BUTE Institute of Mathematics
Department for Analysis

Júlia Réffy

Asymptotics of random unitaries

PhD thesis

Supervisor: Dénes Petz
Professor, Doctor of the Mathematical Sciences

2005

Contents

Acknowledgements

I am deeply indebted to my supervisor Prof. Dénes Petz for drawing my attention to the theory of random matrices and its applications, and for his devoted support.

I am also thankful to Attila Andai for helping me over some technical difficulties.

Finally I am grateful to my whole family for encouraging me all through my studies and work.

Introduction

Random matrices are matrix valued random variables or in other words matrices whose entries are random variables. There are different kind of random matrices depending on the size, the distribution of the elements, and the correlation between the elements.

Wishart was the first who studied random matrices in 1928 ([46]), and he was motivated by multivariate statistics. He considered n pieces of m dimensional independent identically distributed random vectors. The covariance matrix of these random variables is the expectation of an $m \times m$ positive random matrix, what we call Wishart matrix if the components of the random vectors are normally distributed random variables.

Another point of view was given by physics. Wigner obtained some properties of the eigenvalues of complex selfadjoint or real symmetric random matrices in the papers [43, 44, 45]. He used large symmetric random matrices in order to have a model of the energy levels of nuclei.

After finding these motivations to study random matrices, in [11, 12, 13] Dyson established the mathematical foundations of random matrix theory. He made a classification of the random matrices according to their invariancy properties.

The main question was the behaviour of the eigenvalues of the random matrices. The set of eigenvalues in the above cases, when the matrix self-adjoint consist of n identically distributed but not independent real valued random variables. If we have the joint eigenvalue density, then we have all the information about the eigenvalues, but for this we need to know the joint density of the entries, and the invariance of the distribution of the random matrix under unitary conjugation. Therefore, though Wigner in [44] gave the joint eigenvalue density of the selfadjoint random matrices if the entries are Gaussian, but in the general case he studied the mean distribution of the eigenvalues. This means, that he defined the random function for an $n \times n$ random matrix A_n

$$F_n(x) := \frac{\#\{i : \lambda_i(A_n) < x\}}{n}$$

We can find the limit of the expectation of the empirical eigenvalue distribution ([43]) or the convergence of the empirical eigenvalue distribution in probability or almost surely ([1, 27, 32]). Also we can study the rate of convergence in each case ([2, 3, 20]). Others found not only the limit of the empirical eigenvalue distribution but that there is no eigenvalue outside the support of the limit measure with probability 1, i.e. the almos sure limit of the smallest and the largest eigenvalue of the random matrix is the infimum or the supremum of the support respectively.

There are theorems which are valid only in the case of Gaussian matrices and there are some universal theorems, when we need only some properties of the entries. For example the exponential rate of the convergence with some rate function (the so-called large deviation principle, see [6, 19, 23]) holds only for random matrices, where the joint density of the eigenvalues are known. But there are universal theorems, which are independent from the density of the entries. For example for the convergence of the empirical eigenvalue distribution function we need only the finiteness of some moments of the entries, and also the convergence of the smallest and the largest eigenvalues can be proven in similar ways as in the case of Gaussian matrices.

The question of non-selfadjoint matrices is also interesting. For example if all the entries are independent, identically distributed random variables, the we get a random matrix whose eigenvalues are not real. This random matrix defines a whole family of random matrices, if we take any linear combination of the matrix and its adjoint. In the Gaussian case the linear combination is also Gaussian, so it is possible to obtain the joint eigenvalue density, and the rate function for the exponential rate of convergence is found ([33]), but the same universal theorem holds as in the case of selfadjoint random matrices, i.e. the empirical eigenvalue distribution measure of the matrix (which is now a random measure on the complex plane) converges to a determonistic measure, if th forth moment of the entries is finite (see [15, 16, 17, 18]).

The other very important type of random matrices is the unitary random matrices. The construction of a random unitary matrix is different from the above random matrices, since we cannot take independent entries. The set of $n \times n$ unitary matrices is not a subspace of the set of $n \times n$ matrices, as in the previous examples, but it is a group with respect to the matrix multiplication. Therefore the matrix density is considered with respect to the translation invariant measure, the so-called Haar measure of this group, not with respect to the Lebesgue measure. The matrix which is distributed according to this measure, i.e. has uniform distribution on the set of $n \times n$ unitary matrices, is called Haar unitary random matrix. Here the eigenvalues are not real, but they are on the unit circle. By the definition of the Haar unitary, since it is invariant under multiplication by a unitary matrix, clearly it is invariant under unitary conjugation. Therefore it is possible to obtain the joint eigenvalue density function, and the convergence of the empirical eigenvalue distribution. The joint density of the eigenvalues is known, so we can prove the exponential convergence with some rate function. The correlation between the entries converges to zero, as the matrix size goes to infinity, so some kind of central limit theorems can be proven. For example the trace of any power of a Haar unitary is asymptotically normally distributed ([10, 34]), and after standardization the random variable which gives numbers of eigenvalues on a specified arc again converges to the standard normal random variable in distribution as the matrix size goes to infinity ([42]).

