characteristic polynomials	9
4.1.1 Three-term recurrence for the characteristic polynomial	9
4.1.2 Spectral connection and eigenvalues	9
4.2 Stieltjes transform / resolvent	9
4.3 Approach via continued fractions	10
5 Determinantal point processes (discrete)	13
6 Application of determinantal processes to random matrices at $\beta = 2$	14
6.1 Local eigenvalue statistics (bulk and edge scaling limits)	14
6.2 Correlation functions and densities	15
6.3 Poisson process example	15

*Course webpage • Live simulations • TeX Source • Updated at 13:37, Thursday 13th February, 2025

D Problems (due 2025-02-28)	16
D.1 Eigenvalue density of G β E	16
D.2 Chi-square mean and variance	16
D.3 Edge contributions in the tridiagonal moment computation	16
D.4 Hermite polynomials and three-term recurrence	16
D.5	16
D.6 Gap probabilities	16
D.7 Stieltjes transform approach for tridiagonal matrices	16

1 Recap

Note: I did some live random matrix simulations [here](#) and [here](#) — check them out. More simulations to come.

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

$$p(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_n} \prod_{i=1}^n e^{-\frac{1}{2}\lambda_i^2} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j), \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

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, \dots, 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).

2 Tridiagonal random matrices

2.1 Distribution of the tridiagonal form of the GOE

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

Theorem 2.1. *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

$$T = \begin{pmatrix} d_1 & \alpha_1 & 0 & \cdots \\ \alpha_1 & d_2 & \alpha_2 & \ddots \\ 0 & \alpha_2 & d_3 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}, \quad (2.1)$$

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

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

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

Remark 2.2 (Chi-square distributions). The *chi-square distribution* with ν degrees of freedom, denoted by χ_ν^2 , 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, \dots, Z_ν 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., $Q \sim \chi_\nu^2$. In the context of Theorem 2.1, 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

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

Proof of Theorem 2.1. 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 $(a_{2,1}, a_{3,1}, \dots, a_{n,1})$ into a vector parallel to $(1, 0, \dots, 0) \in \mathbb{R}^{n-1}$. Since the Householder reflection is orthogonal, it preserves lengths. So,

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

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. \square

2.2 Dumitriu–Edelman G β E tridiagonal random matrices

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

Definition 2.3. Let $\beta > 0$ be a parameter. The tridiagonal G β E is a random $n \times n$ tridiagonal real symmetric matrix T as in (2.1), where $d_i \sim \mathcal{N}(0, 1)$ are independent standard Gaussians, and

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

are chi-distributed random variables.

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

Proposition 2.4 (Without proof). *For $\beta = 2$, the G β E 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 G β E is the tridiagonal form of the GSE random matrix model.*

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

Theorem 2.5 ([DE02]). *Let T be a G β E matrix as in Definition 2.3. Then the joint eigenvalue density is given by*

$$p(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_{n,\beta}} e^{-\frac{1}{2} \sum_{i=1}^n \lambda_i^2} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta, \quad \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

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].

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 $\exp(-\frac{1}{2} \sum \lambda_j^2)$. 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 2.6 (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 $\mathcal{N}(0, \frac{1}{2}) + i\mathcal{N}(0, \frac{1}{2})$ and each diagonal entry $\mathcal{N}(0, 1)$. The ordered eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$ (or, without ordering, thought of as an unordered set) satisfy the joint probability density*

$$p(\lambda_1, \dots, \lambda_n) = \frac{1}{Z_{n,2}} \prod_{j=1}^n e^{-\frac{1}{2}\lambda_j^2} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j)^2, \quad (2.2)$$

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

$$p(\lambda_1, \dots, \lambda_n) = \text{const} \cdot \det \left[\psi_{j-1}(\lambda_k) e^{-\frac{\lambda_k^2}{4}} \right]_{j,k=1}^n \det \left[\psi_{j-1}(\lambda_k) e^{-\frac{\lambda_k^2}{4}} \right]_{j,k=1}^n \quad (2.3)$$

(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 2.7 (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:

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right), \quad k = 0, 1, \dots, \quad (2.4)$$

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

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

The difference between H_k and ψ_k entering Proposition 2.6 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 $\Delta(\lambda_1, \dots, \lambda_n) = \det [\lambda_k^{j-1}]_{j,k=1}^n$. 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. \square

We will work with Hermite polynomials and the determinantal structure in Proposition 2.6 in the next [Lecture 5](#)).

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 $n \rightarrow \infty$, the empirical spectral distribution (ESD) of W/\sqrt{n} converges weakly almost surely to the Wigner semicircle distribution:

$$\mu_{\text{sc}}(dx) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{|x| \leq 2} dx.$$

We already derived this in [Lecture 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 G β E case (but only Gaussian), and thus it will apply to GUE and GSE.

3.1 Moments for tridiagonal matrices

Consider the rescaled G β E 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

$$\frac{1}{n} \text{Tr}\left(\frac{T}{\sqrt{n}}\right)^k = \frac{1}{n^{1+\frac{k}{2}}} \sum_{i_1, \dots, i_k=1}^n t_{i_1, i_2} \cdots t_{i_k, i_1}, \quad (3.1)$$

where t_{ij} are the non-rescaled entries of T . But now t_{ij} is nonzero only if $|i - j| \leq 1$, 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 $\{1, 2, \dots, n\}$ with edges only between consecutive indices.

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

Now, each term in the sum in (3.1) 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 (3.1). This product involves chi random variables which depend on n , too.

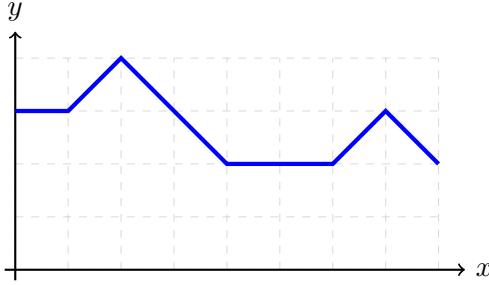


Figure 1: Example of a lattice path starting at height 3.

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

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

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

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

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 3.1 (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 have¹*

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

In particular,

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

Remark 3.2. 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 D.3.

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

¹Here and below, $O_p(\cdot)$ denotes a term that is stochastically bounded at the indicated order as $n \rightarrow \infty$. 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 .

3.3 Completing the proof: global semicircle behavior

Putting the above pieces together, we see that

$$\frac{T}{\sqrt{n}} = \frac{1}{n} \sum_{i_1, \dots, i_k=1}^n \prod_{\ell=1}^k \frac{t_{i_\ell i_{\ell+1}}}{\sqrt{n}}, \quad i_{k+1} = i_1 \text{ by agreement.} \quad (3.2)$$

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 D.3).

If at least one of the t_{ij} 's in (3.2) 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 $(1, \pm 1)$. 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

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

There is an extra factor $1/n$ in front in (3.2), which is interpreted as transforming the sum over i_1, \dots, 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 3.3 (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

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

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

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.

²Not Catalan yet!

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 $\det(T - \lambda I)$.

4.1.1 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:

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

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

Lemma 4.1 (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 D.4.

4.1.2 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 $(T - \lambda I)$ is invertible.

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

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

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

If $z = x + iy$ 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, $\text{Im } G_n(x + i0^+)$ 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

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

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

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

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 $f_u(\theta)$:

$$-zf_u(\theta) + \sqrt{\frac{\beta(1-\theta)}{2}} [f''_u(\theta) + 2f'_u(\theta)] = \delta(\theta - u). \quad (4.1)$$

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 (4.1) to the Stieltjes transform $G(z)$. This would be an excellent topic to explore for a presentation. See Problem D.7.

