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 09:13, Wednesday 22nd 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).
- Tridiagonalization and general beta ensemble.
- 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 $W = (X_{ij})$:

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

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*

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

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:*

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

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

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

Remark 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.

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

2.3 Step A. Joint density of matrix entries

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

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

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

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

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 $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ 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,

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

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

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

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.

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.

[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.

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

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

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

Thus, this action locally means that the infinitesimal b_{ij} is multiplied by $\lambda_i - \lambda_j$, for all $1 \le i < j \le n$. 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.4 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 $\{\lambda_1, \dots, \lambda_n\}$ 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.5.

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 i.i.d. 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 $\exp\big(-\frac{1}{2}\sum \lambda_i\big)$. ☐

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.

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.

3.2.1 Setup

Let X be an $n \times t$ real (or complex) matrix and Y be a $k \times t$ matrix, with $n \le k \le t$. Assume X and Y have i.i.d. 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

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

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

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

These are $t \times t$ projection matrices of ranks n and k, respectively, embedded in a space of dimension t. One checks that P_X and P_Y commute if and only if the row spaces of X and Y are aligned in certain ways. 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 $\operatorname{rank}(P_X P_Y) \le \min(n,k)$, there are at most $\min(n,k)$ nonzero eigenvalues of $P_X P_Y$. In fact, generically there are exactly $\min(n,k)$ nonzero eigenvalues when the subspaces are in "general position."

3.2.2 Jacobi ensemble

Theorem 3.4 (Jacobi/MANOVA/CCA Distribution)**.** *Let X and Y be as above, each having i.i.d. (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.*

Then the nonzero eigenvalues $\lambda_1, \dots, \lambda_n$ of the matrix $P_X P_Y$ lie in the interval $[0,1]$ and have a joint density function on $\{\lambda_i \in [0,1]\}$ of the form

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

up to a normalization constant that depends on n, k, t (but not on $\{\lambda_i\}$). This distribution is called the Jacobi *(or* MANOVA*, or* CCA*) ensemble, and it is also known in statistics as the* multivariate Beta distribution.

Here again $\beta = 1$ for the real case and $\beta = 2$ for the complex case.

Remark 3.5. 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 The General β-ensemble pattern

We have now seen several classical "named" random matrix ensembles:

- **Wigner (GOE/GUE/GSE):** eigenvalues live on the real line $\mathbb{R}$
- **Wishart (Laguerre) ensemble:** eigenvalues live on $[0, \infty)$
- **Jacobi (MANOVA/CCA) ensemble:** eigenvalues live on $[0, 1]$

In all these cases, the **joint distribution of eigenvalues** can be written (up to a normalizing constant) in the universal form

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

where $\beta \in \{1, 2, 4\}$ (real/complex/quaternionic symmetry) and $V(\lambda)$ is some one-variable weight (often of the form $e^{-\lambda^2/2}$, $\lambda^{\alpha} e^{-\lambda}$, $\lambda^{\alpha}(1-\lambda)^{\gamma}$, etc.). This "universal" structure, with the product $\prod_{i<j} |\lambda_i - \lambda_j|^{\beta}$ (the Vandermonde to a power), arises from the **Jacobian** (or measure) factor in the change of variables from the matrix entries to the spectral decomposition.

Such ensembles are often called **β-ensembles** or **log-gases**, because one can interpret

$$-\log\left|\lambda_i - \lambda_j\right|$$

as a pairwise "logarithmic repulsion" potential between charges λ_i.

3.3.1 Concluding remarks

- **Spectral measures.** For large n, the empirical spectral distributions of Wishart and Jacobi ensembles converge to known limits (the Marchenko–Pastur law, and the Wachter or Beta law, respectively). These parallel the semicircle law for Wigner.

- **Applications.** Jacobi/Wishart models arise in various advanced statistical contexts (covariance matrices, canonical correlation analysis, principal components, etc.). Their exact solvability and the presence of the β-ensemble structure make them powerful tools for both theoretical and applied work.

In subsequent subsections, we will see how the tridiagonal approach extends to these ensembles, giving alternative derivations of their spect

4 Tridiagonal (Householder) Form for real symmetric Matrices

We now give a step-by-step procedure (and proof) of how any real symmetric matrix can be orthogonally transformed into a tridiagonal matrix. This is a standard topic in numerical linear algebra (the "Householder reduction") but is also central in random matrix theory (especially the Dumitriu–Edelman approach to the Gaussian ensembles).