Random matrix theory was first used to solve statistical and physical problems, as we mentioned above. Now it play important role in number theory since strong correlation was found between the zeros of Riemann ζ function and the eigenvalues of random unitary matrices ([28]). Random matrices are useful in the noncommutative probability, since every noncommutative random variable can be approximated by a sequence of large random matrices as the matrix size goes to infinity ([39]).

There are still other random matrices to study. For example now we will deal with the $m \times m$ truncation of an $n \times n$ Haar unitary random matrix, which is a random contraction, so the eigenvalues are lying on the unit disc ([34, 35, 48]). Other family of random matrices comes from the modification of Gaussian random matrices, they are the so-called q deformed Gaussian random matrices ([37]), where the random matrix and its adjoint fulfil some commutation relations depending on $0 < q < 1$.

In this dissertation we will study most of the above topics in the following order.

In Section 1 we give an overview of different kind of random matrices. In the case of independent normally distributed entries, it is easy to determine the joint distribution of the entries. As we can see, this joint distribution can be described by the eigenvalues, so if we find the Jacobian of the transformation which transforms the entries into the eigenvalues and some independent parameters, we get the joint density of the eigenvalues. We will show a more detailed version of this calculations, which was first given by Wigner [44] and Mehta [31] in the case of selfadjoint and non-selfadjoint random matrices. Since these matrices are invariant under unitary conjugation, the joint density of the eigenvalues contains all the information about the random matrices. The other important question concerning the random matrices is the limit distribution of the sequence of the empirical eigenvalue distribution as the matrix size goes to infinity. We will consider first the random matrices with independent normally distributed entries, and then we note that some methods work in the case of not normally distributed entries too.

In Section 2 we give an introduction into the large deviation theory. This theory is related to the sequence of random variables with non-random limits, for example in the case of law of large numbers. After recalling the first large deviation theorem of Cramèr, we define the large deviation principle for random matrices. The large deviation theorem for the different kind of Gaussian random matrices mentioned in the Section 1 are also here, as the theorem of Ben Arous and Guionnet [5], and the theorems of Hiai and Petz. Since the rate function in the case of random matrices is some weighted logarithmic energy, and the limit distribution is the so-called equilibrium measure of this functional, we have an overview of the basic notions of potential theory, and some theorems in order to obtain the equilibrium measures of the logarithmic energy with different rate functions.

In Section 3 we give the construction of the so called Haar unitary random matrix, which is a unitary matrix valued random variable with the distribution according to the Haar measure on the set of $n \times n$ unitary matrices. We collect the main properties of this random matrix, as the distribution of the entries, the correlation between any two entries, and the joint eigenvalue density function. We have an elementary proof of the theorem of Diaconis and Shahshahani, which claims that the trace of different powers of the Haar unitary random matrices are asymptotically independent and normally distributed as the matrix size goes to infinity. From this we deduce, that the empirical eigenvalue distribution tends to the uniform distribution on the unit circle. We also prove this for the Haar distributed orthogonal random matrices with the same method. Finally we recall the theorem of Hiai and Petz [25], which proves the large deviation theorem for unitary random matrices.

In Section 4 we consider a new kind of random matrix, the $m \times m$ truncation of an $n \times n$ Haar unitary random matrix. We give a more detailed proof of the theorem of Życzkowski and Sommers which gives the joint eigenvalue density of these random matrices, and then we give the normalization constant [34]. The joint eigenvalue density then helps us to get the main result of the dissertation, which is the large deviation theorem for the empirical eigenvalue distribution of the truncation, as the matrix size goes to infinity, and m/n converges to a constant λ. After minimizing the rate function of this large deviation we get the limit of the empirical eigenvalue distribution.

Finally in Section 5 we point to the connection of the free probability and the random matrix theory. We define the noncommutative probability space, the noncommutative random variables, and random matrix models of different noncommutative random variables, using the random matrices mentioned in the previous sections. We define the Brown measure of a noncommutative random variable, and we study he relationship between the Brown measures of the random variables and the empirical eigenvalue distribution of their random matrix model.

1 Random matrices and their eigenvalues

Random variables which are situated in this special way allows us to examine the behaviour of matrix quantities such as eigenvalues, determinant and trace, or the asymptotic behaviour of the entries and the above quantities as the matrix size $n \to \infty$. Since the trace and the determinant are given as the sum and the product of the eigenvalues, the most important thing is to examine the eigenvalues. In the case of random matrices, the eigenvalues are random variables too, and we can get all the information if we have the joint eigenvalue density of the eigenvalues.

The aim of this section is to give an overview of several kind of random matrices.