Update 2025-02-05: Probably, the limit of α_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 4.3.

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

$$\begin{aligned} p_0(z) &= 1, \\ p_1(z) &= z - a_0, \\ p_{n+1}(z) &= (z - a_n)p_n(z) - b_n^2 p_{n-1}(z), \quad n \geq 1. \end{aligned} \tag{4.2}$$

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

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

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

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

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 unstuck the argument in Section 4.2.

In this homogeneous case the three-term recurrence (4.2) 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

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

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

$$m(z) = \cfrac{1}{z - a_0 - \cfrac{b_1^2}{z - a_1 - \cfrac{b_2^2}{z - a_2 - \cfrac{b_3^2}{z - a_3 - \dots}}}}. \quad (4.3)$$

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

$$m(z) = \cfrac{1}{z - \cfrac{1}{z - \cfrac{1}{z - \cfrac{1}{z - \ddots}}}}. \quad (4.4)$$

Observe that the infinite continued fraction in (4.4) 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

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

Expanding the left-hand side we obtain the quadratic equation

$$m(z)^2 - z m(z) + 1 = 0. \quad (4.5)$$

The quadratic (4.5) 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 ($\text{Im}(z) > 0$) we must have $\text{Im } m(z) > 0$. The proper solution is

$$m(z) = \frac{z - \sqrt{z^2 - 4}}{2}, \quad (4.6)$$

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

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

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

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

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

Thus, from (4.6) we have, for $x \in (-2, 2)$,

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

Taking the imaginary part gives

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

so that

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

This is precisely the celebrated Wigner semicircle law.

5 Determinantal point processes (discrete)

We are now going to start the discussion of the local eigenvalue behavior at $\beta = 2$, started in Section 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 [Lecture 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 $\operatorname{Conf}(\mathfrak{X})$ 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 $\operatorname{Conf}(\mathfrak{X})$.

Definition 5.1 (Determinantal point process). A random point process P on a discrete set \mathfrak{X} is *determinantal* if there exists a kernel function $K : \mathfrak{X} \times \mathfrak{X} \rightarrow \mathbb{C}$ such that for every finite collection of pairwise distinct points $x_1, \dots, x_n \in \mathfrak{X}$,

$$\mathbb{P}\{x_1, \dots, x_n \in X\} = \det[K(x_i, x_j)]_{i,j=1}^n. \tag{5.1}$$

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

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

Correlation functions and the kernel. The condition (5.1) captures all finite-dimensional distributions of P . Equivalently, let

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

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

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

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

$$\mathbb{P}\{X \cap I = \emptyset\} = \det[\mathbf{1} - K_I], \quad (5.2)$$

where K_I is the operator $[K(x, y)]_{x,y \in I}$ (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 $I_1, \dots, I_m \subset \mathfrak{X}$ are disjoint subsets, then the joint event $\{|X \cap I_k| = n_k \text{ for } 1 \leq k \leq m\}$ can be expressed via the determinant $\det[\mathbf{1} - \sum_{k=1}^m z_k K_{I_k}]$ and its derivatives.

Remark 5.2. For any function $\phi: \mathfrak{X} \rightarrow \mathbb{C}$ such that the operator $[(1 - \phi(x))K(x, y)]_{x,y \in \mathfrak{X}}$ is trace class, the exponential generating function for ϕ is

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

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

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 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.

6.1 Local eigenvalue statistics (bulk and edge scaling limits)

Given an $n \times n$ random Hermitian matrix W whose eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ 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).

6.2 Correlation functions and densities

We recall from Section 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

$$\rho_k(x_1, \dots, x_k) dx_1 \cdots dx_k = (\text{probability that there is a particle in each small set } dx_i \text{ near } x_i, \text{ for } 1 \leq i \leq k). \quad (6.1)$$

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

$$\rho_k(x_1, \dots, x_k) = \det[K(x_i, x_j)]_{i,j=1}^k \quad \text{for each } k \geq 1. \quad (6.2)$$

The simplest example is the *Poisson process* (see Section 6.3).

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 $\text{Poisson}(\lambda|I|)$ distribution, and disjoint intervals are filled independently. One can also check that the correlation functions factorize completely:

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

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 $\beta = 2$.

D Problems (due 2025-02-28)

D.1 Eigenvalue density of $\mathbf{G}\beta\mathbf{E}$

Read and understand the main principles of the proof of Theorem 2.5 in [DE02].

D.2 Chi-square mean and variance

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

D.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 (3.2) do not contribute to the limit of the moments.

D.4 Hermite polynomials and three-term recurrence

Show that the monic Hermite polynomials $H_k(x)$ (2.4) satisfy the three-term recurrence relation

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

D.5

Compute the determinant

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

D.6 Gap probabilities

1. Prove identity (5.2) for DPPs.
2. Prove the generalization computing $\{|X \cap I_k| = n_k \text{ for } 1 \leq k \leq m\}$.

D.7 Stieltjes transform approach for tridiagonal matrices

Complete the derivation from Section 4.2 to obtain the limiting Stieltjes transform $G(z)$ for the tridiagonal matrix T/\sqrt{n} .

Remark D.1. This is more of a literature search. It is extensive, and would make an excellent topic for a presentation.

References

- [Bor11] A. Borodin, *Determinantal point processes*, Oxford handbook of random matrix theory, 2011. arXiv:0911.1153 [math.PR]. [↑13](#)
- [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. [↑4, 16](#)
- [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]. [↑12](#)

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

Lectures on Random Matrices (Spring 2025)

Lecture 5: Determinantal Point Processes and the GUE

Leonid Petrov

Wednesday, February 5, 2025*

Contents

1 Recap	2
2 Discrete determinantal point processes	2
2.1 Definition and basic properties	2
3 Determinantal structure in the GUE	3
3.1 Correlation functions as densities with respect to Lebesgue measure	3
3.2 The GUE eigenvalues as DPP	4
3.2.1 Setup	4
3.2.2 Writing the Vandermonde as a determinant	4
3.2.3 Orthogonalization by linear operations	5
3.2.4 Rewriting the density in determinantal form	5
3.3 Christoffel–Darboux formula	8
E Problems (due 2025-03-09)	10
E.1 Gap Probability for Discrete DPPs	10
E.2 Generating Functions for Multiplicative Statistics	10
E.3 Variance	10
E.4 Formula for the Hermite polynomials	10
E.5 Generating function for the Hermite polynomials	10
E.6 Projection Property of the GUE Kernel	11
E.7 Recurrence Relation for the Hermite Polynomials	11
E.8 Differential Equation for the Hermite Polynomials	11
E.9 Norm of the Hermite Polynomials	11
E.10 Existence of Determinantal Point Processes with a Given Kernel	11

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:18, Wednesday 12th February, 2025

1 Recap

In [Lecture 4](#) we discussed global spectral behavior of tridiagonal $G\beta E$ 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.

2 Discrete determinantal point processes

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

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

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

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

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 2.3. The Fredholm determinant might “diverge” (equal to 0 or 1).

Proposition 2.4 (Generating functions). *Let $f : \mathfrak{X} \rightarrow \mathbb{C}$ 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*

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

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

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

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

3 Determinantal structure in the GUE

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 $\rho_n(x_1, \dots, x_n)$ is defined by the relation

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

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

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

Remark 3.1. 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.

