Lectures on Random Matrices (Spring 2025)
Lecture 1: Moments of random variables and random matrices

Leonid Petrov

Monday, January 13, 2025*

Contents

*Course webpage • TeX Source • Updated at 14:35, Monday 27th January, 2025

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

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

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

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

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

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

Convergence in distribution means

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

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:

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

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

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

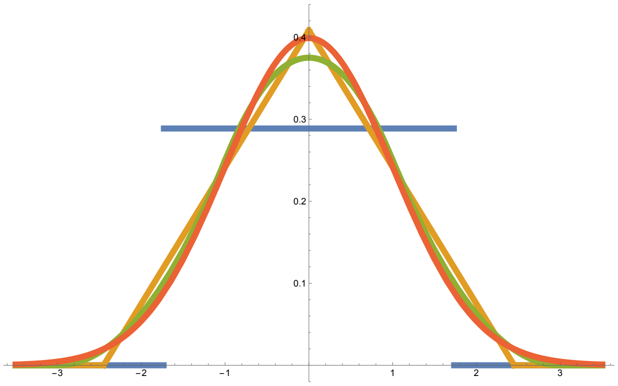

Figure 1: Densities of U_1, U_1+U_2, $U_1+U_2+U_3$ (where U_i are iid uniform on $[0,1]$), and $\mathcal{N}(0,1)$, normalized to have the same mean and variance.

Introduce the normalized quantity

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

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

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

For large n, Stirling's formula gives

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

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

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

Thus,

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

More precisely, one often writes

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

where $p \approx k/n$ thanks to the fact that z (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:*

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

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. □

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:

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

Taking the expectation and using linearity, we have:

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

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

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

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

Example 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 $\mathbb{E}[Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4}] = \mathsf{m}_4$) is to just assign one of the j_p to be 4, this can be done in n ways.

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

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

2.3.2 n-dependent factor

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

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 $n \to \infty$:

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

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

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

Therefore, the moments of Z_n (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:

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

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

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;

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

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

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

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

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 ($n \le m$), 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 $\mathbb{R}$:

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

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

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

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

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

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

$$\nu_n \coloneqq \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \longrightarrow \mu_{\mathrm{sc}}, \tag{3.4}$$

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

$$\mu_{\mathrm{sc}}(dx) \coloneqq \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{|x|\le 2}\,dx. \tag{3.5}$$

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

$$\int_{\mathbb{R}} f(x)\,\nu_n(dx) \longrightarrow \int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx), \qquad n \to \infty, \tag{3.6}$$

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 $\int_{\mathbb{R}} f(x)\,\mu_{\mathrm{sc}}(dx)$.

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 $k \ge 1$, the normalized k-th moment of the empirical spectral distribution of $W_n/\sqrt{n}$ is given by

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

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

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

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

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

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

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:

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

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 $n \to \infty$.

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

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

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

A.2 Convergence in distribution

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

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

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

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

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

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 $\lambda = (\lambda_1, \ldots, \lambda_N)$, $\lambda_1 \geq \ldots \geq \lambda_N$. Let us fix λ to be, say, $\lambda = (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, for some proportion of 1's and 0's (you can start with half ones and half zeros).

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

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

A.8 Invariance of the GOE

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

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

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

*Course webpage • TeX Source • Updated at 03:51, Wednesday 22nd January, 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}:

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

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

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

Remark 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)

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

we compute each term:

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

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

2.2 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:

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

Alternatively, they have the closed form[1]

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

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

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

[1] See 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

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

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

Remark 2.5. 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. ☐

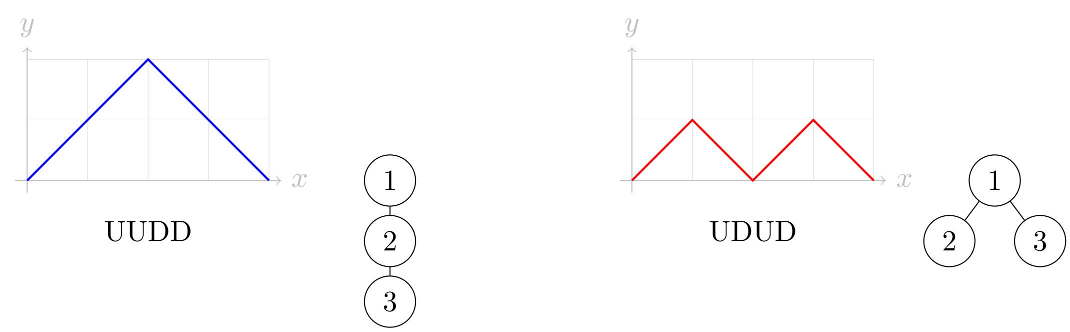

Figure 1: The two possible Dyck paths of semilength $m = 2$ and their corresponding rooted plane trees.