1.1 The standard complex normal variable

In this thesis we mainly study random matrices with Gaussian entries, or random matrices constructed from Gaussian random matrices, so now we will mention some properties of the so-called *standard complex normal variable*

Definition 1.1 *Let ξ be a complex-valued random variable. If $Re\xi$ and $Im\xi$ are independent and normally distributed according to $N(0, 1/2)$, then we call ξ a standard complex normal variable.*

The terminology is justified by the properties $\mathbb{E}(\xi) = 0$ and $\mathbb{E}\left(|\xi|^2|\right) = \mathbb{E}(\xi\overline{\xi}) = 1$.

Lemma 1.2 *Assume that $R \geq 0$ and R^2 has exponential distribution with parameter 1, ϑ is uniform on the interval $[0, 2\pi]$, and assume that R and ϑ are independent. Then $\xi = Re^{i\vartheta}$ is a standard complex normal random variable and*

$$\mathbb{E}(\xi^k\overline{\xi}^\ell) = \delta_{k\ell}k! \qquad (k, \ell \in \mathbb{Z}_+)$$

Proof. Let X and Y be real-valued random variables and assume that $X + iY$ is standard complex normal. For $r > 0$ and $0 \leq \vartheta_0 \leq 2\pi$ set

$$S_{r,\vartheta_0} := \{\rho e^{\mathrm{i}\psi} : 0 \leq \rho \leq r, 0 \leq \psi \leq \vartheta_0\},$$

then

$$\begin{aligned}
\mathbb{P}\left(X + \mathrm{i}Y \in S_{r,\vartheta_0}\right) &= \frac{1}{\pi}\int\int_{\{(s,t):s+\mathrm{i}t\in S_{r,\vartheta_0}\}} e^{-\left(s^2+t^2\right)}dsdt \\
&= \frac{1}{\pi}\int_0^{\vartheta_0} d\psi \int_0^r \rho e^{-\rho^2}d\rho \\
&= \frac{1}{2\pi}\vartheta_0\left(1 - e^{-r^2}\right) = \mathbb{P}\left(\xi \in S_{r,\vartheta_0}\right).
\end{aligned}$$

This proves the first part which makes easy to compute the moments:

$$\mathbb{E}(\xi^k\overline{\xi}^\ell) = \mathbb{E}\left(R^{k+\ell}\right)\mathbb{E}\left(e^{i\vartheta(k-\ell)}\right) = \delta_{k\ell}\mathbb{E}(R^{2k}),$$

so we need the moments of the exponential distribution. We have by partial integration

$$\begin{aligned}\int_0^\infty x^k e^{-x}\,dx &= -\left[x^k e^{-x}\right]_0^\infty + k\int_0^\infty x^{k-1}e^{-x}\,dx = k\int_0^\infty x^{k-1}e^{-x}\,dx \\ &= k(k-1)\int_0^\infty x^{k-2}e^{-x}\,dx = \cdots = k!\int_0^\infty e^{-x}\,dx = k! \qquad (1)\end{aligned}$$

which completes the proof of the lemma. □

Lemma 1.3 *Let ξ and η be independent identically distributed random variables with zero mean and finite variance. Suppose that the distribution of $(\xi+\eta)/\sqrt{2}$ coincides with the distribution of ξ. Then ξ and η are normally distributed.*

Proof. We can assume, that the variance of ξ is 1. If $\varphi(t)$ is the Fourier transform of ξ and φ, i.e.

$$\varphi(t) := \mathbb{E}\left(e^{\mathrm{i}\xi t}\right) = \int e^{\mathrm{i}tx}dF_\xi(x) = \int e^{\mathrm{i}tx}dF_\eta(x),$$

where F_ξ and F_η are the distributions of ξ and η respectively. Then $\varphi(0)=1$,

$$\varphi'(0) = \mathrm{i}\int x dF_\xi(x) = \mathrm{i}\mathbb{E}(\xi) = 0,$$

and

$$\varphi''(0) = \mathrm{i}^2\int x^2 dF_\xi(x) = \mathrm{i}\mathbb{E}\left(\xi^2\right) = -1.$$

If we have the joint distribution $F_{(\xi,\eta)}(x,y) = F_\xi(x)F_\eta(y)$ of ξ and φ then the Fourier transform of $(\xi+\eta)/\sqrt{2}$ is φ, because it has the same distribution. On the other hand

$$\int e^{\mathrm{i}t(x+y)/\sqrt{2}}dF_{(\xi,\eta)} = \int e^{\mathrm{i}tx/\sqrt{2}}dF_\xi(x)\int e^{\mathrm{i}ty/\sqrt{2}}dF_\eta(y) = \varphi^2\left(\frac{1}{\sqrt{2}}\right),$$

so

$$\varphi^2\left(\frac{t}{\sqrt{2}}\right) = \varphi(t). \qquad (2)$$

If $\varphi(t)=0$ for some t, then $\varphi\left(t/2^n\right)=0$, which is impossible, since φ is continuous and $\varphi(0)=1$. For $\psi(t) := \log\varphi(t)$, clearly $\psi(0)=0$,

$$\psi'(0) = \frac{\varphi'(0)}{\varphi(0)} = 0,$$

and

$$\psi''(0) = \frac{\varphi(0)\varphi''(0) - (\varphi'(0))^2}{(\varphi(0))^2} = -1.$$