3.2 The GUE eigenvalues as DPP

3.2.1 Setup

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

$$p(x_1, \dots, x_n) dx_1 \cdots dx_n = \frac{1}{Z_{n,2}} \prod_{j=1}^n e^{-x_j^2/2} \prod_{1 \leq i < j \leq n} (x_i - x_j)^2 dx_1 \cdots dx_n. \quad (3.1)$$

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

$$\rho_k(x_1, \dots, x_k) = \det \left[K_n(x_i, x_j) \right]_{i,j=1}^k, \quad k \geq 1,$$

with the kernel defined as

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

where the functions

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

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.

3.2.2 Writing the Vandermonde as a determinant

The product

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

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

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

Thus, we have

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

3.2.3 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

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

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

$$\phi_j(x) = p_j(x)e^{-x^2/4}, \tag{3.2}$$

and then introduce the orthonormal functions

$$\psi_j(x) = \frac{1}{\sqrt{h_j}} \phi_j(x) = \frac{1}{\sqrt{h_j}} p_j(x)e^{-x^2/4}. \tag{3.3}$$

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 orthonormal basis of the Hilbert space $L^2(\mathbb{R}, dx)$:

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

3.2.4 Rewriting the density in determinantal form

Substituting the determinant form into the joint density (3.1), we have

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

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

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

Thus, the joint density becomes

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

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:

$$\rho_k(x_1, \dots, x_k) = \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_n) dx_{k+1} \cdots dx_n. \quad (3.4)$$

Remark 3.2. 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, \dots, x_k are fixed, and we are not integrating over all permutations of these variables.

Theorem 3.3 (Determinantal structure for squared-determinant densities). *We have*

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

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

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

Expanding the square of the determinant, we have

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

where S_n denotes the symmetric group on n elements.

Next, to obtain the k -point correlation function $\rho_k(x_1, \dots, x_k)$, we integrate out the remaining $n - k$ variables using (3.4). Substituting the expansion of the squared determinant into the expression for ρ_k , we have

$$\begin{aligned} \rho_k(x_1, \dots, x_k) &= \frac{n!}{(n-k)! \tilde{Z}_{n,2}} \sum_{\sigma, \tau \in S_n} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) \\ &\quad \left\{ \prod_{i=1}^k \phi_{\sigma(i)-1}(x_i) \phi_{\tau(i)-1}(x_i) \prod_{j=k+1}^n \int_{\mathbb{R}} \phi_{\sigma(j)-1}(x) \phi_{\tau(j)-1}(x) dx \right\}. \end{aligned} \quad (3.5)$$

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 $j = k+1, \dots, n$, 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 (3.5), we get, using the symmetry over x_1, \dots, x_k :

$$\rho_k(x_1, \dots, x_k) = \frac{1}{(n-k)! \widehat{Z}_{n,2}} \sum_{\substack{\sigma, \tau \in S_n \\ \sigma(k+1) = \tau(k+1), \dots, \sigma(n) = \tau(n)}} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) \prod_{i=1}^k \psi_{\sigma(i)-1}(x_i) \psi_{\tau(i)-1}(x_i). \quad (3.6)$$

Indeed, here we integrated over x_{k+1}, \dots, x_n , and passed from the functions $\phi_0, \phi_1, \dots, \phi_{n-1}$ to $\psi_0, \psi_1, \dots, \psi_{n-1}$. The passage to the orthonormal functions only introduces the constant $h_0 h_1 \dots h_{n-1}$ (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 (3.6), we need two general lemmas.

Lemma 3.4 (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*

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

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

$$\det(zI_p + BA) = z^{p-m} \det(zI_m + AB). \quad (3.7)$$

If we compare the coefficient of z^{p-m} in (3.7), 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 3.5 (Andreief identity). *Let $f_i(x), g_i(x) \in L^1(\mathbb{R})$ for $i = 1, \dots, n$. Then*

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

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, \dots, 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 (3.6), and finish the proof of Theorem 3.3. 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

$$(3.6) = \text{const}_n \sum_{I \subseteq [n], |I|=k} \sum_{\sigma', \tau' \in S(I)} \text{sgn}(\sigma') \text{sgn}(\tau') \prod_{i=1}^k \psi_{\sigma'(i)-1}(x_i) \psi_{\tau'(i)-1}(x_i).$$

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:

$$\det \left[\psi_{\ell_i-1}(x_j) \right]_{i=1}^k \det \left[\psi_{\ell_i-1}(x_j) \right]_{i=1}^k, \quad I = \{\ell_1 < \ell_2 < \dots < \ell_k\}.$$

By Lemma 3.4, we can rewrite the sum (over I) of products of two determinants as a single determinant of the sum. Thus, we have

$$\rho_k(x_1, \dots, x_k) = \text{const} \cdot \det \left[K_n(x_i, x_j) \right]_{i,j=1}^k, \quad (3.8)$$

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 (3.8) is indeed 1 follows from Lemma 3.5. Indeed, once the integral of ρ_n over \mathbb{R}^n is equal to $n!$, the integral over $x_1 > \dots > x_n$ becomes 1 by symmetry, as it should be. This completes the proof of Theorem 3.3. \square

3.3 Christoffel–Darboux formula

Theorem 3.6 (Christoffel–Darboux Formula). *Let $\{p_j(x)\}_{j \geq 0}$ 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

$$K_n(x, y) = \sqrt{w(x)w(y)} \frac{1}{h_{n-1}} \frac{p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)}{x - y}, \quad (3.9)$$

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

$$(x - y)S_n(x, y) = \frac{1}{h_{n-1}} [p_n(x)p_{n-1}(y) - p_{n-1}(x)p_n(y)]. \quad (3.10)$$

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

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

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:

$$\begin{aligned} xp_j(x) &= p_{j+1}(x) + \alpha_j p_j(x) + \beta_j p_{j-1}(x), \\ yp_j(y) &= p_{j+1}(y) + \alpha_j p_j(y) + \beta_j p_{j-1}(y). \end{aligned}$$

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

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

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

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

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

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

This establishes (3.10), and hence the representation (3.9) for $K_n(x, y)$.

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

E Problems (due 2025-03-09)

E.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

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

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

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

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

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

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

E.2 Generating Functions for Multiplicative Statistics

Let $f : \mathfrak{X} \rightarrow \mathbb{C}$ 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

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

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.

E.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 $\text{Var}(I)$ in terms of the kernel $K(x, y)$.

E.4 Formula for the Hermite polynomials

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

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

E.5 Generating function for the Hermite polynomials

Show that

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

E.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 $x, y \in \mathbb{R}$

$$\int_{-\infty}^{\infty} K_n(x, z) K_n(z, y) dz = K_n(x, y).$$

E.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 $p_{-1}(x) = 0$.

E.8 Differential Equation for the Hermite Polynomials

Prove that the monic Hermite polynomials $p_n(x)$ satisfy the second-order differential equation

$$p_n''(x) - x p_n'(x) + n p_n(x) = 0.$$

E.9 Norm of the Hermite Polynomials

Show that

$$h_n = \int_{-\infty}^{\infty} p_n(x)^2 e^{-x^2/2} dx = n! \sqrt{2\pi}.$$

E.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

$$\rho_n(x_1, \dots, x_n) = \det \left[K(x_i, x_j) \right]_{i,j=1}^n.$$

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].