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 $j=0,\ldots,m$. Then the first and the $(2j+1)$-st steps are, respectively, U and D. From 0 to $2j+2$, the path does not return to the x-axis, so we can remove the first and the $(2j+1)$-st steps, and get a proper Dyck path of semilength j. The remainder of the Dyck path is a Dyck path of semilength $m-j$. This yields the desired recurrence. ☐

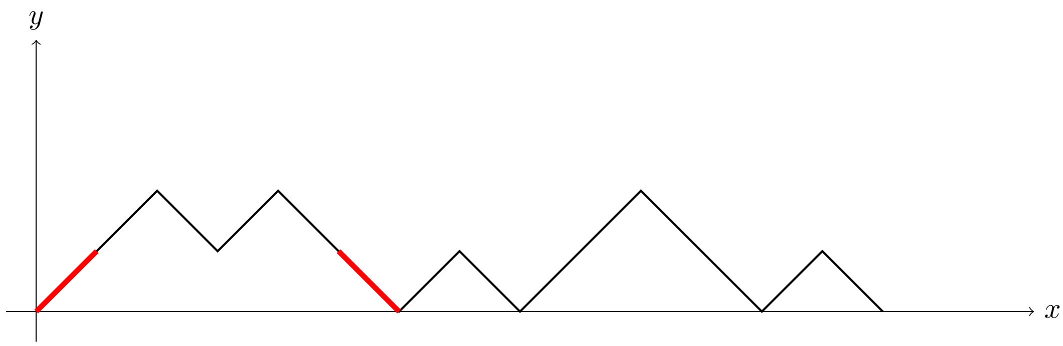

Figure 2: Illustration of a Dyck path decomposition for the proof of Proposition 2.7.

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

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

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

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*

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

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

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

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

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

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

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

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

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

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

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,

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

where in the last step we used Proposition 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 $\mathbb{E}[Y_{n,k}] \to m_k$, we conclude

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

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

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

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

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

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

On the other hand, we have already shown

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

Thus

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

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 $\nu_n \to \nu$ in the weak topology (almost surely). This completes the proof that the ESD of our Wigner matrix (rescaled by $\sqrt{n}$) converges to the semicircle distribution weakly almost surely, modulo Proposition 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

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

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

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

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

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

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

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

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

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

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

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

We split into two cases:

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

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

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

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

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

Theorem 5.2 (Correlated entries [SSB05]). *Let $W_n = \left(\frac{1}{\sqrt{n}} X_{pq}\right)_{1\le p,q\le n}$ be a sequence of $n \times n$ Hermitian random matrices where:*

1. *The entries X_{pq} are complex random variables that are:*
   - *Centered: $\mathbb{E}[X_{pq}] = 0$,*
   - *Unit variance: $\mathbb{E}[|X_{pq}|^2] = 1$,*
   - *Moment bound: $\sup_n \max_{p,q=1,\ldots,n} \mathbb{E}\left[|X_{pq}|^k\right] < \infty$ for all $k \in \mathbb{N}$.*
2. *There exists an equivalence relation $\sim_n$ on pairs of indices (p,q) in $\{1,\ldots,n\}^2$ such that:*
   - *Entries $X_{p_1q_1},\ldots,X_{p_jq_j}$ are independent when $(p_1,q_1),\ldots,(p_j,q_j)$ belong to distinct equivalence classes.*
   - *The relation satisfies the following bounds:*

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

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

     (c) $\#\big\{(p,q,p') \in \{1,\ldots,n\}^3 \mid (p,q) \sim_n (q,p')$ *and* $p \ne p'\big\} = o(n^2)$.
3. *The matrices are Hermitian: $X_{pq} = \overline{X_{qp}}$. In particular, $(p,q) \sim_n (q,p)$, and this is consistent with the conditions on the equivalence relation.*

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

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

Theorem 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 $1 \le i \le n$.*

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 $\mathbb{E}[X_{ij}^2] = 1$.*

[2] For $\beta = 1, 2, 4$, this is the joint eigenvalue density of the Gaussian Orthogonal, Unitary, and Symplectic Ensembles, respectively. For general β, there is no invariant random matrix distribution (while the eigenvalue density (5.1) makes sense), and we can still treat all the β cases in a unified manner.