We have from (2), that

$$\psi(t) = 2\psi\left(\frac{t}{\sqrt{2}}\right),$$

so for all positive n

$$\frac{\psi(t)}{t^2} = \frac{\psi\left(\frac{t}{2^{n/2}}\right)}{\left(\frac{t}{2^{n/2}}\right)^2}.$$

We have that for all t

$$\frac{\psi(t)}{t^2} = \lim_{s\to 0}\frac{\psi(s)}{s^2} = c,$$

so $\psi(t) = ct^2$, and since $\psi''(0) = -1$, $c = -1/2$, so

$$\varphi(t) = e^{-t^2/2},$$

so φ is the Fourier transform of the standard normal distribution. The Fourier transform of a distribution is unique, so ξ is normally distributed. $\square$

1.2 Selfadjoint Gaussian random matrices

Apart from the trivial example of diagonal random matrix with independent entries, the simplest example for random matrix is the following selfadjoint random matrix, which is called *standard selfadjoint Gaussian matrix.* Consider the $n \times n$ random matrix A_n with entries A_{ij}

- $\operatorname{Re} A_{ij}$ and $\operatorname{Im} A_{ij}$ are independent $N\left(0, \frac{1}{2n}\right)$ distributed random variables, if $1 \le i < j \le n$;
- A_{ii} are $N\left(0, \frac{1}{n}\right)$ distributed random variables if $1 \le i \le n$;
- the entries on and above the diagonal are independent;
- $A_{ij} = \overline{A_{ji}}$, for all $1 \le j < i \le n$.

The above matrix is selfadjoint so its eigenvalues are real.

We can obtain a standard selfadjoint Gaussian matrix in the following way. Let X_n be the so called $n \times n$ *standard non-selfadjoint Gaussian matrix* $X_n = (X_{ij})_{1\le i,j\le n}$ such that

- $\operatorname{Re} X_{ij}, \operatorname{Im} X_{ij}$ are independent identically distributed random variables with distribution $N(0, 1/2n)$ for $1 \le i, j \le n$;
- all the entries are independent.

For this matrix

$$A_n := \frac{X_n + X_n^*}{\sqrt{2}} \tag{3}$$

is a standard selfadjoint Gaussian matrix. Clearly A_n is selfadjoint, and the distribution of the entries is normal as the linear combination of normal distributed random variables. Note that the

$$A'_n := \frac{X_n - X_n^*}{\sqrt{2}\mathrm{i}} \tag{4}$$

is a standard selfadjoint Gaussian matrix too. A_n and A'_n are independent, so if we have two independent $n \times n$ standard selfadjoint Gaussian matrices A_n and A'_n, then

$$X_n := \frac{A_n + \mathrm{i}A'_n}{\sqrt{2}} \tag{5}$$

is a standard non-selfadjoint Gaussian random matrix.

The standard non selfadjoint Gaussian random matrices are invariant under the multiplication by a non-random unitary matrix, so we get the following lemma.

Lemma 1.4 *The distribution of A_n is invariant under unitary conjugation, i.e. if $U_n = u_{ij}$ is an $n \times n$ non-random unitary matrix, then A_n and $U_n A_n U_n^*$ have the same distribution.*

Proof. From (3) it is enough to prove, that X_n and $U_n X_n$ has the same distribution where X_n is an $n \times n$ standard non-selfadjoint Gaussian random matrix. The entries ξ_{ij} of $U_n X_n$ are the same as the entries of X_n. Indeed,

$$\xi_{ij} = \sum_{l=1}^{n} u_{il} X_{lj}$$

is normal, since any linear combination of independent normally distributed random variables are normal. But this is not enough, because we need that the joint density of the entries is the same. Indeed, the joint density of the entries of X_n is

$$\frac{n^{n^2}}{\pi^{n^2}} \exp\left(-n \sum_{i,j=1}^{n} x_{ij}^2 + y_{ij}^2\right)$$
$$= \frac{n^{n^2}}{\pi^{n^2}} \exp\left(-n \mathrm{Tr}\, X_n^* X_n\right) = \frac{n^{n^2}}{\pi^{n^2}} \exp\left(-n \mathrm{Tr}\, (X_n U_n)^* X_n U_n\right).$$

Since

$$U_n A_n U_n^* = U_n \left(\frac{X_n + X_n^*}{\sqrt{2}}\right)^* = \frac{U_n X_n U_n^* + U_n X_n^* U_n^*}{\sqrt{2}} = \frac{U_n X_n U_n^* + (U_n X_n U_n^*)^*}{\sqrt{2}},$$

which by (3) has clearly the same distribution as A_n.

