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 12:04, 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 $\mathbb{E}\,\mathrm{Tr}(W^k)$ and counting trees, plus analytic lemmas which ensure that the convergence of expected powers of traces is enough to conclude the convergence (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*

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

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:*

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

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

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

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 $W/\sqrt{2}$:

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

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

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

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

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

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

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

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

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

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.

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:

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

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*

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

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

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 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 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 i.i.d. (real or complex) Gaussian entries of size $n \times t$ and $k \times t$, respectively, with $n \leq k \leq t$. Assume further that X and Y are independent of each other (this is the null hypothesis in statistics).*

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

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

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. We want to apply successive Householder reflections to "zero out" all entries below the first subdiagonal (and by symmetry, above the first superdiagonal), leaving only the tridiagonal part possibly nonzero.

Start with $W^{(0)} = W$.

Step $k = 1$**.** We aim to zero out the entries $W^{(0)}_{2,1}, W^{(0)}_{3,1}, \ldots, W^{(0)}_{n,1}$. Define the vector

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

We then embed x into a vector $\tilde{x} \in \mathbb{R}^n$ by placing 0 in the top coordinate:

$$\tilde{x} \;=\; \big(0,\; W^{(0)}_{2,1},\; W^{(0)}_{3,1},\; \ldots,\; W^{(0)}_{n,1}\big)^\top.$$

To construct a Householder reflection that annihilates all but the first subdiagonal entry, choose

$$v \;=\; \tilde{x} \;+\; \alpha\, e_1, \quad \alpha \;=\; \pm\,\|\tilde{x}\|,$$

picking the sign of α to avoid cancellation. Then define

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

By construction, H_1 is orthogonal and symmetric, and it will force all sub-subdiagonal entries below row 2 in column 1 to vanish when we do $H_1\, W^{(0)}\, H_1$.

Step $k = 2, \ldots, n-2$**.** Inductively, we zero out the $(k+2)$-th through n-th entries in the k-th column (and by symmetry, the same row). Each step uses a smaller Householder matrix H_k acting nontrivially in the lower-right $(n-k+1) \times (n-k+1)$ sub-block. We set

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

End result. After $n-2$ steps, we obtain $W^{(n-2)}$, which is tridiagonal. Define

$$Q \;=\; H_1\, H_2\, \cdots\, H_{n-2}.$$

Since each H_k is orthogonal, $Q \in O(n)$. We see

$$W^{(n-2)} \;=\; (H_{n-2} \cdots H_1)\,\big[\, W \,\big]\,(H_1 \cdots H_{n-2}),$$

so

$$W^{(n-2)} \;=\; Q\, W\, Q^{\top}$$

has the desired tridiagonal form. □

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 $\alpha_j = \sqrt{\frac{\chi^2_{n-j}}{2}}$. This scaling ensures that each α_j^2 has the appropriate variance required for the GOE ensemble, specifically matching the variance structure of the original Wigner matrix. The chi-square distribution's properties, such as its mean being ν and its variance being 2ν, play a crucial role in maintaining the tridiagonal matrix's spectral characteristics that lead to the semicircle law.

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. $\square$

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

5.2 Generalization to β-ensembles

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

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

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

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

C.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

[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

[Gor21] V. Gorin, *Lectures on random lozenge tilings*, Cambridge Studies in Advanced Mathematics. Cambridge University Press (2021). ↑9

[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. ↑9

L. Petrov, University of Virginia, Department of Mathematics, 141 Cabell Drive, Kerchof Hall, P.O. Box 400137, Charlottesville, VA 22904, USA

E-mail: lenia.petrov@gmail.com