Lecture notes of MA6252: Random matrix theory

WANG Dong

April 16, 2013

1 Time reversal symmetry and the three Gaussian ensembles: GUE, GOE and GSE

This section follows mainly [19, Chapter 2]. Another reference is [13, Chapter 1].

1.1 Three kinds of random matrix models corresponding to physical systems with different time reversal properties

Idea from physics How to find a random Hermitian operator to model a generic Hamiltonian operator with discrete spectrum?

In quantum mechanics, the Hamiltonian of a physical system that determines the time-evolution of the system, is represented by a self-adjoint Hamiltonian operator. If it has only discrete spectrum (or if we only care about the discrete part of its spectrum), we would like to use a finitely dimensional Hermitian operator, i.e., an $N \times N$ Hermitian matrix, to model it.

We want to model a generic Hamiltonian, and most special properties of a physical system can be safely ignored. But one property is too fundamental to ignore: the time-reversal symmetry. Later it will be clear that the spectra of Hamiltonians of generic physical systems in different time-reversal invariance classes, and that of generic physical systems without time-reversal invariance, are quite different in the local behaviour.

First we model the generic Hamiltonian without time-reversal invariance. We would like the $N \times N$ random Hermitian matrix to satisfy the following properties.

1. The probability measure $p(H)dH$ where

$$dH = \prod_{j \le k} d\Re H_{jk} \prod_{j<k} d\Im H_{jk} \tag{1.1}$$

   is invariant under the automorphism

$$H \to U^{-1} H U \tag{1.2}$$

   where $U \in \mathrm{U}(N)$ is any unitary matrix.

2. The probability density function $p(H)$ can be written into the form

$$p(H)dH = \prod_{j \le k} f_{jk}^{(0)}(H_{jk}) d\Re H_{jk} \prod_{j<k} f_{jk}^{(1)}(H_{jk}) d\Im H_{jk}. \tag{1.3}$$

Later we will see that these conditions almost define the Gaussian unitary ensemble (GUE).

Time-reversal operator in quantum mechanics In quantum mechanics, the time-reversal operator T is antiunitary, i.e., $\langle T\psi, T\varphi\rangle = \overline{\langle \psi, \varphi\rangle}$. Then T can be decomposed as

$$T = KC, \tag{1.4}$$

where C is the complex conjugation operator such that $C\psi = \bar{\psi}$, and K is a unitary operator, since

$$\langle K\psi, K\varphi\rangle = \langle T\bar{\psi}, T\bar{\varphi}\rangle = \overline{\langle \bar{\psi}, \bar{\varphi}\rangle} = \langle \psi, \varphi\rangle. \tag{1.5}$$

Since applying time-reversal twice one gets the identity transformation, in quantum mechanics

$$T^2 = \alpha I, \quad \text{where} \quad |\alpha| = 1. \tag{1.6}$$

Equivalently,

$$(KC)(KC) = K(CK)C = K(\bar{K}C)C = K\bar{K}(CC) = K\bar{K} = \alpha I, \tag{1.7}$$

where we use the identity that for any ψ

$$CK\psi = \overline{K\psi} = \bar{K}\bar{\psi} = \bar{K}C\psi. \tag{1.8}$$

Since K is a unitary operator with $KK^* = K(\bar{K})^T = I$, we have

$$\bar{K} = \alpha(\bar{K})^T \quad \Leftrightarrow \quad K = \bar{\alpha}K^T. \tag{1.9}$$

By using it twice

$$K = \bar{\alpha}(\bar{\alpha}K^T)^T = \bar{\alpha}^2 K, \tag{1.10}$$

we find

$$\alpha = 1 \quad \text{or} \quad \alpha = -1, \tag{1.11}$$

and K is then symmetric or antisymmetric respectively.

Now we see that there are two cases of time-reversal invariant systems corresponding to symmstric and antisymmetric K, and for some physical reasons we call them even-spin case and odd-spin case respectively.

Remark 1. The original reference for the discussion of time-reversal operator is [27], but modern textbooks on quantum field theory, like [26] may be more accessible.

Even-spin case of time-reversal invariant Hamiltonians Suppose H is an Hermitian operator invariant under time-reversal, then

$$THT^{-1} = H \quad \Leftrightarrow \quad KCHC^{-1}K^{-1} = H. \tag{1.12}$$

Noting that

$$(CHC^{-1})\psi = C(H\bar{\psi}) = \overline{H\bar{\psi}} = \bar{H}\psi, \tag{1.13}$$

we see that (1.12) is equivalent to

$$KH^TK^{-1} = H. \tag{1.14}$$

Since K is a symmetric operator and we assume that it is finitely dimensional, i.e., K is an $N \times N$ matrix, we apply *Takagi's factorization* (proved in Appendix A.1) and write it as

$$K = UDU^T \tag{1.15}$$

where U is unitary and D is a real nonnegative diagonal matrix such that the diagonal elements of D are the nonnegative square roots of the eigenvalues of KK^*. Note that K is also unitary, so $KK^* = I$ and $D = I$. Thus (1.15) implies

$$K = UU^T. \tag{1.16}$$

Taking a unitary transformation of the representation of the states by U^{-1} such that $\psi \mapsto U^{-1}\psi$, where U is that in (1.16), we find that the time-reversal operator T transforms to

$$T \mapsto U^{-1}TU = U^{-1}(UU^TC)U = (U^TC)U = (C\overline{U^T})U = C. \tag{1.17}$$

Thus in the new representation, the operator K becomes the identity matrix. Then the relation (1.14) means that H is a symmetric Hermitian matrix, that is, a real symmetric matrix.

To model the generic Hamiltonian in the even-spin case of time-reversal invariance class, we would like the $N \times N$ random real symmetric matrix to satisfy the following conditions.

1. The probability measure $p(H)dH$ where

$$dH = \prod_{j \le k} dH_{jk} \tag{1.18}$$

is invariant under the automorphism

$$H \to W^{-1}HW \tag{1.19}$$

where $W \in \mathrm{O}(N)$ is any real orthogonal matrix.

2. The probability density function $p(H)$ can be written into the form

$$p(H)dH = \prod_{j \le k} f_{jk}(H_{jk})dH_{jk}. \tag{1.20}$$

Later we show that these conditions almost define the Gaussian orthogonal ensemble (GOE).

Odd-spin case of time-reversal invariant Hamiltonians We consider K as an antisymmetric matrix of dimension N. Then by the theorem on the normal form of an antisymmetric matrix under unitary congruence (proved in Appendix A.1), we have

$$K = U\Sigma U^T \tag{1.21}$$

where $U \in \mathrm{U}(N)$ is a unitary matrix and Σ is of the form

$$\Sigma = \mathrm{diag}\left(\underbrace{\begin{pmatrix} 0 & \sigma_1 \\ -\sigma_1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & \sigma_n \\ -\sigma_n & 0 \end{pmatrix}}_{n}, \underbrace{0, \ldots, 0}_{N-2n}\right), \tag{1.22}$$

where σ_i are all positive numbers. From the unitarity of K, we have $N = 2n$ since K is nonsingular, and $\sigma_1 = \cdots = \sigma_n = 1$. Thus

$$\Sigma = \mathrm{diag}\left(\underbrace{\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}_{n}\right). \tag{1.23}$$

Similar to the derivation in the even-spin case, taking the unitary transformation of the representation of the states by U^{-1} where U is that in (1.21), we have

$$T \mapsto U^{-1}TU = U^{-1}(U\Sigma U^T C)U = \Sigma C, \tag{1.24}$$

and then in the new representation,

$$K = \Sigma. \tag{1.25}$$

By relation (1.14), we see that H in the even-spin case satisfies $\Sigma H^T \Sigma^{-1} = H$, or equivalently

$$\Sigma H^T = H\Sigma. \tag{1.26}$$

People who are familar with classical groups recognise immediately that (1.26) is similar to the formula $\Sigma A^T = -\Sigma A$ that defines the element of $\mathfrak{sp}(2n)$, the symplectic Lie algebra. Another concept related is the unitary symplectic group $\mathrm{Sp}(n)$ whose elements are unitary matrices $W \in \mathrm{U}(N)$ such that

$$\Sigma = W\Sigma W^T. \tag{1.27}$$

Below we analyse (1.26) in an elementary way, without appealing to Lie group/algebra.

Since our H and Σ are both $2n \times 2n$, we think them as block matrices in 2×2 blocks. Each 2×2 complex matrix is the complex linear combination of

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad e_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad e_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \tag{1.28}$$

in the way

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{2}(a+d)1 - \frac{i}{2}(a-d)e_1 + \frac{1}{2}(b-c)e_2 - \frac{i}{2}(b+c)e_3. \tag{1.29}$$

In these notations, we have

$$\Sigma = e_2 I. \tag{1.30}$$

It is better to understand $1, e_1, e_2, e_3$ as the standard basis of quaternion, usually written as $1, i, j, k$. But we do not use these notations in fear of namespace conflicts. Note that one matrix representation of the quaternion $a + bi + cj + dk$ is

$$a + bi + cj + dk \to \begin{pmatrix} a+bi & c+di \\ -c+di & a-bi \end{pmatrix}, \tag{1.31}$$

a 2×2 complex matrix of a special form. General 2×2 complex matrices correspond to a generalisation of quaternions, called the biquaternions (or complexified quaternions), in the sense that a, b, c, d can be complex numbers in the left-hand side of (1.31), where the i along with j, k are different from the imaginary basis for the complex numbers.

Now we write the biquaternion, i.e., 2×2 complex matrix, in the form of

$$q = q^{(0)} + \mathbf{q} \cdot \mathbf{e}, \quad \text{where} \quad \mathbf{q} = (q^{(1)}, q^{(2)}, q^{(3)}),\ \mathbf{e} = (e_1, e_2, e_3). \tag{1.32}$$

The *quaternion conjugate* of q is

$$q^* = q^{(0)} - \mathbf{q} \cdot \mathbf{e}, \quad \text{or equivalently} \quad \begin{pmatrix} a_1 + a_2 i & b_1 + b_2 i \\ c_1 + c_2 i & d_1 + d_2 i \end{pmatrix}^* = \begin{pmatrix} d_1 + d_2 i & -b_1 - b_2 i \\ -c_1 - c_2 i & a_1 + a_2 i \end{pmatrix}; \tag{1.33}$$

the *complex conjugate* of q is $(\bar{\mathbf{q}} = (\overline{q^{(1)}}, \overline{q^{(2)}}, \overline{q^{(3)}}))$

$$q^\star = \overline{q^{(0)}} + \bar{\mathbf{q}} \cdot \mathbf{e}, \quad \text{or equivalently} \quad \begin{pmatrix} a_1 + a_2 i & b_1 + b_2 i \\ c_1 + c_2 i & d_1 + d_2 i \end{pmatrix}^\star = \begin{pmatrix} d_1 - d_2 i & -c_1 + c_2 i \\ -b_1 + b_2 i & a_1 - a_2 i \end{pmatrix}; \tag{1.34}$$

the *Hermitian conjugate* of q is

$$q^\dagger = (q^*)^\star = \overline{q^{(0)}} - \bar{\mathbf{q}} \cdot \mathbf{e}, \quad \text{or equivalently} \quad \begin{pmatrix} a_1 + a_2 i & b_1 + b_2 i \\ c_1 + c_2 i & d_1 + d_2 i \end{pmatrix}^\dagger = \begin{pmatrix} a_1 - a_2 i & c_1 - c_2 i \\ b_1 - b_2 i & d_1 - d_2 i \end{pmatrix}. \tag{1.35}$$

Note that $q^\dagger = q$ if and only if q is represented by a 2×2 Hermitian matrix. These three conjugates satisfies

$$(q_1 q_2)^* = q_2^* q_1^*, \quad (q_1 q_2)^\star = q_1^\star q_2^\star, \quad (q_1 q_2)^\dagger = q_2^\dagger q_1^\dagger. \tag{1.36}$$

Expressing

$$H = (q_{jk})_{j,k=1}^n, \quad \text{where} \quad q_{j,k} = \begin{pmatrix} a_{jk} & b_{jk} \\ c_{jk} & d_{jk} \end{pmatrix} \quad \text{are biquaternions.} \tag{1.37}$$

We see that the Hermitian property $H = H^* = \bar{H}^T$ implies that

$$\begin{pmatrix} a_{kj} & b_{kj} \\ c_{kj} & d_{kj} \end{pmatrix} = \begin{pmatrix} \overline{a_{jk}} & \overline{c_{jk}} \\ \overline{b_{jk}} & \overline{d_{jk}} \end{pmatrix}, \quad \text{or equivalently} \quad q_{kj} = q_{jk}^\dagger. \tag{1.38}$$

On the other hand, if we write H^T into 2×2 blocks, the (j,k) block is

$$\begin{pmatrix} a_{kj} & c_{kj} \\ b_{kj} & d_{kj} \end{pmatrix} = -\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} d_{kj} & -b_{kj} \\ -c_{kj} & a_{kj} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{1.39}$$

and we have

$$(H^T)_{jk} = (-e_2 q_{kj}^* e_2). \tag{1.40}$$

Thus by (1.30), the time-reversal relation $\Sigma H^T = H\Sigma$ is equivalent to the blockwise identity

$$e_2(-e_2 q_{kj}^* e_2) = q_{jk} e_2 \quad \Leftrightarrow \quad q_{kj}^* = q_j k. \tag{1.41}$$

The (bi)quaternion matrix q_{jk} satisfying $q_{kj}^* = q_{jk}$ is said to be self-dual. Thus a finite dimensional Hamiltonian that is time-reversal invariant of the odd-spin class has to be of even dimension, and in the respresentation where $T = \Sigma C$, it is both Hermitian and self-dual if written in the biquaternion form.

Note that the self-dual Hermitian property of H implies

$$q_{jk}^* = q_{jk}^\dagger \quad \Leftrightarrow \quad a_{jk} = \overline{d_{jk}} \quad \text{and} \quad b_{jk} = -\overline{c_{jk}}, \tag{1.42}$$

or equivalently, each $q_{jk} = q_{jk}^{(0)} + q_{jk}^{(1)} + q_{jk}^{(2)} + q_{jk}^{(3)}$ is a *real* quanternion in the sense that $q_{jk}^{(0)}, q_{jk}^{(1)}, q_{jk}^{(2)}, q_{jk}^{(3)}$ are all real numbers. Furthermore we have that $(q_{jk}^{(0)})_{j,k=0}^n$ forms a real symmetric matrix while $(q_{jk}^{(i)})_{j,k=0}^n$ forms a real antisymmetric matrix for $i = 1, 2, 3$. It is straightforward to check that these conditions on $(q_{jk}^{(0)})_{j,k=0}^n$ and $(q_{jk}^{(i)})_{j,k=0}^n$ are equivalent to the self-dual Hermitian condition.

One more remark on self-dual Hermitian matrices is that if H is self-dual Hermitian and $W \in \mathrm{Sp}(n)$ is an element of the unitary symplectic group, then $W^{-1}HW$ is also self-dual Hermitian.

The result we obtained above means that we need to study random self-dual Hermitian matrices along with random real symmetric matrices. We consider the random matrix as follows.

1. The probability measure $p(H)dH$ where

$$dH = \prod_{j\le k} dq_{jk}^{(0)} \prod_{i=1}^{3} \prod_{j<k} dq_{jk}^{(i)} \tag{1.43}$$

is invariant under the automorphism

$$H \to W^{-1}HW \tag{1.44}$$

where $W \in \mathrm{Sp}(n)$ is any unitary symplectic matrix.

2. The probability density function $p(H)$ can be written into the form

$$p(H)dH = \prod_{j\le k} f_{jk}^{(0)}(q_{jk}^{(0)})dq_{jk}^{(0)} \prod_{i=1}^{3} \prod_{j<k} f_{jk}^{(i)}(q_{jk}^{(i)})dq_{jk}^{(i)}. \tag{1.45}$$

Later we will see that these conditions almost define the Gaussian symplectic ensemble (GSE).

1.2 The probability density function of Gaussian orthogonal ensemble (GOE)

The even-spin time-reversal invariant systems are modeled by random real symmetric matrices satisfying two conditions specified on Page 3. Now we give a concrete description of this random matrix model which is called the Gaussian orthogonal ensemble (GOE). It will be clear why it is called "Gaussian".

Before dealing with the probability measure $p(H)dH$, we first show that the measure dH defined in (1.18) is invariant under the orthogonal similarity transformation (1.19). To show this, we recall the Givens rotation in $\mathrm{SO}(N)$

$$G(j,k;\theta) = \begin{pmatrix} \ddots & & & & \\ & \cos\theta & \dots & -\sin\theta & \\ & \vdots & & \vdots & \\ & \sin\theta & \dots & \cos\theta & \\ & & & & \ddots \end{pmatrix}. \tag{1.46}$$

where the $(j,j),(j,k),(k,j),(k,k)$ entries constitute a plane rotation matrix and other (i,j) entries are simply δ_{ij}. It is well known that any rotation matrix in $\mathrm{SO}(N)$ is the product of Givens rotations [14, Section 5.2.3]. On the other hand, the Givens rotation $G(j,k;\theta)$ is the product of $G(1,2;\theta)$ and permutation matrices. For example,

$$\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{1.47}$$

Thus $G(1,2;\theta)$ and permutation matrices generates the orthogonal group $\mathrm{O}(N)$. Hence we need only to show that dH is invariant under the conjugations of (a) the permutation matrices, and (b) $G(1,2;\theta)$. Condition (a) is obvious satisfied, and condition (b) is reduced into a 2×2 matrix problem. Let

$$\begin{pmatrix} u & v \\ v & w \end{pmatrix} := \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x & y \\ y & z \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{1.48}$$

where

$$u = x\cos^2\theta + 2y\cos\theta\sin\theta + d\sin^2\theta, \tag{1.49}$$

$$v = (z-x)\cos\theta\sin\theta + y(\cos^2\theta - \sin^2\theta), \tag{1.50}$$

$$w = x\sin^2\theta - 2y\cos\theta\sin\theta + z\cos^2\theta. \tag{1.51}$$

Then the Jacobian

$$\frac{\partial(u,v,w)}{\partial(x,y,z)} = 1. \tag{1.52}$$

We conclude that dH is invariant under the conjugation of $G(1,2;\theta)$, and prove the claim.

To show $p(H)dH$ is invariant under similarity transformation, we now only need to show the function $p(H) = \prod_{j\le k} f_{jk}(H_{jk})$ is invariant under the conjugation of (a) the permutation matrices, and (b) $G(1,2;\theta)$.

Condition (a) is equivalent to that

$$f_{jj}(x) = f_{kk}(x), \tag{1.53}$$

$$f_{jk}(x) = f_{lm}(x), \qquad \text{where } 1 \le j < k \le N \text{ and } 1 \le l < m \le N. \tag{1.54}$$

Hence we denote $f(x)$ and $g(x)$ as the probability density functions for diagonal entries and off-diagonal entries of H respectively.

Condition (b) is again reduced to a 2×2 matrix problem

$$f(u)g(v)f(w) = f(x)g(y)f(z) \tag{1.55}$$

where u, v, w depend on x, y, z by (1.49)–(1.51).

Consider the case that $\theta = \epsilon$ is infinitesimal. Then (1.55) becomes

$$f(x+2y\epsilon)g(y+(z-x)\epsilon)f(z-2y\epsilon) = f(x)g(y)f(z) + \mathcal{O}(\epsilon^2). \tag{1.56}$$

When $f(x)g(y)f(z) \neq 0$, the identity between ϵ coefficients becomes

$$2y\left(\frac{f'(x)}{f(x)} - \frac{f'(z)}{f(z)}\right) = (x-z)\frac{g'(y)}{g(y)}. \tag{1.57}$$

By separation of variables, we find a constant c such that

$$\frac{1}{2y}\frac{g'(y)}{g(y)} = -c, \tag{1.58}$$

$$\frac{1}{(x-z)}\left(\frac{f'(x)}{f(x)} - \frac{f'(z)}{f(z)}\right) = -c. \tag{1.59}$$

Using separation of variables again to (1.59), we find a canstant b such that

$$\frac{f'(x)}{f(x)} = -cx + b, \tag{1.60}$$

$$\frac{f'(x)}{f(x)} = -cz + b. \tag{1.61}$$

Therefore the probability density functions f and g are solved as

$$f(x) = \sqrt{\frac{c}{2\pi}}e^{-\frac{b^2}{2c}}e^{-\frac{c}{2}x^2+bx}, \tag{1.62}$$

$$g(x) = \sqrt{\frac{c}{\pi}}e^{-cy^2}, \tag{1.63}$$

and the density function $p(H)$ can be written in a more compact form as

$$p(H) = \frac{1}{C_N} \exp(\mathrm{Tr}(-\frac{c}{2}H^2 + bH)). \tag{1.64}$$

After a simple shifting and scaling, we need only to analyse the model where

$$p(H) = \frac{1}{C_N} \exp(-\frac{N}{4} \operatorname{Tr} H^2), \tag{1.65}$$

which is the definition of the Gaussian orthogonal ensemble (GOE). An equivalent definition is that the diagonal entries are in $\mathrm{N}(0, 2N^{-1})$, upper-triangular entries are in $\mathrm{N}(0, N^{-1})$ distributions, and they are all independent.

Exercises

1. Show that the random Hermitian matrix H satisfying the two conditions on page 1 has distribution $p(H)dH = \frac{1}{C_N} \exp(\mathrm{Tr}(-\frac{c}{2}H^2 + bH))dH$. Thus it is essentially equivalent to the *Gaussian unitary ensemble* where

$$p(H) = \frac{1}{C_N} \exp(-\frac{N}{2} \operatorname{Tr} H^2). \tag{1.66}$$

   An equivalent way to describe it is that the diagonal entries are real and in $\mathrm{N}(0, N^{-1})$, upper-triangular entries are complex with both real and imaginary parts in independent $\mathrm{N}(0, \frac{1}{2}N^{-1})$ distributions, and they are all independent.

2. Show that the random self-dual Hermitian matrix H satisfying the two conditions on page 6 has distribution $p(H)dH = \frac{1}{C_N} \exp(\mathrm{Tr}(-\frac{c}{2}H^2 + bH))dH$. Thus it is equivalent to the *Gaussian symplectic ensemble* where

$$p(H) = \frac{1}{C_N} \exp(-\frac{N}{2} \operatorname{Tr} H^2). \tag{1.67}$$

2 Semicircle law and Stieltjes transform

The standard reference for this section is [5, Chapter 2, especially Section 2.3]. The textbooks [3, Chapter 2, especially Section 2.4] and [25, Section 2.4] are also good expositions.

2.1 Statement of the theorem

With the help of newly invented computers, numerical study of high-dimensional random matrices became possible in 1950's. First people were interested in the empirical distribution of eigenvalues of a large random matrix (especially when they are distributed on the real line).

Definition 1. Let M be an $N \times N$ matrix that has N real eigenvalues $\lambda_1^N, \ldots, \lambda_N^N$. The *empirical spectral distribution* (ESD) is

$$F^M(x) := \frac{\text{the number of } j \text{ such that } \lambda_j^N \le x}{N}. \tag{2.1}$$

Define the *semicircle law* as the probability distribution with density function $\sigma(x)$ and cumulative distribution function $F(x)$ given by

$$\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{|x|\le 2}, \quad F(x) = \int_{-\infty}^{x} \sigma(t)dt. \tag{2.2}$$

It was observed that if M_N is the GOE or GUE random matrix, then as $N \to \infty$, the random ESD of M_N converges weakly, in probability, to the semicircle distribution, $F^{M_N}(x) \to F(x)$, in the sense that for any bounded and continuous function f on $\mathbb{R}$, and any $\epsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\int f(x)dF^{M_N}(x) - \int f(x)dF(x)\right| > \epsilon\right) = 0. \tag{2.3}$$

We are going to prove a more general theorem for the *Wigner matrix*. Let $\{Z_{ij}\}_{1\le i<j}$ and $\{Y_i\}_{1\le i}$ be two independent families of independent and identically distributed random variables. We assume that Y_1 is real valued and has zero mean, $Z_{1,2}$ is complex valued, has zero mean, and $\mathrm{E}(Z_{1,2}\bar{Z}_{1,2}) = 1$, and all moments of Y_1 and $|Z_{1,2}|$ are finite. Let X_N be a Hermitian $N \times N$ matrix with entries

$$X_N(i,j) = \begin{cases} \frac{Z_{ij}}{\sqrt{N}} & \text{if } i < j, \\ \frac{Y_i}{\sqrt{N}} & \text{if } i = j. \end{cases} \tag{2.4}$$

We call such a matrix a Wigner matrix. Note that the GUE random matrix is a special case of the Wigner matrix, so is the GOE random matrix (where $\Im Z_{12} = 0$).

Theorem 1. *As $N \to \infty$, the ESD of the Wigner matrix X_N converges weakly, in probability to the semicircle law: $F^{X_N}(x) \to F(x)$.*

2.2 Preliminaries of Stietjes transform

Let μ be a positive, finite measure on $\mathbb{R}$. The *Stieltjes transform* of μ is an analytic function on $\mathbb{C} \setminus \mathbb{R}$ defined as

$$S_\mu(z) = \int \frac{d\mu}{t-z}. \tag{2.5}$$

From the Stieltjes transform, we can reconstruct the measure practically.

Theorem 2. *For any open interval (a,b) where neither a nor b is an atom of μ (i.e., $\mu(\{a\}) = \mu(\{b\}) = 0$), then*

$$\mu((a,b)) = \lim_{\epsilon\to 0^+} \frac{1}{\pi}\int_a^b \Im S_\mu(\lambda + i\epsilon))d\lambda. \tag{2.6}$$

Proof. First we find a lower bound of $\frac{1}{\pi}\int_a^b \Im S_\mu(\lambda+i\epsilon))d\lambda$.

$$\begin{aligned}
\frac{1}{\pi}\int_a^b \Im S_\mu(\lambda+i\epsilon)d\lambda &= \int_a^b \left(\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\mu(x)\right)d\lambda \\
&\ge \int_a^b \left(\int_{a+\sqrt{\epsilon}}^{b-\sqrt{\epsilon}} \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\mu(x)\right)d\lambda \\
&= \int_{a+\sqrt{\epsilon}}^{b-\sqrt{\epsilon}} \left(\int_a^b \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\lambda\right)d\mu(x) \\
&\ge \int_{a+\sqrt{\epsilon}}^{b-\sqrt{\epsilon}} \left(\int_{x+\sqrt{\epsilon}}^{x+\sqrt{\epsilon}} \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\lambda\right)d\mu(x) \\
&= \int_{a+\sqrt{\epsilon}}^{b-\sqrt{\epsilon}} (1+\mathcal{O}(\epsilon))d\mu(x) \\
&= (1+\mathcal{O}(\epsilon))\mu((a+\sqrt{\epsilon}, b-\sqrt{\epsilon})).
\end{aligned} \tag{2.7}$$

On the other hand, we have similarly

$$\begin{aligned}\frac{1}{\pi}\int_a^b \Im S_\mu(\lambda+i\epsilon))d\lambda &= \int_{-\infty}^{\infty}\left(\int_a^b \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\lambda\right)d\mu(x)\\ &= \int_{a-\sqrt{\epsilon}}^{b+\sqrt{\epsilon}} d\mu(x) + \int_{b+\sqrt{\epsilon}}^{\infty} d\mu(x) + \int_{-\infty}^{a-\sqrt{\epsilon}} d\mu(x)\left(\int_a^b \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\lambda\right).\end{aligned} \tag{2.8}$$

Note that

$$\int_a^b \frac{1}{\pi}\frac{\epsilon}{(\lambda-x)^2+\epsilon^2}d\lambda \begin{cases} <1 & \text{for } x\in(a-\sqrt{\epsilon}, b+\sqrt{\epsilon}),\\ =\mathcal{O}(\epsilon) & \text{for } t\le a-\sqrt{\epsilon} \text{ or } t\ge b+\sqrt{\epsilon}.\end{cases} \tag{2.9}$$

We find that

$$\frac{1}{\pi}\int_a^b \Im S_\mu(\lambda+i\epsilon))d\lambda \le \mu((a-\sqrt{\epsilon}, b+\sqrt{\epsilon})) + \mathcal{O}(\epsilon)\mu(\mathbb{R}\setminus(a-\sqrt{\epsilon}, b+\sqrt{\epsilon})). \tag{2.10}$$

Combine (2.7) and (2.10), we finish the proof. □

We need the property of Stieltjes transform that the convergence of measures implies the convergence of their Stieltjes transforms, and vice versa. To be precise, we state the following theorem.

Theorem 3. *Let μ_n be a sequence of probability measures.*

1. *If μ_n converges weakly to a probability measure μ, then $S_{\mu_n}(z)$ converges to $S_\mu(z)$ for each $z\in\mathbb{C}\setminus\mathbb{R}$.*
2. *If $S_{\mu_n}(z)$ converges to S_μ for each $z\in\mathbb{C}\setminus\mathbb{R}$ to a limit $S(z)$, then $S(z)$ is the Stieltjes transform of a sub-probability measure μ (i.e., $\mu(\mathbb{R})\le 1$) and μ_n converges vaguely to μ (i.e., for any continuous function f on $\mathbb{R}$ such that $\lim_{x\to\pm\infty} f(x)=0$, $\int f d\mu_n \to \int f d\mu$).*
3. *If the probability measure μ_n are random and, for each $z\in\mathbb{C}\setminus\mathbb{R}$, $S_{\mu_n}(z)$ converges in probability to a deterministic limit $S(z)$ that is the Stieltjes transform of a probablility measure μ, then μ_n converges weakly in probability to μ.*

Proof. 1 By definition: $\mu_n \xrightarrow{d} \mu$ implies $\frac{1}{2\pi i}\int \frac{d\mu_n(x)}{x-z} \to \frac{d\mu(x)}{x-z}$, since $\frac{1}{2\pi i(x-z)}$ is bounded and continuous in x, if $z\notin\mathbb{R}$.

2 By Helly's selection theorem [9, Section 4.3], a subsequence μ_{n_k} vaguely converges to a sub-probability measure μ. Suppose μ is not the vague limit of μ_n, that is, there is a positive constant ϵ and a function f that is continuous and vanishing at $\pm\infty$, such that for a subsequence $\mu_{n'_k}$, $|\int f d\mu_{n'_k} - \int f d\mu| > \epsilon$. Using Helly's selection theorem again, a subsequence of $\mu_{n'_k}$ converges vaguely to $\mu'\neq\mu$. On the other hand, since $S_{\mu_n}(z)$ converges, we have $S_\mu(z)=S_{\mu'}(z)$, and get a contradiction to Theorem 2.

3 (By Tao [25, Section 4.2]). We show that for each f that is continuous and vanishing at $\pm\infty$, $\int f d\mu_n \xrightarrow{p} \int f d\mu$. Then μ_n converges *vaguely* in probability to μ. Since we assume that μ is a probability measure, the vague convergence is equivalent to weak convergence.