The standard non selfadjoint Gaussian matrix consists of n^2 independent real valued normally distributed random variables (n in the diagonal, and $n(n-1)$ above the diagonal if we consider the real and imaginary parts separately). The joint density of the entries with respect to the Lebesgue measure on $\mathbb{R}^{n^2}$ is the joint density of the above random variables, so can be written in the form

$$\frac{n^{\frac{n^2}{2}}}{2^{\frac{n}{2}}\pi^{\frac{n^2}{2}}}\exp\left(-\frac{n}{2}\left(\sum_{i=1}^{n} x_{ii}^2 + 2\sum_{i<j}\left(x_{ij}^2 + y_{ij}^2\right)\right)\right)$$
$$= \frac{n^{\frac{n^2}{2}}}{2^{\frac{n}{2}}\pi^{\frac{n^2}{2}}}\exp\left(-\frac{n}{2}\mathrm{Tr}\, A_n^2\right) = \frac{n^{\frac{n^2}{2}}}{2^{\frac{n}{2}}\pi^{\frac{n^2}{2}}}\exp\left(-\frac{n}{2}\sum_{i=1}^{n}\lambda_i^2\right). \tag{6}$$

Here $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A_n, so the joint density can be expressed by the eigenvalues. This easily comes from the fact that the distribution of A_n is invariant under unitary conjugation.

In the sequel we will give the joint eigenvalue density of A_n with the transformation of the variables. If we change the variables x_{ij}, y_{ij} into $\lambda_i, \dots \lambda_n$, and $n(n-1)/2$ parameters p_ν, then using the fact that for any normal matrix A there exists a U unitary matrix, and $D := \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, such that

$$A = U^* D U.$$

U is unitary so $U^*U = I$, and therefore

$$\frac{\partial U^*}{\partial p_\nu}U + U^*\frac{\partial U}{\partial p_\nu} = 0,$$

for all $1 \le \nu \le n(n-1)/2$, so we use the notation

$$dS^{(\nu)} := U^*\frac{\partial U}{\partial p_\nu} = -\frac{\partial U^*}{\partial p_\nu}U. \tag{7}$$

U does not depend on the eigenvalues, so

$$\frac{\partial A}{\partial \lambda_\mu} = U^*\frac{\partial D}{\partial \lambda_\mu}U$$

for all $1 \le \mu \le n$, so for the entries

$$\sum_{kl}\frac{\partial A_{kl}}{\partial \lambda_\mu}\overline{U}_{ki}U_{lj} = \frac{\partial D_{ij}}{\partial \lambda_\mu} = \delta_{ij}\delta_{i\mu},$$

and if we separate the real and imaginary parts, we have that since A is selfadjoint, so the diagonal elements are real, and $\mathrm{Re}\, A_{kl} = \mathrm{Re}\, A_{kl}$ and $\mathrm{Im}\, A_{kl} = -\mathrm{Im}\, A_{kl}$ so

$$\sum_{k=1}^{n}\frac{\partial A_{kk}}{\partial \lambda_\mu}\mathrm{Re}\, U_{ki}^* U_{kj} + \sum_{k<l}\frac{\partial \mathrm{Re}\, A_{kl}}{\partial \lambda_\mu}\left(\mathrm{Re}\,(\overline{U}_{ki}U_{lj}) + \mathrm{Re}\,(\overline{U}_{li}U_{kj})\right)$$
$$- \sum_{k<l}\frac{\partial \mathrm{Im}\, A_{kl}}{\partial \lambda_\mu}\left(\mathrm{Im}\,(\overline{U}_{ki}U_{lj}) - \mathrm{Im}\,(\overline{U}_{li}U_{kj})\right) = \delta_{ij}\delta_{i\mu}, \tag{8}$$

and

$$\sum_{k=1}^{n} \frac{\partial A_{kk}}{\partial \lambda_\mu} \mathrm{Im}\, U_{ki}^* U_{kj} + \sum_{k<l} \frac{\partial \mathrm{Re}\, A_{kl}}{\partial \lambda_\mu} \big(\mathrm{Im}\, (\overline{U}_{ki} U_{lj}) + \mathrm{Im}\, (\overline{U}_{li} U_{kj}) \big)$$
$$+ \sum_{k<l} \frac{\partial \mathrm{Im}\, A_{kl}}{\partial \lambda_\mu} \big(\mathrm{Re}\, (\overline{U}_{ki} U_{lj}) - \mathrm{Re}\, (\overline{U}_{li} U_{kj}) \big) = 0, \tag{9}$$

for $1 \le \mu \le n$. Now since D does not depend on p_ν, we have

$$\frac{\partial A}{\partial p_\nu} = \frac{\partial U}{\partial p_\nu} D U^* + U D \frac{\partial U^*}{\partial p_\nu}$$