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

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

Theorem 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*

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

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

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

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

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, Kerchof 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

*Course webpage • TeX Source • Updated at 08:41, Wednesday 29th January, 2025

1 Recap

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

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} = \mu_{\mathrm{sc}}, \tag{1.1}$$

where

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

The convergence in (1.1) is weakly almost sure. The way we got the result is by expanding $\mathbb{E}\operatorname{Tr}(W^k)$ 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}, $1 \le i, j \le n$, such that Y_{ij} are mean zero, variance 1. Then for $1 \le i < j \le n$, we have for the matrix $W = (X_{ij})$:

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

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

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

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

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

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

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

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

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

Both GOE and GUE have real eigenvalues $\lambda_1 \ge \ldots \ge \lambda_n$. We are going to describe the joint distribution of these eigenvalues. Despite the fact that 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 $\lambda_1 \le \cdots \le \lambda_n$ of $W/\sqrt{2}$ have a joint probability density function on $\mathbb{R}^n$ given by:*

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

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

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

[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 $W/\sqrt{2}$:

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

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

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

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

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

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

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

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

2.4 Step B. Spectral decomposition

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

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

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

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

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

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

where dQ is the Haar measure[2] on $O(n)$, and $d\Lambda$ is the Lebesgue measure on $\mathbb{R}^n$. The Lebesgue measure later needs to be restricted to the "Weyl chamber" $\lambda_1 \leq \cdots \leq \lambda_n$ 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*

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

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

Remark 2.5. Equivalently, one often writes

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

is the *Vandermonde determinant*.

We prove Lemma 2.4 in the rest of this subsection.

Consider small perturbations of Λ and Q. Write

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

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

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

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

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

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

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

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

Computing δW. Under an infinitesimal change, say,

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

we have

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

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

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

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

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

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

This completes the proof of Lemma 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:

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

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

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

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 $n \le m$. We can perform the *singular value decomposition* (SVD) of X:

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

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

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

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

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 $\lambda_1, \ldots, \lambda_n \ge 0$ of the $n \times n$ Wishart matrix W have the joint density on $\{\lambda_i \ge 0\}$ proportional to*

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

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

Idea of proof (sketch). The proof is a variant of the derivation for the joint eigenvalue density in the GOE/GUE case (see Section 2.2). One writes down the joint distribution of all entries of X, changes variables to singular values and orthogonal/unitary transformations, and identifies the Jacobian factor as $\prod_{i<j} |s_i^2 - s_j^2|^\beta = \prod_{i<j} |\lambda_i - \lambda_j|^\beta$. The extra factors in front arise from the powers of λ_i (i.e. from $\prod_i s_i$) and the Gaussian exponential $\exp\big(-\frac{1}{2}\sum s_i^2\big)$ when reshaped to $\exp\big(-\frac{1}{2}\sum \lambda_i\big)$. $\square$

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

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

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

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

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

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

Example 3.4. For $n = k = 1$, we have

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

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

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 $n \le k \le t$. Assume further that X and Y are independent of each other (this is the null hypothesis in statistics).*

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

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

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 $n \le t$ and X is in general position. The distribution above reflects the geometry of overlapping projectors in a higher-dimensional space $\mathbb{R}^t$ (or $\mathbb{C}^t$).

3.3 General Pattern and β-Ensembles

We have now seen three classical examples:

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

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

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

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

Remark 3.7. 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:

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

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

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

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

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

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

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

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

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

Remark 5.2 (Chi-square distributions)**.** The *chi-square distribution* with ν degrees of freedom, denoted by χ^2_ν, is a fundamental distribution in statistics and probability theory. It arises naturally as the distribution of the sum of the squares of ν independent standard normal random variables. Formally, if $Z_1, Z_2, \ldots, Z_\nu$ are independent random variables with $Z_i \sim \mathcal{N}(0,1)$, then the random variable

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

follows a chi-square distribution with ν degrees of freedom, i.e., $Q \sim \chi^2_\nu$. In the context of the Dumitriu–Edelman tridiagonal model (Theorem 5.1), the subdiagonal entries α_j are defined as $\alpha_j = \sqrt{\frac{\chi^2_{n-j}}{2}}$. One can call this a *chi random variable*, as this is a square root of a chi-square variable.