References

- [Bor11] A. Borodin, *Determinantal point processes*, Oxford handbook of random matrix theory, 2011. arXiv:0911.1153 [math.PR]. [↑2](#)
- [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]. [↑2](#)
- [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]. [↑2](#)
- [Sos00] A. Soshnikov, *Determinantal random point fields*, Russian Mathematical Surveys **55** (2000), no. 5, 923–975. arXiv:math/0002099 [math.PR]. [↑2](#), [12](#)

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

Lectures on Random Matrices (Spring 2025)

Lecture 6: Double contour integral kernel. Steepest descent and semicircle law

Leonid Petrov

Wednesday, February 19, 2025*

Contents

1 Recap: Determinantal structure of the GUE	2
2 Double Contour Integral Representation for the GUE Kernel	3
2.1 One contour integral representation for Hermite polynomials	3
2.2 Another contour integral representation for Hermite polynomials	4
2.3 Normalization of Hermite polynomials	5
2.4 Double contour integral representation for the GUE kernel	6
2.5 Conjugation of the kernel	7
2.6 Extensions	8
3 Steepest descent — generalities for single integrals	8
3.1 Setup	8
3.2 Saddle points and steepest descent paths	9
3.3 Local asymptotic evaluation near a saddle point	10
4 Steepest descent for the GUE kernel	10
4.1 Scaling	10
4.2 Critical points	11
4.3 Imaginary critical points: $ X < 2$, “bulk”	12
F Problems (due 2025-03-19)	13
F.1 Different global positions	13
F.2 Sine kernel	13
F.3 Discrete sine process	13

*Course webpage • Live simulations • TeX Source • Updated at 03:40, Monday 24th February, 2025

1 Recap: Determinantal structure of the GUE

Last time, we proved the following result:

Theorem 1.1. *The GUE correlation functions are given by*

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

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

$$\begin{aligned} \rho_k(x_1, \dots, x_k) &= \frac{n!}{(n-k)!} \int_{\mathbb{R}^{n-k}} p(x_1, \dots, x_n) dx_{k+1} \cdots dx_n \\ &= \frac{1}{(n-k)! \widehat{Z}_{n,2}} \sum_{\substack{\sigma, \tau \in S_n \\ \sigma(k+1) = \tau(k+1), \dots, \sigma(n) = \tau(n)}} \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau) \prod_{i=1}^k \psi_{\sigma(i)-1}(x_i) \psi_{\tau(i)-1}(x_i) \\ &= \operatorname{const}_n \sum_{I \subseteq [n], |I|=k} \sum_{\sigma', \tau' \in S(I)} \operatorname{sgn}(\sigma') \operatorname{sgn}(\tau') \prod_{i=1}^k \psi_{\sigma'(i)-1}(x_i) \psi_{\tau'(i)-1}(x_i) \\ &= \operatorname{const}_n \sum_{I \subseteq [n], |I|=k} \det [\psi_{i_\alpha}(x_j)]_{\alpha, j=1}^k \det [\psi_{i_\alpha}(x_j)]_{\alpha, j=1}^k, \end{aligned}$$

where $I = \{i_1, \dots, i_k\}$ 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

$$\operatorname{const}_n \cdot \det \left[\sum_{j=0}^{n-1} \psi_j(x_\alpha) \psi_j(x_\beta) \right]_{\alpha, \beta=1}^k,$$

and finally the constant is equal to 1 by Andreief identity.

2 Double Contour Integral Representation for the GUE Kernel

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

$$p_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}. \quad (2.1)$$

Note that the monic Hermite polynomials are uniquely defined by the orthogonality property. We are not proving (2.1) here, it is an exercise.

Lemma 2.1 (Generator function for Hermite polynomials). *We have*

$$\exp\left(xt - \frac{t^2}{2}\right) = \sum_{n \geq 0} p_n(x) \frac{t^n}{n!}.$$

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 2.1:

$$p_n(x) = \frac{n!}{2\pi i} \oint_C \frac{\exp\left(xt - \frac{t^2}{2}\right)}{t^{n+1}} dt, \quad (2.2)$$

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 (2.2) is simply a complex analysis version of the operation of extracting the coefficient of t^n in the Taylor expansion.

Therefore,

$$\psi_n(x) = \frac{1}{\sqrt{h_n}} p_n(x) e^{-x^2/4} = \frac{e^{-x^2/4}}{\sqrt{h_n}} \frac{n!}{2\pi i} \oint_C \frac{\exp\left(xt - \frac{t^2}{2}\right)}{t^{n+1}} dt.$$

2.2 Another contour integral representation for Hermite polynomials

We start with the Fourier transform identity

$$\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2} + itx\right) dt = \sqrt{2\pi} e^{-x^2/2}.$$

Differentiating both sides n times with respect to x yields

$$\frac{d^n}{dx^n} \left(e^{-x^2/2} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (it)^n e^{-t^2/2+itx} dt.$$

Recalling the definition

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

we obtain

$$p_n(x) = \frac{(-1)^n e^{x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (it)^n e^{-t^2/2+itx} dt.$$

Next, perform the change of variable

$$s = it, \quad \text{so that} \quad t = -is, \quad dt = -ids.$$

Under this substitution the factors transform as follows:

$$(it)^n = s^n,$$

and the exponent becomes

$$-\frac{t^2}{2} + itx = -\frac{(-is)^2}{2} + i(-is)x = \frac{s^2}{2} + sx.$$

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,

$$\psi_n(x) = \frac{i e^{x^2/4}}{\sqrt{2\pi h_n}} \int_{-i\infty}^{i\infty} s^n e^{s^2/2-s x} ds.$$

2.3 Normalization of Hermite polynomials

Lemma 2.2. *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

$$\exp\left(xt - \frac{t^2}{2}\right) = \sum_{n \geq 0} p_n(x) \frac{t^n}{n!}$$

with a second copy (with parameter s):

$$\exp\left(xs - \frac{s^2}{2}\right) = \sum_{m \geq 0} p_m(x) \frac{s^m}{m!}.$$

Then,

$$\exp\left(xt - \frac{t^2}{2}\right) \exp\left(xs - \frac{s^2}{2}\right) = \sum_{n,m \geq 0} p_n(x) p_m(x) \frac{t^n s^m}{n! m!}.$$

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

$$\int_{-\infty}^{\infty} e^{-x^2/2} \exp\left(x(t+s) - \frac{t^2 + s^2}{2}\right) dx.$$

Completing the square in x or recalling the standard Gaussian integral yields

$$\sqrt{2\pi} \exp\left(\frac{(t+s)^2 - (t^2 + s^2)}{2}\right) = \sqrt{2\pi} \exp(ts).$$

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

$$\frac{h_n}{(n!)^2} = \frac{\sqrt{2\pi}}{n!} \implies h_n = n! \sqrt{2\pi}.$$

This completes the proof. \square

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):

$$\begin{aligned} K_n(x, y) &= \sum_{k=0}^{n-1} \psi_k(x) \psi_k(y) \\ &= \frac{e^{\frac{x^2-y^2}{4}}}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \exp\left\{-\frac{t^2}{2} + xt + \frac{s^2}{2} - ys\right\} \underbrace{\sum_{k=0}^{n-1} s^k t^{-k-1}}_{\frac{1-(s/t)^n}{t-s}}. \end{aligned} \quad (2.3)$$

Here we used the two contour integral representations for Hermite polynomials, and the explicit norm (Lemma 2.2). 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 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

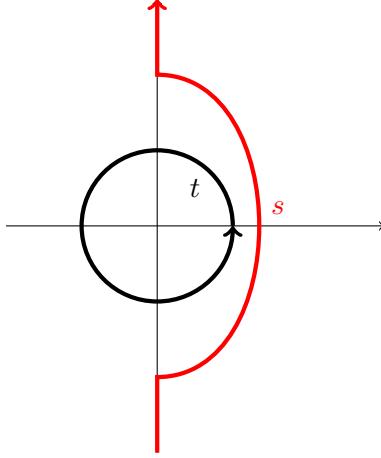


Figure 1: Integration contours for the GUE kernel (2.4).

in (2.3). 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.