so

$$U^* \frac{\partial A}{\partial p_\nu} U \;=\; dS^{(\nu)} D - D\, dS_{(\nu)},$$

and which means for the entries

$$\sum_{k,l=1}^{n} \frac{\partial A_{kl}}{\partial p_\nu} \overline{U}_{ki} U_{lj} \;=\; dS_{ij}^{(\nu)} (\lambda_i - \lambda_j),$$

so by separating the real and imaginary parts we get

$$\sum_{k=1}^{n} \frac{\partial A_{kk}}{\partial p_\nu} \mathrm{Re}\, U_{ki}^* U_{kj} + \sum_{k<l} \frac{\partial \mathrm{Re}\, A_{kl}}{\partial p_\nu} \big(\mathrm{Re}\, (\overline{U}_{ki} U_{lj}) + \mathrm{Re}\, (\overline{U}_{li} U_{kj}) \big) \tag{10}$$
$$- \sum_{k<l} \frac{\partial \mathrm{Im}\, A_{kl}}{\partial p_\nu} \big(\mathrm{Im}\, (\overline{U}_{ki} U_{lj}) - \mathrm{Im}\, (\overline{U}_{li} U_{kj}) \big) \;=\; d\mathrm{Re}\, S_{ij}^{\nu} (\lambda_i - \lambda_j),$$

and

$$\sum_{k=1}^{n} \frac{\partial A_{kk}}{\partial p_\nu} \mathrm{Im}\, U_{ki}^* U_{kj} + \sum_{k<l} \frac{\partial \mathrm{Re}\, A_{kl}}{\partial p_\nu} \big(\mathrm{Im}\, (\overline{U}_{ki} U_{lj}) + \mathrm{Im}\, (\overline{U}_{li} U_{kj}) \big) \tag{11}$$
$$+ \sum_{k<l} \frac{\partial \mathrm{Im}\, A_{kl}}{\partial p_\nu} \big(\mathrm{Re}\, (\overline{U}_{ki} U_{lj}) - \mathrm{Re}\, (\overline{U}_{li} U_{kj}) \big) \;=\; d\mathrm{Im}\, S_{ij}^{\nu} (\lambda_i - \lambda_j).$$

We need the determinant of the $n^2 \times n^2$ matrix

$$J := \begin{pmatrix} \dfrac{\partial A_{ii}}{\partial \lambda_\mu} & \dfrac{\partial \mathrm{Re}\, A_{ij}}{\partial \lambda_\mu} & \dfrac{\partial \mathrm{Im}\, A_{ij}}{\partial \lambda_\mu} \\[2ex] \dfrac{\partial A_{ii}}{\partial p_\nu} & \dfrac{\partial \mathrm{Re}\, A_{ij}}{\partial p_\nu} & \dfrac{\partial \mathrm{Im}\, A_{ij}}{\partial p_\nu} \end{pmatrix}.$$

Here $\partial A_{ii}/\partial \lambda_\mu$ is an $n \times n$ matrix, $\partial \mathrm{Re}\, A_{ij}/\partial \lambda_\mu$ and $\partial \mathrm{Im}\, A_{ij}/\partial \lambda_\mu$ are $n \times n(n-1)/2$ matrices, and we order the columns by lexicographic order of the (i,j) pairs, $\partial A_{ii}/\partial p_\nu$

is an $n(n-1)\times n$ matrix, finally $\partial \mathrm{Re}\, A_{ij}/\partial p_\nu$ and $\partial \mathrm{Im}\, A_{ij}/\partial p_\nu$ are $n(n-1)\times n(n-1)/2$ matrices. Now let

$$V := \begin{pmatrix} \mathrm{Re}\, U^*_{ki}U_{kj} & \mathrm{Im}\, U^*_{ki}U_{kj} \\ \\ \mathrm{Re}\,(\overline{U}_{ki}U_{lj}) + \mathrm{Re}\,(\overline{U}_{li}U_{kj}) & \mathrm{Im}\,(\overline{U}_{ki}U_{lj}) + \mathrm{Im}\,(\overline{U}_{li}U_{kj}) \\ \\ \mathrm{Im}\,(\overline{U}_{li}U_{kj}) - \mathrm{Im}\,(\overline{U}_{ki}U_{lj}) & \mathrm{Re}\,(\overline{U}_{ki}U_{lj}) - \mathrm{Re}\,(\overline{U}_{li}U_{kj}) \end{pmatrix},$$