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

Idea of proof of Theorem 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 $\alpha_j \sim \sqrt{\frac{\chi^2_{n-j}}{2}}$ and $d_i \sim \mathcal{N}(0,1)$. One uses the rotational invariance of Gaussian matrices to ensure at each step that the "residual vector" is isotropic (i.e., its distribution is invariant under orthogonal transformations). The norm of that vector yields the χ^2-type variables. □

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

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

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

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

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 $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal with $(\lambda_1 \ge \cdots \ge \lambda_n)$, 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

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

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 on unordered form):

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

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

Hint: The following identity might be useful:

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

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

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

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

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

Prove that:

1. $\|v\|^2$ follows a χ^2_ν distribution with ν degrees of freedom (determine ν in terms of n).
2. The direction $v/\|v\|$ is uniformly distributed on the unit sphere $\mathbb{S}^{n-1}$ and is independent of $\|v\|$.

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

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 $n \le k \le m$. Consider the matrix

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

1. Prove that it is well-defined (invertible denominator) with probability 1, and that it is symmetric and diagonalizable in $\mathbb{R}^n$.
2. Show that its eigenvalues lie in $[0,1]$ and follow a Jacobi (MANOVA) distribution of parameters $\beta = 1$ and $\big(n,k,m\big)$.
3. Identify explicitly how these parameters match the shape parameters in the standard multivariate Beta / Jacobi pdf

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

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

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

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

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, Kerchof 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 via tridiagonalization.
Orthogonal polynomial ensembles

Leonid Petrov

Wednesday, January 29, 2025*

Contents

*Course webpage • TeX Source • Updated at 10:13, Wednesday 29$^{\text{th}}$ January, 2025

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

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

1.2 Tridiagonalization

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

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

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

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

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

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

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

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

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

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

Remark 2.2 (Chi-square distributions)**.** The *chi-square distribution* with ν degrees of freedom, denoted by χ^2_ν, is a fundamental distribution in statistics and probability theory. It arises naturally as the distribution of the sum of the squares of ν independent standard normal random variables. Formally, if $Z_1, Z_2, \ldots, Z_\nu$ are independent random variables with $Z_i \sim \mathcal{N}(0,1)$, then the random variable

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

follows a chi-square distribution with ν degrees of freedom, i.e., $Q \sim \chi^2_\nu$. In the context of Theorem 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

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

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

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

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

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

2.2 Dumitriu–Edelman GβE 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 GβE 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

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

are chi-distributed random variables.

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

Proposition 2.4 (Without proof). *For $\beta = 2$, the GβE is the tridiagonal form of the GUE random matrix model, which is the random complex Hermitian matrix with Gaussian entries and maximal independence. Similarly, for $\beta = 4$, the GβE is the tridiagonal form of the GSE random matrix model.*

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

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

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

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

2.3 The case $\beta = 2$

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

Proposition 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 $\mathcal{N}(0, \frac{1}{2}) + \mathrm{i}\,\mathcal{N}(0, \frac{1}{2})$ and each diagonal entry $\mathcal{N}(0,1)$. The ordered eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$ (or, without ordering, thought of as an unordered set) satisfy the joint probability density*

$$p(\lambda_1, \ldots, \lambda_n) \;=\; \frac{1}{Z_{n,2}} \prod_{j=1}^n e^{-\frac{1}{2}\lambda_j^2} \prod_{1 \le i < j \le n} (\lambda_i - \lambda_j)^2, \tag{2.2}$$

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

$$p(\lambda_1, \ldots, \lambda_n) \;=\; \text{const} \cdot \det\Big[\psi_{j-1}(\lambda_k) e^{-\frac{\lambda_k^2}{4}}\Big]_{j,k=1}^{n} \det\Big[\psi_{j-1}(\lambda_k) e^{-\frac{\lambda_k^2}{4}}\Big]_{j,k=1}^{n} \tag{2.3}$$

(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:

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

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

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

The difference between H_k and ψ_k entering Proposition 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 $\Delta(\lambda_1, \ldots, \lambda_n) = \det\big[\lambda_k^{j-1}\big]_{j,k=1}^{n}$. Next, the Vandermonde determinant is also equal to the determinant built out of any monic family of polynomials of the corresponding degrees (by linear transformations), and so we get the desired representation. □

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

3 Characteristic Polynomial and Three-Term Recurrence

Once we have $T = (t_{ij})$ tridiagonal, the characteristic polynomial of T takes a well-known form governed by a three-term recurrence. Denote

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

Definition 3.1 (Characteristic Polynomials $p_k(\lambda)$)**.** For $1 \leq k \leq n$, let T_k be the top-left $k \times k$ submatrix of T. Define

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

Moreover, set $p_0(\lambda) := 1$ by convention.