With this extension of the sum, formula (2.3) becomes

$$K_n(x, y) = \frac{e^{(y^2-x^2)/4}}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n. \quad (2.4)$$

Remark 2.3. 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.

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

$$K_n(x, y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n, \quad (2.5)$$

and will use the same notation for it. Throughout the asymptotic analysis in Section 4 below, other conjugation factors may appear, but we can similarly remove them.

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 [Lecture 7](#).

3 Steepest descent — generalities for single integrals

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

$$I(\Lambda) = \int_{\gamma} e^{\Lambda f(z)} \phi(z) dz, \quad (3.1)$$

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 $\Lambda \rightarrow +\infty$.

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 3.1. 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 3.2. 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 Remarks 3.1 and 3.2, 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.

3.2 Saddle points and steepest descent paths

Definition 3.3 (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 3.4 (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., $\operatorname{Im}(f(z)) = \operatorname{Im}(f(z_0))$ for all $z \in \Gamma$), which implies that the real part $\operatorname{Re}(f(z))$ decreases away from z_0 .

In a neighborhood of a saddle point z_0 ,

$$z = z_0 + w, \quad f(z) = f(z_0) + \frac{1}{2}f''(z_0)w^2 + O(w^3).$$

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

$$2\varphi + \theta_0 = \pi \pmod{2\pi}. \tag{3.2}$$

We need the phase π so that the exponent is negative, for the integral to converge.

There are two directions satisfying (3.2) through z_0 , and we use both of them for our contour Γ . Along these directions, one finds that

$$\operatorname{Re}(f(z)) = \operatorname{Re}(f(z_0)) - \frac{1}{2}|f''(z_0)|r^2 + O(r^3),$$

so that $\operatorname{Re}(f(z))$ is maximal at $z = z_0$ and decays quadratically as one moves away from z_0 along the steepest descent paths.

3.3 Local asymptotic evaluation near a saddle point

Assume now that the contour γ in (3.1) 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

$$z = z_0 + w,$$

so the local contribution of a neighborhood of z_0 to the integral is

$$I_{z_0}(\Lambda) = e^{\Lambda f(z_0)} \phi(z_0) \int_{-\infty}^{\infty} e^{\Lambda \frac{1}{2} f''(z_0) w^2} dw \left(1 + O\left(\frac{1}{\Lambda}\right)\right). \quad (3.3)$$

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 (3.2), we have $\operatorname{Re}(f''(z_0)w^2) = -|f''(z_0)|r^2$.) Then the Gaussian integral evaluates to

$$\int_{-\infty}^{\infty} e^{-\Lambda \frac{|f''(z_0)|}{2} r^2} dr = \sqrt{\frac{2\pi}{\Lambda |f''(z_0)|}}.$$

Hence, we arrive at the following fundamental result.

Theorem 3.5 (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 $\Lambda \rightarrow +\infty$, the contribution of a small neighborhood of z_0 to the integral (3.1) is given by*

$$I_{z_0}(\Lambda) \sim e^{\Lambda f(z_0)} \phi(z_0) \sqrt{\frac{2\pi}{\Lambda |f''(z_0)|}}, \quad \Lambda \rightarrow +\infty. \quad (3.4)$$

Moreover, the behavior (3.4) captures the full asymptotic behavior of the integral (3.1) 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 (3.3) to higher order in w . (See, e.g., [Olv74] for a systematic treatment.)

4 Steepest descent for the GUE kernel

4.1 Scaling

Let us now consider the GUE kernel (2.5),

$$K_n(x, y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s - t} \left(\frac{s}{t}\right)^n,$$

where the integration contours are as in Figure 1.

We know from the Wigner semicircle law (established for real symmetric matrices with general iid entries in in [Lecture 2](#), and for the GUE in [Lecture 4](#)) that the eigenvalues live on the scale \sqrt{n} . This means that to capture the local asymptotics, we need to scale

$$x = X\sqrt{n} + \frac{\Delta x}{\sqrt{n}}, \quad y = Y\sqrt{n} + \frac{\Delta y}{\sqrt{n}}, \quad \Delta x, \Delta y \in \mathbb{R}. \quad (4.1)$$

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 [F.1](#). In what follows, we take $Y = X$.

Let us also make a change of the integration variables:

$$t = z\sqrt{n}, \quad s = w\sqrt{n}.$$

The integration contours for z and w look the same as in Figure 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 1.

We thus have:

$$\begin{aligned} K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n}) \\ = \frac{\sqrt{n}}{(2\pi)^2} \oint_C dz \int_{-i\infty}^{i\infty} dw \frac{\exp \left\{ n \left(\log w - \log z + \frac{w^2}{2} - \frac{z^2}{2} + X(z-w) + \frac{z\Delta x - w\Delta y}{n} \right) \right\}}{w-z}. \end{aligned} \quad (4.2)$$

Remark 4.1. 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 (4.2) in the context of determinantal point processes was pioneered in [[Oko02](#), Section 3].

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:

$$S'(z) = z + \frac{1}{z} - X = 0, \quad z_{cr} = \frac{X \pm \sqrt{X^2 - 4}}{2}. \quad (4.3)$$

Depending on whether $|X| < 2$, there are three cases. Unless $|X| = 2$, equation (4.3) has a single root, and thus $S''(z_{cr}) \neq 0$.

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 [Lecture 7](#), we discuss limits of the correlation functions.

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 2 for an illustration in the case $X = \frac{1}{2}$.

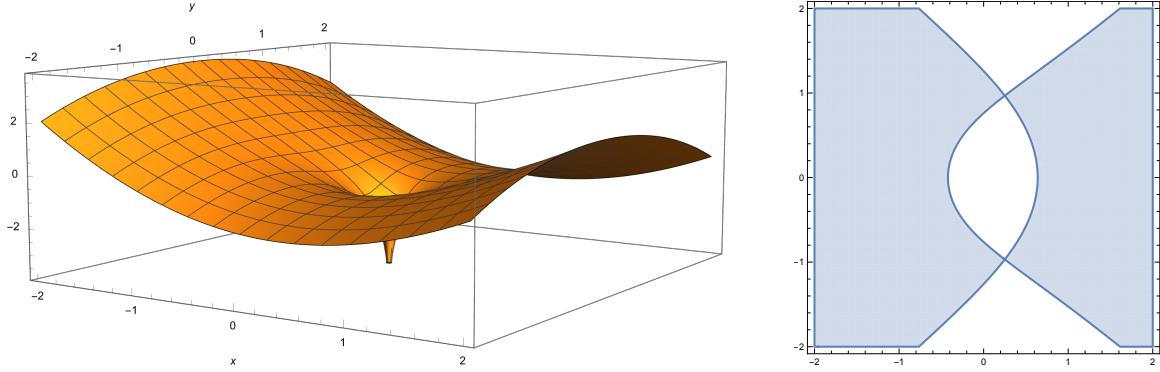


Figure 2: A 3D plot and a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ is positive (highlighted) or negative, in the case $X = \frac{1}{2}$. In this case, $z_{cr} \approx 0.25 + 0.96i$.