To show that

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\int f d\mu_n - \int f d\mu\right| > \epsilon\right) = 0, \tag{2.11}$$

we need only to find an $\tilde{f}$ such that

$$\max_{x\in\mathbb{R}}|f(x)-\tilde{f}(x)|<\frac{\epsilon}{3} \tag{2.12}$$

and

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\int\tilde{f}d\mu_n-\int\tilde{f}d\mu\right|>\epsilon\right)=0. \tag{2.13}$$

Such an $\tilde{f}$ can be chosen as (see Exercise 1)

$$\tilde{f}(x)=\sum_{k=1}^{m}c_k\Im\frac{1}{x-z_k}=\sum_{k=1}^{m}\frac{c_kb_k}{(x-a_k)^2+b_k^2}\quad\text{where}\quad z_k=a_k+ib_k. \tag{2.14}$$

Then (2.13) becomes

$$\lim_{n\to\infty}\mathbb{P}\left(\left|\sum_{k=1}^{m}c_k(S_{\mu_n}(z_k)-S_\mu(z_i)\right|>\frac{\epsilon}{3}\right)=0, \tag{2.15}$$

which is a consequence of $S_{\mu_n}(z_k)\xrightarrow{p}S_\mu(z_k)$.

□

For an $N\times N$ Hermitian matrix X, define the Stieltjes transform of its ESD as

$$S_X(z)=S_{dF^X}(z)=\int\frac{dF^X(z)}{x-z}=\frac{1}{N}\sum_{k=1}^{N}\frac{1}{\lambda_k^N-z}. \tag{2.16}$$

Note that

$$S_X(z)=\frac{1}{N}\operatorname{Tr}(X-zI)^{-1}. \tag{2.17}$$

For the analysis of Wigner matrix, we state a linear algebraic result for Hermitian matrix.

Lemma 1. *Let X be an $N\times N$ symmetric matrix, and let $\mathbf{x}_k$ denote the k-th column of X with the entry x_{kk} removed. Let the $(N-1)\times(N-1)$ matrix X_{kk} be the minor of X obtained by removing the k-th column and k-th row. Then for $z\in\mathbb{C}\setminus\mathbb{R}$, the (k,k) entry of $(X-zI)^{-1}$ is*

$$(X-zI)^{-1}(k,k)=\frac{1}{x_{kk}-z-\bar{\mathbf{x}}_k^T(X_{kk}-zI_{N-1})^{-1}\mathbf{x}_k}. \tag{2.18}$$

Therefore (2.17) can be written as

$$S_X(z)=\frac{1}{N}\sum_{k=1}^{N}\frac{1}{x_{kk}-z-\bar{\mathbf{x}}_k^T(X_{kk}-zI_{N-1})^{-1}\mathbf{x}_k}. \tag{2.19}$$

Proof of Lemma 1. Without loss of generality, we consider the $k=1$ case. From basis linear algebra,

$$(X-zI)^{-1}(1,1)=\frac{\det(X_{1,1}-zI)}{\det(X-zI)}=\frac{1}{\det(X-zI)/\det(X_{1,1}-zI)}. \tag{2.20}$$

Computing $\det(X_{1,1}-zI)$ by Laplace's formula along the first row $(x_{1,1}-z,\bar{\mathbf{x}}_1^T)$, and then divide it by $\det(X_{1,1}-zI)$, we have

$$\frac{\det(X-zI)}{\det(X_{1,1}-zI)}=(x_{1,1}-z)-\sum_{k=2}^{N}(-1)^k\bar{x}_{k,1}\frac{\det(X_{k,1}-zI_{k,1})}{\det(X_{1,1}-zI)}. \tag{2.21}$$

The form of $\frac{\det(X_{k,1}-zI_{k,1})}{\det(X_{1,1}-zI)}$ give us reminiscence of Cramer's rule, and we have that $(-1)^k\frac{\det(X_{k,1}-zI_{k,1})}{\det(X_{1,1}-zI)} = a_k$, where $a_2,\dots,a_N$ are solution to

$$(X_{1,1}-zI)\begin{pmatrix} a_2 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} x_{2,1} \\ \vdots \\ x_{N,1} \end{pmatrix}. \tag{2.22}$$

Then (2.21) becomes

$$\frac{\det(X_{1,1}-zI)}{\det(X_{1,1}-zI)} = (x_{1,1}-z) - \bar{\mathbf{x}}_1^T\cdot(X_{1,1}-zI)\mathbf{x}_1, \tag{2.23}$$

and we prove (2.18) when $k=1$. □

2.3 Limiting ESD of truncated Wigner matrices

Before giving the proof of Theorem 1, we consider the truncated version of Wigner matrix. In this subsection, we assume the random variables Y_i and Z_{ij} that define the Wigner matrix satisfy the follows.

Truncation conditions

1. $Y_1 = 0$.
2. The support of Z_{11} is compact, i.e., there exists $C>0$ such that $\mathbb{P}(|Z_{11}|>C)=0$.

We prove Theorem 1 under these two conditions.

Here and later, for notational simplicity we write X_N as X if there is no confusion, i.e., $S_{X_N}(z)$ is written as $S_X(z)$.

Since the diagonal entries are 0, (2.19) becomes

$$S_X(z) = \frac{-1}{N}\sum_{k=1}^{N}\frac{1}{z+\bar{\mathbf{x}}_k^T(X_{kk}-zI_{N-1})^{-1}\mathbf{x}_k}. \tag{2.24}$$

Two heuristic observation

1. The ESD of X_{kk} is close to the ESD of X, due to the Cauchy interlacing inequality [16, Section 4.3], [10, Section 1.4]. Suppose eigenvalues of X are $\{\lambda_j^N\}$ and those of X_{kk} are $\{\lambda_j^{N-1}\}$, both in increasing order, then

$$\lambda_1^N \le \lambda_1^{N-1} \le \lambda_2^N \le \lambda_2^{N-1} \le \cdots \le \lambda_{N-1}^N \le \lambda_{N-1}^{N-1} \le \lambda_N^N. \tag{2.25}$$

   Thus $S_{X_{kk}}(z) \sim S_X(z)$.

2. Note that $X_{kk}-zI_{N-1}$ is a *normal* matrix, such that

$$\begin{aligned}(X_{kk}-zI)^*(X_{kk}-zI) &= (X_{kk}-\bar{z}I)(X_{kk}-zI) = X_{kk}^2 - zX_{kk} - \bar{z}X_{kk} + |z|^2 I \\ &= (X_{kk}-zI)(X_{kk}-zI)^*.\end{aligned} \tag{2.26}$$

   Then $(X_{kk}-zI_{N-1})^{-1}$ is also normal. Let $\lambda_1^{(N-1)},\dots,\lambda_{N-1}^{(N-1)}$ be eigenvalues of X_{kk} and denote

$$z_j = \frac{1}{\lambda_j^{(N-1)}-z}, \quad \text{for } j=1,\dots,N-1, \tag{2.27}$$

then $z_1, \ldots, z_{N-1}$ are eigenvalues of $(X_{kk} - zI_{N-1})^{-1}$. By the property of normal matrices [16, Section 2.5], there exists a unitary matrix $U \in \mathrm{U}(N-1)$ such that

$$(X_{kk} - zI_{N-1})^{-1} = U^* D U, \quad \text{where} \quad D = \mathrm{diag}(z_1, \ldots, z_{N-1}), \tag{2.28}$$

and

$$\mathbf{x}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k = \bar{\mathbf{y}}_k^T D \mathbf{y}_k, \quad \text{where} \quad \mathbf{y}_k = U \mathbf{x}_k. \tag{2.29}$$

In the special case that all components of $\mathbf{x}_k$ are in i.i.d. $\mathrm{N}(0, \frac{1}{2}) + i\mathrm{N}(0, \frac{1}{2})$ distribution, then components of $\mathbf{y}_k$ are also in i.i.d. $\mathrm{N}(0, \frac{1}{2}) + i\mathrm{N}(0, \frac{1}{2})$ distribution (exercise). In this special case, by (a stretch of) law of large numbers,

$$\bar{\mathbf{y}}_k^T D \mathbf{y}_k = \sum_{j=1}^{N-1} \overline{\mathbf{y}_k(j)} \mathbf{y}_k(j) z_j \sim \frac{1}{N} \sum_{j=1}^{N-1} z_j = S_{X_{kk}}(z) \sim S_X(z), \tag{2.30}$$

where in the last approximate identity we use the heuristic observation 1.

The idea of the following transformation is then clear: To approximate $\mathbf{x}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k$ by $S_X(z)$.

$$\begin{aligned} S_X(z) &= \frac{-1}{N} \sum_{k=1}^{N} \frac{1}{z + S_X(z) + (\bar{\mathbf{x}}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k - S_X(z))} \\ &= -\frac{1}{z + S_X(z)} + \delta_N(z), \end{aligned} \tag{2.31}$$

where

$$\delta_N(z) = \frac{1}{N} \sum_{k=1}^{N} \frac{\epsilon_{k,N}}{(z + S_X(z))(z + S_X(z) + \epsilon_{k,N})}, \quad \text{with} \quad \epsilon_{k,N} = \bar{\mathbf{x}}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k - S_X(z). \tag{2.32}$$

The next task os to show that for any $z \notin \mathbb{C} \setminus \mathbb{R}$,

$$\delta_N(z) \xrightarrow{p} 0. \tag{2.33}$$

Suppose this is true, then

$$S_X(z) + \frac{1}{z + S_X(z)} = \delta_N(z) \xrightarrow{p} 0. \tag{2.34}$$

Note that for any $z \notin \mathbb{C} \setminus \mathbb{R}$, the equation in w

$$w + \frac{1}{z + w} = 0 \tag{2.35}$$

has two solutions

$$w_1(z) = \frac{-z + \sqrt{z^2 - 4}}{2}, \quad w_2(z) = \frac{-z - \sqrt{z^2 - 4}}{2}. \tag{2.36}$$

So for any $\epsilon > 0$

$$\lim_{N \to \infty} \mathbb{P}(S_X(z) \in N_\epsilon(w_1(z)) \cup N_\epsilon(w_2(z))) = 1. \tag{2.37}$$

From the definition of Stieltjes transform, it is clear that $\Im S_X(z)$ has the same sign as $\Im z$. Therefore $S_X(z) \xrightarrow{p} w_1(z)$, which is the Stieltjes transform of the semicircle law $F(z)$. Verification:

$$\lim_{\epsilon \to 0^+} \Im w_1(x + i\epsilon) = \begin{cases} \frac{1}{2\pi} \sqrt{4 - x^2} & \text{if } x \in (-2, 2), \\ 0 & \text{if } x \in \mathbb{R} \setminus (-2, 2). \end{cases} \tag{2.38}$$

Then by Theorem 3, we prove Theorem 1 under the two trancation conditions.

To show $\delta_N(z) \xrightarrow{p} 0$, note that for any z in the upper half plane, $\Im z > 0$ and $\Im S_X(z) > 0$, and then

$$|z + S_X(z)| \geq |\Im z + \Im S_X(z)| > |\Im z|. \tag{2.39}$$

For z in the lower half plane, the result of (2.39) also holds. Therefore if we can show that

$$\max_{1\leq k\leq N} |\epsilon_{k,N}| \xrightarrow{p} 0, \tag{2.40}$$

then the absolute value of the denominator of each summand of $\delta_N(z)$ is bounded when $|\epsilon_{k,N}|$ is small, and then (2.33) is proved.

To estimate $|\epsilon_{k,N}|$, we write it as

$$\epsilon_{k,N} = \epsilon_{k,N}(1) + \epsilon_{k,N}(2), \tag{2.41}$$

where

$$\begin{aligned} \epsilon_{k,N}(1) &= \bar{\mathbf{x}}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k - \frac{N-1}{N} S_{X_{kk}}(z) \\ &= \bar{\mathbf{x}}_k^T (X_{kk} - zI_{N-1})^{-1} \mathbf{x}_k - \frac{1}{N} \operatorname{Tr}(X_{kk} - zI_{N-1})^{-1}, \end{aligned} \tag{2.42}$$

and

$$\epsilon_{k,N}(2) = \frac{N-1}{N} S_{X_{kk}}(z) - S_X(z) = \frac{1}{N} \left(\operatorname{Tr}(X_{kk} - zI_{N-1})^{-1} - \operatorname{Tr}(X - zI_N)^{-1} \right). \tag{2.43}$$

To estimate $\epsilon_{k,N}(1)$, we take short-handed notations

$$\sqrt{N}\mathbf{x}_k = (\xi_1, \ldots, \xi_{N-1}) \quad \text{and} \quad (X_{kk} - zI_{N-1})^{-1} = B. \tag{2.44}$$

Note that ξ_j are independent and has the same distribution as Z_{12}, and the ξ_j's are independent to B. Further write

$$\epsilon_{k,N}(1) = \epsilon_{k,N}(3) + \epsilon_{k,N}(4), \quad \text{where} \quad \begin{cases} \epsilon_{k,N}(3) = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N-1} (\bar{\xi}_j \xi_j - 1) B(j,j), \\ \epsilon_{k,N}(4) = \dfrac{1}{N} \displaystyle\sum_{\substack{j,l=1 \\ j\neq l}}^{N-1} \bar{\xi}_j \xi_l B(j,l). \end{cases} \tag{2.45}$$

By the property of ξ_j, $\mathrm{E}(\bar{\xi}_j \xi_j - 1) = 0$. Since $B(j,j)$ is independent to ξ_j, we see that each summand of $\epsilon_{k,N}(3)$ has mean 0, and then $\mathrm{E}(\epsilon_{k,N}(3)) = 0$.

Now we consider the conditional expectation of $|\epsilon_{k,N}(3)|^2$ when B is fixed.

$$\begin{aligned} \mathrm{E}(|\epsilon_{k,N}(3)|^2 \mid B = B_0) &= \frac{1}{N^2} \mathrm{E}\left(\left(\sum_{j=1}^{N-1} (\bar{\xi}_j \xi_j - 1) B(j,j) \right) \left(\sum_{j=1}^{N-1} (\bar{\xi}_j \xi_j - 1) \overline{B(j,j)} \right) \right) \\ &= \frac{1}{N^2} \sum_{j=1}^{N-1} \mathrm{E}((\bar{\xi}_j \xi_j - 1)^2) |B_0(j,j)|^2, \end{aligned} \tag{2.46}$$

where in the second identity we use the independence of ξ_j's. We know that $\mathrm{E}((\bar{\xi}_j \xi_j - 1)^2) = \mathrm{E}((\bar{Z}_{12} Z_{12} - 1)^2)$ is a constant, and need to estimate $|B(j,j)|^2$.

Recall the facterisation $B = U^*DU$ where $U \in \mathrm{U}(N-1)$ and $D = \mathrm{diag}(z_1, \dots, z_{N-1})$ and z_j are defined in (2.27). Let the j-th column of U be $\mathbf{u}_j$ that is a unit vector. Then

$$B(j,j) = \mathbf{u}^* D \mathbf{u}, \tag{2.47}$$

and

$$|B(j,j)| \le \|D\| = \max_{1 \le j \le N-1} |z_j|. \tag{2.48}$$

(Recall the operator norm $\|D\| = \max_{|\mathbf{v}|=1} |D\mathbf{v}|$.) From the definition of z_j, it is clear that

$$|z_j| = \frac{1}{|z - \lambda_j^{(N-1)}|} = \frac{1}{|\Im z + (\Re z - \lambda_j^{(N-1)})|} \le \frac{1}{|\Im z|}. \tag{2.49}$$

Hence $B(j,j)$ is bounded by $|\Im z|^{-1}$. We conclude that $\mathrm{E}(|\epsilon_{k,N}(3)|^2 \mid B = B_0)$ is bounded by cN^{-1} where c is a positive constant independent to B and k. We further conclude that

$$\mathrm{E}(|\epsilon_{k,N}(3)|^2) < cN^{-1}, \tag{2.50}$$

with the same c independent to k.

This result is good, and by $\mathrm{E}(|\epsilon_{k,N}(3)|^2)$ we directly find $\epsilon_{k,N}(3) \xrightarrow{p} 0$. But this is not enough, since our goal is to estimate $\max_{1 \le k \le N} \epsilon_{k,N}$, so we need

$$\max_{1 \le k \le N} \epsilon_{k,N}(3) \xrightarrow{p} 0. \tag{2.51}$$

This cannot be proved by the estimate of second moment (2.50). How to remedy it? Keep computing higher moments!

$$\begin{aligned}
&\mathrm{E}(|\epsilon_{k,N}(3)|^4 \mid B = B_0) \\
&= \frac{1}{N^4} \mathrm{E}\left(\left(\sum_{j=1}^{N-1} (\bar{\xi}_j \xi_j - 1) B(j,j) \right)^2 \left(\sum_{j=1}^{N-1} (\bar{\xi}_j \xi_j - 1) \overline{B(j,j)} \right)^2 \right) \\
&= \frac{1}{N^4} \left(\sum_{j=1}^{N-1} \mathrm{E}((\bar{\xi}_j \xi_j - 1)^4) |B_0(j,j)|^4 + \sum_{1 \le j < l \le N-1} \mathrm{E}((\bar{\xi}_j \xi_j - 1)^2) \mathrm{E}((\bar{\xi}_l \xi_l - 1)^2) \right. \\
&\quad \left. \times (2B_0(j,j)^2 \overline{B_0(l,l)}^2 + 2\overline{B_0(j,j)}^2 B_0(l,l)^2 + 8|B_0(j,j)|^2 |B_0(j,j)|^2) \right) \\
&< \frac{c}{N^2}.
\end{aligned} \tag{2.52}$$

and by arguments the same as in the discussion of the second moment, we see that $\mathrm{E}(|\epsilon_{k,N}(3)|^2) < cN^{-1}$ for a constant independent of k. Then by Chebyshev's inequality [9, Section 3.2], we find that for any $\epsilon > 0$,

$$\mathbb{P}(|\epsilon_{k,N}(3)| > \epsilon) = \mathcal{O}(N^{-2}), \tag{2.53}$$

and prove (2.51) by

$$\mathbb{P}(\max_{1 \le k \le N} |\epsilon_{k,N}(3)| > \epsilon) = \mathcal{O}(N^{-1}). \tag{2.54}$$

Similarly, we can show that (exercise)

$$\mathrm{E}(|\epsilon_{k,N}(4)|^4) < \frac{c}{N^2}. \tag{2.55}$$

for some c independent of k, and then show

$$\max_{1\le k\le N} \epsilon_{k,N}(4) \xrightarrow{p} 0. \tag{2.56}$$

Therefore we conlude from (2.51) and (2.56) that $\max_{1\le k\le N} \epsilon_{k,N}(1) \xrightarrow{p} 0$.

Next we need to estimate $\epsilon_{k,N}(2)$. Denote $\hat{X}_{kk}$ as the $N\times N$ Hermitian matrix with entries in the k-th column and k-th row 0 and all other entries the same as those of X. The eigenvalues of $\hat{X}_{kk}$ are those of X_{kk}, i.e., $\lambda_1^{(N-1)},\dots,\lambda_{N-1}^{(N-1)}$ and 0. Then

$$\mathrm{Tr}((X_{kk}-zI_{N-1})^{-1}) = \sum_{j=1}^{N-1}\frac{1}{\lambda_j^{(N-1)}-z} = \mathrm{Tr}((\hat{X}_{kk}-zI_N)^{-1}) - \frac{1}{z}, \tag{2.57}$$

and

$$\epsilon_{k,N}(2) = \frac{1}{N}\left(\mathrm{Tr}((\hat{X}_{kk}-zI_N)^{-1}) - \mathrm{Tr}((\hat{X}_{kk}-zI_N)^{-1}) - \frac{1}{z}\right). \tag{2.58}$$

To estimate the $\mathrm{Tr}(\hat{X}_{kk}-zI_N)-\mathrm{Tr}(\hat{X}_{kk}-zI_N)$, we apply the following linear algebraic theorem.

Theorem 4 (Hoffman-Wielandt)**.** *Let A,B be $N\times N$ Hermitian matrices, with eigenvalues $\lambda_1^A\le\lambda_2^A\le\cdots\le\lambda_N^A$ and $\lambda_1^B\le\lambda_2^B\le\cdots\le\lambda_N^B$. Then*

$$\sum_{j=1}^{N}|\lambda_j^A-\lambda_j^B|^2 \le \mathrm{Tr}((A^*-B^*)(A-B)). \tag{2.59}$$

The proof of Theorem 4 is given in Appendix A.2. Here we note that For any $A=(a_{jk})_{j,k=1}^n$, $\mathrm{Tr}(A^*A)=\sum_{j,k=1}^n|a_{jk}|^2$. Denoting the eigenvalues of $\hat{X}_{kk}$ as $\lambda_1\le\cdots\le\lambda_N$ and the eigenvalues of X as $\lambda_1^{(N)}\le\cdots\le\lambda_N^{(N)}$, we have

$$\begin{aligned}
\left|\mathrm{Tr}((\hat{X}_{kk}-zI_N)^{-1}) - \mathrm{Tr}((\hat{X}_{kk}-zI_N)^{-1})\right| &= \left|\sum_{j=1}^{N}\left(\frac{1}{\lambda_j-z}-\frac{1}{\lambda_j^{(N)}-z}\right)\right| \\
&= \left|\sum_{i=1}^{N}\frac{\lambda_j^{(N)}-\lambda_j}{(\lambda_j-z)(\lambda_j^{(N)}-z)}\right| \\
&\le \sum_{j=1}^{N}\frac{|\lambda_j^{(N)}-\lambda_j|}{|\lambda_j-z||\lambda_j^{(N)}-z|} \\
&\le \frac{1}{|\Im z|^2}\sum_{j=1}^{N}|\lambda_j^{(N)}-\lambda_j| \\
&\le \frac{1}{|\Im z|^2}\left(\sum_{j=1}^{N}|\lambda_j^{(N)}-\lambda_j|^2\right)^{\frac12}\left(\sum_{j=1}^{N}1\right)^{\frac12} \\
&\le \frac{\sqrt{N}}{|\Im z|^2}\left(\sum_{j=1}^{N}|\lambda_j^{(N)}-\lambda_j|^2\right)^{\frac12}.
\end{aligned} \tag{2.60}$$

By Hoffman-Wielandt theorem,

$$\sum_{j=1}^{N}|\lambda_j^{(N)}-\lambda_j|^2 \le 2\sum_{j=1}^{N}|X(j,k)|^2 \le 2C^2, \tag{2.61}$$

where C is the upper bound of Z_{12}. Thus we find that $\epsilon_{k,N}(2)$ converges to 0 surely. Now we finish the proof of Theorem 1 under the two trancation conditions.

2.4 Limiting ESD for general Wigner matrices

In this subsection, we show that the semicircle law proved for trancated Wigner matrices also holds for general Wigner matrices.

Restore the truncation on diagonal entries First, let X_N be the trancated Wigner matrix considered in last subsection, and $\tilde{X}_N$ be the Wigner matrix that satisfies Condition 2 but not Condition 1 for truncated Wigner matrix. Namely, the diagonal entries of $\tilde{X}_N$ are given by i.i.d. random variables $N^{-1/2}Y_i$ where Y_1 has mean 0 and every moment finite, and the off-diagonal entries of $\tilde{X}_N$ are the same as those of X_N. Then a direct application of the Hoffman-Wielandt theorem shows that

$$\sum_{i=1}^{N}|\lambda_i^{(N)} - \tilde{\lambda}_i^{(N)}|^2 \leq \mathrm{Tr}((\tilde{X}_N^* - X_N^*)(\tilde{X}_N - X_N)) = \frac{1}{N}\sum_{i=1}^{N}|Y_i|^2, \tag{2.62}$$

where $\lambda_i^{(N)}$ and $\tilde{\lambda}_i^{(N)}$ are eigenvalues of X_N and $\tilde{X}_N$ respectively, and are sorted in increasing orders.

By the law of large numbers, we see that $\frac{1}{N}\sum_{i=1}^{N}|Y_i|^2$ converges to $\mathrm{E}(|Y_1|^2)$ almost surely (and also in probability). Since we are interested in the weak, inprobability convergence of the random ESD measures, we apply the following theorem.

Theorem 5 ([9, Exercise 9 in Section 4.4])**.** *Let μ_n be a sequence of probability measures with cumulative functions $F_n(x) = \int_{-\infty}^{x} 1 d\mu_n$. Then μ_n converges weakly to a probability measure μ if and only if $\rho(F_n, F) \to 0$, where F is the cumulative distribution function of μ and ρ is the Lévy distance*

$$\rho(F, G) = \inf\{\epsilon \mid F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon \text{ for all } x\}. \tag{2.63}$$

The proof of this theorem is an exercise. Note that if $\rho(F, G) > \epsilon$, without loss of generality, we can assume $F(x_0 - \epsilon) - \epsilon > G(x_0)$ for an $x_0 \in \mathbb{R}$ and then

$$\int_{-\infty}^{\infty}|F(x) - G(x)|dx \geq \int_{x_0-\epsilon}^{x_0} F(x) - G(x)dx \geq \epsilon^2. \tag{2.64}$$

Geometrically, it means that we can put a square with size ϵ between the graphs of $F(x)$ and $G(x)$.

Now we consider $\rho(F^{\tilde{X}_N}, F^{X_N})$ where $F^{\tilde{X}_N}$ and F^{X_N} are the ESDs of $\tilde{X}_N$ and X_N.

$$\rho(F^{\tilde{X}_N}, F^{X_N})^2 \leq \int_{-\infty}^{\infty}|F^{\tilde{X}_N} - F^{X_N}|dx = \frac{1}{N}\sum_{i=1}^{N}|\tilde{\lambda}_i^{(N)} - \lambda_i^{(N)}| \leq \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}(\tilde{\lambda}_i^{(N)} - \lambda_i^{(N)})^2\right)^{\frac{1}{2}}, \tag{2.65}$$

where we use the Cauchy-Schwarz inequality. Now it is clear that $\rho((F^{\tilde{X}_N}, F^{X_N}) \xrightarrow{p} 0$ (actually the convergence is almost sure). We know as well that $\rho((F^{X_N}, F) \xrightarrow{p} 0$ from the result in last subsection, where F is the ESD of the semicircle law. Hence $\rho((F^{\tilde{X}_N}, F) \xrightarrow{p} 0$, i.e., $F^{\tilde{X}_N}$ converges weakly, in probability to the seimicircle law, and we prove Theorem 1 for $\tilde{X}_N$ whose off-diagonal entries are truncated but diaagonal entries are general.

Restore the truncation on off-diagonal entries Now we consider the general Wigner matrix, and denote it as $\hat{X}_N$ for fear of namespace conflict. Then the diagonal and off-diagonal entries of $\hat{X}_N$ are given by Y_i and Z_{ij} by (2.4) with X_N substituted by $\hat{X}_N$. We compare $\hat{X}_N$ with $\tilde{X}_N$ analysed above. To make them related, below we define $\tilde{X}_N$ by $\hat{X}_N$ as

$$\tilde{X}_N(i,i) = \hat{X}_N(i,i), \quad \text{for } i = 1, \dots, N, \tag{2.66}$$

$$\tilde{X}_N(i,j) = \hat{X}_N(i,j) - \mathbf{1}_{\sqrt{N}|\hat{X}_N(i,j)|\le C} - \mathrm{E}(\hat{X}_N(i,j) - \mathbf{1}_{\sqrt{N}|\hat{X}_N(i,j)|\le C}), \quad \text{for } 1 \le i < j \le N, \tag{2.67}$$

where C is a large but finite number independent of N.

Using the Hoffman-Wielandt theorem again (now $\hat{\lambda}_i^{(N)}$ are eigenvalues of $\hat{X}_N$ in increasing order), we have

$$\sum_{i=1}^{N} |\hat{\lambda}_i^{(N)} - \tilde{\lambda}_i^{(N)}|^2 = \mathrm{Tr}((\hat{X}_N^* - \tilde{X}_N^*)(\hat{X}_N - \tilde{X}_N)) = \frac{2}{N} \sum_{1\le i<j\le N} |\hat{Z}_{12}|^2, \tag{2.68}$$

where

$$\hat{Z}_{ij} = Z_{ij}\mathbf{1}_{|Z_{ij}|>C} - \mathrm{E}(Z_{ij}\mathbf{1}_{|Z_{ij}|>C}). \tag{2.69}$$

The expectation $\mathrm{E}(Z_{ij}\mathbf{1}_{|Z_{ij}|>C})$ depends on C, and we denote is as $\epsilon(C)$. It is not difficult to see (exercise) that

$$\lim_{C\to\infty} \epsilon(C) = 0. \tag{2.70}$$

By the law of large numbers,

$$\frac{1}{N}\sum_{i=1}^{N} |\hat{\lambda}_i^{(N)} - \tilde{\lambda}_i^{(N)}|^2 \xrightarrow{p} \epsilon(C). \tag{2.71}$$

For any $\epsilon > 0$, there is a C such that $\epsilon(C) < \epsilon^2/4$. Then (2.71) implies that

$$\lim_{N\to\infty} \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^{N} |\hat{\lambda}_i^{(N)} - \tilde{\lambda}_i^{(N)}|^2 > \frac{\epsilon^2}{4}\right) = 0. \tag{2.72}$$

Analogous to (2.65), we have

$$\rho(F^{\hat{X}_N}, F^{\tilde{X}_N}) \le \int_{-\infty}^{\infty} |F^{\hat{X}_N} - F^{\tilde{X}_N}| dx = \frac{1}{N}\sum_{i=1}^{N} |\tilde{\lambda}_i^{(N)} - \lambda_i^{(N)}| \le \left(\frac{1}{N}\sum_{i=1}^{N} (\tilde{\lambda}_i^{(N)} - \lambda_i^{(N)})^2\right)^{\frac{1}{2}}, \tag{2.73}$$

and by (2.72),

$$\lim_{N\to\infty} \mathbb{P}\left(\rho(F^{\hat{X}_N}, F^{\tilde{X}_N}) > \frac{\epsilon}{2}\right) = 0. \tag{2.74}$$

On the other hand, the weak, inprobability convergence of the ESD of $\tilde{X}_N$ implies (via Theorem 5)

$$\lim_{N\to\infty} \mathbb{P}\left(\rho(F^{\tilde{X}_N}, F) > \frac{\epsilon}{2}\right) = 0. \tag{2.75}$$

Thus

$$\lim_{N\to\infty} \mathbb{P}\left(\rho(F^{\hat{X}_N}, F) > \epsilon\right) = 0. \tag{2.76}$$

By Theorem 5 and the arbitrariness of ϵ, we see that the ESD of $\hat{X}_N$ converges weakly, in probability to the semicircle law. Finally we prove Theorem 1 in the general case.