Lemma 3.2 (Three-Term Recurrence)**.** *Let $p_k(\lambda)$ be as above, with $p_1(\lambda) = d_1 - \lambda$. Then for $k \geq 1$,*

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

Idea of proof. One checks the base case $k = 1, 2$ directly by computing determinants of 1×1 and 2×2 blocks. For the general step, expand $\det(T_{k+1} - \lambda I_{k+1})$ by minors along the last row or column. The block structure ensures exactly the claimed recurrence. □

Remark 3.3. This is analogous to recurrences in orthogonal polynomials, e.g. the three-term recursion for polynomials orthonormal with respect to certain measures. In fact, the polynomial $p_n(\lambda)$ can be viewed as an orthogonal polynomial if $\alpha_j > 0$. This interplay is vital in random matrix theory.

4 Semicircle Law via the Tridiagonal Form

We now present a fairly detailed sketch (or roadmap) for how the tridiagonal form yields the Wigner semicircle law. Recall we aim to show:

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}} \;\longrightarrow\; \mu_{\mathrm{sc}}(dx) \;=\; \frac{1}{2\pi}\sqrt{4 - x^2}\; 1_{|x|\leq 2} \quad \text{almost surely.}$$

4.1 The Scaling and Law of Large Numbers in α_j

Write

$$T = \begin{pmatrix} d_1 & \alpha_1 & & \\ \alpha_1 & d_2 & \alpha_2 & \\ & \alpha_2 & d_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix}.$$

In the Dumitriu–Edelman setting, $\alpha_j^2 = \frac{1}{2}\,\chi^2_{(n-j)}$, so $E[\alpha_j^2] = (n-j)/2$ and α_j^2 is tightly concentrated around its mean for large n. Precisely, for any $\varepsilon > 0$,

$$\Pr\Bigl(\bigl|\alpha_j^2 - \tfrac{n-j}{2}\bigr| > \varepsilon(n-j)\Bigr) \;\leq\; e^{-c(n-j)}$$

for some $c > 0$, or a similar bound from large deviation estimates. Hence, with overwhelming probability,

$$\alpha_j \;\approx\; \sqrt{\frac{n-j}{2}}.$$

4.2 Diagonal vs. Subdiagonal Entries

When we examine $\frac{1}{\sqrt{n}}T$, the diagonal entries become $\frac{d_i}{\sqrt{n}}$. Since $d_i \sim \mathcal{N}(0,1)$ i.i.d., with probability going to 1 these entries lie within $O(n^{-1/2})$, so they vanish in the large-n limit. Meanwhile,

$$\frac{\alpha_j}{\sqrt{n}} \approx \sqrt{\frac{n-j}{2n}} \approx \sqrt{\frac{1-j/n}{2}}.$$

In the bulk region (i.e. $j \approx \theta n$ for $\theta \in (0,1)$), we have $\alpha_j/\sqrt{n} \approx \sqrt{\frac{1-\theta}{2}}$. In short, the subdiagonal elements (scaled by $1/\sqrt{n}$) remain of order 1, while the diagonal elements vanish.

4.3 Characteristic Polynomial and Recurrence Analysis

Denote

$$p_n(\lambda) = \det\Big(\tfrac{1}{\sqrt{n}}T - \lambda I\Big).$$

Equivalently, λ is an eigenvalue of $\frac{1}{\sqrt{n}}T$ if and only if $\mu = \sqrt{n}\,\lambda$ is an eigenvalue of T. We want to understand the distribution of the roots $\lambda_i(\frac{1}{\sqrt{n}}T)$ as $n \to \infty$. The three-term recurrence for $p_n(\lambda)$ can be viewed in the limit as $n \to \infty$, turning into an integral equation for the Stieltjes transform of the limiting measure.