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 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

$$\oint_{\text{old contours}} = \oint_{\text{new contours}} + \int_{\overline{z_{cr}}}^{z_{cr}} 2\pi i \operatorname{Res}_{w=z} dz, \quad (4.4)$$

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 (4.4). Recall that the correlation functions are defined relative to a reference measure, and the right object to scale is

$$K_n(x, y) dy = \frac{1}{\sqrt{n}} K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n}) d(\Delta y).$$

The extra factor $n^{-1/2}$ compensates the prefactor \sqrt{n} in (4.2).

The single integral takes the form

$$\frac{-i}{2\pi} \int_{\overline{z_{cr}}}^{z_{cr}} dz = \frac{\sin(\arg z_{cr})}{\pi}. \quad (4.5)$$

The double integral in (4.4) 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 $n \rightarrow +\infty$. This is because the main term comes from the single integral, which does not vanish.

Note that

$$z_{cr} = \frac{X \pm \sqrt{X^2 - 4}}{2}, \quad \sin(\arg z_{cr}) = \frac{\sqrt{4 - X^2}}{2}.$$

This again establishes the *Wigner semicircle law* for the GUE kernel.

Remark 4.2. 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 [Lecture 7](#).

F Problems (due 2025-03-19)

F.1 Different global positions

Show that if in (4.1) we take $X \neq Y$, then $K_n(x, y)$ vanishes as $n \rightarrow +\infty$. Moreover, establish the rate of decay in n . Is it power-law or exponential?

F.2 Sine kernel

Compute the integral (4.5).

F.3 Discrete sine process

Define the discrete sine kernel on \mathbb{Z} by

$$K_{\text{dsine}}(x, y) := \begin{cases} \frac{\sin \rho(x - y)}{\pi(x - y)}, & x \neq y, \\ \frac{\rho}{\pi}, & x = y, \end{cases}$$

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:

$$\mathbb{P}\left(\underbrace{\circ \circ \dots \circ}_{n \text{ times}} \underbrace{\bullet \bullet \dots \bullet}_{n \text{ times}}\right), \quad \mathbb{P}\left(\underbrace{\circ \bullet \circ \bullet \dots \circ \bullet}_{2n \text{ points}}\right),$$

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?

References

- [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]. [↑8](#)
- [JN06] K. Johansson and E. Nordenstam, *Eigenvalues of GUE minors*, Electron. J. Probab. **11** (2006), no. 50, 1342–1371. arXiv:math/0606760 [math.PR]. [↑8](#)
- [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]. [↑8](#)
- [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. [↑8](#)
- [Oko02] A. Okounkov, *Symmetric functions and random partitions*, Symmetric functions 2001: Surveys of developments and perspectives, 2002. arXiv:math/0309074 [math.CO]. [↑11](#)
- [Olv74] F. W. J. Olver, *Asymptotics and Special Functions*, Academic Press, 1974. [↑10](#)

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

Lectures on Random Matrices (Spring 2025)

Lecture 7: Double contour integral kernel. Steepest descent and local statistics

Leonid Petrov

Monday, February 24, 2025*

Contents

1	Steepest descent for the GUE kernel	1
1.1	Recap	1
1.2	Scaling	3
1.3	Critical points	4
1.4	Imaginary critical points: $ X < 2$, “bulk”	4
1.5	Real critical points: $ X > 2$, “large deviations”	6
1.6	Double critical point: $ X = 2$, “edge”	7
1.7	Airy kernel, Tracy–Widom distribution, and convergence of the maximal eigenvalue	8
1.8	Remark: what happens for general β ?	9
2	Cutting corners: setup	9
3	Corners of Hermitian matrices	10
3.1	Principal corners	10
3.2	Interlacing	10
G	Problems (due 2025-03-25)	11
G.1	General bulk case	11
G.2	Large deviations	11
G.3	Airy kernel	12
G.4	Interlacing proof	12

1 Steepest descent for the GUE kernel

1.1 Recap

We continue the asymptotic analysis of the GUE kernel.

*Course webpage • Live simulations • TeX Source • Updated at 04:35, Monday 24th February, 2025

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

$$\exp\left(xt - \frac{t^2}{2}\right) = \sum_{n \geq 0} p_n(x) \frac{t^n}{n!},$$

one obtains by Cauchy's integral formula

$$p_n(x) = \frac{n!}{2\pi i} \oint_C \frac{\exp\left(xt - \frac{t^2}{2}\right)}{t^{n+1}} dt,$$

which leads to

$$\psi_n(x) = \frac{e^{-x^2/4}}{\sqrt{h_n}} \frac{n!}{2\pi i} \oint_C \frac{\exp\left(xt - \frac{t^2}{2}\right)}{t^{n+1}} dt.$$

Starting from the Fourier transform identity

$$\int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2} + itx\right) dt = \sqrt{2\pi} e^{-x^2/2},$$

and differentiating with respect to x , then changing variables, one obtains

$$\psi_n(x) = \frac{i e^{x^2/4}}{\sqrt{2\pi h_n}} \int_{-i\infty}^{i\infty} s^n e^{s^2/2 - sx} ds.$$

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 $1/(s-t)$):

$$K_n(x, y) = \frac{1}{(2\pi)^2} \oint_C dt \int_{-i\infty}^{i\infty} ds \frac{\exp\left\{\frac{s^2}{2} - sy - \frac{t^2}{2} + tx\right\}}{s-t} \left(\frac{s}{t}\right)^n.$$

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.

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 [Lecture 2](#), and for the GUE in [Lecture 4](#)) that the eigenvalues live on the scale \sqrt{n} . This means that to capture the local asymptotics, we need to scale

$$x = X\sqrt{n} + \frac{\Delta x}{\sqrt{n}}, \quad y = Y\sqrt{n} + \frac{\Delta y}{\sqrt{n}}, \quad \Delta x, \Delta y \in \mathbb{R}. \quad (1.1)$$

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 $Y = X$.

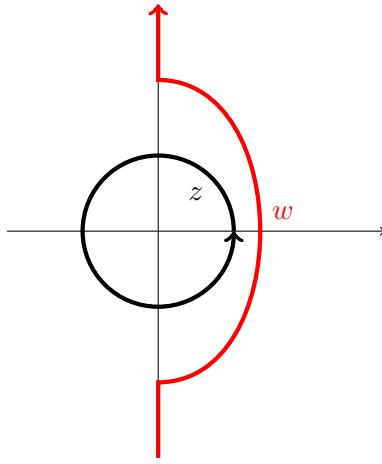


Figure 1: Integration contours for the GUE kernel.

Let us also make a change of the integration variables:

$$t = z\sqrt{n}, \quad s = w\sqrt{n}.$$

The integration contours for z and w look the same as for t and s , up to a rescaling ([Figure 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 1](#).

We thus have:

$$\begin{aligned} K_n(X\sqrt{n} + \Delta x/\sqrt{n}, X\sqrt{n} + \Delta y/\sqrt{n}) \\ = \frac{\sqrt{n}}{(2\pi)^2} \oint_C dz \int_{-i\infty}^{i\infty} dw \frac{\exp\left\{n\left(\log w - \log z + \frac{w^2}{2} - \frac{z^2}{2} + X(z-w) + \frac{z\Delta x - w\Delta y}{n}\right)\right\}}{w-z}. \end{aligned} \quad (1.2)$$

Remark 1.1. 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 (1.2) in the context of determinantal point processes was pioneered in [Oko02, Section 3].