Exercises

1. Let $f(x)$ be a continuous function on $\mathbb{R}$ and $\lim_{x\to\pm\infty} f(x) = 0$.

   (a) Define
   $$f_\delta(x) = \frac{1}{\pi}\int_{\mathbb{R}} \frac{\delta}{(t-x)^2+\delta^2} f(t)dt. \tag{2.77}$$
   Prove that as $\delta \to 0$, $f_\delta(x)$ converges uniformly to $f(x)$.

   (b) By result 1a, show the existence of the function $\tilde{f}(x)$ that is in the form of (2.14) and satisfies (2.12).

2. Prove Theorem 5.

3 Local properties of the distribution of eigenvalues in GUE

In this section we focus on random matrices in Gaussian Unitary Ensemble (GUE), or GUE matrix for short. The definition of GUE matrix is given in Exercise 1, and it is a special case of the Wigner matrix defined by (2.4) where Y_1 has distribution $\mathrm{N}(0,1)$ and Z_{12} has distribution $\mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2})$.

GUE matrix is a very special Wigner matrix in the sense that it is invariant under unitary conjugation, while all other Wigner matrices, except for those equivalent to the GUE matrix up to scaling and translation, are not. See Exercise 1. Some powerful techniques involving orthogonal polynomials can be applied to GUE matrix so that we can analyse the local properties of the distributions of its eigenvalues, expecially the asymptotic behaviours as $N \to \infty$. Many of these properties can be generalised to Wigner matrices with $\Re Y_{12}$ and $\Im Y_{12}$ in identical distribution, but the proofs are much more difficult.

3.1 Joint distribution of eigenvalues of a GUE matrix

In this subsection we prove the following theorem:

Theorem 6. *Let $H = H_N$ be the $N \times N$ random matrix in GUE. The distribution of its eigenvalues $\lambda_1, \ldots, \lambda_N$ is*

$$P(\lambda_1, \ldots, \lambda_N) = \frac{1}{C_N}\Delta(\lambda)^2 \prod_{i=1}^N e^{\frac{N}{2}\lambda_i^2}, \tag{3.1}$$

where $\Delta(\lambda) = \prod_{1\le i<j\le N}(\lambda_j - \lambda_i)$ is the Vandermonde determinant and C_N is a constant; or equivalently, if we assume $\lambda_1, \ldots, \lambda_N$ are in increasing order,

$$P(\lambda_1, \ldots, \lambda_N) = \frac{1}{C'_N}\Delta(\lambda)^2 \prod_{i=1}^N e^{\frac{N}{2}\lambda_i^2} \mathbf{1}_{\lambda_1\le\cdots\le\lambda_N}, \tag{3.2}$$

where $C'_N = (N!)^{-1}C_N$.

Note that the distribution of $\lambda_1, \ldots, \lambda_N$ is a *marginal* distribution of X. Recall the standard method to compute the marginal distribution function from the joint distribution function:

$$P(x) = \int P(x,y)dy. \tag{3.3}$$

The question is: What is y?

The idea of the proof of Theorem 6 is that

$$H = UDU^*, \quad \text{where} \quad D = \operatorname{diag}(\lambda_1, \dots, \lambda_N), \quad U = (u_{ij})_{i,j=1}^N \in \mathrm{U}(N). \tag{3.4}$$

We show that except for a measure zero part of the probability space of H, which can be assumed to be $\mathbb{R}^{N^2}$, the decomposition above is unique and U is parametrised smoothly by $N(N-1)$ variables, say $\theta_1, \dots, \theta_{N(N-1)}$, and then obtain the marginal distribution of $\lambda_1, \dots, \lambda_N$ by integrating out the variables $\theta_1, \dots, \theta_{N(N-1)}$. However, we only know that the probability distribution function of $X = (h_{ij})_{i,j=1}^n$, given in coordinates h_{ii} $(1 \le i \le N)$, $\Re h_{jk}$ and $\Im h_{jk}$ $(1 \le j < k \le N)$ is

$$\begin{aligned} P(h_{ii}, \Re h_{jk}, \Im h_{jk}) &= \prod_{i=1}^N \frac{\sqrt{N}}{\sqrt{2\pi}} e^{-\frac{N}{2}x_{ii}^2} \prod_{1 \le j < k \le N} \frac{N}{\pi} e^{-N(\Re h_{jk})^2} e^{-N(\Im h_{jk})^2} \\ &= 2^{-\frac{N}{2}} \left(\frac{N}{\pi}\right)^{\frac{N^2}{2}} e^{-\frac{N}{2}\operatorname{Tr} H^2} \\ &= 2^{-\frac{N}{2}} \left(\frac{N}{\pi}\right)^{\frac{N^2}{2}} \prod_{i=1}^N e^{-\frac{N}{2}\lambda_i^2}. \end{aligned} \tag{3.5}$$

To obtain the probability distribution function of H in coordinates $\lambda_1, \dots, \lambda_N, \theta_1, \dots, \theta_{N(N-1)}$, we need

$$P(\lambda_1, \dots, \lambda_N, \theta_1, \dots, \theta_{N(N-1)}) = P(h_{ii}, \Re h_{jk}, \Im h_{jk}) \frac{\partial(\lambda_1, \dots, \lambda_N, \theta_1, \dots, \theta_{N(N-1)})}{\partial(h_{ii}, \Re h_{jk}, \Im h_{jk})}, \tag{3.6}$$

where the Jacobian

$$\frac{\partial(\lambda_1, \dots, \lambda_N, \theta_1, \dots, \theta_{N(N-1)})}{\partial(h_{ii}, \Re h_{jk}, \Im h_{jk})} = \det \mathbf{H}, \quad \text{and} \quad \mathbf{H} = \det \begin{pmatrix} \frac{\partial h_{ii}}{\partial \lambda_k} & \frac{\partial h_{ii}}{\partial \theta_k} \\ \frac{\partial \Re h_{jl}}{\partial \lambda_k} & \frac{\partial \Re h_{jl}}{\partial \theta_k} \\ \frac{\partial \Im h_{jl}}{\partial \lambda_k} & \frac{\partial \Im h_{jl}}{\partial \theta_k} \end{pmatrix} \tag{3.7}$$

is the determinant of an $N^2 \times N^2$ matrix, whose six blocks are of size $N \times N$, $N(N-1)/2 \times N$, $N \times N(N-1)$ and $N(N-1)/2 \times N(N-1)$. The computation of the Jacobian in (3.6) turns out to be the trickiest step.

To prove the "bad" part has zero probability, we need an apparently obvious result.

Lemma 2. *Given any nonzero polynomial from $\mathbb{R}^n$ to $\mathbb{R}$, its zero set has Lebesgue measure zero.*

The proof is left as an exercise. Below we identify the (real) linear space of $N \times N$ Hermitian matrices as $\mathbb{R}^{N^2}$ with coordinates $h_{ii}, \Re h_{jk}, \Im h_{jk}$, in order to apply the lemma above.

Lemma 3. *In the space of $N \times N$ Hermitian matrices, the set of matrices with repeated eigenvalues has Lebesgue measure zero.*

Proof. The *discriminant* of an $N \times N$ matrix $H = (h_{ij})_{i,j=1}^N$ is defined as $\prod_{1 \le j < k \le N} (\lambda_k - \lambda_j)^2$ where $\lambda_1, \dots, \lambda_N$ are eigenvalues of H. Actually it is the discriminant of the characteristic polynomial of H. We need the property that the discriminant of H is a polynomial in h_{ij} ([22, Exercise 19 in Section 2]) and hence a polynomial in $h_{ii}, \Re h_{jk}, \Im h_{jk}$, and the zero set of the discriminant consists of matrices with repeated eigenvalues. A direct application of Lemma 2 finishes the proof. □

Suppose H is an $N \times N$ Hermitian matrix with distinct eigenvalues, then in the decomposition (3.4),

$$U = (\mathbf{u}_1, \dots, \mathbf{u}_N), \tag{3.8}$$

where $\mathbf{u}_k$ are eigenvectors such that $H\mathbf{u}_k = \lambda_k \mathbf{u}_k$. All u_k are linear independent, and they are unique up to a factor $e^{i\theta_k}$. Then if one U in decomposition (3.4) has all its entries nonzero, then all possible U, given that the eigenvalues of H are distinct, consist of nonzero entries. For H that satisfies 1. all eigenvalues are distinct, and 2. in one (and hence all) U, all entries are nonzero, it has a *good* decomposition

$$H = U \operatorname{diag}(\lambda_1, \dots, \lambda_N) U^*, \quad \text{where} \quad \lambda_1 < \cdots < \lambda_N \quad \text{and} \quad u_{jj} \in \mathbb{R}_+ \text{ for } j = 1, \dots, N. \tag{3.9}$$

The following lemma shows that we need only to consider Hermitian matrices with good decomposition.

Lemma 4. *In the space of $N \times N$ Hermitian matrices, the set*

$$\{H \mid \text{ in at least one decomposition } H = UDU^*, U \text{ has a zero entry}\} \tag{3.10}$$

has Lebesgue measure zero.

Proof. Based on last lemma, we need to consider only the Hermitian matrices with distinct eigenvalues.

For each $k = 1, \dots, N$, we find a condition that $\mathbf{u}_k$ has all components nonzero. Note that $A := H - \lambda_k$ is a matrix with corank 1. Hence its adjoint matrix A^{adj} (whose (i, j) entry is $(-1)^{i+j} \det(A_{ji})$ where A_{ji} is the $(N-1) \times (N-1)$ matrix by removing the i-th row and j-th column of A) is a nonzero matrix, and $AA^{\mathrm{adj}} = 0$. It means that each column of A^{adj} is a multiple of $\mathbf{u}_k$. Hence if for all $j = 1, \dots, N$, $A^{\mathrm{adj}}(j, j) = \det(H_{jj} - \lambda_k I)$ are nonzero, then all components of $\mathbf{u}_k$ are nonzero.

Therefore, if for all $j, k = 1, \dots, N$, $\det(H_{jj} - \lambda_k I) \neq 0$, then all entries of U are nonzero. Or equivalently, for all $j = 1, \dots, N$, the *resultant* of H_{jj} and H, which means the resultant of the characteristic polynomials of the pair of matrices, does not vanish, then all entries of U are nonzero. Because of the fact that the resultant of A and B can be written as a polynomial in entries of A and B (exercise, which you can solve in a similar way as in [22, Exercise 19 in Section 2] for resultant), we prove this lemma by Lemma 2 and the conclusion that all Hermitian matrices in the set considered are zeros of a polynomial in $h_{ii}, \Re h_{jk}, \Im h_{jk}$. □

Among the good Hermitian matrices, we define the *very good* Hermitian matrices as those whose U in decomposition (3.9) has all its minors nonzero. Similarly, we call a unitary matrix $U \in \mathrm{U}(N)$ *very good* if all its minors are non zero, that is, for each $k = 1, \dots, N$, and for any pair of $I, J \subseteq \{1, \dots, k\}$ with $|I| = |J|$, $\det(u_{ij})_{i \in I, j \in J} \neq 0$. We define $T : \mathrm{U}(N)^{\mathrm{vg}} := \{U \in \mathrm{U}(N) \mid U \text{ is very good}\} \to \mathbb{C}^{N(N-1)/2}$

$$T(U) = \left(\frac{u_{21}}{u_{11}}, \dots, \frac{u_{N1}}{u_{11}}, \frac{u_{32}}{u_{22}}, \dots, \frac{u_{N2}}{u_{22}}, \dots, \frac{u_{N,N-1}}{u_{N-1,N-1}} \right). \tag{3.11}$$

Lemma 5. *T is injective with smooth inverse, and $\mathbb{C}^{N(N-1)/2} \backslash T(\mathrm{U}(N)^{\mathrm{vg}})$ has Lebesgue measure zero.*

Proof. We define for each $k = 1, \dots, N$ the mapping T_k, whose domain is the array of mutually orthogonal vectors $(\mathbf{v}_1, \dots, \mathbf{v}_k)$, with $v_i(i) = 1$, all components nonzero, and for any pair of

$I, J \subseteq \{1, \dots, k\}$ with $|I| = |J|$, $\det(\mathbf{v}_i(j))_{i\in I, j\in J} \neq 0$, and whose value is in $\mathbb{C}^{(N-1)+(N-2)+\cdots+(N-k)}$, such that

$$T_k(\mathbf{v}_1, \dots, \mathbf{v}_k) = (\mathbf{v}_1(2), \dots, \mathbf{v}_1(N), \mathbf{v}_2(3), \dots, \mathbf{v}_2(N), \dots, \mathbf{v}_k(k+1), \dots, \mathbf{v}_k(N)). \tag{3.12}$$

Note that T_N is equivalent to T. Given $U = (\mathbf{u}_1, \dots, \mathbf{u}_N) \in \mathrm{U}(N)^{\mathrm{vg}}$ and let $\mathbf{v}_i = \mathbf{u}_i/u_{ii}$, then $T_N(\mathbf{v}_1, \dots, \mathbf{v}_N) = T(U)$; on the other hand, given $(\mathbf{v}_1, \dots, \mathbf{v}_N)$ in the domain of T_N, $U = (\mathbf{v}_1/\|\mathbf{v}_1\|, \dots, \mathbf{v}_1/\|\mathbf{v}_1\|)$ is in $\mathrm{U}(N)^{\mathrm{vg}}$, and $T(U) = T_N(\mathbf{v}_1, \dots, \mathbf{v}_N)$.

T_1 is obviously injective, and its range is all points in $\mathbb{C}^{N-1}$ with all coordinates nonzero. Furthermore $T_1^{-1}(z_1, \dots, z_{N-1}) = (1, z_1, \dots, z_{N-1})^\perp$.

To show that T_2 is injective, we only need to show that given any pair of vectors $(\mathbf{v}_1, \mathbf{v}_2)$ in the domain of T_2, if $\mathbf{v}_1(1) = 1, \mathbf{v}_1(2), \dots, \mathbf{v}_1(N), \mathbf{v}_2(2) = 1, \mathbf{v}_2(3), \dots, \mathbf{v}_2(N)$ are fixed (by the value of $T_2(\mathbf{v}_1, \mathbf{v}_2)$), then $\mathbf{v}_2(1)$ is unique. This is clear, since by the orthogonality, $\overline{\mathbf{v}_1(1)}\mathbf{v}_2(1) + \sum_{j=2}^N \overline{\mathbf{v}_1(j)}\mathbf{v}_2(j) = 0$ and $\overline{\mathbf{v}}_1(1) = 1$, $\mathbf{v}_2(1)$ is uniquely solved. We have

$$T_2^{-1}(z_2^{(1)}, \dots, z_N^{(1)}, z_3^{(2)}, \dots, z_N^{(2)}) = ((1, z_2^{(1)}, \dots, z_N^{(1)})^\perp, (-(\overline{z_2^{(1)}} + \sum_{j=3}^N \overline{z_j^{(1)}} z_j^{(2)}), 1, z_3^{(2)}, \dots, z_N^{(2)})^\perp). \tag{3.13}$$

To show T_k is injective for $k = 3, \dots, N$, given that T_{k-1} is injective, it suffices to show that given $(\mathbf{v}_1, \dots, \mathbf{v}_k)$ in the domain of T_k, we can recover $\mathbf{v}_k(1), \dots, \mathbf{v}_k(k-1)$ from the other components of $\mathbf{v}_k$ and the components of $\mathbf{v}_1, \dots, \mathbf{v}_{k-1}$. By the orthogonality, the $k-1$ components satisfied $k-1$ linear equations

$$\begin{cases} \overline{\mathbf{v}_1(1)}\mathbf{v}_k(1) + \cdots + \overline{\mathbf{v}_1(k-1)}\mathbf{v}_k(k-1) = -\sum_{j=k}^N \overline{\mathbf{v}_1(j)}\mathbf{v}_k(j), \\ \vdots \\ \overline{\mathbf{v}_{k-1}(1)}\mathbf{v}_k(1) + \cdots + \overline{\mathbf{v}_{k-1}(k-1)}\mathbf{v}_k(k-1) = -\sum_{j=k}^N \overline{\mathbf{v}_{k-1}(j)}\mathbf{v}_k(j). \end{cases} \tag{3.14}$$

By the condition that $\det(\mathbf{v}_i(j))_{i,j=1}^{k-1} \neq 0$, we find that $\mathbf{v}_k(1), \dots, \mathbf{v}_k(k-1)$ are uniquely determined. By induction, we find that T_N is injective. Although we are not going to write the explicit formula of T_N^{-1}, inductively we find that it is smooth like T_2^{-1} in (3.13).

On the other hand, we set up the necessary and sufficient conditions on $(z_2^{(1)}, \dots, z_N^{(1)}, \dots, z_3^{(2)}, \dots, z_N^{(2)}, \dots, z_N^{(N-1)})$ that it is in the range of T_N.

(1) All of them are nonzero.

(2) $\mathbf{v}_2(1)$ as shown in (3.13) is nonzero.

(k) (where k runs from 3 to N repeatedly)

- (k1) $\mathbf{v}_k(1), \dots, \mathbf{v}_k(k-1)$ solved by (3.14) are nonzero.
- (k2) All $\det(\mathbf{v}_i(j))_{i\in I, j\in J}$ are nonzero, for $I, J \subseteq 1, \dots, k$ and $|I| = |J|$, where $\mathbf{v}_i(j) = z_j^{(i)}$ if $j < i$; $\mathbf{v}_i(j) = 1$ if $j < i$; $\mathbf{v}_i(j)$ is solved in step (2) or (i1) if $j < i$.

The points that do not satisfy all of the conditions are in the union of zero sets of polynomials in $\Re z_j^{(i)}$ and $\Im z_j^{(i)}$, so they are of Lebesgue measure zero. □

The following lemma shows that most Hermitian matrices are very good.

Lemma 6. *The $N \times N$ Hermitian matrices that are not very good have Lebesgue measure zero.*

Proof. For $k = 2, \dots, N$, we consider the Hermitian matrices H such that in at least one of its decomposition $H = UDU^*$ as in (3.4), U has a zero k-minor, i.e., there is a pair of $I, J \subseteq \{1, \dots, N\}$ with $|I| = |J| = k$ and the minor $U_{IJ} = \det(u_{ij})_{i \in I, j \in J}$ of U is zero. If we can show for each k the set of such Hermitian matrices have Lebesgue measure zero, the lemma is proven. We note that the $k = 1$ case of the result is exactly Lemma 4.

Below we prove the $k = 2$ case. For any $N \times N$ matrix X, we define the $N(N-1)/2 \times N(N-1)/2$ matrix

$$\wedge^2 X := (X_{IJ})_{\substack{I,J \subseteq \{1,\dots,N\} \\ |I|=|J|=2}}, \quad \text{where} \quad X_{IJ} = \begin{vmatrix} x_{i_1,j_1} & x_{i_1,j_2} \\ x_{i_2,j_1} & x_{i_2,j_2} \end{vmatrix}. \tag{3.15}$$

Although the order of $I = (i_1, i_2)$ is not important, we may take the lexicographic order.

It is obvious that $\wedge^2 I_N = I_{N(N-1)/2}$. It is also not difficult to see that for any X,

$$\wedge^2 X^* = (\wedge^2 X)^*. \tag{3.16}$$

So if H is Hermitian, then $\wedge^2 H$ is also Hermitian, and if $U \in \mathrm{U}(N)$, then $\wedge^2 U \in \mathrm{U}(N(N-1)/2)$. A not so easy result is that

$$\wedge^2 (XY) = (\wedge^2 X)(\wedge^2 Y). \tag{3.17}$$

To see it, we note that

$$\begin{aligned}
(XY)_{(i_1,i_2)(j_1,j_2)} &= \begin{vmatrix} (XY)_{i_1,j_1} & (XY)_{i_1,j_2} \\ (XY)_{i_2,j_1} & (XY)_{i_2,j_2} \end{vmatrix} \\
&= \begin{vmatrix} \sum_{k=1}^N x_{i_1,k} y_{k,j_1} & \sum_{l=1}^N x_{i_1,l} y_{l,j_2} \\ \sum_{k=1}^N x_{i_2,k} y_{k,j_1} & \sum_{l=1}^N x_{i_2,l} y_{l,j_2} \end{vmatrix} \\
&= \sum_{k=1}^N \sum_{l=1}^N \begin{vmatrix} x_{i_1,k} y_{k,j_1} & x_{i_1,l} y_{l,j_2} \\ x_{i_2,k} y_{k,j_1} & x_{i_2,l} y_{l,j_2} \end{vmatrix} \\
&= \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^N \begin{vmatrix} x_{i_1,k} y_{k,j_1} + x_{i_1,l} y_{l,j_1} & x_{i_1,k} y_{k,j_2} + x_{i_1,l} y_{l,j_2} \\ x_{i_2,k} y_{k,j_1} + x_{i_2,l} y_{l,j_1} & x_{i_2,k} y_{k,j_2} + x_{i_2,l} y_{l,j_2} \end{vmatrix} \\
&= \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^N \begin{vmatrix} x_{i_1,k} & x_{i_1,l} \\ x_{i_2,k} & x_{i_2,l} \end{vmatrix} \begin{vmatrix} y_{k,j_1} & y_{k,j_2} \\ y_{l,j_1} & y_{l,j_2} \end{vmatrix} \\
&= \sum_{1 \le k < l \le N} X_{(i_1,i_2)(k,l)} Y_{(k,l),(j_1,j_2)}.
\end{aligned} \tag{3.18}$$

Hence we have for a Hermitian matrix $H = UDU^*$ as in (3.4),

$$\wedge^2 H = (\wedge^2 U)(\wedge^2 D)(\wedge^2 U)^* \tag{3.19}$$

is also a decomposition in the form of (3.4). Note that U has vanishing 2-minor is equivalent to that $\wedge^2 U$ has vanishing entries. As shown in the proof of Lemma (4), $\wedge^2 U$ has zero entries only if the enties of $\wedge^2 H$ satisfies a polynomial equation, which implies that the entries of H satisfies a polynomial equation. Thus by Lemma (2), such H have Lebesgue measure zero.

The $k = 3, \dots, N$ cases are left for exercise. The method is similar to the $k = 2$ case, but we need the Cauchy-Binet formula [16, Section 0.8.7] in place of (3.18). Thus we prove the lemma. □

Now we are ready to prove Theorem 6. Since Hermitian matrices that are not very good has Lebesgue measure zero, they have zero probability as the density function is given by (3.5). Thus we need only to consider the very good Hermitian matrices such that they have the unique decomposition (3.9) and the unitary matrix U is determined uniquely and smoothly as $U = T^{-1}(\mathbf{z})$, where $\mathbf{z} = (z_2^{(1)}, \ldots, z_N^{(1)}, \ldots, z_3^{(2)}, \ldots, z_N^{(2)}, \ldots, z_N^{(N-1)})$. To parametrise U in real parameters, we define $\theta_1, \ldots, \theta_{N(N-1)}$ as $\theta_{2k-1} = \Re(\text{the } k\text{-th component of } \mathbf{z})$ and $\theta_{2k} = \Im(\text{the } k\text{-th component of } \mathbf{z})$. Then we write $U = T^{-1}(\theta_1, \ldots, \theta_{N(N-1)})$.

Proof of Theorem 6. For each θ_k where $k = 1, \ldots, N(N-1)$, we consider $\frac{\partial U}{\partial \theta_k}$. Since $U^*U = I$, we have

$$\frac{\partial U^*}{\partial \theta_k} U + U^* \frac{\partial U}{\partial \theta_k} = 0. \tag{3.20}$$

Denote

$$S^{(k)} := U^* \frac{\partial U}{\partial \theta_k} = -\frac{\partial U^*}{\partial \theta_k} U, \tag{3.21}$$

we have that $S^{(k)}$ is anti-Hermitian, i.e., $(S^{(k)})^* = -S^{(k)}$. By differentiation on the identity $H = U \operatorname{diag}(\lambda_1, \ldots, \lambda_N) U^*$, we have

$$\frac{\partial H}{\partial \theta_k} = \frac{\partial U}{\partial \theta_k} \operatorname{diag}(\lambda_1, \ldots, \lambda_N) U^* + U \operatorname{diag}(\lambda_1, \ldots, \lambda_N) \frac{\partial U^*}{\partial \theta_k}, \tag{3.22}$$

and then

$$\begin{aligned} U^* \frac{\partial H}{\partial \theta_k} U &= U^* \frac{\partial U}{\partial \theta_k} \operatorname{diag}(\lambda_1, \ldots, \lambda_N) + \operatorname{diag}(\lambda_1, \ldots, \lambda_N) \frac{\partial U^*}{\partial \theta_k} U \\ &= S^{(k)} \operatorname{diag}(\lambda_1, \ldots, \lambda_N) - \operatorname{diag}(\lambda_1, \ldots, \lambda_N) S^{(k)}. \end{aligned} \tag{3.23}$$

In terms of entries,

$$\sum_{m,n=1}^{N} \left(\frac{\partial H}{\partial \theta_k} \right)_{mn} \bar{u}_{mj} u_{nl} = \left(U^* \frac{\partial H}{\partial \theta_k} U \right)_{jl} = (\lambda_l - \lambda_j) S_{jl}^{(k)}. \tag{3.24}$$

Similarly, for any $k = 1, \ldots, N$,

$$\sum_{m,n=1}^{N} \left(\frac{\partial H}{\partial \lambda_k} \right)_{mn} \bar{u}_{mj} u_{nl} = \left(U^* \frac{\partial H}{\partial \lambda_k} U \right)_{jl} = \delta_{jk} \delta_{lk}. \tag{3.25}$$

Note that we use $h_{ii}, \Re h_{jk}, \Im h_{jk}$ as coordinates of H, and

$$\left(\frac{\partial H}{\partial \mu} \right)_{mn} = \begin{cases} \frac{\partial \Re h_{mn}}{\partial \mu} + i \frac{\partial \Im h_{mn}}{\partial \mu} & \text{if } m < n, \\ \frac{\partial h_{mm}}{\partial \mu} & \text{if } m = n, \\ \frac{\partial \Re h_{nm}}{\partial \mu} - i \frac{\partial \Im h_{nm}}{\partial \mu} & \text{if } m > n, \end{cases} \quad \text{where } \mu \text{ stands for } \theta_k \text{ or } \lambda_k. \tag{3.26}$$

From (3.26), we see that if we arrange $\frac{\partial h_{ii}}{\partial \mu}, \frac{\partial \Re h_{jl}}{\partial \mu}, \frac{\partial \Im h_{jl}}{\partial \mu}$ as a vector and $\frac{\partial h_{mn}}{\partial \mu}$ as another vector, both of dimension N^2, there is a linear transformation, expressed by an $N^2 \times N^2$ matrix V_1, such that

$$V_1 \begin{pmatrix} \frac{\partial h_{ii}}{\partial \mu} \\ \frac{\partial \Re h_{jl}}{\partial \mu} \\ \frac{\partial \Im h_{jl}}{\partial \mu} \end{pmatrix} = \left(\frac{\partial h_{mn}}{\partial \mu} \right). \tag{3.27}$$

Although we do not write down V_1 explicitly, it is clear that $\det(V_1) \neq 0$ is a constant. Similarly, from (3.24) and (3.25), we find a linear transformation, expressed by an $N^2 \times N^2$ matrix V_2, such that

$$V_2\left(\frac{\partial h_{mn}}{\partial \theta_k}\right) = \left((\lambda_l - \lambda_j)S_{jl}^{(k)}\right), \quad V_2\left(\frac{\partial h_{mn}}{\partial \lambda_k}\right) = \left(\delta_{jk}\delta_{lk}\right). \tag{3.28}$$

Note that V_2 depends on u_{ij} and hence depends on θ_k but not λ_k. Also V_2 is unitary (exercise) and so $\det(V_2) \neq 0$.

Recall the matrix $\mathbf{H}$ in (3.7). We have

$$V_2 V_2 \mathbf{H} = \left(\delta_{jk}\delta_{lk} \quad (\lambda_l - \lambda_j)S_{jl}^{(k)}\right), \tag{3.29}$$

where then index k runs to the right and the index (jl) runs to the bottom. Note that each factor $(\lambda_l - \lambda_j)$ appears in one row on the right-hand side of (3.29), (3.29) implies

$$\det(V_2)\det(V_1)\det(\mathbf{H}) = \prod_{1\le j<l\le N}(\lambda_l - \lambda_j)^2 \det(\mathbf{S}), \tag{3.30}$$

where entries of $\mathbf{S}$ are given by 0, 1 and $S_{jl}^{(k)}$. Note that $S_{jl}^{(k)}$ are expressed in θ_k, and so is $\det \mathbf{S}$.

Therefore

$$\det(\mathbf{H}) = \prod_{1\le j<l\le N}(\lambda_l - \lambda_j)^2 f(\theta_1, \dots, \theta_{N(N-1)}), \tag{3.31}$$

and by (3.5), (3.6) and (3.7), we have

$$P(\lambda_1, \dots, \lambda_N, \theta_1, \dots, \theta_{N(N-1)}) = 2^{-\frac{N}{2}}\left(\frac{N}{\pi}\right)^{\frac{N^2}{2}} \prod_{i=1}^{N} e^{-\frac{N}{2}\lambda_i^2} \prod_{1\le j<l\le N}(\lambda_l - \lambda_j)^2 f(\theta_1, \dots, \theta_{N(N-1)}). \tag{3.32}$$

After intergration with respect to $\theta_1, \dots, \theta_{N(N-1)}$, we derive (3.2), and then (3.1) is equivalent. □

3.2 Cumulative probability distribution of the largest eigenvalue in GUE

We have derived the joint probability density function of the eigenvalues in the $N \times N$ GUE in Theorem 6. The next question is: What is the distribution of the largest eigenvalue, especially as $N \to \infty$?

The semicircle law suggests that as $N \to \infty$, the largest eigenvalue is at $\sqrt{2}$. But it is only a speculation. Actually it is true for all Wigner matrices considered in Section 2. We prove it for GUE and furthermore derive the limiting distribution of the largest eigenvalue after scaling.