Stieltjes Transform Argument (Sketch). Set

$$G_n(z) = \frac{1}{n}Tr\Big(\big(\tfrac{1}{\sqrt{n}}T - z\big)^{-1}\Big),$$

the Stieltjes transform. As n grows, the main input is that $\alpha_j \approx \sqrt{\frac{1-j/n}{2}}$; substituting this approximate profile into the recursion for Green's functions (a linear difference equation akin to orthogonal polynomials) yields a limiting functional equation. Solving that equation for $G(z)$ leads to

$$G(z) = \frac{z \pm \sqrt{z^2-4}}{2},$$

with the appropriate branch cut. This is the well-known Stieltjes transform for the semicircle law on $[-2,2]$. (See advanced texts, e.g. *Deift* or *Tao's* books on RMT for a full rigorous derivation.)

Moment Argument (Sketch). One can also proceed by computing or bounding the moments $E\big[\frac{1}{n}Tr(\frac{1}{\sqrt{n}}T)^k\big]$ and showing that they match the semicircle moments. Indeed, as $n \to \infty$, the diagonal part becomes negligible, while the subdiagonal structure essentially forces closed loops in the sum expansions, reproducing the Catalan numbers that appear in the Wigner moment method (Lectures 1–2). The advantage of tridiagonalization is a simpler combinatorial interpretation of the entries used in each closed loop.

4.4 Conclusion: Wigner's Semicircle Law

Putting these ingredients together, we conclude that with probability 1,

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i(\frac{1}{\sqrt{n}}W)} \;\longrightarrow\; \mu_{\rm sc}(dx) \;=\; \frac{1}{2\pi}\sqrt{4-x^2}\;1_{|x|\le 2}$$

This completes the proof by tridiagonal methods.

Remark 4.1 (Universality)**.** The argument here specifically used the Gaussian Wigner distribution for off-diagonal entries to get an explicit χ^2 structure. However, the final result (semicircle law) remains true under vastly weaker assumptions on the entries. The *universality principle* states that the global spectral behavior is insensitive to fine details of the distribution (e.g. it depends primarily on the first two moments).

5 Exercises (Due 2025-02-28)

Below are problems elaborating on the main concepts. They range from verifying computations to exploring deeper aspects of tridiagonalization and simulations.

1. Detailed Householder Steps and Reflection Properties

(a) **Constructing a Reflection that Maps One Vector to Another.** Let $x, y \in \mathbb{R}^n$ be nonzero. Show how to pick v so that $Hx = y$, where $H = I - 2\frac{v\,v^\top}{\|v\|^2}$ is a Householder reflection. Why does H remain orthogonal?

(b) **Eliminating Below-Diagonal Entries.** In the first step of the tridiagonalization algorithm, pick a reflection H_1 that modifies the subspace spanned by $a_{21}, \ldots, a_{n1}$. Show explicitly how H_1 zeroes out all these subdiagonal entries while keeping a_{11} unchanged (aside from possibly a sign).

(c) **Symmetry and Superdiagonal.** Why does $H_1(A)H_1^\top$ also have the corresponding superdiagonal entries zeroed out in row 1? (Hint: Use A is symmetric: $A_{ij} = A_{ji}$.)

2. Three-Term Recurrence Warmup

(a) **Base Cases.** Compute $p_1(\lambda)$ and $p_2(\lambda)$ for a 2×2 matrix

$$\begin{pmatrix} d_1 - \lambda & \alpha_1 \\ \alpha_1 & d_2 - \lambda \end{pmatrix}.$$

Verify that $p_2(\lambda) = (d_2 - \lambda)\,p_1(\lambda) - \alpha_1^2\,p_0(\lambda)$.

(b) **Inductive Step.** Using determinant expansion along the $(k+1)$-st row or column, outline why $p_{k+1}(\lambda) = (d_{k+1} - \lambda)p_k(\lambda) - \alpha_k^2 p_{k-1}(\lambda)$ for $k \ge 1$.

3. Tridiagonal Model for GOE and GUE

(a) **Real Case (GOE).** Starting with a real Wigner matrix W ($X_{ij} \sim \mathcal{N}(0,1)$ for $i < j$, $X_{ii} \sim \mathcal{N}(0,2)$), show that the Householder steps produce a tridiagonal T whose diagonal entries are $d_i \sim \mathcal{N}(0,1)$ and subdiagonals $\alpha_j^2 = \frac{1}{2}\chi^2_{(n-j)}$.