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:

$$S'(z) = z + \frac{1}{z} - X = 0, \quad z_{cr} = \frac{X \pm \sqrt{X^2 - 4}}{2}. \quad (1.3)$$

Depending on whether $|X| < 2$, there are three cases. Unless $|X| = 2$, equation (1.3) has a single root, and thus $S''(z_{cr}) \neq 0$. We will consider the three cases in Sections 1.4 to 1.6 below.

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 G.1. We have

$$z_{cr} = i, \quad S''(z_{cr}) = 2.$$

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 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 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, 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

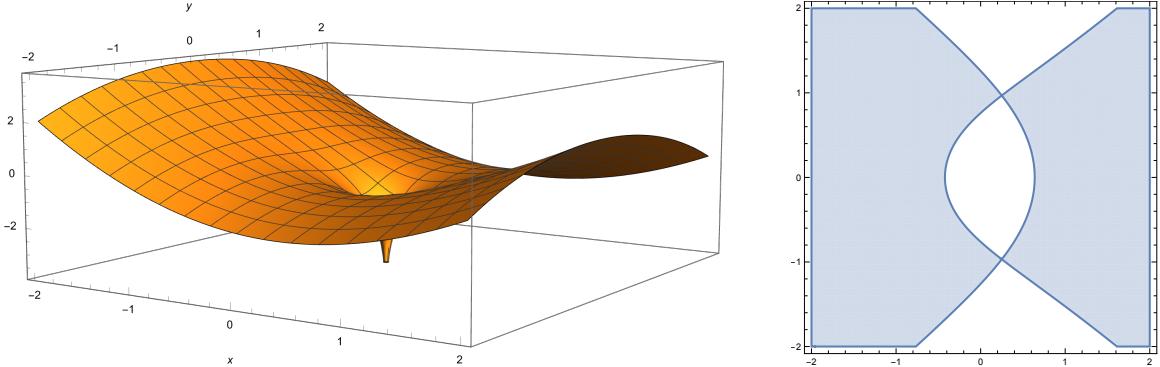


Figure 2: A 3D plot and a region plot of the regions where $\operatorname{Re} S(z) - \operatorname{Re} S(z_{cr})$ is positive (highlighted) or negative, in the case $X = \frac{1}{2}$. In this case, $z_{cr} \approx 0.25 + 0.96i$.

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

$$\oint_{\text{old contours}} = \oint_{\text{new contours}} + \int_{-i}^i 2\pi i \operatorname{Res}_{w=z} dz, \quad (1.4)$$

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 (1.4). Recall that the correlation functions are defined relative to a reference measure, and the right object to scale is

$$K_n(x, y) dy = \frac{1}{\sqrt{n}} d(\Delta y).$$

The extra factor $n^{-1/2}$ compensates the prefactor \sqrt{n} in (1.2).

The single integral takes the form

$$\frac{-i}{2\pi} \int_{-i}^i e^{z(\Delta x - \Delta y)} dz = \frac{\sin(\Delta x - \Delta y)}{\pi(\Delta x - \Delta y)}, \quad \Delta x, \Delta y \in \mathbb{R}. \quad (1.5)$$

Definition 1.2. The *sine kernel* is defined as

$$K_{\text{sine}}(x, y) := \begin{cases} \frac{\sin(x - y)}{\pi(x - y)}, & x \neq 0, \\ \frac{1}{\pi}, & x = 0. \end{cases}$$

(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

$$\operatorname{const} \cdot \frac{e^{n(\operatorname{Re} S(z_{cr}) - \operatorname{Re} S(z_{cr}))}}{\sqrt{n}} \sim \frac{\operatorname{const}}{\sqrt{n}}.$$

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 $n \rightarrow +\infty$. This is because the main term comes from the single integral, which does not vanish.

We have established the following result:

Proposition 1.3 (Bulk asymptotics at $X = 0$). *The correlation kernel K_n of the GUE has the following asymptotics close to zero as $n \rightarrow +\infty$:*

$$\lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} K_n \left(\frac{\Delta x}{\sqrt{n}}, \frac{\Delta y}{\sqrt{n}} \right) = K_{\text{sine}}(\Delta x, \Delta y), \quad \Delta x, \Delta y \in \mathbb{R}.$$

Consequently, the eigenvalues of the GUE converge to the sine process determined by the sine kernel (Definition 1.2), in the sense of finite-dimensional distributions.

Remark 1.4. 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 G.1.

1.5 Real critical points: $|X| > 2$, “large deviations”

For $X^2 > 4$, both solutions (1.3) are real. Let us assume $X > 2$, the case $X < -2$ is similar. For $X > 2$, both solutions are positive. Label these solutions as

$$z_+ = \frac{X + \sqrt{X^2 - 4}}{2}, \quad z_- = \frac{X - \sqrt{X^2 - 4}}{2}, \quad \text{so that } z_+ z_- = 1.$$

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

$$\exp \{n (\operatorname{Re} S(z_+) - \operatorname{Re} S(z_-))\} = O(e^{-\delta(X)n}), \quad |X| > 2. \quad (1.6)$$

Here $\delta(X) > 0$ for $|X| > 2$, and $\delta(X) \rightarrow 0$ when $|X| \rightarrow 2$.

The outcome (1.6) 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 G.2.

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 (1.3) merge into a double critical point $z_{cr} = 1$. We have

$$S'(1) = 0, \quad S''(1) = 0, \quad S'''(1) = 2.$$

Thus, the usual quadratic approximation fails and one must expand to third order. Writing

$$z = 1 + u, \quad w = 1 + v,$$

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

$$u = \frac{U}{n^{1/3}}, \quad v = \frac{V}{n^{1/3}},$$

so that

$$n[S(1 + v) - S(1 + u)] = \frac{V^3 - U^3}{3} + O(n^{-1/3}).$$

Moreover, the correct edge scaling for the spatial variables is obtained by writing

$$x = 2\sqrt{n} + \frac{\xi}{n^{1/6}}, \quad y = 2\sqrt{n} + \frac{\eta}{n^{1/6}}, \quad \xi, \eta \in \mathbb{R}.$$

We have

$$n(S(w) - S(z)) = n^{1/3}(\xi - \eta) + \frac{V^3 - U^3}{3} + \xi U - \eta V + O(n^{-1/3}).$$

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 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.

Thus, we have shown that under the rescaling, the GUE correlation kernel $K_n(x, y) dy$ converges to a new kernel.

Definition 1.5. Define the *Airy kernel* on \mathbb{R} by

$$K_{\text{Ai}}(\xi, \eta) = \frac{1}{(2\pi i)^2} \int_{e^{-\frac{\pi i}{3}}\infty}^{e^{\frac{\pi i}{3}}\infty} dV \int_{e^{-\frac{2\pi i}{3}}\infty}^{e^{2\frac{\pi i}{3}}\infty} dU \frac{\exp\left\{\frac{V^3 - U^3}{3} + U\xi - V\eta\right\}}{V - U}.$$

For another formula for the Airy kernel which does not involve integrals, see Problem G.3.