Write

$$P(\lambda_1, \dots, \lambda_N) = \frac{1}{C_N}\begin{vmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_N \\ \vdots & & \vdots \\ \lambda_1^{N-1} & \dots & \lambda_N^{N-1} \end{vmatrix}^2 \prod_{i=1}^{N} e^{\frac{N}{2}\lambda_i^2}. \tag{3.33}$$

It is clear that if $p_j(x)$ and $q_j(x)$ are polynomials of degree j and the leading coefficients nonzero, then

$$P(\lambda_1, \dots, \lambda_N) = \frac{1}{C'_N}\begin{vmatrix} p_0(\lambda_1) & \dots & p_0(\lambda_N) \\ p_1(\lambda_1) & \dots & p_1(\lambda_N) \\ \vdots & & \vdots \\ p_{N-1}(\lambda_1) & \dots & p_{N-1}(\lambda_N^{N-1}) \end{vmatrix}\begin{vmatrix} q_0(\lambda_1) & \dots & q_0(\lambda_N) \\ q_1(\lambda_1) & \dots & q_1(\lambda_N) \\ \vdots & & \vdots \\ q_{N-1}(\lambda_1) & \dots & q_{N-1}(\lambda_N^{N-1}) \end{vmatrix} \prod_{i=1}^{N} e^{\frac{N}{2}\lambda_i^2}. \tag{3.34}$$

The following identity was discovered by Andréief in 1886 [4]:

Theorem 7. *Let $f_1(x), \dots, f_n(x)$ and $g_1(x), \dots, g_n(x)$ be L^2 functions on $\mathbb{R}$, then*

$$\int_{\mathbb{R}^n} \det(f_i(x_j))_{i,j=1}^n \det(g_i(x_j))_{i,j=1}^n dx_1 \cdots dx_n = n! \det \left(\int_{\mathbb{R}} f_i(x) g_j(x) dx \right)_{i,j=1}^n . \tag{3.35}$$

Proof. By the Leibniz formula for determinants,

$$\begin{aligned}
&\int_{\mathbb{R}^n} \det(f_i(x_j))_{i,j=1}^n \det(g_i(x_j))_{i,j=1}^n dx_1 \cdots dx_n \\
&= \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} \int_{\mathbb{R}^n} \begin{vmatrix} f_1(x_1) & \dots & f_1(x_n) \\ \vdots & & \vdots \\ f_n(x_1) & \dots & f_n(x_n) \end{vmatrix} \prod_{j=1}^n g_{\sigma(j)}(x_j) dx_1 \cdots dx_n \\
&= \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} \int_{\mathbb{R}^n} \begin{vmatrix} f_1(x_1) g_{\sigma(1)}(x_1) & \dots & f_1(x_n) g_{\sigma(n)}(x_n) \\ \vdots & & \vdots \\ f_n(x_1) g_{\sigma(1)}(x_1) & \dots & f_n(x_n) g_{\sigma(n)}(x_n) \end{vmatrix} dx_1 \cdots dx_n \\
&= \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} \begin{vmatrix} \int_{\mathbb{R}} f_1(x_1) g_{\sigma(1)}(x_1) dx_1 & \dots & \int_{\mathbb{R}} f_1(x_n) g_{\sigma(n)}(x_n) dx_n \\ \vdots & & \vdots \\ \int_{\mathbb{R}} f_n(x_1) g_{\sigma(1)}(x_1) dx_1 & \dots & \int_{\mathbb{R}} f_n(x_n) g_{\sigma(n)}(x_n) dx_n \end{vmatrix} \\
&= \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} \det \left(\int_{\mathbb{R}} f_i(x) g_{\sigma(j)}(x) dx \right)_{i,j=1}^n \\
&= \sum_{\sigma \in S_n} (-1)^{\mathrm{sgn}(\sigma)} (-1)^{\mathrm{sgn}(\sigma)} \det \left(\int_{\mathbb{R}} f_i(x) g_j(x) dx \right)_{i,j=1}^n \\
&= n! \sum_{\sigma \in S_n} \det \left(\int_{\mathbb{R}} f_i(x) g_j(x) dx \right)_{i,j=1}^n .
\end{aligned} \tag{3.36}$$

□

We want to consider the probability $\mathbb{P}(\lambda_{\max} = \max_{1 \le j \le N}(\lambda_j) < a)$ for any $a \in \mathbb{R}$. It is equal to

$$\begin{aligned}
&\mathbb{P}(\lambda_1 < a, \dots, \lambda_N < a) \\
&= \frac{1}{C'_N} \int_{\mathbb{R}^N} \det(p_{i-1}(x_j) e^{-\frac{N}{4} x_j^2} \mathbf{1}_{(-\infty,a)}(x_j))_{i,j=1}^N \det(q_{i-1}(x_j) e^{-\frac{N}{4} x_j^2} \mathbf{1}_{(-\infty,a)}(x_j))_{i,j=1}^N dx_1 \cdots dx_N \\
&= \frac{N!}{C'_N} \det \left(\int_{\mathbb{R}} p_{i-1}(x) q_{j-1}(x) e^{-\frac{N}{2} x^2} \mathbf{1}_{(-\infty,a)}(x) dx \right)_{i,j=1}^N .
\end{aligned} \tag{3.37}$$

To eliminate the annoying constant C'_N, we note that let $a = +\infty$, (3.37) becomes

$$1 = \frac{N!}{C'_N} \det \left(\int_{\mathbb{R}} p_{i-1}(x) q_{j-1}(x) e^{-\frac{N}{2} x^2} dx \right)_{i,j=1}^N \tag{3.38}$$

and then

$$\mathbb{P}(\lambda_{\max} < a) = \frac{\det \left(\int_{\mathbb{R}} p_{i-1}(x) q_{j-1}(x) e^{-\frac{N}{2} x^2} \mathbf{1}_{(-\infty,a)}(x) dx \right)_{i,j=1}^N}{\det \left(\int_{\mathbb{R}} p_{i-1}(x) q_{j-1}(x) e^{-\frac{N}{2} x^2} dx \right)_{i,j=1}^N}. \tag{3.39}$$

To simplify (3.39), we assume $q_j(x) = p_j(x)$, and let them be the *orthogonal* polynomials such that

$$\int_{\mathbb{R}} p_i(x)p_j(x)e^{-\frac{N}{2}x^2}dx = \delta_{ij}. \tag{3.40}$$

Then (3.39) becomes

$$\mathbb{P}(\lambda_{\max} < a) = \det\left(\int_{\mathbb{R}} p_{i-1}(x)q_{j-1}(x)e^{-\frac{N}{2}x^2}\mathbf{1}_{(-\infty,a)}(x)dx\right)_{i,j=1}^{N}. \tag{3.41}$$

Below we will see that the orthogonality can bring us more advantages.

Now let $A : L^2(\mathbb{R}) \mapsto \ell^2(N)$ and $B : \ell^2(N) \mapsto L^2(\mathbb{R})$ be the operators defined by the kernels

$$A(s,i) = p_{i-1}(s)e^{-\frac{N}{4}s^2}, \quad B(j,s) = q_{j-1}(s)e^{-\frac{N}{4}s^2}, \quad i,j = 1,\dots,N, \tag{3.42}$$

such that for any $f(s) \in L^2(\mathbb{R})$, $A(f) \in \ell^2(N)$ is

$$A(f) = \left(\int_{\mathbb{R}} A(s,1)f(s)ds, \dots, \int_{\mathbb{R}} A(s,N)f(s)ds\right)^{\perp} \tag{3.43}$$

and for any $\mathbf{k} = (k_1,\dots,k_N)^{\perp} \in \ell^2(N)$, $B(\mathbf{k}) \in L^2(\mathbb{R})$ is

$$B(\mathbf{k}) = \sum_{j=1}^{N} k_j B(j,s). \tag{3.44}$$

Then AB is a linear transformation from $\ell^2(N)$ to itself, represented by the $N \times N$ matrix

$$AB(i,j) = \int_{\mathbb{R}} A(s,i)B(j,s)ds, \tag{3.45}$$

and BA is a linear transformation from $L^2(\mathbb{R})$ to itself, represented by the kernel

$$BA(s,t) = \sum_{j=1}^{N} B_j(s)A_j(t). \tag{3.46}$$

Similarly, we define $A_{\pm} : L^2(\mathbb{R}) \mapsto \ell^2(N)$ and $B_{\pm} : \ell^2(N) \mapsto L^2(\mathbb{R})$ by

$$\begin{aligned} A_-(s,i) &= A(s,i)\mathbf{1}_{(-\infty,a)}(s), \quad A_+(s,i) = A(s,i)\mathbf{1}_{(a,\infty)}(s), \\ B_-(j,s) &= B(j,s)\mathbf{1}_{(-\infty,a)}(s), \quad B_+(j,s) = B(j,s)\mathbf{1}_{(a,\infty)}(s). \end{aligned} \tag{3.47}$$

Then $A_-B_- : \ell^2(N) \mapsto \ell^2(N)$ and $B_-A_- : L^2(\mathbb{R}) \mapsto L^2(\mathbb{R})$ are represented by the matrices

$$A_-B_-(i,j) = \int_{\mathbb{R}} A(s,i)B(j,s)\mathbf{1}_{(-\infty,a)}(s)ds \tag{3.48}$$

and the kernel

$$B_-A_-(s,t) = \mathbf{1}_{(-\infty,a)}(s)\left(\sum_{j=1}^{N} B_j(s)A_j(t)\right)\mathbf{1}_{(-\infty,a)}(t), \tag{3.49}$$

and A_+B_+ and B_+A_+ are represented anlogously.

Noting that

$$A_-B_- + A_+B_+ = AB = I, \tag{3.50}$$

the probability $\mathbb{P}(\lambda_{\max} < a)$ becomes

$$\mathbb{P}(\lambda_{\max} < a) = \det(A_- B_-) = \det(I - A_+ B_+). \tag{3.51}$$

A heuristic idea to simplify (3.51) is to use the identity $\det(A_- B_-) = \det(B_- A_-)$ and transform the determiant of a large matrix into the determinant of an integral operator. This approach does not work, since $\det(XY) = \det(YX)$ holds only if X and Y are square matrices, which is not true in our case. (A and B can be regarded as $N \times \infty$ and $\infty \times N$ matrices respectively.) Nevertheless, a similar identity that does not require squareness saves us:

Theorem 8. *Let X and Y be $m \times n$ and $n \times m$ matrices respectively. Then*

$$\det(I_{m\times m} + XY) = \det(I_{n\times n} + YX). \tag{3.52}$$

Proof. Exercise. Hint: Cauchy-Binet formula. □

The theorem above serves as an inspiration for us. The real technical tool is

Theorem 9. *Let X be a linear transformation from $L^2(\mathbb{R})$ to $\ell^2(n)$ and Y be a linear transformation from $\ell^2(n)$ to $L^2(\mathbb{R})$. Then*

$$\det(I + XY) = \det(1 + YX), \tag{3.53}$$

where on the left-hand side the determinant is the usual one for $n \times n$ matrices, and on the right-hand side the determinant os the Fredholm determinant *for trace class operators.*

Before giving the proof of the theorem, we need to explain what a Fredholm determinant is. The abstract definition for general trace class operators is given in the appendix, and a special definition below suffices for us.

Definition 2. Let K be an integral operator on $\mathbb{R}$ either with a piecewise continuous and rapidly decreasing kernel or of finte rank. Then the Fredholm determinant of $1 + K$ is

$$\det(1 + K) = 1 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^n} \det(K(x_i, x_j))_{i,j=1}^n dx_1 \cdots dx_n. \tag{3.54}$$

The convergence of the series in (3.54) can be verified directly when $K(x, y)$ is rapidly decreasing, i.e., as $\max(|x|, |y|) \to \infty$, $|K(x, y)| \to 0$ faster than any power function. It will be shown that if K is of rank $m < \infty$, i.e., $K = \sum_{i=1}^m f_i(x) g_i(y)$, then the series terminates after m terms.

Proof of Theorem 9. Expressing X and Y like we did for A and B in (3.42), we have

$$\begin{aligned}
\det(1 + YX) &= \det\left(1 + \sum_{l=1}^n Y(l, s) X(t, l)\right) \\
&= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathbb{R}^k} \det\left(\sum_{l=1}^n Y(l, x_i) X(x_j, l)\right)_{i,j=1}^k dx_1 \cdots dx_k \\
&= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{l_1,\ldots,l_k=1}^n \int_{\mathbb{R}^k} \det(Y(l_j, x_i) X(x_j, l_j))_{i,j=1}^k dx_1 \cdots dx_k \qquad (3.55) \\
&= 1 + \sum_{k=1}^{n} \frac{1}{k!} \sum_{l_1,\ldots,l_k=1}^n \int_{\mathbb{R}^k} \det(Y(l_j, x_i) X(x_j, l_j))_{i,j=1}^k dx_1 \cdots dx_k \\
&= 1 + \sum_{k=1}^{n} \frac{1}{k!} \sum_{l_1,\ldots,l_k=1}^n \det\left(\int_{\mathbb{R}^k} Y(l_i, x) X(x, l_j) dx\right)_{i,j=1}^k.
\end{aligned}$$

Note that in the last step we use the fact that the series terminates when $k > n$. The reason is that if $l_1, \dots, l_k$ are not distinct, then the integral of $\det(Y(l_j, x_i)X(x_j, l_j))_{i,j=1}^k$ vanishes due to identical rows.

On the other hand, we have

$$\det(I + XY) = \det \begin{pmatrix} 1 + \int_{\mathbb{R}} X(s,1)Y(1,s)ds & \int_{\mathbb{R}} X(s,1)Y(2,s)ds & \dots & \int_{\mathbb{R}} X(s,1)Y(n,s)ds \\ \int_{\mathbb{R}} X(s,2)Y(1,s)ds & 1 + \int_{\mathbb{R}} X(s,2)Y(2,s)ds & \dots & \int_{\mathbb{R}} X(s,2)Y(n,s)ds \\ \vdots & \vdots & & \vdots \\ \int_{\mathbb{R}} X(s,n)Y(1,s)ds & \int_{\mathbb{R}} X(s,n)Y(2,s)ds & \dots & 1 + \int_{\mathbb{R}} X(s,n)Y(n,s)ds \end{pmatrix}. \tag{3.56}$$

To compare this determinant with the Fredholm determinant expressed in (3.55), we use the following expansion of determinant. Let $M = (m_{ij})_{i,j=1}^n$ be an $n \times n$ matrix and t be a variable, then

$$\det(I + tM) = \begin{vmatrix} 1 + tm_{11} & tm_{12} & \dots & tm_{1n} \\ tm_{21} & 1 + tm_{22} & \dots & tm_{2n} \\ \vdots & \vdots & & \vdots \\ tm_{n1} & 1 + tm_{n2} & \dots & 1 + tm_{nn} \end{vmatrix} \tag{3.57}$$

is a degree n polynomial in t. Computing its coefficient of each t^k $(k = 0, \dots, n)$, we have

$$\begin{aligned} \det(I + tM) &= 1 + \sum_{k=1}^n t^k \sum_{1 \le l_1 < \dots < l_k \le n} \det(m_{l_i, l_j})_{i,j=1}^k \\ &= 1 + \sum_{k=1}^n t^k \frac{1}{k!} \sum_{l_1, \dots, l_k = 1}^n \det(m_{l_i, l_j})_{i,j=1}^k. \end{aligned} \tag{3.58}$$

Let $t = 1$ and $M = XY$, we have

$$\det(I + XY) = 1 + \sum_{k=1}^n \frac{1}{k!} \sum_{l_1, \dots, l_k = 1}^n \det \left(\int_{\mathbb{R}} X(x, l_i) Y(l_j, x) dx \right)_{i,j=1}^k. \tag{3.59}$$

Comparing (3.59) with (3.55), we prove the theorem. □

Now applying Theorem 9 to (3.51), we have (see (3.42), (3.47) and (3.49) and note that $p_j = q_j$)

$$\mathbb{P}(\lambda_{\max} < a) = \det(1 - B_+ A_+) = \det \left(1 - \mathbf{1}_{(a,\infty)}(s) \left(\sum_{j=1}^N p_j(s) p_j(t) e^{-\frac{N}{4}(s^2+t^2)} \right) \mathbf{1}_{(a,\infty)}(t) \right). \tag{3.60}$$

To further simplify (3.60), we use a property of orthogonal polynomials.

Theorem 10. *Let $p_n(x)$ be orthogonal polynomials with respect to a measure μ such that*

$$\int_{\mathbb{R}} p_i(x) p_j(x) d\mu(x) = \delta_{ij}. \tag{3.61}$$

Denote the constants

$$a_n := \int_{\mathbb{R}} x p_n^2(x) d\mu(x), \quad b_n := \int_{\mathbb{R}} x p_n(x) p_{n+1}(x) d\mu(x). \tag{3.62}$$

Then

$$xp_n(x) = \begin{cases} b_n p_{n+1}(x) + a_n p_n(x) + b_{n-1} p_{n-1}(x) & \text{if } n \geq 1, \\ b_0 p_1(x) + a_0 p_0(x) & \text{if } n = 0. \end{cases} \tag{3.63}$$

Proof. Without loss of generality, we assume $n > 0$ and

$$xp_n(x) = c_0 p_0(x) + c_1 p_1(x) + \cdots + c_{n+1} p_{n+1}(x). \tag{3.64}$$

Then for any $j \leq n+1$,

$$\int_{\mathbb{R}} xp_n(x) p_j(x) d\mu(x) = \sum_{i=0}^{n+1} c_i \int_{\mathbb{R}} p_i(x) p_j(x) d\mu(x) = c_j. \tag{3.65}$$

For $j < n-1$, we have

$$c_j = \int_{\mathbb{R}} p_n(x)(xp_j(x)) d\mu(x) = 0, \tag{3.66}$$

since $xp_j(x)$ is a polynomial of degree $j+1 < n$, and $p_n(x)$ is orthogonal to all polynomials of degree less than n. Similarly we find $c_{n-1} = b_{n-1}$, $c_n = a_n$ and $c_{n+1} = b_n$. □

Then we have the celebrated Christoffel-Darboux formula.

Theorem 11. *Let $p_n(x)$, a_n, b_n be defined as in Theorem 10. Then*

$$\sum_{i=0}^{n-1} p_i(x) p_i(y) = b_{n-1} \frac{p_n(x) p_{n-1}(y) - p_{n-1}(x) p_n(y)}{x - y}. \tag{3.67}$$

Proof. Take the difference between the two formulas

$$\begin{aligned} x \sum_{i=0}^{n-1} p_i(x) p_i(y) = & \\ & a_0 p_0(x) p_0(y) + b_0 p_0(x) p_1(y) \\ & + b_0 p_0(x) p_1(y) + a_1 p_1(x) p_1(y) + b_1 p_1(x) p_2(y) \\ & + b_0 p_0(x) p_1(y) + a_1 p_1(x) p_1(y) + \ddots \\ & \ddots \quad \ddots \\ & + a_{n-1} p_{n-1}(x) p_{n-1}(y) \\ & + b_{n-1} p_n(x) p_{n-1}(y) \end{aligned} , \tag{3.68}$$

$$\begin{aligned} y \sum_{i=0}^{n-1} p_i(x) p_i(y) = & \\ a_0 p_0(x) p_0(y) & + b_0 p_0(x) p_1(y) \\ + b_0 p_0(x) p_1(y) & + a_1 p_1(x) p_1(y) + b_1 p_1(x) p_2(y) \\ & + b_0 p_0(x) p_1(y) + a_1 p_1(x) p_1(y) + \ddots \\ & \ddots \quad \ddots \\ & + a_{n-1} p_{n-1}(x) p_{n-1}(y) + b_{n-1} p_n(x) p_{n-1}(y) \end{aligned} , \tag{3.69}$$

and divide both sides by $x - y$. □

The orthogonal polynomial associated to GUE turns out to be quite special: They are Hermite polynomials up to scaling. The standard reference book of Hermite polynomial is [23], while Wikipedia is also a reliable source. Hermite polynomials (as used by probabilists) are defined by the formula

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}. \tag{3.70}$$

$H_n(x)$ is a monic polynomial of degree n, and satisfies the orthogonality

$$\int_{\mathbb{R}} H_m(x) H_n(x) e^{-\frac{x^2}{2}} dx = \sqrt{2\pi} n! \delta_{mn}. \tag{3.71}$$

We also have (for $n \geq 1$)

$$xH_n(x) = H_{n+1}(x) + nH_{n-1}(x). \tag{3.72}$$

It is straightforward to verify that if we define

$$p_n(x) := \frac{N^{\frac{1}{4}}}{(2\pi)^{\frac{1}{4}}} (n!)^{-\frac{1}{2}} H_n(\sqrt{N}x), \tag{3.73}$$

$p_n(x)$ satisfies the orthogonality condition (3.40), and

$$xp_n(x) = \sqrt{\frac{n+1}{N}} p_{n+1}(x) + \sqrt{\frac{n}{N}} p_{n-1}(x). \tag{3.74}$$

The conclusion of this subsection is:

Theorem 12. *In the $N \times N$ GUE random matrix, the distribution of the largest eigenvalue is given by the cumulative distribution function*

$$\mathbb{P}(\lambda_{\max} < a) = \det(1 - \mathbf{1}_{(a,\infty)}(x) K_N(x,y) \mathbf{1}_{(a,\infty)}(y)), \tag{3.75}$$

where

$$K_N(x,y) = \frac{p_N(x) e^{-\frac{N}{4}x^2} p_{N-1}(y) e^{-\frac{N}{4}y^2} - p_{N-1}(x) e^{-\frac{N}{4}x^2} p_N(y) e^{-\frac{N}{4}y^2}}{x-y}, \tag{3.76}$$

and $p_n(x)$ is expressed in Hermite polynomials by (3.73) *and* (3.70).

3.3 Limiting distribution of the largest eigenvalue in GUE

In this subsection we do asymptotic analysis, and assume N, the dimension of the GUE, to be sufficiently large.

Hermite polynomial $H_n(x)$ has a contour integral formula

$$H_n(x) = \frac{n!}{2\pi i} \oint \frac{e^{xt - \frac{t^2}{2}}}{t^{n+1}} dt, \tag{3.77}$$

where the contour encloses the origin.

Proof. Since $\exp(xt - t^2/2)$ is an analytic function in t, the Cauchy integral formula implies, with the help of the change of variable $u = x - t$,

$$\begin{aligned}\frac{n!}{2\pi i}\oint \frac{e^{xt-\frac{t^2}{2}}}{t^{n+1}}dt &= \frac{d^n}{dt^n}e^{xt-\frac{t^2}{2}}\bigg|_{t=0} \\ &= e^{\frac{x^2}{2}}\frac{d^n}{dt^n}e^{-\frac{1}{2}(t-x)^2}\bigg|_{t=0} \\ &= e^{\frac{x^2}{2}}(-1)^n\frac{d^n}{du^n}e^{-\frac{u^2}{2}}\bigg|_{u=x} \\ &= e^{\frac{x^2}{2}}(-1)^n\frac{d^n}{dx^n}e^{-\frac{x^2}{2}} = H_n(x).\end{aligned} \tag{3.78}$$

□

Thus our $p_N(x)$ and $p_{N-1}(x)$ have contour integral formulas, with the change of variable $z = N^{-1/2}t$,

$$p_N(x)e^{-\frac{N}{4}x^2} = \frac{C_N}{2\pi i}\oint_\Gamma \frac{e^{N(xz-\frac{z^2}{2}-\frac{x^2}{4})}}{z^{N+1}}dz, \quad p_{N-1}(x)e^{-\frac{N}{4}x^2} = \frac{C_N}{2\pi i}\oint_\Gamma \frac{e^{N(xz-\frac{z^2}{2}-\frac{x^2}{4})}}{z^{N}}dz, \tag{3.79}$$

where Γ is a contour enclosing 0 and

$$C_N = \frac{\sqrt{N!}}{N^{\frac{N}{2}-\frac{1}{2}}(2\pi)^{\frac{1}{4}}}. \tag{3.80}$$

Note that by Stirling's formula, as $N \to \infty$,

$$C_N = \sqrt{N}e^{-\frac{N}{2}}(1 + \mathcal{O}(N^{-1})). \tag{3.81}$$

Next we apply the *steepest-descent method* to estimate $p_N(x)e^{-Nx^2/2}$ and $p_{N-1}(x)e^{-Nx^2/2}$ for $x = \sqrt{2} + \mathcal{O}(N^{-2/3})$. The strategy of the steepest-descent method can be explained briefly as below: To estimate the contour integral

$$\oint_C e^{nh(z)}g(z)dz, \tag{3.82}$$

where $h(z)$ and $g(z)$ are "usual" analytic functions and n is a large parameter, we take the steps

1. Find the critical points $z_1, \ldots, z_m$ where $h'(z_i) = 0$ for $i = 1, \ldots, m$.
2. Deform the contour C such that it passes through one or more critical points and $\Re h(z)$ attains its global maximum $h_{\max}$, on those critical points, say $z_1, \ldots, z_k$ on the contour.
3. Estimate
$$\oint_{C\setminus N_\delta(z_1)\cup\cdots\cup N_\delta(z_k)} e^{nh(z)}g(z)dz \tag{3.83}$$
where $N_\delta(z_i)$ is the neighbourhood of z_i with radius δ, and show that the integral is of the order $\mathcal{O}(\exp(nh_{\max} - n\epsilon))$, where ϵ depends on δ and $h(z)$.
4. Evaluate for each $z_1, \ldots, z_k$
$$\oint_{C\cap N_\delta(z_i)} e^{nh(z)}g(z)dz \tag{3.84}$$
and find that the integral equals $C_i N^{-\alpha_i} g(z_i)\exp(nh_{\max})(1+o(1))$ where α_i and C_i depend on the local behavior of $h(z)$ at z_i.

5. Take the sum of the parts of the contour integral evaluated in steps 3 and 4 The integral over $C \setminus N_\delta(z_1) \cup \cdots \cup N_\delta(z_k)$ ie negligible since the exponential function part vanishes faster than the power function parts as $N \to \infty$.

In practice, sometimes we modify some steps.

To apply the steepest-descent method to $p_N(x)e^{-Nx^2/2}$ and $p_{N-1}(x)e^{-Nx^2/2}$ in (3.85), we write

$$p_N(x)e^{-\frac{N}{4}x^2} = \frac{C_N}{2\pi i}\oint_\Gamma e^{N(xz-\frac{z^2}{2}-\log z-\frac{x^2}{4})}\frac{1}{z}dz, \quad p_{N-1}(x)e^{-\frac{N}{4}x^2} = \frac{C_N}{2\pi i}\oint_\Gamma e^{N(xz-\frac{z^2}{2}-\log z-\frac{x^2}{4})}dz. \tag{3.85}$$

First we evaluate $p_N(x)$. We have

$$h(z) = xz - \frac{z^2}{2} - \log z - \frac{x^2}{4}, \quad g(z) = z^{-1}. \tag{3.86}$$

According to (3.75) and the suggestion of the semicircle law, we need to consider $x > a$ where a is around 2. We start by the special case $x = 2$. Hence

$$h'(z) = 2 - z - \frac{1}{z}, \tag{3.87}$$

and $h'(z)$ has a double zero $z = 1$, such that

$$h(1) = \frac{1}{2}, \quad h'(1) = h''(1) = 0, \quad h'''(2) = -2 < 0. \tag{3.88}$$

To find the correct shape of Γ, we first consider the contour Γ^{pre} that differs from Γ only locally around 1. See Figure 1.

$$\Gamma^{\text{pre}} = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3, \quad \text{where} \begin{cases} \Gamma_1 \text{ is the line segment from } i\sqrt{3} \text{ to } 1, \\ \Gamma_2 \text{ is the line segment from } 1 \text{ to } -i\sqrt{3}, \\ \Gamma_3 \text{ is the semicircle with center } 0, \text{ radius } \sqrt{3} \\ \quad \text{from } -i\sqrt{3} \text{ to } i\sqrt{3} \text{ counterclockwise.} \end{cases} \tag{3.89}$$

We show that $\Re h(z)$ attains its global maximum on Γ^{pre} at 1 by explicit computation. For $z \in \Gamma_1$, we write $z = (1-t) + i\sqrt{3}t$ where $t \in [0,1]$. Then

$$\Re h(z) = 2(1-t) - \frac{1}{2}(1 - 2t - 2t^2) - \log\sqrt{(1-t)^2 + 3t^2} - 1 \tag{3.90}$$

and for $t \in (0,1)$

$$\frac{d}{dt}\Re h(z) = \frac{8t^3 - 8t^2}{1 - 2t + 4t^2} = \frac{8t^2(t-1)}{(1-t)^2 + 3t^2} < 0. \tag{3.91}$$

Hence $\Re h(z)$ decreases as z moves from 1 to $\sqrt{3}i$ along Γ_1. Since Γ_2 is symmetric to Γ_1 about the real axis and $h(z)$ is a real analytic function, $\Re h(z)$ decreases as z moves from 1 to $-i\sqrt{3}$ along Γ_2. For $z \in \Gamma_3$, denote $z = \sqrt{3}(\cos\theta + i\sin\theta)$ where $\theta \in [\pi/2, 3\pi/2]$. Then for $z \in \Gamma_3$

$$\begin{aligned} \Re h(z) &= \frac{1}{2} - 3\cos^2\theta + 2\sqrt{3}\cos\theta - \log\sqrt{3} \\ &= \frac{1}{2} + 1 - (1 - \sqrt{3}\cos\theta)^2 - \log\sqrt{3} \\ &\le \frac{1}{2} - \log\sqrt{3} < \Re h(1). \end{aligned} \tag{3.92}$$

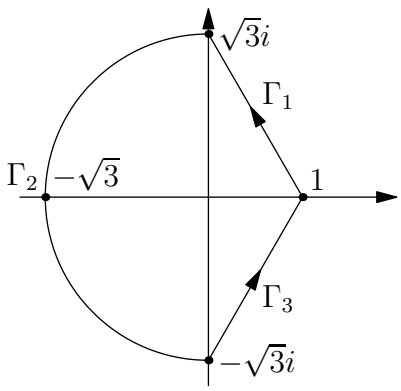

Figure 1: The shape of Γ^{pre}. It differs from Γ only locally around 1.

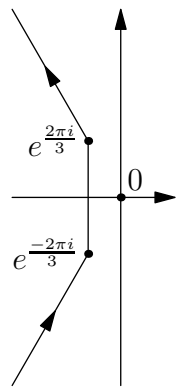

Figure 2: The shapes of $\tilde{\Gamma}$ and $\tilde{\Gamma}^{\infty}$ around 0. For $\tilde{\Gamma}^{\infty}$ the contour extents to infinity, and for $\tilde{\Gamma}$ it extends to $N^{1/30}e^{\pm 2\pi i/3}$. $\tilde{\Gamma}$ has the same shape as $\Gamma \cap N_{\delta_N}(1)$ up to scaling.

Thus Γ^{pre} satisfies the requirement of the steepest-descent contour in Step 2. Γ is defined by deforming Γ^{pre} such that the two points $1 + N^{-1/3}e^{\pm 2\pi i/3}$ are joint by a vertical line segment instead of a polygonal chain through 1. See Figure 2 for this local part of Γ maginified.