Here $\mathrm{Re}\, U^*_{ki}U_{kj}$ and $\mathrm{Im}\, U^*_{ki}U_{kj}$ are $n\times n(n-1)/2$ matrices, where k is fixed in a row, and the pairs (i,j) are ordered lexicographically. The submatrices $\mathrm{Re}\,(\overline{U}_{ki}U_{lj})+\mathrm{Re}\,(\overline{U}_{li}U_{kj})$, $\mathrm{Im}\,(\overline{U}_{ki}U_{lj}) + \mathrm{Im}\,(\overline{U}_{li}U_{kj})$, $\mathrm{Im}\,(\overline{U}_{li}U_{kj}) - \mathrm{Im}\,(\overline{U}_{ki}U_{lj})$ and $\mathrm{Re}\,(\overline{U}_{ki}U_{lj}) - \mathrm{Re}\,(\overline{U}_{li}U_{kj})$ are $n(n-1)/2\times n(n-1)/2$ matrices, so V is again an $n^2\times n^2$ matrix, and by the previous equations

$$JV = \begin{pmatrix} \delta_{ij}\delta_{i\mu} & 0 \\ d\mathrm{Re}\, S^\nu_{ij}(\lambda_i-\lambda_j) & d\mathrm{Im}\, S^\nu_{ij}(\lambda_i-\lambda_j) \end{pmatrix}$$

where the (i,j) pair is fixed in one row, so we have for the determinants of the above matrices

$$\det J \det V = \prod_{i<j}(\lambda_i-\lambda_j)^2 \det \begin{pmatrix} \delta_{ij}\delta_{i\mu} & 0 \\ d\mathrm{Re}\, S^\nu_{ij} & d\mathrm{Im}\, S^\nu_{ij} \end{pmatrix}$$

From this we get the Jacobian

$$C\prod_{i<j}(\lambda_i-\lambda_j)^2, \tag{12}$$

for some constant C, since the matrix on the right hand side of the above equation, and the matrix V does not depend on the eigenvalues.

Finally we got the joint density of the eigenvalues

$$C_n^{sa}\exp\left(-\frac{n}{2}\sum_{i=1}^n \lambda_i^2\right)\prod_{i<j}(\lambda_i-\lambda_j)^2 \tag{13}$$

with the normalization constant

$$C_n^{sa} := C\frac{n^{\frac{n^2}{2}}}{2^{\frac{n}{2}}\pi^{\frac{n^2}{2}}} = \frac{n^{\frac{n^2}{2}}}{(2\pi)^{\frac{n}{2}}\prod_{j=1}^n j!}. \tag{14}$$

Now consider the asymptotic behaviour of the empirical eigenvalue distribution which is defined by

$$F_n(x) := \frac{1}{n}\#\left\{\lambda_i : \lambda_i \le x\right\}, \tag{15}$$

so this is a random distribution function.

In fact Wigner studied more general random matrices, the so-called *Wigner matrices*, which are selfadjoint random matrices with independent, identically distributed entries on and above the diagonal, where all the moments of the entries are finite. The first theorem of Wigner about the empirical eigenvalue distribution concerned only the expectation of F_n, and found that this sequence of distribution functions converges to the so-called semicircle distribution. which has the density

$$w(x) := \frac{1}{2\pi}\sqrt{4-x^2}\chi_{[-2,2]}. \tag{16}$$

This is the Wigner semicircle law, since the first form of this theorem, which concerned only the expectation of the empirical eigenvalue distribution was proven by Wigner in [43].

The almost sure weak convergence of the sequence of random distribution functions was proven by Arnold in[1]. He proved the almost sure convergence for general Wigner matrices, with the assumption that some moments of the entries are finite.

Wigner's and Arnold's proofs are based on the fact, that the moments of F_n converge to the α_k moments of the semicircle distribution. These are the so-called *Catalan numbers*

$$\alpha_k := \begin{cases} 0, & \text{if} \quad k = 2m+1 \\ \\ \dfrac{1}{m+1}\dbinom{2m}{m}, & \text{if} \quad k = 2m \end{cases}$$

By the Carleman criterion (in VII. 3 of [14]) for a real valued random variable the moments γ_k determine the distribution uniquely if

$$\sum_{k\in\mathbb{N}} \gamma_{2k}^{-\frac{1}{k}} = \infty.$$

This holds for the Catalan numbers, so it is enough to prove, that

$$\int x^k dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^k = \frac{1}{n}\mathrm{Tr}\,(A_n^k) \overset{n\to\infty}{\longrightarrow} \alpha_k.$$

This trace is a sum of products of matrix element, and we have to take the summation over the terms, which are not asymptotically small. The number of this terms can be obtained by combinatorial methods.