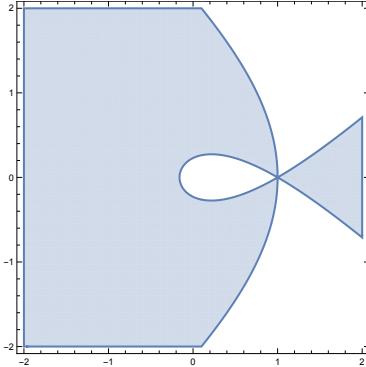


Figure 3: The plot of the region $\operatorname{Re} S(z) - \operatorname{Re} S(1) > 0$ for $X = 2$.

Proposition 1.6. *We have*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1/6}} K_n \left(2\sqrt{n} + \frac{\xi}{n^{1/6}}, 2\sqrt{n} + \frac{\eta}{n^{1/6}} \right) \rightarrow K_{\text{Ai}}(\xi, \eta).$$

Consequently, the eigenvalue statistics at the edge of the spectrum converge to the Airy point process, in the sense of fine-dimensional distributions.

1.7 Airy kernel, Tracy–Widom distribution, and convergence of the maximal eigenvalue

Let us make a few remarks on the asymptotic results of Propositions 1.3 and 1.6. 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 α_1), and that λ_{\max} converges to α_1 under appropriate rescaling:

$$\lim_{n \rightarrow \infty} n^{1/6} (\lambda_{\max} - 2\sqrt{n}) = \alpha_1. \quad (1.7)$$

This is indeed the case, but to show (1.7), one needs to show the convergence in distribution:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(n^{1/6} (\lambda_{\max} - 2\sqrt{n}) \leq x \right) \rightarrow \mathbb{P}(\alpha_1 \leq x). \quad (1.8)$$

Both events (1.8) are so-called *gap probabilities*, for example,

$$\mathbb{P}(\alpha_1 \leq x) = \mathbb{P}(\text{there are no eigenvalues in the interval } (x, \infty)),$$

which is expressed as the Fredholm determinant

$$\det(1 - K_{\text{Ai}})_{(x, \infty)} = 1 + \sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \int_x^{\infty} dy_1 \int_x^{\infty} dy_2 \cdots \int_x^{\infty} dy_m \det_{i,j=1}^m K_{\text{Ai}}(y_i, y_j). \quad (1.9)$$

Thus, to get (1.8)), 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 (1.9) defines a cumulative probability distribution function in x :

$$F_2(x) = \det(1 - K_{\text{Ai}})_{(x, \infty)}. \quad (1.10)$$

The distribution (1.10) 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].

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].

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 $k = 1, \dots, n$. 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 $1 \leq k \leq n$) have jointly distributed eigenvalues that

exhibit a determinantal structure. We will employ the technique of *polynomial equations* and then *loop equations* to study limits.

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 $O(n^2)$.

3 Corners of Hermitian matrices

3.1 Principal corners

Let H be an $n \times n$ Hermitian matrix. For each $1 \leq k \leq n$, define the *top-left $k \times k$ corner* $H^{(k)}$ by

$$H^{(k)} = [H_{ij}]_{1 \leq i,j \leq k}.$$

Since H is Hermitian, each $H^{(k)}$ is also Hermitian. Let

$$\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \dots \geq \lambda_k^{(k)}$$

denote the eigenvalues of $H^{(k)}$. Then the collection

$$\{\lambda_j^{(k)} : 1 \leq j \leq k \leq n\}$$

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 $\{1, \dots, n\} \times \mathbb{R}$.

3.2 Interlacing

A fundamental feature of Hermitian matrices is that the eigenvalues of corners interlace with the eigenvalues of the full matrix:

Proposition 3.1. *If $\nu_1 \geq \dots \geq \nu_n$ are the eigenvalues of H itself (i.e., the full $n \times n$ matrix), and $\mu_1 \geq \dots \geq \mu_{n-1}$ are the eigenvalues of $H^{(n-1)}$, then we have:*

$$\nu_1 \geq \mu_1 \geq \nu_2 \geq \mu_2 \geq \dots \geq \mu_{n-1} \geq \nu_n.$$

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 $\nu_1 \geq \nu_2 \geq \dots \geq \nu_n$, 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 G.4. \square

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.

G Problems (due 2025-03-25)

G.1 General bulk case

Perform the asymptotic analysis of the correlation kernel as in Section 1.4, but in the general case $-2 < X < 2$.

G.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

$$E_{n,\eta} := \left\{ \exists i \in \{1, \dots, n\} \text{ such that } W_{ii} \geq \eta \right\}.$$

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 $\mathbb{P}(E_{n,\eta})$.

G.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

$$K_{Ai}(\xi, \eta) = \frac{Ai(\xi) Ai'(\eta) - Ai(\eta) Ai'(\xi)}{\xi - \eta}.$$

Note that this expression is parallel to the sine kernel,

$$\frac{\sin(x-y)}{\pi(x-y)} = \frac{\sin x \cos y - \cos x \sin y}{\pi(x-y)}, \quad \cos x = (\sin x)'.$$

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 $K_{Ai}(\xi, \eta)$ from Definition 1.5.

G.4 Interlacing proof

Finish the proof of Proposition 3.1.

References

- [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]. [↑11](#)
- [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. [↑10](#)
- [For93] P. J. Forrester, *The spectrum edge of random matrix ensembles*, Nuclear Physics B **402** (1993), 709–728. [↑9](#)
- [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]. [↑9](#)
- [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]. [↑9](#)
- [GXZ24] V. Gorin, J. Xu, and L. Zhang, *Airy $_\beta$ line ensemble and its Laplace transform*, arXiv preprint (2024). arXiv:2411.10829 [math.PR]. [↑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. [↑12](#)
- [Oko02] A. Okounkov, *Symmetric functions and random partitions*, Symmetric functions 2001: Surveys of developments and perspectives, 2002. arXiv:math/0309074 [math.CO]. [↑4](#)
- [RRV11] J. Ramirez, B. Rider, and B. Virág, *Beta ensembles, stochastic Airy spectrum, and a diffusion*, Jour. AMS **24** (2011), no. 4, 919–944. arXiv:math/0607331 [math.PR]. [↑9](#)

- [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. ↑[9](#)
- [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. ↑[9](#)
- [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]. ↑[9](#)

L. PETROV, UNIVERSITY OF VIRGINIA, DEPARTMENT OF MATHEMATICS, 141 CABELL DRIVE, KERCHOV HALL, P.O. Box 400137, CHARLOTTESVILLE, VA 22904, USA
E-mail: lenia.petrov@gmail.com

Lectures on Random Matrices (Spring 2025)

Lecture 8: Loop equations for corners

Leonid Petrov

Wednesday, February 26, 2025*

H Problems (due 2025-03-25)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 9: Title TBD

Leonid Petrov

DATE, 2025*

I Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 10: Title TBD

Leonid Petrov

DATE, 2025*

J Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 11: Title TBD

Leonid Petrov

DATE, 2025*

K Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 12: Title TBD

Leonid Petrov

DATE, 2025*

L Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 13: Title TBD

Leonid Petrov

DATE, 2025*

M Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 14: Title TBD

Leonid Petrov

DATE, 2025*

N Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025

Lectures on Random Matrices (Spring 2025)

Lecture 15: Title TBD

Leonid Petrov

DATE, 2025*

O Problems (due DATE)

References

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

*[Course webpage](#) • [Live simulations](#) • [TeX Source](#) • Updated at 07:19, Wednesday 12th February, 2025