By Taylor expansion around 1,

$$h(z) = h(1) + \frac{1}{6}h'''(1)(z-1)^3 + \mathcal{O}((z-1)^4) = \frac{1}{2} - \frac{(z-1)^3}{3} + \mathcal{O}((z-1)^4). \tag{3.93}$$

Consider $\delta_N = N^{-3/10}$. The two points $1 + \delta_N e^{\pm 2\pi i/3}$ are on Γ, and

$$h(1 + \delta_N e^{\pm 2\pi i/3}) = \frac{1}{2} - \frac{1}{3}N^{-\frac{9}{10}} + \mathcal{O}(N^{-\frac{6}{5}}) \tag{3.94}$$

and then

$$\Re h(1 + \delta_N e^{\pm 2\pi i/3}) < \frac{1}{2} - \epsilon N^{-\frac{9}{10}} \tag{3.95}$$

for an $\epsilon > 0$. From the construction of Γ, we can verify that

$$\Re h(z) < \frac{1}{2} - \epsilon N^{-\frac{9}{10}} \tag{3.96}$$

for all $z \in \Gamma \setminus N_{\delta_N}(1)$. Hence

$$\begin{aligned} &\left| \frac{1}{2\pi i} \oint_{\Gamma \setminus N_{\delta_N}(1)} e^{Nh(z)} g(z) dz \right| \\ \leq &\left| \frac{1}{2\pi i} \oint_{\Gamma \setminus N_{\delta_N}(1)} e^{N(\frac{1}{2} - \epsilon N^{-\frac{9}{10}})} |g(z)||dz| \right| \\ = &\, e^{-\epsilon N^{-\frac{9}{10}}} e^{\frac{1}{2}N} \left| \frac{1}{2\pi i} \oint_{\Gamma \setminus N_{\delta_N}(1)} e^{N(\frac{1}{2} - \epsilon N^{-\frac{9}{10}})} |g(z)||dz| \right| = C e^{-\epsilon N^{-\frac{9}{10}}} e^{\frac{1}{2}N} \end{aligned} \tag{3.97}$$

where C depends on the behavior of $g(z)$ on Γ.

On the other hand, on $\Gamma \cap N_{\delta_N}(1)$, we take the change of variable $z = 1 + N^{-1/3}w$ and have

$$e^{Nh(z)} = e^{N(\frac{1}{2} - \frac{1}{3}(z-1)^3 + \mathcal{O}((z-1)^4))} = e^{\frac{N}{2}} e^{-\frac{w^3}{3}} e^{\mathcal{O}(N^{-\frac{1}{5}})} = e^{\frac{N}{2}} e^{-\frac{w^3}{3}} (1 + \mathcal{O}(N^{-\frac{1}{5}})), \tag{3.98}$$

where we use the estimate that for $z \in N_{\delta_N}(1)$, $(z-1)^4 = \mathcal{O}(N^{-6/5})$. Thus

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\Gamma\cap N_{\delta_N(1)}} e^{Nh(z)}g(z)dz &= N^{-\frac{1}{3}}\frac{1}{2\pi i}\oint_{\tilde{\Gamma}} e^{\frac{N}{2}}e^{-\frac{w^3}{3}}(1+\mathcal{O}(N^{-\frac{1}{5}}))dw \\ &= N^{-\frac{1}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\frac{1}{2\pi i}\oint_{\tilde{\Gamma}^\infty} e^{-\frac{w^3}{3}}dw \\ &= N^{-\frac{1}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\,\mathrm{Ai}(0).\end{aligned} \tag{3.99}$$

Here the contours $\tilde{\Gamma}$ and $\tilde{\Gamma}^\infty$ are described in Figure 2, and we use the Airy function to denote the integral in (3.99). The Airy function is a special function with many applications. See [1] for its various properties. The general formula of the Airy function is, in its contour integral formula,

$$\mathrm{Ai}(x) = \frac{1}{2\pi i}\oint_{\tilde{\Gamma}^\infty} e^{-\frac{w^3}{3}+xw}dw. \tag{3.100}$$

It is easy to see that the integrand in (3.100), although defined on an infinite contour, converges for all $x \in \mathbb{C}$.

Now we consider more generally $x = 2 + N^{-2/3\xi}$ where ξ is in a compact subset of $\mathbb{C}$. Although we are interested mostly in real value of x, it is harmless and actualy useful to consider complex x. Then

$$h(z) = 2z - \frac{z^2}{2} - \log z - 1 + N^{-\frac{2}{3}}\xi z - N^{-\frac{2}{3}}\xi - \frac{1}{4}N^{-\frac{4}{3}}\xi^2, \tag{3.101}$$

and for $z \in N_{\delta_N}(1)$, after the change of variable $z = 1 + N^{-1/3}w$, we have like (3.98)

$$e^{Nh(z)} = e^{\frac{N}{2}}e^{-\frac{w^3}{3}+\xi w}(1+\mathcal{O}(N^{-\frac{1}{5}})). \tag{3.102}$$

Thus like (3.99)

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\Gamma\cap N_{\delta_N(1)}} e^{Nh(z)}g(z)dz &= N^{-\frac{1}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\frac{1}{2\pi i}\oint_{\tilde{\Gamma}^\infty} e^{-\frac{w^3}{3}+\xi w}g(1)dw \\ &= N^{-\frac{1}{3}}g(1)e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\,\mathrm{Ai}(\xi).\end{aligned} \tag{3.103}$$

On the other hand, since the ingegrand is only slightly changed, the estimate of the integral over $\Gamma \setminus N_{\delta_N}(1)$, which is not as sensitive as the integral near the critical point, barely changes, and we still have like (3.97)

$$\left|\frac{1}{2\pi i}\oint_{\Gamma\setminus N_{\delta_N(1)}} e^{Nh(z)}g(z)dz\right| < Ce^{-\epsilon N^{-\frac{9}{10}}}e^{\frac{N}{2}}. \tag{3.104}$$

The detail is left to the reader.

The asymptotics (3.103) and the estimate (3.104), together with the asymptotics (3.81), imply that as $N \to \infty$, for $x = 2 + N^{-2/3}\xi$ and ξ is in a compact subset of $\mathbb{R}$,

$$p_N(x)e^{-\frac{N}{4}x^2} = N^{\frac{1}{6}}(1+\mathcal{O}(N^{-\frac{1}{5}})\,\mathrm{Ai}(\xi). \tag{3.105}$$

As $x \to +\infty$, the Airy function $\mathrm{Ai}(x) \to 0$ exponentially. Actually

$$\mathrm{Ai}(x) = \frac{e^{-\frac{2}{3}x^{\frac{3}{2}}}}{2\sqrt{\pi}x^{\frac{1}{4}}}, \quad \text{as } x \to +\infty. \tag{3.106}$$

It suggests that the asymptotic formula for $p_N(x)e^{-Nx^2/4}$ can be extrapolated so show that it vanishes exponentially as x increases. However, we need always to be cautious to do extrapolation. Below we give an expenential decay result, which is not the optimal one, but suffices for our purpose.

We distinguish $h(z)$ with different parameters x, and write

$$h_\xi(z) = h(z)\big|_{x=2+N^{-\frac{2}{3}}\xi} = 2z - \frac{z^2}{2} - \log z + N^{-\frac{2}{3}}\xi z - 1 - N^{-\frac{2}{3}}\xi - \frac{1}{4}N^{-\frac{4}{3}}\xi^2. \tag{3.107}$$

Then we have for all $\xi \in \mathbb{R}_+$ and all $z \in \Gamma$

$$\Re(h_\xi(z) - h_0(z)) = N^{-\frac{2}{3}}\xi\left(\Re z - 1 - \frac{1}{4}N^{-\frac{2}{3}}\xi\right) < -N^{-1}\frac{\xi}{2}, \tag{3.108}$$

where we use that $\Re z \le 1 - N^{-1/3}/2$ for all $z \in \Gamma$. Thus

$$|e^{Nh_\xi(z)}| < e^{-\frac{\xi}{2}}|e^{Nh_0(z)}| \tag{3.109}$$

for all $z \in \Gamma$.

Recall the computation of $p_N(x)e^{-Nx^2/4}$ for $x = 2$, especially (3.97) and (3.99). We see that we can use almost the same computation to derive that

$$\int_\Gamma |e^{Nh_0(z)}||g(z)||dz| < CN^{-\frac{1}{3}}e^{\frac{N}{2}}, \tag{3.110}$$

where $C = \mathcal{O}(1)$ is a positive constant. Hence (3.85) and (3.109) imply

$$p_N(x)e^{-\frac{N}{4}x^2}\Big|_{x=2+N^{-\frac{2}{3}}\xi} < \frac{1}{2\pi}\int_\Gamma |e^{Nh_\xi(z)}||g(z)||dz| = \frac{C_N}{2\pi}e^{-\frac{\xi}{2}}CN^{-\frac{1}{3}}e^{\frac{N}{2}} = e^{-\frac{\xi}{2}}\frac{C}{2\pi}N^{-\frac{1}{6}}(1+\mathcal{O}(N^{-1})), \tag{3.111}$$

where the $\mathcal{O}(N^{-1})$ term comes from the Stirling formula of C_N and is independent of ξ. We conclude that as $x = 2 + N^{-\frac{2}{3}\xi}$ and $\xi > 0$, $p_N(x)e^{-Nx^2/4}$ decays exponentially with respect to ξ. A slight stretch of the computation shows that the exponential decay holds if ξ is complex, with its imaginary part bounded.

Now we have asymptotics of $p_N(x)e^{-Nx^2/4}$. The method can be applied to $p_{N-1}(x)e^{-Nx^2/4}$ verbatim, since the only difference between their contour integral formulas is from the $g(z)$ factor, but all our derivation does not use any specific property of $g(z)$. Similar to (3.105), we have that for $x = 2 + N^{-2/3}\xi$ and ξ is in a compact subset of $\mathbb{C}$

$$p_{N-1}(x)e^{-\frac{N}{4}x^2} = N^{\frac{1}{6}}(1 + \mathcal{O}(N^{-\frac{1}{5}})\,\mathrm{Ai}(\xi), \tag{3.112}$$

and we also have the exponential decay property as $\xi \to +\infty$. However, if we substitute formulas (3.105) and (3.112) into (3.76), we get no meaningful result.

The situation can be saved in a simple way. Write

$$K_N(x,y) = \frac{(p_N(x) - p_{N-1}(x))e^{-\frac{N}{4}x^2}p_{N-1}(y)e^{-\frac{N}{4}y^2} - p_{N-1}(x)e^{-\frac{N}{4}x^2}(p_N(y) - p_{N-1}(y))e^{-\frac{N}{4}y^2}}{x-y}, \tag{3.113}$$

and express

$$(p_N(x) - p_{N-1}(x))e^{-\frac{N}{4}x^2} = \frac{C_N}{2\pi i}\oint_\Gamma e^{Nh(z)}g(z)dz, \tag{3.114}$$

where

$$h(z) = xz - \frac{z^2}{2} - \log z - \frac{x^2}{4}, \quad g(z) = z^{-1} - 1. \tag{3.115}$$

The estimation of the integral over $\Gamma \setminus N_{\delta_N}(1)$ can be done similarly to (3.97), and the same result can be obtained. But since the saddle point 1 is a zero of $g(z)$, the integral over $\Gamma \cap N_{\delta_N(1)}$ need to be modified a little. With the change of variable $z = 1 + N^{-1/3}w$, we have similar to (3.99) and (3.103)

$$\begin{aligned}
\frac{1}{2\pi i}\oint_{\Gamma\cap N_{\delta_N(1)}} e^{Nh(z)}g(z)dz &= N^{-\frac{1}{3}}\frac{1}{2\pi i}\oint_{\tilde{\Gamma}} e^{\frac{N}{2}}e^{-\frac{w^3}{3}+\xi w}(-N^{-\frac{1}{3}}w)(1+\mathcal{O}(N^{-\frac{1}{5}}))dw \\
&= -N^{-\frac{2}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\frac{1}{2\pi i}\oint_{\tilde{\Gamma}_\infty} e^{-\frac{w^3}{3}+\xi w}wdw \\
&= -N^{-\frac{2}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\frac{d}{d\xi}\left(\frac{1}{2\pi i}\oint_{\tilde{\Gamma}_\infty} e^{-\frac{w^3}{3}+\xi w}dw\right) \\
&= -N^{-\frac{2}{3}}e^{\frac{N}{2}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\,\mathrm{Ai}'(\xi),
\end{aligned} \tag{3.116}$$

and then

$$(p_N(x) - p_{N-1}(x))e^{-\frac{N}{4}x^2} = -N^{-\frac{1}{6}}(1+\mathcal{O}(N^{-\frac{1}{5}}))\,\mathrm{Ai}'(\xi). \tag{3.117}$$

Furthermore we know that it decays exponentially with respect to ξ as $\xi \to +\infty$.

We define the kernel function

$$K_{\mathrm{Airy}}(\xi,\eta) = \begin{cases} \dfrac{\mathrm{Ai}(\xi)\,\mathrm{Ai}'(\eta) - \mathrm{Ai}'(\xi)\,\mathrm{Ai}(\eta)}{\xi-\eta} & \text{if } \xi \neq \eta, \\ \mathrm{Ai}'(\eta)^2 - \mathrm{Ai}''(\eta)\,\mathrm{Ai}(\eta) & \text{if } \xi = \eta. \end{cases} \tag{3.118}$$

Note that $K(\xi,\eta)$ is analytic in both ξ and η. The asymptotics (3.105), (3.112), (3.117) and the identity (3.113) immdiately give as that as $x = 2 + N^{-2/3}\xi$ and $y = 2 + N^{-2/3}\eta$, where ξ, η are in a compact subset of $\mathbb{R}$ and $\xi \neq \eta$,

$$\lim_{N\to\infty} N^{-\frac{2}{3}}K_N(x,y) = K_{\mathrm{Airy}}(\xi,\eta). \tag{3.119}$$

By a little complex analysis, we can generalise it to the $\xi = \eta$ case. For any η, we consider

$$\xi_\theta = \eta + e^{i\theta}, \quad \text{and} \quad x_\theta = 2 + \xi_\theta. \tag{3.120}$$

Then the Cauchy integral formula implies that

$$\lim_{N\to\infty} N^{-\frac{2}{3}}K_N(y,y) = \lim_{N\to\infty}\int_0^{2\pi} N^{-\frac{2}{3}}K_N(x_\theta,y)d\theta = \frac{1}{2\pi}\int_0^{2\pi} K_{\mathrm{Airy}}(\xi_\theta,\eta)d\theta = K_{\mathrm{Airy}}(\eta,\eta). \tag{3.121}$$

Now we have the pointwise convergence that $N^{-2/3}K_N(x,y) \to K_{\mathrm{Airy}}(\xi,\eta)$. We also know that as $\xi,\eta \to \pm\infty$, $|N^{-2/3}K_N(x,y)|$ vanishes uniformly. Then by argument in functional analysis, we have that if $a = 2 + N^{-2/3}T$, then

$$\lim_{N\to\infty}\det(1-\mathbf{1}_{(a,\infty)}(x)K_N(x,y)\mathbf{1}_{(a,\infty)}(y)) = F_{\mathrm{TW}}(T) := \det(1-\mathbf{1}_{(T,\infty)}(\xi)K_{\mathrm{Airy}}(\xi,\eta)\mathbf{1}_{(T,\infty)}(\eta)). \tag{3.122}$$

Here the F_{TW} is the cumulative distribution function of the celebrated Tracy-Widom distribution. Finally we have the result for the limiting distribution of the largest eigenvalue in GUE:

Theorem 13. *As the dimension $N \to \infty$, the largest eigenvalue $\lambda_{\max}$ in GUE is almost surely at 2, and its fluctuation is of order $N^{-2/3}$, given by the Tracy-Widom distribution that*

$$\lim_{N\to\infty}\mathbb{P}(\lambda_{\max} < 2 + N^{-2/3}T) = F_{\mathrm{TW}}(T). \tag{3.123}$$

We are not going to give the proof of this theorem, unless we have enough time left in the end of this semester.

Exercise

1. Complete the proof of Lemma 6 for the $k = 3, \ldots, N$ case. (Hint: Cauchy-Binet formula is an important linear algebraic identity that you can learn from `http://en.wikipedia.org/wiki/CauchyBinet_formula`, for example.

2. The *Laguerre polynomials* (see [23] and `http://en.wikipedia.org/wiki/Laguerre_polynomials`)

$$L_n^{(\alpha)}(x) = \frac{1}{2\pi i} \oint \frac{e^{\frac{xt}{1-t}}}{(1-t)^{\alpha+1} t^{n+1}} dt \tag{3.124}$$

   where $\alpha > -1$ is a parameter, and the contour encloses 0 but not 1.

   Find an asymptotic formula involving Airy function of $e^{-x/2} L_n^{(0)}(x)$ for $x = 4n + 2 + 2(2n/3)^{1/3} t$ where $n \to \infty$ and t is in a compact subset of $\mathbb{R}$.

4 Limiting distribution of the largest eigenvalue in Gaussian β ensemble

We have derived the Tracy-Widom distribution for the limiting distribution of the largest eigenvalue in the GUE random matrices. For Gaussian orthogonal ensemble (GOE) and Gaussian symplectic ensemble (GSE), similar results hold, but the derivation becomes more sophisticated. Although the limiting distribution of the largest eigenvalues is a problem in probability, the solution for GUE as shown in Section 3 and these for GOE and GSE rely on the exploit of the symmetry of the models and analytical techniques not commonly seen in probability literature. For people inclining to study probability, we present a unified solution for the limiting distribution of the largest eigenvalue in GUE, GOE and GSE. The drawback of this approach is that the result is stated in a rather abstract way.

This section follows closely to [3, Section 4.5]. Another good reference is the Ph.D. thesis of Alex Bloemendal [8]. The interested readers are refered to the original paper by Ramírez, Rider and Virág [20] that contains more results about the β ensembles.

Remark 2. In this section, the terms "Gaussian unitary ensemble" and "Gaussian orthogonal ensemble" are defined slightly different from those in Section 2. For GUE, we define it as a random Hermitian matrix such that the diagonal entries are real and in $\mathrm{N}(0,1)$, upper-triangular entries are complex with both real and imaginary parts in independent $\mathrm{N}(0, \frac{1}{2})$ distributions, and they are all independent. So it is the GUE in Section 2 scaled up by N. The GOE (as well as GSE) is redefined by the same scaling.

4.1 Tridiagonal matrix models and the Gaussian β ensemble

We are going to define a tridiagonal random matrix model that has a parameter $\beta > 0$, such that the distribution of the eigenvalues in the tridiagonal random matrix model with $\beta = 1, 2, 4$ is identical to that of the eigenvalues in GOE, GUE and GSE respectively. Thus we call this tridiagonal random matrix model *Gaussian β ensemble* (GβE). The limiting distributions of the largest eigenvalue in GOE, GUE and GSE become those of the largest eigenvalue in the GβE for the corresponding β.

First we recall the χ_t distribution that is a positive continuous probability distribution given by the density function

$$f_t(x) = \frac{2^{1-\frac{t}{2}} x^{t-1} e^{-\frac{x^2}{2}}}{\Gamma(\frac{t}{2})} \mathbf{1}_{(0,\infty)}(x), \tag{4.1}$$

where the real parameter $t > 0$. When t is an integer, χ_t distribution is related to the normal distribution in the way that if $X_1, \ldots, X_t$ are i.i.d. random variables in distribution $\mathrm{N}(0,1)$, then

$$\left(\sum_{i=1}^{t} X_i^2\right)^{\frac{1}{2}} \sim \chi_t(x). \tag{4.2}$$

If t is an integer, then it is a direct consequence of (4.2) and the central limit theorem that as $t \to \infty$

$$2(\chi_t(x) - \sqrt{t}) \xrightarrow{d} \mathrm{N}(0,1). \tag{4.3}$$

A straightforward computation confirms that (4.3) holds for real t.

We define the Gaussian β ensemble as follows. Let $\xi_1, \xi_2, \ldots, Y_1, Y_2, \ldots$ be independent random variables such that $\xi_i \sim N(0,1)$ and $Y_i \sim \xi_{i\beta}$. For any dimension N, the N dimensional symmetric matrix H_N in GβE is given by

$$h_{ij} = 0 \quad \text{if} \quad |i-j| > 1, \tag{4.4}$$

$$h_{ii} = \sqrt{2/\beta}\xi_i, \tag{4.5}$$

$$h_{i,i+1} = h_{i+1,i} = Y_{N-i}/\sqrt{\beta}. \tag{4.6}$$

Then we have

Theorem 14. *For any β and N, let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of H_N. Then the joint probability density function of $\lambda_1, \ldots, \lambda_N$ is*

$$P_\beta(\lambda_1, \ldots, \lambda_N) = \frac{1}{C_{N,\beta}} \Delta(\lambda)^\beta e^{-\frac{\beta}{4}\sum_{i=1}^N \lambda_i^2}. \tag{4.7}$$

When $\beta = 2$, Theorem 14 shows that the GβE has the same spectral property as GUE. In Exercise 1 we will see that the GβE with $\beta = 1$ has the same spectral property as GOE. The relation between GβE with $\beta = 4$ and GSE is analogous, but the proof is a little too complicated to be an exercise.

Proof of Theorem 14 for $\beta = 2$. First we show the eqivalence between the distribution of eigenvalues of GβE with $\beta = 2$ and GUE, and prove the theorem for $\beta = 2$. For any N, let the $N \times N$ random Hermitian matrix $X_N = (x_{ij})_{i,j=1}^N$ be a GUE random matrix, such that

$$X_N = \begin{pmatrix} \mathrm{N}(0,1) & \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) & \ldots & \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) \\ \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) & \mathrm{N}(0,1) & \ldots & \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) \\ \vdots & \vdots & & \vdots \\ \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) & \mathrm{N}(0,\frac{1}{2}) + i\mathrm{N}(0,\frac{1}{2}) & \ldots & \mathrm{N}(0,1) \end{pmatrix} \tag{4.8}$$

where all the upper diagonal entries are independent. Since any unitary similarity transformation does not change the spectrum of a matrix, we define a random unitary operator $U(1) \in \mathrm{U}(N-1)$ that depends on X_N such that

1. Let $\mathbf{x}_1 = (x_{21}, x_{31}, \ldots, x_{N1})^T$. Then

$$U(1)\mathbf{x}_1 = \|\mathbf{x}_1\|\mathbf{e}_1 = (\|\mathbf{x}_1\|, 0, \ldots, 0)^T. \tag{4.9}$$

2. $U(1)\mathbf{x} = \mathbf{x}$ if $\mathbf{x}$ is orthogonal to $\mathbf{x}_1$ and $\mathbf{e}_1$.

This unitary transformation is the *Householder transformation* [16, Section 2.2.4]

$$U(1) = I - 2\frac{\mathbf{u}\mathbf{u}^*}{\|\mathbf{u}\|^2}, \quad \text{where} \quad \mathbf{u} = \mathbf{x}_1 - \|\mathbf{x}_1\|\mathbf{e}_1. \tag{4.10}$$

Then we define the unitary matrix $\tilde{U}(1) \in \mathrm{U}(N)$ by

$$\tilde{U}(1) = I_{1\times 1} \oplus U(1), \quad \text{or equivalently} \quad \tilde{U}(1) = \begin{pmatrix} 1 & 0 \\ 0 & U(1) \end{pmatrix}. \tag{4.11}$$

Then

$$\tilde{U}(1)X_N\tilde{U}(1)^{-1} = \begin{pmatrix} x_{11} & \|\mathbf{x}_1\| & 0 & \dots & 0 \\ \|\mathbf{x}_1\| & & & & \\ 0 & & & & \\ \vdots & & X_{N-1} & & \\ 0 & & & & \end{pmatrix}, \tag{4.12}$$

where X_{N-1} is a random matrix defined as

$$X_{N-1} = U(1)(X_N)_{11}U(1)^{-1}, \tag{4.13}$$

where $(X_N)_{11}$ is the $(N-1)$ dimensional matrix obtained by removing the first row and first column of X_N.

Note that $x_{11} \sim \mathrm{N}(0,1)$ and

$$\|\mathbf{x}_1\| = \left(\sum_{i=2}^{N} (\Re x_{i2})^2 + (\Im x_{i2})^2\right)^{\frac{1}{2}} \sim \frac{1}{\sqrt{2}}\chi_{2(N-1)} \tag{4.14}$$

since both $\Re x_{ij}$ and $\Im x_{ij}$ are in $\mathrm{N}(0, \frac{1}{2})$ distribution for $i \neq j$. Thus $\tilde{U}(1)X_N\tilde{U}(1)^{-1}$ agrees with the tridiagonal matrix H_N in the first column and the first row.

$(X_N)_{11}$ is a random Hermitian matrix different from the $(N-1)$ dimensional GUE only by scaling. Thus by Exercise 1 in Section 1, its distribution is invariant under unitary similarity transformation. Note that the result in the exercise is valid when the unitary similarity transformation is fixed. Here $U(1)$ is random, but by its construction, $U(1)$ depends only on entries of X_N that are not in $(X_N)_{11}$, so $U(1)$ is independent of $(X_N)_{11}$, and then the result of the exercise still holds. Also we have that X_{N-1} is independent of $U(1)$, and if we write $X_{N-1} = (\tilde{x}_{ij})_{i,j=1}^{N-1}$, all $\tilde{x}_{ij}$ are independent of x_{11} and $\|\mathbf{x}_1\|$.

Let $U(2) \in \mathrm{U}(N-2)$ be a Householder transformation sending $\mathbf{x}_2 = (\tilde{x}_{21}, \tilde{x}_{31}, \dots, \tilde{x}_{N-1,1})^T$ into $\|\mathbf{x}_2\|\mathbf{e}_1$ and keeping vectors orthogonal to $\mathbf{x}_2$ and $\mathbf{e}_1$ invariant. Then denoting $\tilde{U}(2) = I_{2\times 2} \oplus U(2)$, we have

$$\tilde{U}(2)\tilde{U}(1)X_N\tilde{U}(1)^{-1}\tilde{U}(2)^{-1} = \begin{pmatrix} x_{11} & \|\mathbf{x}_1\| & 0 & 0 & \dots & 0 \\ \|\mathbf{x}_1\| & \tilde{x}_{11} & \|\mathbf{x}_2\| & 0 & & 0 \\ 0 & \|\mathbf{x}_2\| & & & & \\ 0 & 0 & & & & \\ \vdots & \vdots & & X_{N-2} & & \\ 0 & 0 & & & & \end{pmatrix}, \tag{4.15}$$

which agrees with H_N in the first two columns and the first two rows, where X_{N-2} is defined analogously to X_{N-1}. Repeat the construction $N-1$ times, we have finally that $\tilde{U}(N-1)\cdots\tilde{U}(2)\tilde{U}(1)X_N\tilde{U}(1)^{-1}\tilde{U}(2)^{-1}\cdots\tilde{U}(N-1)^{-1}$ is the tridiagonal random matrix H_N that has the same spectral property as X_N. □

Now we prove the general case. Denote the $N \times N$ tridiagonal matrix

$$J_N = \begin{pmatrix} a_0 & b_1 & & & \\ b_1 & a_1 & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & a_{N-2} & b_{N-1} \\ & & & b_{N-1} & a_{N-1} \end{pmatrix}. \tag{4.16}$$

Without loss of generality, we assume that $b_1, \dots, b_{N-1}$ are all strictly positive. Suppose $\lambda_1 \geq \cdots \geq \lambda_N$ are eigenvalues of J_N and $\mathbf{v}_1, \dots, \mathbf{v}_N$ are corresponding eigenvectors. Note that if λ_i's are distinct, then $\mathbf{v}_i$'s are fixed up to a scalar multiple.

Lemma 7. *The eigenvalues $\lambda_1, \dots, \lambda_N$ are distinct. Furthermore, writing $\mathbf{v}_i = (v_i(1), \dots, v_i(N))^T$, we have $v_i(1) \neq 0$.*

Proof. The identity $J_N \mathbf{v}_i = \lambda_i \mathbf{v}_i$ is expressed as

$$\begin{pmatrix} a_0 - \lambda_i & b_1 & & & \\ b_1 & a_1 - \lambda_i & b_2 & & \\ & b_2 & \ddots & \ddots & \\ & & \ddots & a_{N-2} - \lambda_i & b_{N-1} \\ & & & b_{N-1} & a_{N-1} - \lambda_i \end{pmatrix} \begin{pmatrix} v_i(1) \\ v_i(2) \\ \vdots \\ v_i(N-1) \\ v_i(N) \end{pmatrix} = \mathbf{0}, \tag{4.17}$$

or componentwise

$$\begin{cases} (a_0 - \lambda_i) v_i(1) + b_1 v_i(2) = 0, \\ b_1 v_i(1) + (a_1 - \lambda_i) v_i(2) + b_2 v_i(3) = 0, \\ \vdots \\ b_{N-2} v_i(N-2) + (a_{N-2} - \lambda_i) v_i(N-1) + b_{N-1} v_i(N) = 0, \\ b_{N-1} v_i(N-1) + (a_{N-1} - \lambda_i) v_i(N) = 0. \end{cases} \tag{4.18}$$

If $v_i(1) = 0$, then iteratively we find all components to be zero, and then the vector is a zero vector. For any nonzero $v_i(1)$, iteratively we can uniquely solve $v_i(2), \dots, v_i(N)$ by the first, $\dots$, $(N-1)$-th equations in (4.18). Then the eigenspace associated to λ_i is 1-dimensional, and $\lambda_1, \dots, \lambda_N$ are distinct. □

Converse to Lemma 7, we have

Lemma 8. *Given a diagonal matrix*

$$D = \operatorname{diag}(\lambda_1, \dots \lambda_N) \quad \textit{where} \quad \lambda_1 > \cdots > \lambda_N. \tag{4.19}$$

and a unit row vector $\mathbf{u} = (u_1, \dots, u_N)$ such that all its components are strictly positive and $\sum_{i=1}^N u_i^2 = 1$, there exists a tridiagonal matrix J_N expressed as in (4.16) such that $b_1, \dots, b_{N-1}$ are strictly positive, and

$$J_N = ODO^{-1} \tag{4.20}$$

where $O = (o_{ij})_{i,j=1}^N \in \mathrm{O}(N)$ with the first row $o_{1j} = u_j$.

Proof. We relate tridiagonal matrices with orthogonal polynomials. Given a measure μ on $\mathbb{R}$, let $p_0(x), p_1(x), \dots$ be orthogonal polynomials of degree $0, 1, \dots$ such that

$$\int p_i(x)p_j(x)d\mu(x) = \delta_{ij}. \tag{4.21}$$

Then the three term-recurrence formula is (assuming $a_{-1} = 0$)

$$xp_n(x) = b_n p_{n+1}(x) + a_n p_n(x) + b_{n-1}p_{n-1}(x), \tag{4.22}$$

or equivalently (note that the a_i, b_i in (4.22) and (4.23) are not consistent with those in (4.16))

$$x \begin{pmatrix} p_0(x) \\ p_1(x) \\ \vdots \\ \vdots \end{pmatrix} = \begin{pmatrix} a_0 & b_1 & & \\ b_1 & a_1 & b_2 & \\ & b_2 & \ddots & \ddots \\ & & \ddots & \end{pmatrix} \begin{pmatrix} p_0(x) \\ p_1(x) \\ \vdots \\ \vdots \end{pmatrix}. \tag{4.23}$$

Truncate the $N \times N$ upper left block of the operator on the right-hand side of (4.23), we get a tridiagonal matrix. Note that $b_i > 0$ if we choose all the orthogonal polynomials $p_i(x)$ to have positive leading coefficient.