(b) **Complex Case (GUE).** For a complex Hermitian W, the off-diagonal entries are $\mathcal{N}(0, \frac{1}{2}) + i\,\mathcal{N}(0, \frac{1}{2})$. Sketch how the same approach yields a complex Hermitian tridiagonal form with d_i real normal and α_j drawn from appropriate χ distributions. (You do not need a fully rigorous proof; just highlight the changes in dimension counting for real vs. complex parts.)

4. Semicircle Law via Stieltjes Transform

(a) **Defining the Green's Function.** For $z \in C \setminus \mathbb{R}$, let

$$G_n(z) = \frac{1}{n}\, Tr\Big(\big(\tfrac{1}{\sqrt{n}}T - zI\big)^{-1}\Big).$$

Argue (informally) that $G_n(z)$ converges to a limit $G(z)$ which must satisfy an algebraic equation derived from the tridiagonal structure and the typical size of α_j.

(b) **Solving for $G(z)$.** Show that $G(z)$ satisfies

$$G(z)^2 + z\,G(z) + 1 = 0,$$

and deduce that

$$G(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

Hence identify the imaginary part of $G(z)$ on the real interval $(-2, 2)$, concluding that the limiting distribution is the semicircle law.

5. Simulation of Tridiagonal vs. Dense Wigner

Write a small program (in Python, MATLAB, or another language):

(a) Generate a dense GOE matrix W of size $n = 1000$ and scale by $\frac{1}{\sqrt{n}}$. Compute eigenvalues and plot a histogram.

(b) Generate the corresponding tridiagonal matrix T from the Dumitriu–Edelman approach (diagonal $d_i \sim \mathcal{N}(0,1)$, subdiag $\alpha_j = \sqrt{\frac{1}{2}\,\chi^2_{n-j}}$). Compute eigenvalues of $\frac{1}{\sqrt{n}}T$ and compare histograms. They should match well with the semicircle shape for sufficiently large n.

(c) (Optional) Investigate how large n must be before the histogram looks convincingly semicircular.

6. Wishart and MANOVA Exercises

(a) **Wishart from Data Matrix.** Generate an $n \times m$ data matrix X with iid $\mathcal{N}(0,1)$. Form $W = X\,X^\top$. Plot the normalized eigenvalues λ_i vs. the Marchenko–Pastur distribution μ_{MP} (with aspect ratio m/n). Discuss approximate agreement for moderate n, m.

(b) **Jacobi (MANOVA).** Let X be $n \times t$ and Y be $k \times t$ independent $\mathcal{N}(0,1)$. Form the matrix $M = (X\,X^\top + Y\,Y^\top)^{-1}(X\,X^\top)$ and find its eigenvalues in $[0,1]$. Plot their histogram and compare with the Jacobi Beta distribution.

6 Further Reading and Next Steps

- **Local Laws and Universality.** After establishing the global semicircle distribution, one may delve into local spectral laws (e.g. the sine kernel or GOE Tracy–Widom distribution at the edge). The *local semicircle law* refines the analysis of the Green's function in small intervals.

- **Dyson Brownian Motion.** Another approach interprets the eigenvalues as particles with logarithmic repulsion and uses stochastic differential equations to show that equilibrium distributions converge to the β-ensembles.

- **Other Ensembles.** Beyond Wigner, Wishart, and Jacobi, one finds many integrable and combinatorial random matrices (e.g. discrete random partitions, polynomial ensembles, random tilings). Orthogonal polynomial techniques remain central in these broader contexts.

- **References.** - I. Dumitriu and A. Edelman, *Matrix models for beta ensembles*, J. Math. Phys., 2002. - T. Tao, *Topics in Random Matrix Theory*, 2012. - P. Deift, *Orthogonal Polynomials and Random Matrices*, 2000. - M. Mehta, *Random Matrices*, 3rd ed., Elsevier, 2004. - T. Anderson, *An Introduction to Multivariate Statistical Analysis*, for Wishart and MANOVA.

End of Lecture 4. In this expanded lecture, we detailed how any real symmetric matrix is tridiagonalized, derived the three-term recurrence for the characteristic polynomial, and showed how this structure underlies a clean proof of Wigner's semicircle law for random Wigner matrices. We also saw how these ideas extend to Wishart and Jacobi ensembles, bridging us toward the broader β-ensemble world. Upcoming lectures will further explore local eigenvalue statistics, universality results, and connections with integrable probability.