4.1 Statement

Theorem 4.1 (real symmetric Tridiagonalization)**.** *Any real symmetric matrix $A \in \mathbb{R}^{n\times n}$ can be represented as*

$$A = Q^\top T\, Q, \quad \textit{where } Q \in O(N) \textit{ and } T \textit{ is real symmetric tridiagonal.}$$

That is, T has nonzero entries only on the main diagonal and the first sub- and super-diagonals:

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

4.2 Householder Reflections: A Detailed Algorithm

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

$$H = I - 2\,\frac{v\,v^\top}{\|v\|^2},$$

where $v \in \mathbb{R}^N$ is nonzero. One can check:

$$H^\top = H, \quad H^2 = I, \quad H \text{ is orthogonal, i.e. } H^\top H = I.$$

Geometrically, H reflects vectors across the hyperplane orthogonal to v.

Goal. We want to apply successive Householder reflections to "zero out" all sub-subdiagonal (and super-subdiagonal by symmetry) entries of A, leaving only the main diagonal and the first super-/sub-diagonal possibly nonzero.

1. **Start with** $A^{(0)} = A$.

2. **Step** $k = 1$**.** We aim to zero out entries $A^{(0)}_{2,1}, A^{(0)}_{3,1}, \ldots, A^{(0)}_{N,1}$, except for one to remain on the first subdiagonal if needed. Specifically, define the vector

$$x = (A^{(0)}_{2,1},\, A^{(0)}_{3,1},\, \ldots, A^{(0)}_{N,1})^\top \;\in\; \mathbb{R}^{N-1}.$$

   We want a Householder H_1 such that

$$H_1\, A^{(0)}\, H_1 = A^{(1)}$$

has zeros in the first column (and row, by symmetry) except possibly $A^{(1)}_{2,1}$.

Concretely, embed x into $\tilde{x} \in \mathbb{R}^N$ by placing a 0 in the top slot:

$$\tilde{x} = (\,0,\, A^{(0)}_{2,1}, \ldots, A^{(0)}_{N,1}\,)^\top.$$

Choose

$$v = \tilde{x} \;+\; \alpha e_1 \;\in\; \mathbb{R}^N,$$

with α chosen so that $\|v\| \neq 0$ and $(I - 2vv^\top/\|v\|^2)\,\tilde{x}$ is a scalar multiple of e_1. A common choice is

$$\alpha = \pm\,\|\tilde{x}\|,$$

picking a sign that avoids cancellation. Define

$$H_1 = I - 2\,\frac{v\,v^\top}{\|v\|^2}.$$

Then H_1 is an orthogonal, symmetric matrix that kills the sub-subdiagonal entries in column 1.

3. **Step $k = 2, \ldots, N-2$.** Inductively, we zero out the $(k+2)$-th to N-th entries in the k-th column (and by symmetry, in the k-th row). Each step uses a smaller Householder reflection H_k acting nontrivially in the lower-right $(N-k+1) \times (N-k+1)$ submatrix. Then set

$$A^{(k)} = H_k\, A^{(k-1)}\, H_k.$$

4. **End result.** After $N-2$ steps, we get $A^{(N-2)}$, which is tridiagonal, and

$$A^{(N-2)} = (H_{N-2} \cdots H_1)\, A\, (H_1 \cdots H_{N-2}).$$

Define

$$Q = H_1 \cdots H_{N-2}.$$

Since each H_k is orthogonal, $Q \in O(N)$. Moreover,

$$A^{(N-2)} = Q\, A\, Q^\top$$

has the desired tridiagonal form.

Remark 4.2. This procedure is also used in numerical methods for eigenvalue computations: once you reduce to tridiagonal form, one can apply specialized algorithms (like the QR algorithm) more efficiently.

Proof of Theorem 4.1. It is essentially just the algorithmic outline above. Each step is valid because Householder transformations preserve symmetry: if B is symmetric, then

$$(HBH)_{ij} = \sum_{r,s} H_{ir} B_{rs} H_{sj}.$$

But since H is symmetric itself, (HBH) remains symmetric. Also, each step zeroes out the sub-subdiagonal entries in the appropriate column and row, thus eventually forcing a tridiagonal shape. Finally, the product of all Householder reflections used is an orthogonal matrix. This completes the argument. □

5 Wigner's Semicircle Law via Tridiagonalization

We now present a *detailed* outline of how one proves the Wigner semicircle law for the GOE by using its *random tridiagonal model*. This method is due to Dumitriu and Edelman (2002) and is often considered more direct than Wigner's original moment method.