Suppose μ is a measure on $\mathbb{R}$ supported on N points $\lambda_1, \dots, \lambda_N$, such that

$$\mu(x) = \sum_{i=1}^{N} c_i \delta_{\lambda_i}(x), \tag{4.24}$$

where we assume $c_i \neq 0$. Then the space $L^2(\mu)$ is an N dimensional space. Obviously $L^2(\mu)$ is spanned by $\{\mathbf{1}_{\lambda_1}(x), \dots, \mathbf{1}_{\lambda_N}(x)\}$, and it is also spanned by $\{1, x, \dots, x^{N-1}\}$ by the Lagrange interpolation formula. Given that the orthogonal polynomials $p_0(x), \dots, p_{N-1}(x)$ exist, (the existence of the orthogonal polynomials can be proved and is left as an exercise), then $\{p_0(x), \dots, p_{N-1}(x)\}$ is also a basis.

Furthermore, the basis

$$B_1 = \{\frac{1}{c_i}\mathbf{1}_{\lambda_i}(x) \overset{\text{a.e. in } \mu}{=} \frac{1}{c_i} \prod_{k=1,\dots,N,k\neq i} \frac{x - \lambda_k}{\lambda_i - \lambda_k} \mid i = 1, \dots, N\} \tag{4.25}$$

is an orthonormal basis of $L^2(\mu)$ for which the multiplication operator x has the matrix representation D in (4.19), while the basis

$$B_2 = \{p_{i-1}(x) \mid i = 1, \dots, N\} \tag{4.26}$$

is another orthogonal basis of $L^2(\mu)$ for which the multiplication operator x has the matrix representation J_N in (4.16).

The identity $J_N = ODO^{-1}$ means that

$$\begin{pmatrix} p_0(x) \\ \vdots \\ p_{N-1}(x) \end{pmatrix} = O \begin{pmatrix} c_1^{-1}\mathbf{1}_{\lambda_1}(x) \\ \vdots \\ c_N^{-1}\mathbf{1}_{\lambda_N}(x) \end{pmatrix}, \tag{4.27}$$

and especially

$$\left(\sum_{i=1}^{N} c_i^2\right)^{-\frac{1}{2}} = p_0(x) \overset{\text{a.e. in } \mu}{=} o_{11}\frac{1}{c_1}\mathbf{1}_{\lambda_1}(x) + \dots + o_{1N}\frac{1}{c_N}\mathbf{1}_{\lambda_N}(x). \tag{4.28}$$

This identity is satisfied when $c_i = o_{1i}$. Thus we conclude that the matrix constructed from the three-term recurrence relation for the orthogonal polynomials $p_i(x)$ defined by (4.21) with μ given by

$$\mu(x) = \sum_{i=1}^{N} u_i \delta_{\lambda_i}(x), \tag{4.29}$$

and the property that $b_i > 0$ are satisfied if we require all $p_i(x)$ to have positive leading coefficients. □

Lemmas 7 and 8 define a bijection between the space of tridiagonal matrices H_N with strictly positive b_i as denoted in (4.16) and the space of $(D, \mathbf{u})$ where the diagonal matrix D and the unit vector $\mathbf{u}$ satisfy the conditions in Lemma 8. We denote this map from H_N to $(D, \mathbf{u})$ as

$$F : \Delta_N \times S_+^{N-1} \to \mathbb{R}^N \times (0, \infty)^{N-1}, \tag{4.30}$$

where

$$\Delta_N = \{(x_1, \dots, x_N) \in \mathbb{R}^N \mid x_1 > \cdots > x_N\}, \tag{4.31}$$

$$S_+^{N+1} = \{(x_1, \dots, x_{N-1}) \in (0, \infty)^{N-1} \mid 0 < \sum_{i=1}^{N-1} x_i^2 < 1\}, \tag{4.32}$$

and

$$F(\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1}) = (a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1}). \tag{4.33}$$

Obviously F is differentiable, and we have

Lemma 9. *The Jacobian of F is*

$$\frac{\partial(a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1})}{\partial(\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1})} = C \frac{\Delta(\lambda)}{\prod_{i=1}^{N-1} b_j^{N-j-1}}. \tag{4.34}$$

The proof of Lemma 9 relies on the following technical result

Lemma 10. *Any element in the set of $N \times N$ Hermitian matrices, except for those in a subset of Lebesgue measure 0 (or equivalently, probability 0 in GUE), can be written uniquely as*

$$X = O^{-1} D O, \tag{4.35}$$

where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ and $\lambda_1 < \lambda_2 < \cdots < \lambda_N$, and $O = (o_{ij})_{i,j=1}^N \in \mathrm{O}(N)$ such that $o_{11}, o_{21}, \dots, o_{N1}$ are all strictly positive.

We are not going to prove this lemma, since its proof is parallel to those of Lemmas 3, 4 and 5 in Section 3.1.

Proof of Lemma 9. By Exercise 1, we have that given an $N \times N$ random real symmetric matrix X_N in GOE (as defined in this section), by the tridiagonalisation procedure we get a random tridiagonal matrix whose distribution is exactly the same as H_N with $\beta = 1$. The tridiagonal procedure is (analogous to the proof of Theorem 14 for $\beta = 2$)

$$X_N \to \tilde{O}(N-1) \cdots \tilde{O}(2) \tilde{O}(1) X_N \tilde{O}(1)^{-1} \tilde{O}(2)^{-1} \cdots \tilde{O}(N-1)^{-1}, \tag{4.36}$$

where

$$\tilde{O}(k) = \begin{pmatrix} I_k & 0 \\ 0 & O(k) \end{pmatrix} \quad \text{and} \quad O(k) \in \mathrm{O}(N-k). \tag{4.37}$$

Suppose X_N has the unique decomposition (4.35), then the tridiagonalisation procedure becomes

$$O^{-1}DO \to \tilde{O}^{-1}D\tilde{O}, \quad \text{where} \quad \tilde{O} = (\tilde{o}_{ij})_{i,j=1}^N = O\tilde{O}(1)^{-1}\cdots\tilde{O}(N-1)^{-1}. \tag{4.38}$$

We observe that the first column of $\tilde{O}$ is identical to that of O. Therefore by Lemma 8, we conclude that the tridiagonalisation procedure transforms real symmetric matrices X_N (except for those in a measure 0 subset) whose unique decomposition (4.35) has eigenvalues $\lambda_1, \ldots, \lambda_N$ and first column of O as $o_{i1} = u_1$ $(i = 1, \ldots, N)$ to the tridiagonal matrix J_N denoted in (4.16) such that

$$a_0, \ldots, a_{N-1}; b_1, \ldots, b_{N-1} = F(\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_{N-1}). \tag{4.39}$$

Suppose we have the marginal distribution of $\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_{N-1}$, then by the bijectivity results in Lemmas 7 and 8, the distribution of $a_0, \ldots, a_{N-1}; b_1, \ldots, b_{N-1}$ is

$$P(a_0, \ldots, a_{N-1}; b_1, \ldots, b_{N-1}) = P(\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_{N-1}) \frac{\partial(a_0, \ldots, a_{N-1}; b_1, \ldots, b_{N-1})}{\partial(\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_{N-1})}. \tag{4.40}$$

We take a special choice of X_N such that X_N is in the GOE distribution. Then we have the marginal distribution

$$P(\lambda_1, \ldots, \lambda_N; u_1, \ldots, u_{N-1}) = \frac{1}{C}\Delta(\lambda) \prod_{i=1}^N e^{\frac{1}{4}jj\lambda_i^2}, \tag{4.41}$$

where C does not depend on λ_i or u_i. (The proof is similar to the arguments in Section 3.1, and we are not going to give any detail. On the other hand, we have by Exercise 1 that

$$P(a_0, \ldots, a_{N-1}; b_1, \ldots, b_{N-1}) = \frac{1}{C} \prod_{i=0}^{N-1} e^{-\frac{1}{4}a_i^2} \prod_{j=1}^{N-1} e^{-\frac{1}{2}b_j^2} b_j^{N-j-1}. \tag{4.42}$$

Noting that

$$\prod_{i=0}^{N-1} e^{-\frac{1}{4}a_i^2} \prod_{j=1}^{N-1} e^{-\frac{1}{2}b_j^2} = e^{-\frac{1}{4}\operatorname{Tr}(J_N^2)} = \prod_{i=1}^N e^{-\frac{1}{4}\lambda_i^2}, \tag{4.43}$$

we derive (4.34) from (4.40), (4.41) and (4.42). □

For the proof of Theorem 14, we need another technical result

Lemma 11. *With notations the same as in Lemma 9, we have*

$$\prod_{j=1}^{N-1} b_j^{N-j} = \Delta(\lambda) \prod_{i=1}^N u_i. \tag{4.44}$$

Proof. Let $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)^T$. We have

$$J_N\mathbf{e}_1 = (a_0, b_1, 0, \ldots, 0), \tag{4.45}$$

$$J_N^2\mathbf{e}_1 = (a_0^2 + b_1b_0, a_0b_1 + b_1a_1, b_1b_2, \ldots, 0), \tag{4.46}$$

$$\vdots$$

Then

$$\det(\mathbf{e}_1, J_N\mathbf{e}_1, J_N^2\mathbf{e}_1, \dots, J_N^{N-1}\mathbf{e}_1) = \begin{pmatrix} 1 & * & * & \dots & * \\ 0 & b_1 & * & \dots & * \\ 0 & 0 & b_1b_2 & & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & b_1b_2\cdots b_{N-1} \end{pmatrix} = \prod_{j=1}^{N-1} b_j^{N-j}. \tag{4.47}$$

On the other hand, using the decomposition $J_N = O^{-1}DO$ where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ and λ_i are increasing, and the first column of O is given by $\mathbf{u} = (u_1, \dots, u_N)^T$, we have

$$\begin{aligned} \det(\mathbf{e}_1, J_N\mathbf{e}_1, J_N^2\mathbf{e}_1, \dots, J_N^{N-1}\mathbf{e}_1) &= \det(O^{-1}D^0O\mathbf{e}_1, O^{-1}D^1O\mathbf{e}_1, \dots, O^{-1}D^{N-1}O\mathbf{e}_1) \\ &= \det(O^{-1}(D^0O\mathbf{e}_1, D^1O\mathbf{e}_1, \dots, D^{N-1}O\mathbf{e}_1)) \\ &= \det(O^{-1})\det(D^0\mathbf{u}, D^1\mathbf{u}, \dots, D^{N-1}\mathbf{u})) \\ &= \pm \begin{vmatrix} u_1 & \lambda_1u_1 & \dots & \lambda_1^{N-1}u_1 \\ u_1 & \lambda_2u_1 & \dots & \lambda_2^{N-1}u_1 \\ \vdots & \vdots & \dots & \vdots \\ u_1 & \lambda_Nu_1 & \dots & \lambda_N^{N-1}u_1 \end{vmatrix} \\ &= \pm\,\Delta(\lambda)\prod_{i=1}^N u_i. \end{aligned} \tag{4.48}$$

The sign ambiguity can be solved since $\Delta(\lambda)$ is positive. Then we prove (4.44). □

Proof of Theorem 14 for $\beta > 0$. The results in Lemmas 9 and 11 together imply that

$$\frac{\partial(a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1})}{\partial(\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1})} = \frac{\prod_{j=1}^{N-1} b_j}{\prod_{i=1}^N u_i}. \tag{4.49}$$

For the N dimensional random tridiagonal matrix H_N defined in the beginning of this subsection, in probability 1 the entries of $h_{i,i+1} = h_{i+1,i}$ are strictly positive, so we identify it with J_N in (4.16) and use $a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1}$ to denote its nontrivial entries. From the distribution of $a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1}$ and the Jacobian (4.34), we compute the joint probability density of $\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1}$ as

$$\begin{aligned} P(\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1}) &= P(a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1})\frac{\partial(a_0, \dots, a_{N-1}; b_1, \dots, b_{N-1})}{\partial(\lambda_1, \dots, \lambda_N; u_1, \dots, u_{N-1})} \\ &= C\prod_{i=0}^{N-1} e^{-\frac{\beta}{4}a_i^2}\prod_{j=1}^{N-1} b_j^{\beta(N-j)-1}e^{-\frac{\beta}{2}b_j^2}\frac{\prod_{j=1}^{N-1} b_j}{\prod_{i=1}^N u_i} \\ &= Ce^{-\frac{\beta}{4}\mathrm{Tr}\,H_N^2}\frac{\prod_{j=1}^{N-1} b_j^{\beta(N-j)}}{\prod_{i=1}^N u_i} \\ &= C\Delta(\lambda)^\beta\prod_{i=1}^N e^{-\frac{\beta}{4}\lambda_i^2}\left(\prod_{i=1}^N u_i\right)^{\beta-1}. \end{aligned} \tag{4.50}$$

Integrating out $u_1, \dots, u_{N-1}$, we obtain the distribution of $\lambda_1, \dots, \lambda_N$ as in Theorem 14. □

4.2 Limiting distribution of the largest eigenvalue in Gaussian β ensemble

Before the statement of the result on the limiting distribution of the largest eigenvalue of H_N, we give a heuristic argument to motivate the theorem and its proof. From our result for the largest eigenvalue in GUE that is equivalent to the special case of H_N with $\beta = 2$, we conjecture that the largest eigenvalue is around $2\sqrt{N}$ with fluctuation $\mathcal{O}(N^{-1/6})$. Write

$$\begin{aligned}
\tilde{H}_N &= N^{\frac{1}{6}}(H_N - 2\sqrt{N}I) \\
&= \begin{pmatrix} -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_1 & \frac{1}{\sqrt{\beta}}N^{\frac{1}{6}}Y_{N-1} & & \\ \frac{1}{\sqrt{\beta}}N^{\frac{1}{6}}Y_{N-1} & -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_2 & \frac{1}{\sqrt{\beta}}N^{\frac{1}{6}}Y_{N-2} & \\ & \frac{1}{\sqrt{\beta}}N^{\frac{1}{6}}Y_{N-2} & -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix} \\
&= \begin{pmatrix} -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_1 & N^{\frac{1}{6}}\sqrt{N-1} + \frac{1}{\sqrt{2\beta}}N^{\frac{1}{6}}\xi_1' & & \\ N^{\frac{1}{6}}\sqrt{N-1} + \frac{1}{\sqrt{2\beta}}N^{\frac{1}{6}}\xi_1' & -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_2 & N^{\frac{1}{6}}\sqrt{N-2} + \frac{1}{\sqrt{2\beta}}N^{\frac{1}{6}}\xi_2' & \\ & N^{\frac{1}{6}}\sqrt{N-2} + \frac{1}{\sqrt{2\beta}}N^{\frac{1}{6}}\xi_2' & -2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix},
\end{aligned} \tag{4.51}$$

where $\xi_i' = \sqrt{2}(Y_{N-i} - \sqrt{(N-i)\beta})$, and from (4.3) we have that the distribution of ξ_i' is close to $\mathrm{N}(0,1)$ if i is not close to N.

The matrix $\tilde{H}_N$ acts on N-dimensional vector $(f_1, f_2, \ldots, f_N)^T$. Define a function $f(x)$ on $[0, N^{2/3}]$ such that

$$f(x) = f_i \quad \text{for} \quad x \in [(i-1)N^{-\frac{1}{3}}, iN^{-\frac{1}{3}}). \tag{4.52}$$

The function $f(x)$ is not continuous, let alone differentiable. But we assume that $f(x)$ is a "discretised" continuous, or even smooth, function with step length $N^{-1/3}$. One way to think of it is to let $\tilde{f}(x)$ be a smooth function, and $f(x) = \tilde{f}((i-1)N^{-1/3})$ where i is the smallest possible number such that $iN^{-1/3} \geq x$. Then heuristically (as commonly used in numerical analysis of smooth functions)

$$f_{i+1} - f_i \approx N^{-\frac{1}{3}} f'(iN^{-\frac{1}{3}}), \tag{4.53}$$

$$f_{i+1} - 2f_i + f_{i-1} \approx N^{-\frac{1}{3}} f'(iN^{-\frac{1}{3}}) - N^{-\frac{1}{3}} f'((i-1)N^{-\frac{1}{3}}) \approx N^{-\frac{2}{3}} f''(iN^{-\frac{1}{3}}). \tag{4.54}$$

For $i = xN^{1/3}$ where $x \in (0, \infty)$, the vector identity

$$\tilde{H}_N \begin{pmatrix} \vdots \\ f_i \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ (N^{\frac{1}{6}}\sqrt{N-i+1} + \frac{N^{\frac{1}{6}}}{\sqrt{2\beta}}\xi'_{i-1})f_{i-1} + (-2N^{\frac{2}{3}} + \sqrt{\frac{2}{\beta}}N^{\frac{1}{6}}\xi_i)f_i \\ +(N^{\frac{1}{6}}\sqrt{N-i} + \frac{N^{\frac{1}{6}}}{\sqrt{2\beta}}\xi'_i)f_{i+1} \\ \vdots \end{pmatrix}$$

$$= \begin{pmatrix} \vdots \\ N^{\frac{2}{3}}(f_{i-1} + f_i + f_{i+1}) + (N^{\frac{1}{6}}\sqrt{N-i+1} - N^{\frac{2}{3}})f_{i-1} + (N^{\frac{1}{6}}\sqrt{N-i} - N^{\frac{2}{3}})f_{i+1} \\ +N^{\frac{1}{6}}\left(\frac{1}{\sqrt{2\beta}}\xi'_{i-1}f_{i-1} + \sqrt{\frac{2}{\beta}}\xi_i f_i + \frac{1}{\sqrt{2\beta}}\xi'_i f_{i+1}\right) \\ \vdots \end{pmatrix} \tag{4.55}$$

can be approximated, by the identities

$$N^{\frac{1}{6}}\sqrt{N-i+1} - N^{\frac{2}{3}} \approx -\frac{x}{2}, \quad N^{\frac{1}{6}}\sqrt{N-i} - N^{\frac{2}{3}} \approx -\frac{x}{2} \tag{4.56}$$

and $f_{i-1} \approx f_i \approx f_{i+1}$, as

$$\tilde{H}(f)(x) \approx f''(x) - xf(x) + \frac{2}{\sqrt{\beta}}N^{\frac{1}{6}}\left(\frac{1}{2\sqrt{2}}\xi'_{i-1} + \frac{1}{\sqrt{2}}\xi_i + \frac{1}{2\sqrt{2}}\xi'_{i+1}\right)f(x). \tag{4.57}$$

Note that as $N \to \infty$,

$$\frac{1}{2\sqrt{2}}\xi'_{i-1} + \frac{1}{\sqrt{2}}\xi_i + \frac{1}{2\sqrt{2}}\xi'_{i+1} \xrightarrow{d} \mathrm{N}(0,1). \tag{4.58}$$

We relate this factor in normal distribution to Brownian motion B_x. It is well known that B_x is not differentiable. But if we consider the discretised derivative of B_x with step length $N^{-1/3}$, we have

$$\frac{B_{x+N^{-\frac{1}{3}}} - B_x}{N^{-\frac{1}{3}}} \sim N^{\frac{1}{6}}\mathrm{N}(0,1). \tag{4.59}$$

Thus we can write (4.57) as

$$H_N(f)(x) \approx \mathbf{H}_\beta(f)(x), \tag{4.60}$$

where the stochastic differential operator

$$\mathbf{H}_\beta = \frac{d^2}{dx^2} - x + B'_x. \tag{4.61}$$

Remark 3. When we approximate $N^{1/6}(\frac{1}{2\sqrt{2}}\xi'_{i-1} + \frac{1}{\sqrt{2}}\xi_i + \frac{1}{2\sqrt{2}}\xi'_{i+1})$ by B'_x, we violate the martingale property of B_x, that is, B'_x and $B'_{x+N^{-1/3}}$ should be independent. To solve this discrepency, we aproximate $N^{1/6}\xi_i$ by $B_x(1)'$ and approximate $N^{1/6}\xi_i$ by $B_x(2)'$ where $B_x(1)$ and $B_x(2)$ are independent Brownian motions. Then $N^{1/6}(\frac{1}{2\sqrt{2}}\xi'_{i-1} + \frac{1}{\sqrt{2}}\xi_i + \frac{1}{2\sqrt{2}}\xi'_{i+1})$ is approximated by $\frac{1}{2\sqrt{2}}B_{x-N^{-1/3}}(2)' + \frac{1}{\sqrt{2}}B_x(1)' + \frac{1}{2\sqrt{2}}B_x(2)'$, and this is in some sense close to a Brownian motion.

The next question is, how to make sense of the stochastic differential operator (that is sometimes called *Stochastic Airy operator*). The B'_x term does not make sense, but it makes sense in distribution: If $h(x)$ is a function such that it has compact support and $h'(x)$ exists

and is continuous, then for any continuous but not necessarily differentiable function f, we can talk about its derivative in the sense that

$$\int f'(x)h(x)dx = -\int f(x)h'(x)dx. \tag{4.62}$$

We use this idea, together with that the path of B_x is continuous, to rigorously define $\mathbf{H}_\beta$.

For any $f, g \in C_0^\infty(0,\infty)$, i.e., smooth functions of compact support in $(0,\infty)$, we define the inner product

$$\langle f, g\rangle_* = \int_0^\infty f'(x)g'(x)dx + \int_0^\infty (1+x)f(x)g(x)dx, \tag{4.63}$$

and then define $\mathcal{L}^*$ as the Hilbert space that is the completion of $C_0^\infty(0,\infty)$ with respect to the inner product $\langle\cdot,\cdot\rangle_*$, and denote $\|f\|_* = \langle f, f\rangle_*$ for the norm of $f \in \mathcal{L}^*$. Readers familiar with Sobolev spaces can see immediately the similarity between our $\mathcal{L}^*$ and $H_0^1(0,\infty)$, which is defined as the Hilbert space obtained as the completion of smooth functions of compact support in $(0,\infty)$ with respect to the inner product

$$\langle f, g\rangle_{H^1(0,\infty)} = \int_0^\infty f'(x)g'(x)dx + \int_0^\infty f(x)g(x)dx. \tag{4.64}$$

The norm $\|f\|_{H^1(0,\infty)}$ is also defined similar to $\|f\|_*$. Below we give some basic properties of the space $\mathcal{L}^*$.

Lemma 12. *The point set of $\mathcal{L}^*$ can be realised as a subset of Hölder $\frac{1}{2}$-continuous functions on $(0,\infty)$. Then for any $f \in \mathcal{L}^* \in C^{0,\frac{1}{2}}(0,\infty)$, we have $f(0) := \lim_{x\to 0^*} f(x) = 0$ and for all $x > 1$*

$$(x+1)^{\frac{1}{4}}|f(x)| \le \sqrt{2}\|f\|_*. \tag{4.65}$$

Proof. Comparing $\langle\cdot,\cdot\rangle_*$ with $\langle\cdot,\cdot\rangle_{H^1(0,\infty)}$, we see that any $f \in \mathcal{L}^*$ satisfies $\|f\|_{H^1(0,\infty)} \le \|f\|_*$ and then $\mathcal{L}^*$ can be embedded into $H_0^1(0,\infty)$ naturally. It is a standard result that $H_0^1(0,\infty)$ can be embedded into $C^{0,\frac{1}{2}}(0,\infty)$ [2, Theorem 4.12, Part II]. Hence we have the embedding result from $\mathcal{L}^*$ to the space of Hölder $\frac{1}{2}$-continuous functions. Furthermore, the Sobolev embedding theorem also imply that on any finite closed interval $I = [a,b] \in [0,\infty)$, there is a constant C depending on $b-a$ only such that

$$\begin{aligned}\|f\|_{C^{0,\frac{1}{2}}(I)} &:= \max_{x\in I}(|f(x)|) + \max_{x_1,x_2\in I}\frac{f(x_1)-f(x_2)}{\sqrt{|x_1-x_2|}} \\ &< C\left(\int_a^b f(x)^2 + f'(x)^2dx\right)^{\frac{1}{2}} \le C\|f\|_{H_0^1(0,\infty)} \le \|f\|_*.\end{aligned} \tag{4.66}$$

Let $f \in \mathcal{L}^* \subset C^{0,\frac{1}{2}}(0,\infty)$ and $f = \lim_{n\to\infty} f_n$ where $f_n \in C_0^\infty(0,\infty)$. Since $f_n(0) = 0$, we have $f(0) = 0$ by the convergence in Hölder norm (4.66). Below we prove the inequality (4.65). By the convergence in Hölder norm, we only need to consider $f \in C_0^\infty(0,\infty)$. To prove (4.65), We show that for $f \in C_0^\infty(0,\infty)$,

$$f(x)^2 \le 4\|f\|_{L^2(x,\infty)}\|f'\|_{L^2(x,\infty)}, \quad \text{where} \quad \|\phi\|_{L^2(x,\infty)} = \left(\int_x^\infty \phi(t)^2dt\right)^{\frac{1}{2}} \quad \text{for } \phi = f, f'. \tag{4.67}$$

Suppose (4.67) does not hold for f, without loss of generality we assume that there is $x > 0$ such that $f(x) > 0$ and $f(x)^2 > 4\|f\|_{L^2(x,\infty)}\|f'\|_{L^2(x,\infty)}$. Then in the set $A_x = \{y : 0 < y - x <$

$\frac{1}{4}f(x)^2/\|f'\|^2_{L^2(x,\infty)}\}$, because

$$|f(y)-f(x)| = \left|\int_x^y f'(t)dt\right| < \left(\int_x^y f'(t)^2dt\right)^{\frac{1}{2}}\sqrt{y-x} \le \|f'\|_{L^2(x,\infty)}, \tag{4.68}$$

we have

$$f(y) \ge f(x) - \sqrt{|y-x|}\|f'\|_{L^2(x,\infty)} \ge f(x) - \frac{1}{2}\frac{f(x)}{\|f'\|_{L^2(x,\infty)}}\|f'\|_{L^2(x,\infty)} = \frac{1}{2}f(x), \tag{4.69}$$

and get the contradiction that

$$\|f\|^2_{L^2(x,\infty)} \ge \int_{A_x} f(t)^2dt > \left(\frac{1}{2}f(x)\right)^2 \cdot \frac{1}{4}\frac{f(x)^2}{\|f'\|^2_{L^2(x,\infty)}} > \|f\|^2_{L^2(x,\infty)}. \tag{4.70}$$

Then we prove (4.65) by combining (4.67) with the inequality

$$\begin{aligned}\|f\|^2_* &\ge \int_x^\infty (1+t)f(t)^2dt + \int_x^\infty f'(t)^2dt \\ &\ge (x+1)\|f\|^2_{L^2(x,\infty)} + \|f'\|^2_{L^2(x,\infty)} \\ &\ge 2\sqrt{x+1}\|f\|_{L^2(x,\infty)}\|f'\|_{L^2(x,\infty)}.\end{aligned} \tag{4.71}$$

□

Due to the embedding property of $\mathcal{L}^*$ into $C^{0,\frac{1}{2}}[0,\infty)$, we assume all elements of $\mathcal{L}^*$ as continuous functions.

Lemma 13. *Let $\{f_n\}$ be a bounded sequence in $\mathcal{L}^*$, then we can choose a subsequence $\{f_{n_k}\} \subseteq \{f_n\}$ and a $f \in \mathcal{L}^*$ and have*

1. *L^2 convergence: $f_{n_k} \to f$ in $L^2(0,\infty)$.*
2. *Weak L^2 convergence in derivative: $f'_{n_k} \to f'$ weakly in $L^2(0,\infty)$, that is, for any $h \in L^2(0,\infty)$, $\langle f'_{n_k}, h\rangle_2 \to \langle f', h\rangle_2$.*
3. *Locally uniform convergence: On any compact subset of $(0,\infty)$, $f_{n_k} \to f$ uniformly.*
4. *Weak $\mathcal{L}^*$ convergence: $f_{n_k} \to f$ weakly in $\mathcal{L}^*$, that is, for any $h \in \mathcal{L}^*$, $\langle f_{n_k}, h\rangle_* \to \langle f, h\rangle_*$.*

Proof. Property 4 is direct consequence of the Banach-Alaoglu theorem. Noting that the boundedness of $\|f\|_*$ implies the boundedness of $\|f'\|_2$, we find property 2 to be the sequence of the Banach-Alaoglu theorem as well. By the inequality (4.66), the boundedness in $\mathcal{L}^*$ norm implies the uniform boundedness and equicontinuous over a compact interval, so property 3 is proved by the Arzelá-Ascoli theorem. Property 3 immediately implies that given any $C > 0$, a subsequence $\{f_{n_k}\}$ converges in $L^2(0,C)$ if we consider these functions on $(0,C)$. The condition that $\int_0^\infty (1+x)f_n(x)^2dx \le \|f\|^2_*$ is bounded implies that the result holds when $C = \infty$. The detail is left as an exercise. □

We call a pair $(f,\lambda) \in \mathcal{L}^* \times \mathbb{R}$ an *eigenvector-eigenvalue pair* of $\mathbf{H}_\beta$, if $\|f\|_2 := \|f\|_{L^2(0,\infty)} = 1$ and $\mathbf{H}_\beta f = \lambda f$ in the sense of Schwarz distribution, i.e., for any test function $\phi(x) \in C_0^\infty(0,\infty)$,

$$\lambda\int_0^\infty f(x)\phi(x)dx = \int_0^\infty \left[f''(x) - xf(x) + \frac{2}{\sqrt{\beta}}B'_x f(x)\right]\phi(x)dx, \tag{4.72}$$

or equivalently

$$\lambda \int_0^\infty f(x)\phi(x)dx = \int_0^\infty f(x)\left[\phi''(x) - x\phi(x) - \frac{2}{\sqrt{\beta}}B_x\phi'(x)\right] - \frac{2}{\sqrt{\beta}}f'(x)B_x\phi(x)dx. \quad (4.73)$$

Now we can state the main result in this section

Theorem 15. *Fix $\beta > 0$. Let $\lambda_N^N \geq \lambda_{N-1}^N \geq \cdots$ be eigenvalues of H_N. We consider any $k \in \mathbb{Z}_+$.*

1. *For almost any Brownian motion path B_x, $\mathbf{H}_\beta$ has well defined largest, second largest, $\ldots$, k-th largest eigenvalues $\lambda_1, \ldots, \lambda_k$ and all of them are simple eigenvalues.*
2. *The random vector $N^{1/6}(\lambda_N^N - 2\sqrt{N}, \ldots, \lambda_{N-k+1}^N - 2\sqrt{N})$ converges in distribution to $(\lambda_1, \ldots, \lambda_k)$ as $N \to \infty$.*

To prove part 1 of Theorem 15, we introduce a bilinear form on $\mathcal{L}^*$ associated with $\mathbf{H}_\beta$. Suppose $f, g \in C_0^\infty(0,\infty) \subset \mathcal{L}^*$ are smooth functions, we intepreter $\mathbf{H}_\beta g$ as a distribution, and define

$$\begin{aligned}
\langle f, g\rangle_{\mathbf{H}_\beta} &= -\int_0^\infty (\mathbf{H}_\beta g)(x)f(x)dx \\
&= -\int_0^\infty f(x)\left(g''(x) - xg(x) + \frac{2}{\sqrt{\beta}}B_x' g(x)\right)dx \\
&= \int_0^\infty f'(x)g'(x)dx + \int_0^\infty xf(x)g(x)dx + \frac{2}{\sqrt{\beta}}\int_0^\infty B_x(f(x)g(x))'dx \\
&= \langle f, g\rangle_* - \langle f, g\rangle_2 + \frac{2}{\sqrt{\beta}}\int_0^\infty B_x(f'(x)g(x) + f(x)g'(x))dx.
\end{aligned} \quad (4.74)$$

Since $\langle\cdot,\cdot\rangle_{\mathbf{H}_\beta}$ is symmetric, we only need to consider the $\langle f, f\rangle_{\mathbf{H}_\beta}$, since $\langle f, g\rangle_{\mathbf{H}_\beta} = \frac{1}{4}(\langle f+g, f+g\rangle_{\mathbf{H}_\beta} - \langle f-g, f-g\rangle_{\mathbf{H}_\beta})$.