D Problems (due 2025-02-28)

D.1 Eigenvalue density of GβE

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

Notes for the lecturer

Add / finish up the discussion about tridiagonalization. The notes for L3 are updated, but mention it here.

The reflection is mapping any vector into any other vector (unit vectors); Also, we apply it in $(n-1)$-dimensional space in the first pass.

maybe make a simulation for Wishart and MANOVA in L4

D.2 Characteristic Polynomial and Three-Term Recurrence

Consider $p_n(\lambda) = \det(T - \lambda I)$. Because T is tridiagonal, we have the classical three-term recurrence for these characteristic polynomials:

$$p_0(\lambda) := 1, \quad p_1(\lambda) := d_1 - \lambda,$$
$$p_{k+1}(\lambda) \ = \ (d_{k+1} - \lambda)\, p_k(\lambda) \ - \ \alpha_k^2\, p_{k-1}(\lambda), \quad (k = 1, \ldots, n-1).$$

The eigenvalues of T are precisely the roots of $p_n(\lambda)$.

D.3 Sketch of the Semicircle Limit Proof

We want to show that the empirical distribution

$$L_n \ = \ \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}$$

(where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T) converges weakly to the semicircle law

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

as $n \to \infty$. A typical outline:

1. **Law of Large Numbers for α_j.** Since $\alpha_j^2 = \frac{1}{2}\, \chi^2_{n-j}$ has mean $\frac{n-j}{2}$, it is typically of order $n/2$. More precisely, for large n, $\alpha_j \approx \sqrt{\frac{n-j}{2}}$ with high probability.

2. **Scaling by $\sqrt{n}$.** One rescales T by $\frac{1}{\sqrt{n}}$. This gives subdiagonal entries

$$\frac{\alpha_j}{\sqrt{n}} \ \approx \ \sqrt{\frac{n-j}{2n}} \ \approx \ \sqrt{\frac{1 - j/n}{2}},$$

   while the diagonal entries become $\frac{d_i}{\sqrt{n}}$, which vanish in the large-n limit. So effectively, the subdiagonal structure drives the main spectral behavior in the bulk, producing the semicircle shape in the limit.

3. **Orthogonal Polynomial / Recurrence Analysis.** The polynomial $p_n(\lambda)$ satisfies a discrete three-term recurrence whose "continuum limit" yields a certain integral equation (specifically the Stieltjes transform for the measure) whose solution is precisely the semicircle distribution. In more detailed treatments, one shows that the moments or the Cauchy transform of L_n converge to that of μ_{sc}. The relevant PDE or integral equation is exactly solvable, producing the semicircle.

Hence, with probability 1, as $n \to \infty$, the empirical spectrum of $\frac{1}{\sqrt{n}}W$ converges to the semicircle distribution on $[-2, 2]$. This precisely recovers *Wigner's semicircle law.*

Remark D.1 (Extensions)**.** A very similar approach works for the Gaussian Unitary Ensemble ($\beta = 2$), leading to a random *complex Hermitian* tridiagonal matrix. For $\beta = 4$, there is a quaternionic block-tridiagonal model. All of these point toward the same semicircle law for the global spectral distribution.

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. ↑4, 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 5: Title TBD

Leonid Petrov

DATE, 2025*

E 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 • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 6: Title TBD

Leonid Petrov

DATE, 2025*

F 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 • TeX Source • Updated at 13:44, Wednesday 15th January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 7: Title TBD

Leonid Petrov

DATE, 2025*

G 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 • TeX Source • Updated at 13:44, Wednesday 15^{th} January, 2025

Lectures on Random Matrices (Spring 2025)
Lecture 8: Title TBD

Leonid Petrov

DATE, 2025*

H 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 • TeX Source • Updated at 13:44, Wednesday 15th January, 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 • TeX Source • Updated at 13:44, Wednesday 15th January, 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 • TeX Source • Updated at 13:44, Wednesday 15th January, 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 • TeX Source • Updated at 13:45, Wednesday 15th January, 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 • TeX Source • Updated at 13:45, Wednesday 15^{th} January, 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 • TeX Source • Updated at 13:45, Wednesday 15th January, 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 • TeX Source • Updated at 13:45, Wednesday 15th January, 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 • TeX Source • Updated at 13:45, Wednesday 15th January, 2025