5.1 Dumitriu–Edelman Tridiagonal Model

Theorem 5.1 (Tridiagonal Representation of GOE)**.** *Let M be an $N \times N$ GOE matrix (real symmetric) with variance chosen so that the off-diagonal entries have variance $\frac{1}{2}$ and diagonal entries have variance 1. Then there exists an orthogonal matrix Q such that*

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

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 $\{d_i, \alpha_j\}$ are mutually independent with

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

where χ^2_ν is a chi-square distribution with ν degrees of freedom, and equivalently $\sqrt{\frac{\chi^2_\nu}{2}}$ is half the norm of a Gaussian vector in $\mathbb{R}^\nu$.

Remark 5.2. - In short, the diagonal entries d_i are i.i.d. $\mathcal{N}(0,1)$. - The subdiagonal entries $\alpha_1, \ldots, \alpha_{N-1}$ are independent with each α_j distributed like $\sqrt{\frac{\chi^2_{N-j}}{2}}$. - Off-diagonal entries above the first superdiagonal are all zero, so T has only $2N-1$ nontrivial entries (the N diagonal + $(N-1)$ sub-/super-diagonal).

Sketch of Construction. This is essentially a specialized version of the Householder procedure (Section 4), carefully arranged so that each step ends up with exactly the distributions described for α_j and d_i. One uses the fact that a Gaussian matrix is rotationally invariant in a suitable sense, ensuring that each step's "residual vector" has an isotropic Gaussian distribution. Then the norm of that vector yields χ^2 variables. Full details appear in [?DumitriuEdelman2002] or advanced RMT texts. □

Thus, to study the eigenvalues of the GOE matrix M, we can equivalently study the eigenvalues of the (much sparser) tridiagonal matrix T.

5.2 Characteristic Polynomial and Three-Term Recurrence

Consider $p_N(\lambda) = \det(T - \lambda I)$. Since T is tridiagonal, one has the well-known three-term recurrence:

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

The roots of $p_N(\lambda)$ are precisely the eigenvalues $\lambda_1, \dots, \lambda_N$ of T.

5.3 Outline of the Semicircle Limit Proof

We now want to show that the empirical measure

$$L_N \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$$

converges weakly (almost surely) to the semicircle distribution

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

A typical route has these ingredients:

1. **Law of Large Numbers for α_j.** Notice that $\alpha_j^2 = \frac{1}{2}\chi^2_{N-j}$ has mean $\frac{N-j}{2}$. For large N, it is typically of order N. More precisely, $\alpha_j \approx \sqrt{\frac{N-j}{2}}$ in a probabilistic sense as $N \to \infty$.

2. **Scale invariance.** One usually rescales T by $\sqrt{N}$. That is, consider $\frac{1}{\sqrt{N}}\, T$. Its subdiagonal entries become
$$\frac{\alpha_j}{\sqrt{N}} \;\approx\; \sqrt{\frac{N-j}{2N}} \;\approx\; \sqrt{\frac{1 - j/N}{2}} \quad \text{(for large } N\text{)}.$$
Meanwhile, the diagonal entries become $\frac{d_i}{\sqrt{N}}$, which are $\mathcal{O}(\frac{1}{\sqrt{N}})$. Hence the subdiagonal terms set the main scale for the "bulk" of the spectrum, while the diagonal is negligible in the large N limit.

3. **Asymptotic Analysis of Recurrence.** A known fact from orthogonal polynomial theory (or from direct PDE-like arguments on the discrete recurrence) is that the location of the roots of $p_N(\lambda)$ concentrate where the effective continuum limit of the recurrence matches a certain "Stieltjes equation" whose solution is the semicircle density.

   In more elementary terms, one can check that the *moment generating function* or *Stieltjes transform* of the measure L_N converges to that of μ_{sc}. Alternatively, one can do a direct argument on the polynomials $p_k(\lambda)$ by bounding their growth and linking it to an integral equation reminiscent of
$$g(z) = \int \frac{1}{x - z}\, d\mu_{\mathrm{sc}}(x),$$
   which leads to a quadratic equation solved by the semicircle's Cauchy transform.

   For details, see [?DumitriuEdelman2002] or [?TaoTopics], as the full proof is somewhat technical but completely rigorous.

The net result is that, *with probability 1*, as $N \to \infty$, the empirical spectral measure of $\frac{1}{\sqrt{N}} M$ (equivalently of $\frac{1}{\sqrt{N}} T$) converges to the semicircle distribution on $[-2, 2]$:

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

This is precisely *Wigner's semicircle law.*

Remark 5.3 (Extensions)**.** A very similar approach works for the Gaussian Unitary Ensemble ($\beta = 2$), yielding a random *complex Hermitian* tridiagonal (or banded) matrix. And for $\beta = 4$, there is an analogous construction with quaternionic entries, usually leading to a block-tridiagonal matrix. All roads lead to the semicircle law for the limiting global spectrum.

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 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.5 Normalization for GOE

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

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

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

C.6 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.

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