To extend the definition of $\langle f, f\rangle_{\mathbf{H}_\beta}$ to $f \in \mathcal{L}^*$, we need to make the integral $\int_0^\infty B_x f(x)f'(x)dx$ meaningful. It is not a problem if we consider the integral on a compact interval $[0, C]$. Since $f'(x) \in L^2(0,\infty)$, $\sqrt{1+x}f(x) \in L^2(0,\infty)$ and B_x is bounded on $[0, C]$,

$$\begin{aligned}
\left|\int_0^C B_x f(x)f'(x)dx\right| &\leq \|(1+x)^{-\frac{1}{2}}B_x\|_{L^\infty(0,C)}\|\sqrt{1+x}f(x)\|_{L^2(0,C)}\|f'(x)\|_{L^2(0,C)} \\
&\leq \sup_{0\leq x\leq C}|(1+x)^{-\frac{1}{2}}B_x|\|\sqrt{1+x}f(x)\|_2\|f'(x)\|_2 \\
&\leq \sup_{0\leq x\leq C}|(1+x)^{-\frac{1}{2}}B_x|\|f(x)\|_*^2.
\end{aligned} \quad (4.75)$$

But as $C \to \infty$, almost surely

$$\sup_{0\leq x\leq\infty}|(1+x)^{-\frac{1}{2}}B_x| = \infty. \quad (4.76)$$

So the well-definedness of $\langle f, f\rangle_{\mathbf{H}_\beta}$ is not obvious. It is an interesting question whether a straightforward definition

$$\int_0^\infty B_x f(x)f'(x)dx = \lim_{C\to\infty}\int_0^C B_x f(x)f'(x)dx \quad (4.77)$$

leads to a satisfactory definition of $\langle f, f\rangle_{\mathbf{H}_\beta}$ such that it is almost surely a bounded operator. Below we are going to take an indirect way to solve the problem.

Consider the smoothed Brownian motion

$$\bar{B}_x = \int_x^{x+1} B_t dt. \tag{4.78}$$

Then $\bar{B}_x$ is differentiable with $\bar{B}'_x = B_{x+1} - B_x$. Denoting

$$Q_x = \bar{B}'_x = B_{x+1} - B_x, \quad \text{and} \quad R_x = B_x - \bar{B}_x, \tag{4.79}$$

for $f \in C_0^\infty(0,\infty)$, we have

$$\begin{aligned} \int_0^\infty B_x f(x) f'(x) dx &= \int_0^\infty \bar{B}_x f(x) f'(x) dx + \int_0^\infty (B_x - \bar{B}_x) f(x) f'(x) dx \\ &= -\frac{1}{2}\int_0^\infty Q_x f(x)^2 dx + \int_0^\infty R_x f(x) f'(x) dx. \end{aligned} \tag{4.80}$$

Inspired by this transform, we define for all $f \in \mathcal{L}^*$

$$\langle f, f\rangle_{\mathbf{H}_\beta} = \|f\|_*^2 - \|f\|_2^2 - \frac{2}{\sqrt{\beta}}\int_0^\infty Q_x f(x)^2 dx + \frac{4}{\sqrt{\beta}}\int_0^\infty R_x f(x) f'(x) dx. \tag{4.81}$$

Below we show that this definition is valid.

Lemma 14. *Given any $\epsilon > 0$, almost surely for all paths of B_x, there exists a constant C_1, depending on β, ϵ and the path, such that*

$$\frac{2}{\sqrt{\beta}} \sup_{x\geq 0} \frac{|Q_x|}{C_1 + \sqrt{x}} < \frac{\epsilon}{2}, \quad \textit{and} \quad \frac{4}{\sqrt{\beta}} \sup_{x\geq 0} \frac{|R_x|}{C_1 + \sqrt{x}} < \epsilon. \tag{4.82}$$

Proof. For each $k \in \mathbb{Z}_+$, we define

$$Z_k = \max_{0\leq t\leq 1} |B_{k+t} - B_k|. \tag{4.83}$$

Then we have

$$\begin{aligned} |R_x| &\leq \max_{x\leq y\leq x+1} |B_y - B_x| \\ &\leq \max \begin{pmatrix} \max\limits_{x\leq y\leq \lfloor x\rfloor+1} (|B_x - B_{\lfloor x\rfloor}| + |B_y - B_{\lfloor x\rfloor}|), \\ \max\limits_{\lfloor x\rfloor+1\leq y\leq x+1} (|B_x - B_{\lfloor x\rfloor}| + |B_{\lfloor x+1\rfloor} - B_{\lfloor x\rfloor}| + |B_y - B_{\lfloor x\rfloor+1}|) \end{pmatrix} \\ &\leq 2Z_{\lfloor x\rfloor} + Z_{\lfloor x\rfloor+1}. \end{aligned} \tag{4.84}$$

Similarly we also have $|Q_x| \leq 2Z_{\lfloor x\rfloor} + Z_{\lfloor x\rfloor+1}$. Note that Z_k are independently and identically distributed. The distribution of Z_0 can be explicitly computed, but the following estimate suffices for us. The distribution of $M_t = \sup_{0\leq s\leq t} B_s$ is the same as $|B_t|$ by the reflection principle [11, Example 8.4.1]. Then for $t = 1$ and $x \geq 0$,

$$\mathbb{P}(M_1 > x) = \frac{2}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{t^2}{2}} dt = 2 - 2\Phi(x). \tag{4.85}$$

By the symmetry of B_t about 0, for any $x \geq 0$

$$\mathbb{P}(Z_0 > x) \leq 2\mathbb{P}(M_1 > x) = 4(1 - \Phi(x)). \tag{4.86}$$

Then for each $k \in Z_+$, if $x > 3$,

$$\begin{aligned}\mathbb{P}(2Z_k + Z_{k+1} > x) &\le \mathbb{P}(Z_k > \frac{x}{3}) + \mathbb{P}(Z_{k+1} > \frac{x}{3}) \\ &\le 8(1 - \Phi(\frac{x}{3}) = \frac{8}{\sqrt{2\pi}} \int_{\frac{x}{3}}^{\infty} e^{-\frac{t^2}{2}} dt \\ &\le \frac{8}{\sqrt{2\pi}} \int_{\frac{x}{3}}^{\infty} t e^{-\frac{t^2}{2}} dt = \frac{8}{\sqrt{2\pi}} e^{-\frac{x^2}{18}}.\end{aligned} \tag{4.87}$$

Thus for any $\epsilon > 0$,

$$\sum_{k=0}^{\infty} \mathbb{P}\left(2Z_k + Z_{k+1} > \frac{\sqrt{\beta}}{4}\sqrt{k}\right) \le \underbrace{1 + 1 + \cdots + 1}_{L\text{'s } 1} + \frac{8}{\sqrt{2\pi}} \sum_{k=L+1}^{\infty} e^{-\frac{\beta\epsilon}{288}k} < \infty, \tag{4.88}$$

where L is the smallest number such that $\sqrt{\beta k}\epsilon/4 \ge 3$. By the Borel-Cantelli lemma [9, Section 4.2], [11, Section 2.3], we have that in probability 1 there are only finitely many $k \in \mathbb{Z}_+$ such that

$$2Z_k + Z_{k+1} > \frac{\sqrt{\beta}}{4}\sqrt{k}. \tag{4.89}$$

For any path of B_x such that (4.89) holds for only finitely many k, there exists a C_1 such that for all $k \in \mathbb{Z}_+$

$$2Z_k + Z_{k+1} > \frac{\sqrt{\beta}}{4}(C_1 + \sqrt{k}). \tag{4.90}$$

Thus for all $x \in (0, \infty)$

$$\frac{4}{\sqrt{\beta}} \frac{|R_x|}{C + \sqrt{x}} < \frac{4}{\sqrt{\beta}} \frac{2Z_{\lfloor x \rfloor} + Z_{\lfloor x \rfloor + 1}}{C + \sqrt{x}} < \epsilon, \tag{4.91}$$

and we prove the lemma 14 for R_x. The proof for Q_x is identical. □

Lemma 15. *The quadratic form $\langle \cdot, \cdot \rangle_{\mathbf{H}_\beta}$ defined in* (4.81) *extends to a continuous symmetric bilinear form on $\mathcal{L}^* \times \mathcal{L}^*$ almost surely for all paths of B_x, such that there exists constants C_2, C_2' depending on the path and*

$$\frac{1}{2}\|f\|_*^2 - C_2\|f\|_2^2 \le \langle f, f \rangle_{\mathbf{H}_\beta} \le C_2'\|f\|_*^2. \tag{4.92}$$

Proof. We need to consider only $\langle f, f \rangle_{\mathbf{H}_\beta}$ for $f \in \mathcal{L}^*$. The first two terms on the right-hand side of (4.81) are well defined, so we only need to consider the last two terms. Suppose for the path of B_x there exists C_1 to make the inequality (4.82) hold with $\epsilon = \frac{1}{4}$. (The paths that are qualified are in probability 1.) Then

$$\begin{aligned}\left| \int_0^{\infty} \frac{2}{\sqrt{\beta}} Q_x f^2(x) dx \right| &\le \frac{1}{8} \int_0^{\infty} (C_1 + \sqrt{x}) f^2(x) dx \\ &\le \frac{1}{8} \left(\int_0^{\infty} (1 + x) f^2(x) dx + C_1 \int_0^{\infty} f^2(x) dx \right) \\ &\le \frac{1}{8}\|f\|_*^2 + \frac{C_1}{8}\|f\|_2^2,\end{aligned} \tag{4.93}$$

where we use the inequality $1+x>\sqrt{x}$, and

$$
\begin{aligned}
&\left|\int_0^\infty \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx\right| \\
\leq & \frac{1}{4}\left|\int_0^\infty (C_1+\sqrt{x}) f(x) f'(x) dx\right| \\
\leq & \frac{1}{4}\left|\int_0^\infty \sqrt{1+x} f(x) f'(x) dx + C_1 \int_0^\infty f(x) f'(x) dx\right| \\
\leq & \frac{1}{4}\left[\left(\int_0^\infty (1+x) f(x)^2 dx\right)^{\frac{1}{2}} \left(f'(x)^2 dx\right)^{\frac{1}{2}} + C_1 \left(f(x)^2 dx\right)^{\frac{1}{2}} \left(f'(x)^2 dx\right)^{\frac{1}{2}}\right] \\
\leq & \frac{1}{4}[\|f\|_* \|f\|_* + C_1 \|f\|_2 \|f\|_*] \\
\leq & \frac{1}{4}\left[\|f\|_*^2 + \left(\frac{1}{2}\|f\|_*^2 + \frac{C_1^2}{2}\|f\|_2^2\right)\right] \\
= & \frac{3}{8}\|f\|_*^2 + \frac{C_1^2}{8}\|f\|_2^2,
\end{aligned}
\tag{4.94}
$$

where we use the Cauchy-Schwarz inequality. Summing up all terms together, we see that the $\langle f, f\rangle_{\mathbf{H}_\beta}$ is well defined by (4.81) as long as C_1 exists, and C_2, C_2' in (4.92) are determined by C_1 as $C_2 = 1 + (C_1 + C_1^2)/8$ and $C_2' = 3/2 + (C_1 + C_1^2)/8$, where in the calculation of C_2' we use that $\|f\|_2 \leq \|f\|_*$. □

Lemma 16. *In probability* 1*, the infimum*

$$
\Lambda_0 = \inf_{g \in \mathcal{L}^*,\ \|f\|_2 = 1} \langle g, g\rangle_{\mathbf{H}_\beta} \tag{4.95}
$$

exists, and is achieved at some $f \in \mathcal{L}^*$. Furthermore, $(f, -\Lambda_0)$ is an eigenvector-eigenvalue pair for $\mathbf{H}_\beta$ with $-\Lambda_0 = \lambda_0$.

Proof. We consider a path of B_x such that C_2, C_2' in Lemma 15 exist. The set of these paths is of probability 1.

By the result of Lemma 15, we see that for all $f \in \mathcal{L}^*$ with $\|f\|_2 = 1$,

$$
\langle f, f\rangle_{\mathbf{H}_\beta} > \frac{1}{2}\|f\|_*^2 - C_2\|f\|_2^2 \geq (\frac{1}{2} - C_2)\|f\|_2^2 = \frac{1}{2} - C_2. \tag{4.96}
$$

Thus the infimum on the right-hand side of (4.95) exists.

Let $\{f_n\} \subset \mathcal{L}^*$ be a sequence such that

$$
\|f_n\|_2 = 1 \quad \text{and} \quad \lim_{n\to\infty} \langle f_n, f_n\rangle_{\mathbf{H}_\beta} = \Lambda_0. \tag{4.97}
$$

Then $\|f_n\|_*$ are bounded, since by Lemma 15 again,

$$
\frac{1}{2}\|f\|_*^2 < \langle f, f\rangle_{\mathbf{H}_\beta} + C_2\|f\|_2^2 = \langle f, f\rangle_{\mathbf{H}_\beta} + C_2. \tag{4.98}
$$

Below we denote $\|f_n\|_* < K$. Using the result of Lemma 12, we see that by passing to a subsequence if necessary, there exists $f \in \mathcal{L}^*$ that is the limit of $\{f_n\}$ in the sense of (1) L^2, (2) weak L^2 in derivative, (3) locally uniform and (4) weak $\mathcal{L}^*$. But the desired convergence in the sense of $\mathcal{L}^*$ still need to be proved. Some direct consequences of the convergence results above are:

- Due to the L^2 convergence,

$$\|f\|_2 = \lim_{n\to\infty} \|f_n\|_2 = 1. \tag{4.99}$$

- Due to the weak $\mathcal{L}^*$ convergence,

$$\|f\|_* \le \liminf_{n\to\infty} \|f_n\|_* < K. \tag{4.100}$$

By inequality (4.82) in Lemma 14, in probability 1 there exists a constant C_3 depending only on the path of B_x such that

$$\frac{2}{\sqrt{\beta}} Q_x < \frac{\epsilon}{2}(1+x), \quad \text{and} \quad \frac{4}{\sqrt{\beta}} R_x < \frac{\epsilon}{2}\sqrt{1+x}. \tag{4.101}$$

Then for all $g \in \mathcal{L}^*$, by arguments similar to those in (4.93) and (4.94), we have

$$\begin{aligned} &\left| -\int_{C_3}^{\infty} \frac{2}{\sqrt{\beta}} Q_x g(x)^2 dx + \int_{C_3}^{\infty} \frac{4}{\sqrt{\beta}} R_x g(x) g'(x) dx \right| \\ \le &\frac{\epsilon}{2} \left(\int_{C_3}^{\infty} (1+x) g(x)^2 dx + \left| \int_{C_3}^{\infty} \sqrt{1+x} g(x) g'(x) dx \right| \right) \\ \le &\frac{\epsilon}{2} (\|g\|_*^2 + \|g\|_*^2) = \epsilon \|g\|_*^2. \end{aligned} \tag{4.102}$$

Hence

$$\begin{aligned} \lim_{n\to\infty} -\int_0^{\infty} \frac{2}{\sqrt{\beta}} Q_x f_n(x)^2 dx + \int_0^{\infty} \frac{4}{\sqrt{\beta}} R_x f_n(x) f_n'(x) dx = \\ -\int_0^{\infty} \frac{2}{\sqrt{\beta}} Q_x f(x)^2 dx + \int_0^{\infty} \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx, \end{aligned} \tag{4.103}$$

since for any $\epsilon > 0$ and C_3 determined by (4.101), by L^2 convergence,

$$\lim_{n\to\infty} \int_0^{C_3} \frac{2}{\sqrt{\beta}} Q_x f_n(x)^2 dx = \int_0^{C_3} \frac{2}{\sqrt{\beta}} Q_x f(x)^2 dx, \tag{4.104}$$

by L^2 convergence and weak L^2 convergence in derivative,

$$\lim_{n\to\infty} \int_0^{C_3} \frac{4}{\sqrt{\beta}} R_x f_n(x) f_n'(x) dx = \int_0^{C_3} \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx, \tag{4.105}$$

and by the boundedness $\|f\|, \|f_n\| < K$ together with (4.102), for n large enough

$$\begin{aligned} &\left| \begin{aligned} &\left(-\int_0^{\infty} \frac{2}{\sqrt{\beta}} Q_x f_n(x)^2 dx + \int_0^{\infty} \frac{4}{\sqrt{\beta}} R_x f_n(x) f_n'(x) dx \right) \\ &\quad - \left(\int_0^{\infty} \frac{2}{\sqrt{\beta}} Q_x f(x)^2 dx + \int_0^{\infty} \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx \right) \end{aligned} \right| \\ \le &\left| \int_0^{C_3} \frac{2}{\sqrt{\beta}} Q_x f_n(x)^2 dx - \int_0^{C_3} \frac{2}{\sqrt{\beta}} Q_x f(x)^2 dx \right| \\ &+ \left| \int_0^{C_3} \frac{4}{\sqrt{\beta}} R_x f_n(x) f_n'(x) dx - \int_0^{C_3} \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx \right| \\ &+ \left| -\int_{C_3}^{\infty} \frac{2}{\sqrt{\beta}} Q_x f_n(x)^2 dx + \int_{C_3}^{\infty} \frac{4}{\sqrt{\beta}} R_x f_n(x) f_n'(x) dx \right| \\ &+ \left| -\int_{C_3}^{\infty} \frac{2}{\sqrt{\beta}} Q_x f(x)^2 dx + \int_{C_3}^{\infty} \frac{4}{\sqrt{\beta}} R_x f(x) f'(x) dx \right| \\ < &\epsilon + \epsilon + \epsilon K + \epsilon K = \epsilon(2 + 2K). \end{aligned} \tag{4.106}$$

Substituting the limit formulas (4.99), (4.100) and (4.103) into the definition formula (4.81) for $\langle f, f\rangle_{\mathbf{H}_\beta}$, We write

$$\begin{aligned}\langle f, f\rangle_{\mathbf{H}_\beta} &\leq \liminf_{n\to\infty}\|f_n\|_*^2 - \lim_{n\to\infty}\|f\|_2^2 + \lim_{n\to\infty}\left(-\int_0^{C_3}\frac{2}{\sqrt{\beta}}Q_x f(x)^2dx + \int_0^{C_3}\frac{4}{\sqrt{\beta}}R_x f(x)f'(x)dx\right)\\ &= \liminf_{n\to\infty}\langle f_n, f_n\rangle_{\mathbf{H}_\beta} = \Lambda_0.\end{aligned} \tag{4.107}$$

Thus we prove the first part of the lemma.

To prove the second part, we consider any $\phi \in C_0^\infty(0,\infty)$ and ϵ an arbitrary small real number. Then

$$f^{\phi,\epsilon} = \frac{1}{\|f+\epsilon\phi\|_2}(f+\epsilon\phi) \tag{4.108}$$

is a function in $\mathcal{L}^*$ with $\|f^{\phi,\epsilon}\|_2 = 1$. Since we assume that f solves the minimisation problem (4.95), we have

$$\begin{aligned}0 &\leq \langle f^{\phi,\epsilon}, f^{\phi,\epsilon}\rangle_{\mathbf{H}_\beta} - \langle f, f\rangle_{\mathbf{H}_\beta}\\ &= \int_0^\infty (f^{\phi,\epsilon})'(x)^2 - f'(x)^2dx + \int_0^\infty x(f^{\phi,\epsilon}(x)^2 - f(x)^2)dx\\ &\quad - \frac{2}{\sqrt{\beta}}\int_0^\infty Q_x(f^{\phi,\epsilon}(x)^2 - f(x)^2)dx + \frac{4}{\sqrt{\beta}}\int_0^\infty R_x(f^{\phi,\epsilon}(x)(f^{\phi,\epsilon})'(x) - f(x)f'(x))dx\\ &= 2\epsilon\left[-\langle f, f\rangle_{\mathbf{H}_\beta}\int_0^\infty f(x)\phi(x)dx\right.\\ &\quad \left.+ \int_0^\infty f'(x)\phi'(x) + xf(x)\phi(x) - \frac{2}{\sqrt{\beta}}Q_x f(x)\phi(x) + \frac{2}{\sqrt{\beta}}(f'(x)\phi(x) + f(x)\phi'(x))dx\right] + \mathcal{O}(\epsilon^2).\end{aligned} \tag{4.109}$$

By the standard argument for variational problem, noting that $\langle f, f\rangle_{\mathbf{H}_\beta} = \Lambda_0$ and the relation (4.79) between Q_x, R_x and B_x, we have

$$\begin{aligned}&-\Lambda_0\int_0^\infty f(x)\phi(x)dx\\ &= \int_0^\infty f(x)\left[\phi''(x) - x\phi(x) - \frac{2}{\sqrt{\beta}}R_x\phi'(x)\right] - \frac{2}{\sqrt{\beta}}(R_x f'(x)\phi(x) - Q_x f(x)\phi'(x))dx\\ &= \int_0^\infty f(x)\left[\phi''(x) - x\phi(x) - \frac{2}{\sqrt{\beta}}B_x\phi'(x)\right] - \frac{2}{\sqrt{\beta}}f'(x)B_x\phi(x)\\ &\quad - \frac{2}{\sqrt{\beta}}\int_0^\infty f'(x)\bar{B}_x\phi(x) + f(x)\bar{B}_x'\phi(x) + f(x)\bar{B}_x\phi(x)'dx\\ &= \int_0^\infty f(x)\left[\phi''(x) - x\phi(x) - \frac{2}{\sqrt{\beta}}B_x\phi'(x)\right] - \frac{2}{\sqrt{\beta}}f'(x)B_x\phi(x).\end{aligned} \tag{4.110}$$

Then $(f(x), -\Lambda_0)$ is an eigenvector-eigenvalue pair of $\mathbf{H}_\beta$.

At last we show that $-\Lambda_0 = \lambda_0$, i.e., $-\Lambda_0$ is the largest eigenvalue of $\mathbf{H}_\beta$. Suppose (g, λ) is an eigenvector-eigenvalue pair. Since $g \in \mathcal{L}^*$, there is a sequence $\phi_n \to g$ in $\mathcal{L}^*$ where $\{\phi_n\} \subset C_0^\infty(0,\infty)$. By the definition (4.72), (4.73), for all ϕ_n,

$$\lambda\int_0^\infty g(x)\phi_n(x)dx = \int_0^\infty (\mathbf{H}_\beta g)(x)\phi_n(x)dx, \tag{4.111}$$

or equivalently,

$$\lambda\langle g, \phi_n\rangle_2 = -\langle g, \phi_n\rangle_{\mathbf{H}_\beta}. \tag{4.112}$$

By Lemma 15, we can take the limit on both sides of (4.112) and get

$$\lambda\langle g, g\rangle_2 = -\langle g, g\rangle_{\mathbf{H}_\beta}, \tag{4.113}$$

or equivalently, by $\|g\|_2 = 1$, $\langle g, g\rangle_{\mathbf{H}_\beta} = -\lambda$. Thus $\lambda \leq -\Lambda_0$, and we finish the proof. □

We have considered $\lambda_0 = -\Lambda_0$, the largest eigenvalue of $\mathbf{H}_\beta$. Now define

$$\mathcal{H}_0 = \text{span of } \{f \in \mathcal{L}^* \mid \|f\|_2 = 1 \text{ and } \langle f, f\rangle_{\mathbf{H}_\beta} = \Lambda_0\}, \tag{4.114}$$

and consider the minimisation problem

$$\Lambda_1 = \inf_{f\in\mathcal{L}^*, \|f\|_2=1, f\in\mathcal{H}_0} \langle f, f\rangle_{\mathbf{H}_\beta}. \tag{4.115}$$

By the same arguments as the proof of Lemma 16, there exists an f that achieves the minimisation problem, (f, Λ_0) is an eigenvector-eigenvalue pair of $\mathbf{H}_\beta$, and $\lambda_1 = -\Lambda_1$. Recursively we can find $\lambda_2, \lambda_3, \ldots$.

At last, we show that the eigenvalues of $\mathbf{H}_\beta$ are all simple.

Lemma 17. *For each path of B_x and $\lambda \in \mathbb{R}$ where $\mathbf{H}_\beta f = \lambda f$ has a nontrivial solution $f \in \mathcal{L}^*$, the solution space is one-dimensional.*

Proof. Write the identity $(\mathbf{H}_\beta - \lambda)f = 0$ as

$$\int_0^\infty f(x)\left[\phi''(x) - (x+\lambda)\phi(x) - \frac{2}{\sqrt{\beta}}B_x\phi'(x)\right] - \frac{2}{\sqrt{\beta}}f'(x)B_x\phi(x)dx = 0, \tag{4.116}$$

or equivalently,

$$\int_0^\infty \left[-f'(x) + \int_0^x (t+\lambda)f(t)dt - \frac{2}{\sqrt{\beta}}B_x f(x) + \frac{2}{\sqrt{\beta}}\int_0^x B_t f'(t)dt\right]\phi'(x)dx, \tag{4.117}$$

for all $\phi \in C_0^\infty(0,\infty)$. The identity (4.117) holds for all ϕ if and only if the formula in the bracket is a constant. Using the identity

$$\int_0^x (t+\lambda)f(t)dt = \frac{1}{2}(x+\lambda)^2\int_0^x f'(t)dt - \int_0^x \frac{1}{2}(t+\lambda)^2 f'(t)dt, \tag{4.118}$$

we express the condition that the formula in the bracket in (4.117) as

$$f'(x) = C + \int_0^x \left(\frac{1}{2}((x+\lambda)^2 - (t+\lambda)^2) - \frac{2}{\sqrt{\beta}}(B_x - B_t)\right)f'(t)dt. \tag{4.119}$$

To show that the solution space of the integral equation (4.119) is one-dimensional, we only need to show that for $C = 0$, (4.119) has only the trivial solution $f(x) = 0$. The (4.119) shows that $f'(x)$ is continuous (if it is the solution to $\mathbf{H}_\beta f = \lambda f$). For any $T > 0$, there is an upper bound C' for $\frac{1}{2}((x+\lambda)^2 - (t+\lambda)^2) - \frac{2}{\sqrt{\beta}}(B_x - B_t)$ with $x, t \in [0, T]$. Thus for all $x \in [0, T]$,

$$|f'(x)| \leq \int_0^x C'|f'(t)|dt. \tag{4.120}$$

This is an integral form of Gronwell inequality [6], and it implies that $f'(x) = 0$ for all $x \in [0, T]$. Since T is arbitrary, it shows that $f'(x) = 0$ for $x \geq 0$. □

Below we prove part 2 of Theorem 15. Recall that H_N is a tridiagonal matrix whose diagonal entries $h_{i,i}$ are in normal distribution and off-diagonal entries $h_{i,i+1}$ are in χ distribution, as defined in (4.5) and (4.6). Define

$$y_{N,1}(x) = N^{-\frac{1}{6}} \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} h_{i,i} = N^{-\frac{1}{6}} \sqrt{\frac{2}{\beta}} \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} \xi_i, \tag{4.121}$$

$$y_{N,2}(x) = 2N^{-\frac{1}{6}} \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} (\sqrt{N} - h_{i,i+1}) = N^{-\frac{1}{6}} \frac{2}{\sqrt{\beta}} \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} (\sqrt{\beta N} - Y_{N-i}). \tag{4.122}$$

Lemma 18. *There exists a probability space supporting the processes $y_{N,1}(x), y_{N,2}(x)$ for all $N \in \mathbb{Z}_+$ and two independent Brownian motion $B_x(1), B_x(2)$ such that with respect to the Skorokhod topology [12, Section3.5] on $D_{\mathbb{R}}[0,\infty)$,*

$$y_{N,1} \to \sqrt{\frac{2}{\beta}} B_x(1), \quad y_{N,2} \to \sqrt{\frac{2}{\beta}} B_x(2) + \frac{x^2}{2}. \tag{4.123}$$

Sketch of the proof. The convergence of $y_{N,1}$ to Brownian motion is Donsker's theorem, and the convergence of $y_{N,2}$ can be proved similarly, since each term in the summation (4.122) converges in distribution to normal distribution as $N \to \infty$. □

As $N \to \infty$, $y_{N,1}(x)$ converges to a Brownian motion and $y_{N,2}(x)$ converges to a Brownian motion with a drift. We separate the Brownian motion part and the drift part as

$$y_{N,2}(x) = \bar{y}_{N,2}(x) + \tilde{y}_{N,2}(x), \tag{4.124}$$

where

$$\bar{y}_{N,2}(x) = \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} \eta_i, \quad \tilde{y}_{N,2}(x) = \sum_{i=1}^{\lfloor xN^{\frac{1}{3}} \rfloor} \gamma_i, \tag{4.125}$$

and

$$\eta_i = N^{-\frac{1}{6}} \frac{2}{\sqrt{\beta}} (\sqrt{\beta N} - \mathrm{E}Y_{N-i}), \quad \gamma_i = N^{-\frac{1}{6}} \frac{2}{\sqrt{\beta}} (\mathrm{E}Y_{N-i} - Y_{N-i}). \tag{4.126}$$

We can compute $\sqrt{\beta N} - \mathrm{E}Y_{N-i}$ explicitly, but for our purpose we are satisfied with the inequality

$$\frac{i}{\kappa\sqrt{N}} - \kappa \le \eta_i \le \frac{\kappa i}{\sqrt{N}} + \kappa \tag{4.127}$$

for some constant κ. Like the estimate in Lemma 14, we have that for any $\epsilon > 0$ there is a tight sequence of random variables $\kappa_{N,\epsilon}$ such that

$$\sup_{x \le t \le x+1} |Y_{N,1}(t) - Y_{N,1}(x)| \le \epsilon x + \kappa_{N,\epsilon}, \qquad \sup_{x \le t \le x+1} |\tilde{Y}_{N,2}(t) - \tilde{Y}_{N,2}(x)| \le \epsilon x + \kappa_{N,\epsilon}. \tag{4.128}$$

We define a norm $\|\cdot\|_{N,*}$ on the space of N-dimensional vector $\mathbb{R}^N$, such that (denoting $v_{N+1} = 0$)

$$\|v\|_{N,*}^2 = \sum_{i=1}^{N} (N^{\frac{1}{3}}(v_{i+1} - v_i))^2 N^{-\frac{1}{3}} + \sum_{i=1}^{N} (1 + N^{-\frac{1}{3}} i) v_i^2 N^{-\frac{1}{3}}. \tag{4.129}$$

We identify an N-dimensional vector as a step function on $[0, N^{2/3}]$ as in (4.52), and then $\|\cdot\|_{N,*}$ is a norm defined on the space of the step functions that we denote by $L^{N,2}$. This norm

gives a symmetric bilinear form $\langle\cdot,\cdot\rangle_{N,*}$ on $L^{N,2}$ as the norm $\|\cdot\|_*$ gives $\langle,\cdot,\cdot\rangle_*$ on $\mathcal{L}^*$. We note that as $N\to\infty$, $\|\cdot\|_{N,*}$ converges to $\|\cdot\|_*$ in some sense. Also $L^{N,2}$ can be viewed as a subspace of $L^2(0,\infty)$, where $f(x)\in L^{N,2}$ is piecewise constant for $x<N^{2/3}$ and $f(x)=0$ for $x\geq N^{2/3}$. Thus $L^{N,2}$ has the natural norm $\|\cdot\|_2$ and inner product $\langle\cdot,\cdot\rangle_2$ inherited from $L^2(0,\infty)$.

From the heuristic argument in the beginning of this subsection, we expect that $\hat{H}_N\to\mathbf{H}_\beta$ as $N\to\infty$. We interpret the convergence in the associated bilinear forms. Analogous to $\langle\cdot,\cdot\rangle_{\mathbf{H}_\beta}$ defined by $\mathbf{H}_\beta$, we consider the bilinear form $-\langle f,\hat{H}_N g\rangle_2$ for $f,g\in L^{N,2}$. Since $L^{N,2}=\mathbb{R}^N$ is finite dimensional, the bilinear form is clearly well defined. Since $\hat{H}_N$ is a symmetric matrix, it is well known that its top eigenvalues are characterised recursively by

$$\text{the } k\text{-th largest eigenvalue of } \hat{H}_N = \max_{\substack{f\in L^{N,2},\ f\perp\text{eigenvectors}\\ \text{associated to the top } k-1 \text{ eigenvalues}}} \frac{\langle f\hat{H}_N f\rangle_2}{\langle f,f\rangle_2}. \tag{4.130}$$

Analogous to Lemma 15, we have

Lemma 19. *There exists a tight sequence of random variables $c_i=c_i(N)$, $i=1,2,3$, such that*

$$c_1\|v\|_{N,*}^2-c_2\|v\|_2^2\leq-\langle v,\hat{H}_N v\rangle_2\leq c_3\|v\|_{N,*}^2. \tag{4.131}$$

Sketch of the proof. Write

$$-\langle v,\hat{H}_N v\rangle_2=S_1+S_2-S_3+S_4, \tag{4.132}$$

where

$$\begin{aligned} S_1&=N^{\frac{1}{3}}\sum_{i=1}^N(v_{i+1}-v_i)^2, \quad S_2=2N^{-\frac{1}{6}}\sum_{i=1}^N\eta_i v_i v_{i+1},\\ S_3&=\sqrt{\frac{2}{\beta}}N^{-\frac{1}{6}}\sum_{i=1}^N v_i^2\xi_i, \qquad S_4=2N^{-\frac{1}{6}}\sum_{i=1}^N\gamma_i v_i v_{i+1}. \end{aligned} \tag{4.133}$$

Then apply the estimates (4.127) and (4.128). □

The tightness of the random variables $c_i(N)$ in Lemma 19 implies that any subsequence $\{N_l\}$ of $\{N\}$ contains a further subsequence $\{N_{l_i}\}$ such that the estimate (4.131) holds with random variables c_i, independent of N_{l_i}.

For each N define the projection operator $P_N:L^2\to L^{N,2}$ such that

$$(P_N)f(x)=\begin{cases}\text{the average of } f(x) \text{ on } [iN^{-\frac{1}{3}},(i+1)N^{-\frac{1}{3}}) & \text{if } x\in[iN^{-\frac{1}{3}},(i+1)N^{-\frac{1}{3}}) \text{ and } x<N^{\frac{2}{3}},\\ 0 & \text{otherwise.}\end{cases} \tag{4.134}$$

Then the operator $\hat{H}_N P_N$ is an operator from $L^2(0,\infty)$ to $L^{N,2}\subset L^2(0,\infty)$. The convergence of $\hat{H}_N$ to $\mathbf{H}_\beta$ is described by the following two lemmas. In the statement of Lemma 20, we use the difference operator in $L^{N,2}$

$$D_N f(x)=N^{\frac{1}{3}}(f(x+N^{-\frac{1}{3}})-f(x)). \tag{4.135}$$

Lemma 20. *Let $f\in\mathcal{L}^*$, $f_N\in L^{N,2}$ and suppose $f_N\to f$ weakly in L^2 and $D_N f_N\to f'$ weakly in L^2. Then for all $\phi\in C_0^\infty(0,\infty)$,*

$$-\langle\phi,\hat{H}_N f_N\rangle_2\to\langle\phi,f\rangle_{\mathbf{H}_\beta}, \tag{4.136}$$

and in particular

$$-\langle P_N\phi,\hat{H}_N P_N\phi\rangle_2\to\langle\phi,\hat{H}_N P_N\phi\rangle_2\to\langle\phi,\phi\rangle_{\mathbf{H}_\beta}. \tag{4.137}$$

Lemma 21. *Let $f_N \in L^{N,2}$ with $\|f_N\|_2 = 1$ and $\|f_N\|_{N,*} \le c$. Then there exists a subsequence $\{N_l\}$ and $f \in \mathcal{L}^*$ such that $f_{N_l} \to f$ in $L^2(0,\infty)$ and for all $\phi \in C_0^\infty(0,\infty)$,*

$$-\langle \phi, \hat{H}_N f_{N_l}\rangle_2 \to \langle \phi, f\rangle_{\mathbf{H}_\beta}. \tag{4.138}$$

Below we prove part 2 of Theorem 15. Write the k top eigenvalues of $\hat{H}_N$ as $\eta_{N,i} = N^{1/6}(\lambda_{N-i}^N - 2\sqrt{N})$, $i = 0, \dots, k-1$. Let $v_{N,i} \in \mathbb{R}^N$ be the associated eigenvectors such that $\|v_{N,i}\|_2 = 1$.

First we show that

$$\eta_i^+ := \limsup_{N\to\infty} n_{N,i} \le \lambda_i, \quad i = 0, \dots, k-1. \tag{4.139}$$

Without loss of generality, we assume $\eta_i^+ > -\infty$ for all $i < k$. Then for each i, there is a subsequence N_l such that $(\eta_{N_l,0}, \dots, \eta_{N_l,i}) \to (\xi_0, \dots, \xi_i = \eta_i^+)$. By Lemma 19, for $j = 0, \dots, i$, $\|v_{N_l,j}\|_{N,*}$ are uniformly bounded, and then a subsequence of $v_{N_l,j}$ converges in L^2 to a limit function $g_j \in \mathcal{L}^*$. For notational simplicity, we denote the also subsequence by $v_{N_l,j}$. Then we have that $(g_0, \xi_1), \dots, (g_i, \xi_i = n_i^+)$ are eigenvector-eigenvalue pairs of $\mathbf{H}_\beta$ where $\xi_0 \ge \cdots \ge \xi_i$. Since for each N_l, $v_{N_l,0}, \cdots, v_{N_l,i}$ are orthogonal to each other in $L^2(0,\infty)$, we have that $g_0, \dots, g_i$ are orthogonal in $L^2(0,\infty)$ to each other. Then $\xi_0, \dots, \xi_i$ are distinct since the eigenspaces of $\mathbf{H}_\beta$ are 1-dimensional. Hence $\xi_0 \le \lambda_0, \dots, \lambda_i = \eta_i^+ \le \lambda_i$, and we prove (4.139).

Next we show that

$$\eta_i^- := \liminf_{N\to\infty} n_{N,i} \ge \lambda_i., \quad i = 0, \dots, k-1. \tag{4.140}$$

and if the signs of $v_{N,i}$ and f_i properly, we have the L^2 convergence as $N \to \infty$

$$v_{N,i} \to f_i, \quad , i = 0, \dots, k-1. \tag{4.141}$$

We prove (4.140) and (4.141) inductively in i. Suppose for $j < i$, (4.140) and (4.141) hold with i replaced by j. (Note that if $i = 0$, then the inductive assumption is void.) For any $\epsilon > 0$, let $f_i^\epsilon \in C_0^\infty(0,\infty)$ satisfy

$$\|f_i^\epsilon - f_i\|_* \le \epsilon, \tag{4.142}$$

and set

$$f_{N,k} = P_N f_k^\epsilon - \sum_{j=0}^{i-1} \langle v_{N,j}, P_N f_i^\epsilon\rangle_2 v_{N,j}. \tag{4.143}$$

In (4.143), $\|v_{N,j}\|_{N,*}$ is bounded by Lemma 19. Also by the orthogonality of f_i and f_j in $L^2(0,\infty)$,

$$\begin{aligned} |\langle v_{N,j}, P_N f_i^\epsilon\rangle_2| &= |\langle f_j + (v_{N,j} - f_j), f_i + (f_i^\epsilon - f_i)\rangle_2| \\ &= |\langle f_j, (f_i^\epsilon - f_i)\rangle_2| + |\langle (v_{N,j} - f_j), f_i\rangle_2| + |\langle (v_{N,j} - f_j), (f_i^\epsilon - f_i)\rangle_2| \\ &\le \|f_i^\epsilon - f_i\|_2 + \|v_{N,j} - f_j\|_2 + \|f_i^\epsilon - f_i\|_2 \|v_{N,j} - f_j\|_2. \end{aligned} \tag{4.144}$$

By the assumption 4.142 and the convergence (4.141) for j, we have that if N is large enough, $|\langle v_{N,j}, P_N f_i^\epsilon\rangle_2| < 2\epsilon$. Thus we have that for N large enough,

$$\|f_{N,i} - P_N f_i^\epsilon\|_{N,*} < c\epsilon, \tag{4.145}$$

where c is a random constant independent of N. By (4.130), $\eta_{N,i} \ge \langle f_{N,i}, f_{N,i}\rangle_2^{-1} \langle f_{N,i} \hat{H}_N f_{N,i}\rangle_2$, so by Lemma 20

$$\eta_i^- \ge \liminf_{N\to\infty} \frac{\langle f_{N,i} \hat{H}_N f_{N,i}\rangle_2}{\langle f_{N,i}, f_{N,i}\rangle_2} = \liminf_{N\to\infty} \frac{\langle P_N f_i^\epsilon, \hat{H}_N P_N f_i^\epsilon\rangle_2}{\langle P_N f_i^\epsilon\rangle_2} + s(\epsilon), \tag{4.146}$$

where $s(\epsilon) \to 0$ as $\epsilon \to 0$. By (4.137),

$$\lim_{N\to\infty} \langle P_N f_i^\epsilon, \hat{H}_N P_N f_i^\epsilon \rangle_2 = \langle f_k^\epsilon, f_k^\epsilon \rangle_{\mathbf{H}_\beta} = \langle f_i, f_i \rangle_{\mathbf{H}_\beta} + s'(\epsilon), \tag{4.147}$$

where $s'(\epsilon) \to 0$ as $\epsilon \to 0$. Also it is not difficult to see that

$$\langle P_N f_i^\epsilon \rangle_2 \to \langle f_i, f_i \rangle_2 = 1 \tag{4.148}$$

as $N \to \infty$. Thus by (4.146), (4.147) and (4.148), we obtain (4.140) after taking $\epsilon \to 0$.

To prove (4.141), note that each subsequence of $v_{N,i}$ has a further subsequence, say $v_{N_l,i}$, converging in L^2 to $g \in \mathcal{L}^*$ such that

$$\langle \phi, g \rangle_{\mathbf{H}_\beta} = \lim_{l\to\infty} \langle \phi, \hat{H}_{N_l} v_{N_l,i} \rangle_2 = \lim_{l\to\infty} n_{N_l,i} \langle \phi, v_{N_l,i} \rangle_2 = \lambda_i \langle \phi, g \rangle_2, \tag{4.149}$$

Hence $g = f_i$, and we prove (4.141).

Exercises

1. Prove that the joint probability density function of the eigenvalues of the random tridiagonal matrix H_N with $\beta = 1$ is the same as that of the eigenvalues of an $N \times N$ GOE random matrix whose diagonal entries are in $\mathrm{N}(0, 2)$ and upper triangular entries in $\mathrm{N}(0, 1)$.

   Hint: Follow the argument in the proof for $\beta = 2$.

A Some related results in linear algebra

A.1 Normal forms under the unitary congruence group for a complex symmetric matrix (Takagi's factorisation) and for an complex anti-symmetric matrix

In this appendix, we show that

1. Given any $n \times n$ complex symmetric matrix M, there is a unitary matrix $U \in \mathrm{U}(n)$ and a diagonal matrix D such that
$$M = UDU^T \tag{A.1}$$
and the diagonal entries of D are square roots of the eigenvalues of $M\bar{M}$.

2. Given any $n \times n$ complex anti-symmetric matrix M, there is a unitary matrix $U \times \mathrm{U}(n)$ and a block diagonal matrix J such that
$$M = UJU^T \tag{A.2}$$
and
$$J = \mathrm{diag}(\begin{pmatrix} 0 & \sqrt{a_1} \\ -\sqrt{a_1} & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & \sqrt{a_r} \\ -\sqrt{a_r} & 0 \end{pmatrix}, 0, \ldots, 0), \tag{A.3}$$
where a_i are eigenvalues of $-M\bar{M}$.

The result in the symmetric case is called Takagi's factorisation, first discovered by Takagi [24] in 1925, and reproved by Jacobson [18], Hua [17] and Schur [21], and The text book [16] by Horn and Johnson is a good reference. In the anti-symmetric case, the original reference can be traced back to [17]. An elementary proof for Youla decomposition, which generalises the symmetric and antisymmetric cases to general complex square matrix case, was provided by Youla [28]. Here our proofs are based on the approach in [16, Section 4.4].

Proof of the symmetric case. First, $M\bar{M} = MM^*$ is a nonnegative definite Hermitian matrix, so it has nonnegative eigenvalues $a_1, \ldots, a_n$. Let $\mathbf{v} \in \mathbb{C}^n$ be an eigenvector of $M\bar{M}$ such that $M\bar{M}\mathbf{v} = a_1\mathbf{v}$. Then $\bar{M}M(\bar{M}\mathbf{v}) = a_1\bar{M}\mathbf{v}$, i.e., $\bar{M}\mathbf{v}$ is an eigenvector of $\bar{M}M$. Taking complex conjugate on both sides, we have

$$\overline{\bar{M}M(\bar{M}\mathbf{v})} = a_1\overline{\bar{M}\mathbf{v}}, \quad \text{or equivalently} \quad M\bar{M}(M\bar{\mathbf{v}}) = a_1 M\bar{\mathbf{v}}, \tag{A.4}$$

and see that $M\bar{\mathbf{v}}$ is an eigenvector of $M\bar{M}$ with eigenvalue a_1. If $M\bar{\mathbf{v}}$ is dependent on $\mathbf{v}$, i.e., $M\bar{\mathbf{v}} = \lambda\mathbf{v}$, then $\bar{M}\mathbf{v} = \bar{\lambda}\bar{\mathbf{v}}$, and $M\bar{M}\mathbf{v} = M\bar{\lambda}\bar{\mathbf{v}} = |\lambda|^2\mathbf{v}$. We find that $\lambda = \sqrt{a_1}e^{i\theta}$. Let the unit vector $\mathbf{u} = e^{-i\theta/2}\mathbf{v}/\|\mathbf{v}\|$, then

$$M\bar{\mathbf{u}} = \sqrt{a_1}\mathbf{u}. \tag{A.5}$$

Otherwise if $M\bar{\mathbf{v}} = \mathbf{w}$ is independent of $\mathbf{v}$, we have

$$M\bar{\mathbf{w}} = M\bar{M}\mathbf{v} = a_1\mathbf{v}. \tag{A.6}$$

Let the unit vector $\mathbf{u} = w + \sqrt{a_1}\mathbf{v}$, we have

$$M\bar{\mathbf{u}} = M(\bar{\mathbf{w}} + \sqrt{a_1}\bar{\mathbf{v}}) = a_1\mathbf{v} + \sqrt{a_1}\mathbf{w} = \sqrt{a_1}\mathbf{u}. \tag{A.7}$$

In either case, there is a unitary matrix U_1 such that $U_1\mathbf{e}_1 = \mathbf{u}$. Then

$$M = U_1 M' U_1^T, \quad \text{where} \quad M' = \begin{pmatrix} 1 & * & \ldots & * \\ 0 & & & \\ \vdots & & M_2 & \\ 0 & & & \end{pmatrix}. \tag{A.8}$$

Note that $M' = U_1^{-1}M(U_1^{-1})^T$ is symmetric, and so its first row is $(1, 0, \ldots, 0)$, and M_2 is symmetric. Furthermore,

$$M'\bar{M}' = U_1^{-1}M(U_1^{-1})^T\bar{U}_1^{-1}\bar{M}\overline{(U_1^{-1})^T} = U_1^{-1}M\bar{U}_1(\bar{U})_1^{-1}\bar{M}U_1 = U_1^{-1}M\bar{M}U_1, \tag{A.9}$$

and has the same spectrum as $M\bar{M}$. Hence $M_2\bar{M}_2$ has eigenvalues $a_2, \ldots, a_n$.

Repeating this procedure, we have

$$M = U_n \cdots U_2 U_1 \begin{pmatrix} \sqrt{a_1} & & \\ & \ddots & \\ & & \sqrt{a_n} \end{pmatrix} U_1^T U_2^T \cdots U_n^T, \tag{A.10}$$

and prove Tagaki's factorization, the result in the symmetric case. □

Proof of the anti-symmetric case. First, note that $M\bar{M} = -MM^*$ is a nonpositive definite Hermitian matrix. Without loss of generality, suppose it has a negative eigenvalue $-a_1$. If $\mathbf{v}$ is an eigenvector of $M\bar{M}$ such that $M\bar{M}\mathbf{v} = -a_1\mathbf{v}$, then $M\bar{\mathbf{v}}$ cannot be equal to $\lambda\mathbf{v}$, otherwise by the argument in the proof of the symmetric case, we have $|\lambda|^2 = -a_1$, a contradiction. Thus $\mathbf{w} := M\bar{\mathbf{v}}$ is independent to $\mathbf{v}$, and by the same argument as in the proof of the symmetric case we see that $\mathbf{w}$ is also an eigenvector of $M\bar{M}$ associated to the eigenvalue $-a_1$.

Let $\mathbf{u}_1 = \mathbf{v}/\|\mathbf{v}\|$ and $\mathbf{u}_2$ be a unit vector in $\text{span}(\mathbf{v}, \mathbf{w})$ that is orthogonal to $\mathbf{u}_1$. Note that there is a angular parameter not specified for $\mathbf{u}_2$, and we are going to fix it later. Both $\mathbf{u}_1$ and $\mathbf{u}_2$ are eigenvectors of $M\bar{M}$ with eigenvalue $-a_1$, and we denote

$$M\bar{\mathbf{u}}_i = c_{1i}\mathbf{u}_1 + c_{2i}\mathbf{u}_2, \quad \text{for } i = 1, 2. \tag{A.11}$$

Suppose U_1 is a unitary matrix such that $U_1\mathbf{e}_1 = \mathbf{u}_1, U_1\mathbf{e}_2 = U\mathbf{u}_2$. Then

$$M = U_1 M' U_1^T, \quad \text{where} \quad M' = \begin{pmatrix} c_{11} & c_{12} & * & \dots & * \\ c_{21} & c_{22} & * & \dots & * \\ 0 & 0 & & & \\ \vdots & \vdots & & M_2 & \\ 0 & 0 & & & \end{pmatrix}. \tag{A.12}$$

Since M is anti-symmetric, $M' = U_1^{-1} M (U_1^{-1})^T$ is also anti-symmetric, and so its first two rows become

$$\begin{pmatrix} c_{11} & c_{12} & * & \dots & * \\ c_{21} & c_{22} & * & \dots & * \end{pmatrix} = \begin{pmatrix} 0 & \lambda & 0 & \dots & 0 \\ -\lambda & 0 & 0 & \dots & 0 \end{pmatrix}, \tag{A.13}$$

and M_2 is anti-symmetric. Note that (A.12) and (A.13) imply

$$M\bar{\mathbf{u}}_1 = -\lambda \mathbf{u}_2, \quad M\bar{\mathbf{u}}_2 = \lambda \mathbf{u}_1. \tag{A.14}$$

Using (A.14), we see that $M\bar{M}\mathbf{u}_1 = M(-\lambda\mathbf{u}_2) = -|\lambda|^2\mathbf{u}_1$, and then obtain $\lambda = \sqrt{a_1}e^{i\theta}$ by the property $M\bar{M}\mathbf{u}_1 = -a_1\mathbf{u}_1$. Finally by a change of variable $\mathbf{u}_2 \mapsto e^{i\theta}\mathbf{u}_2$ and use the new $\mathbf{u}_2$ instead of the original one used above, we find the factorisation

$$M = U_1 \begin{pmatrix} 0 & \sqrt{a_1} & 0 & \dots & 0 \\ -\sqrt{a_1} & 0 & 0 & \dots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & M_2 & \\ 0 & 0 & & & \end{pmatrix}. \tag{A.15}$$

Repeating this argument, we prove the result in anti-symmetric case. □

A.2 Proof of Hoffman-Wielandt theorem

In the lecture notes we only consider the theorem of Hoffman and Wielandt for Hermitian matrices. The original theorem discovered by these two authors [15] is more general, for normal matrices.

Theorem 16 (Hoffman-Wielandt)**.** *Let A and B be $n \times n$ matrices with eigenvalues $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ respectively. Then there is a permutation $\sigma \in S_n$, the symmetric group of $\{1, \dots, n\}$, such that*

$$\sum_{k=1}^{n} |\alpha_k - \beta_{\sigma(k)}|^2 \le \|A - B\|^2, \tag{A.16}$$

where $\|\cdot\|$ is the Frobenius norm (aka Hilbert-Schmidt norm) of a matrix such that

$$\|(x_{ij})_{i,j=1}^n\| = \left(\sum_{i,j=1}^{n} |x_{ij}|^2 \right)^2. \tag{A.17}$$

Remark 4. For any square matrix X,

$$\|X\|^2 = \mathrm{Tr}(X^* X). \tag{A.18}$$

Remark 5. If A and B are Hermitian, i.e., α_j's and β_j's are all real, then the minimum of $\sum_{k=1}^n |\alpha_k - \beta_{\pi(k)}|^2$ is obtained when both α_j's and β_j's are in increasing order. Thus Theorem 4 becomes a special case of 16.

Proof. Since A and B are normal matrices, they can be written as

$$A = U_1^* D_1 U_1, \quad B = U_2^* D_2 U_2, \tag{A.19}$$

where $D_1 = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$, $D_2 = (\beta_1, \ldots, \beta_n)$ and $U_1, U_2 \in \mathrm{U}(n)$ are unitary matrices.

With the help of (A.18), we write (A.16) as

$$\sum_{k=1}^n |\alpha_k|^2 + \sum_{k=1}^n |\beta_k|^2 + \sum_{k=1}^n (\bar{\alpha}_k \beta_{\sigma(k)} + \alpha_k \bar{\beta}_{\sigma(k)}) \leq \mathrm{Tr}\, A^* A + \mathrm{Tr}\, B^* B + \mathrm{Tr}(A^* B + A B^*). \tag{A.20}$$

It is clear that

$$\sum_{k=1}^n |\alpha_k|^2 = \mathrm{Tr}\, D_1^* D_1 = \mathrm{Tr}(U_1^* D_1^* U_1 U_1^* D_1 U_1) = \mathrm{Tr}\, A^* A, \tag{A.21}$$

and similarly $\sum_{k=1}^n |\beta_k|^2 = \mathrm{Tr}\, B^* B$. Then we need only to show that there is a $\sigma \in S_n$ such that

$$\sum_{k=1}^n (\bar{\alpha}_k \beta_{\sigma(k)} + \alpha_k \bar{\beta}_{\sigma(k)}) \geq \mathrm{Tr}(A^* B + A B^*). \tag{A.22}$$

Define $V = U_2 U_1^*$, suppose $V = (v_{ij})_{i,j=1}^n$ and $|v_{ij}|^2 = w_{ij}$. We have

$$\begin{aligned} Tr(A^* B + A B^*) &= \mathrm{Tr}(U^* D_1^* U_1 U_2^* D_2 U_2 + U^* D_1 U_1 U_2^* D_2^* U_2) \\ &= \mathrm{Tr}\, U_1^* (D_1^* V^* D_2 V + D_1 V^* D_2 V) U_1 \\ &= \mathrm{Tr}(D_1^* V^* D_2 V + D_1 V^* D_2 V) \\ &= \sum_{i,j=1}^n (\bar{\alpha}_i \beta_j + \alpha_i \bar{\beta}_j) |v_{ij}|^2 \\ &= \sum_{i,j=1}^n (\bar{\alpha}_i \beta_j + \alpha_i \bar{\beta}_j) w_{ij}. \end{aligned} \tag{A.23}$$

Note that $(w_{ij})_{i,j=1}^n$ is a doubly stochastic matrix in the sense that

$$\sum_{i=1}^n w_{ij} = 1, \quad \sum_{j=1}^n w_{ij} = 1, \quad w_{ij} \geq 0, \quad \text{for } i, j = 1, \ldots, n. \tag{A.24}$$

A theorem of Birkhoff [16, Section 8.7], [7] shows that the set of all $n \times n$ doubly stochastic matrices form an $(n-1)^2$ dimensional polytope in the n^2 dimensional space of all real $n \times n$ matrices, denoted as $\mathfrak{X}_n$, and this polytope has permutation matrices as its vertices.

Consider the right-hand side of (A.23) as a linear function of w_{ij} on the domain $\mathfrak{X}_n$, the set of doubly stochastic matrices (actually $\{(w_{ij})_{i,j=1}^n\}$ is a subset of $\mathfrak{X}_n$). It is an (almost) obvious and well known (at least in linear programming) that a linear function on a polytope attains its maximum at a vertex. In our case, it implies that

$$\min_{(w_{ij})_{i,j=1}^n \in \mathfrak{X}_n} \sum_{i,j=1}^n (\bar{\alpha}_i \beta_j + \alpha_i \bar{\beta}_j) w_{ij} = \sum_{k=1}^n (\bar{\alpha}_k \beta_{\sigma(k)} + \alpha_k \bar{\beta}_{\sigma(k)}), \tag{A.25}$$

where the permutation σ corresponds to the vertex that is a permutation matrix. Then we prove Theorem 16 with the σ in (A.16) the same as the σ in (A.25). □

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[3] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[4] C. Andréief. Note sur une relation entre les intégrales définines des produits des fonctions. *Mém. Soc. Sci. Phys. Nat. Bordeaux (3)*, 2:1–14, 1886.

[5] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[6] Richard Bellman. *Stability theory of differential equations*. McGraw-Hill Book Company, Inc., New York-Toronto-London, 1953.

[7] Garrett Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A.*, 5:147–151, 1946.

[8] Alexander Bloemendal. *Finite Rank Perturbations of Random Matrices and Their Continuum Limits*. ProQuest LLC, Ann Arbor, MI, 2011. Thesis (Ph.D.)–University of Toronto (Canada).

[9] Kai Lai Chung. *A course in probability theory*. Academic Press Inc., San Diego, CA, third edition, 2001.

[10] R. Courant and D. Hilbert. *Methods of mathematical physics. Vol. I*. Interscience Publishers, Inc., New York, N.Y., 1953.

[11] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[12] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[13] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[14] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[15] A. J. Hoffman and H. W. Wielandt. The variation of the spectrum of a normal matrix. *Duke Math. J.*, 20:37–39, 1953.

[16] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[17] Loo-Keng Hua. On the theory of automorphic functions of a matrix variable. I. Geometrical basis. *Amer. J. Math.*, 66:470–488, 1944.

[18] N. Jacobson. Normal Semi-Linear Transformations. *Amer. J. Math.*, 61(1):45–58, 1939.

[19] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[20] José A. Ramírez, Brian Rider, and Bálint Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *J. Amer. Math. Soc.*, 24(4):919–944, 2011.

[21] Issai Schur. Ein Satz ueber quadratische Formen mit komplexen Koeffizienten. *Amer. J. Math.*, 67:472–480, 1945.

[22] Denis Serre. *Matrices*, volume 216 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2010. Theory and applications.

[23] Gábor Szegő. *Orthogonal polynomials*. American Mathematical Society, Providence, R.I., fourth edition, 1975. American Mathematical Society, Colloquium Publications, Vol. XXIII.

[24] Teiji Takagi. On an algebraic problem related to an analytic theorem of Caratheodory and Fejer and on an allied theorem of Landau. *Japan. J. Math.*, 1(1):83–93, 1924.

[25] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[26] Steven Weinberg. *The quantum theory of fields. Vol. I.* Cambridge University Press, Cambridge, 2005. Foundations.

[27] Eugene P. Wigner. *Group theory: And its application to the quantum mechanics of atomic spectra*. Expanded and improved ed. Translated from the German by J. J. Griffin. Pure and Applied Physics. Vol. 5. Academic Press, New York, 1959.

[28] D. C. Youla. A normal form for a matrix under the unitary congruence group. *Canad. J. Math.*, 13:694–704, 1961.