Wigner proved that for an A_n sequence of $n \times n$ Wigner matrices

$$\lim_{n\to\infty} \mathbb{E}\mathrm{Tr}\, A_n^k = \alpha_k.$$

Arnold's proof contained more about the convergence. By the Chebyshev inequality he obtained

$$\mathbb{P}\left(|\mathrm{Tr}\, A_n^k - \mathbb{E}\mathrm{Tr}\, A_n^k| > \varepsilon\right) \le O\left(n^{-\frac{3}{2}}\right),$$

so by the Borel-Cantelli lemma it implies the almost sure convergence of F_n. As we mentioned, these proofs did not use the exact distribution of the entries. For standard selfadjoint Gaussian matrix Haagerup and Thorbjørnsen had another proof for the convergence. Their method based on the fact, that the mean density of the eigenvalues (i.e. the density of the arithmetic mean of the eigenvalues) is

$$\frac{1}{n}\sum_{k=0}^{n-1}\varphi_k(x)^2,$$

where

$$\varphi_k(x) := e^{-\frac{x^2}{2}} H_k(x),$$

with the kth Hermite polynomial H_k. In their paper [22] they proved moreover, that there is no eigenvalue outside the interval $[-1, 1]$ with probability one, i. e. if $\lambda_{\max}^{(n)}$ and $\lambda_{\min}^{(n)}$ denote the largest and smallest eigenvalue of A_n respectively, then

$$\lambda_{\max}^{(n)} \xrightarrow{n\to\infty} 2,$$

and

$$\lambda_{\min}^{(n)} \xrightarrow{n\to\infty} -2,$$

and the convergence is almost sure in both case.

The Wigner semicircle law holds for symmetric Gaussian random matrices with real entries, where the distribution of the entries on and above the diagonal are independent real $N(0, 1/n)$ distributed random variables. Here, the density of the matrix is with respect to the Lebesgue measure on $\mathbb{R}^{\frac{(n+1)n}{2}}$

$$C_1 \exp\left(-n\left(\sum_{i\le j}^{n} x_{ij}^2\right)\right) = C_1 \exp\left(-n\mathrm{Tr}\, A_n^2\right) = C_1 \exp\left(-n\sum_{i=1}^{n}\lambda_i^2\right), \tag{17}$$

In this case of symmetric matrices the Jacobian will be

$$\prod_{i<j}|\lambda_i - \lambda_j|, \tag{18}$$

similarly to the complex case, but here the imaginary parts are zero, so the size of transformation matrix is smaller. Therefore the joint density of the eigenvalues will be

$$C_{symm} \exp\left(-n\sum_{i=1}^{n}\lambda_i^2\right)\prod_{i<j}|\lambda_i - \lambda_j|. \tag{19}$$

The Wigner theorem can be proven for these matrices in the same way by means of the method of moments.

1.3 Wishart type random matrices

The matrices defined below called Wishart matrices, since they were introduced by Wishart in 1928 [46]. He used these matrices in multivariate statistics, so he studied matrices with real entries. Suppose, that we have an X $p \times n$ Gaussian random matrix, such that X_{ij} independent random variables with distribution $N(0, 1/n)$ for all $1 \leq i \leq p$ and $1 \leq j \leq n$. Then the $p \times p$ matrix $W_p := XX^*$ is the so called *Wishart matrix*. It has very important role in the multivariate statistics. This matrix W_p is not only selfadjoint, but positive, so the eigenvalues lie on $\mathbb{R}^+$.

If $p > n$, then the rank of W_p is at most p, so it has $n - p$ zero eigenvalues. Moreover, if λ is a non-zero eigenvalue of W_p, then there exists an $v \in \mathbb{R}^p$ such that

$$W_p v = XX^* v = \lambda v.$$

Then

$$X^* X X^* v = \lambda X^* v,$$

so λ is an eigenvalue of X^*X too with eigenvector X^*v. So all the non-zero eigenvalues of W_p coincide with the eigenvalues of X^*X, therefore, it is enough to deal with the $p \leq n$ case.

The Jacobian of the transformation which maps the entries into the set of eigenvalues is the same as in the case of symmetric Gaussian matrices, since the Wishart matrix is symmetric too, so we can transform into a diagonal matrix by unitary conjugation. Similarly to the case of Wigner matrices, the joint density of the eigenvalues can be derived the joint density of the entries and the Jacobian in (18), and it can be written in the form

$$C_{n,p}^{wish} \left(\prod_{i=1}^{p} \lambda_i \right)^{\frac{n-p-1}{2}} \left(\prod_{i<j} |\lambda_i - \lambda_j| \right) \exp \left(-\frac{1}{2} \sum_{i=1}^{p} \lambda_i \right),$$

supported on $(\mathbb{R}^+)^n$. Again, this contains all the information about the matrix, since it is invariant under unitary conjugation.

For the asymptotic behaviour of the empirical eigenvalue distribution we must find some relation between the number of the rows and columns, so $p := p(n) \leq n$. If

$$\frac{p(n)}{n} \xrightarrow{n \to \infty} \lambda > 0,$$

then we can state a similar to the Wigner semicircle law, i.e. the random measure sequence of the empirical eigenvalue distribution has a non-random limit distribution function, but the density of the function is different. The first form of the theorem below was proven by Marchenko and Pastur in [30] and the distribution was named after them. (It is also called free Poisson distribution, cf [23]).

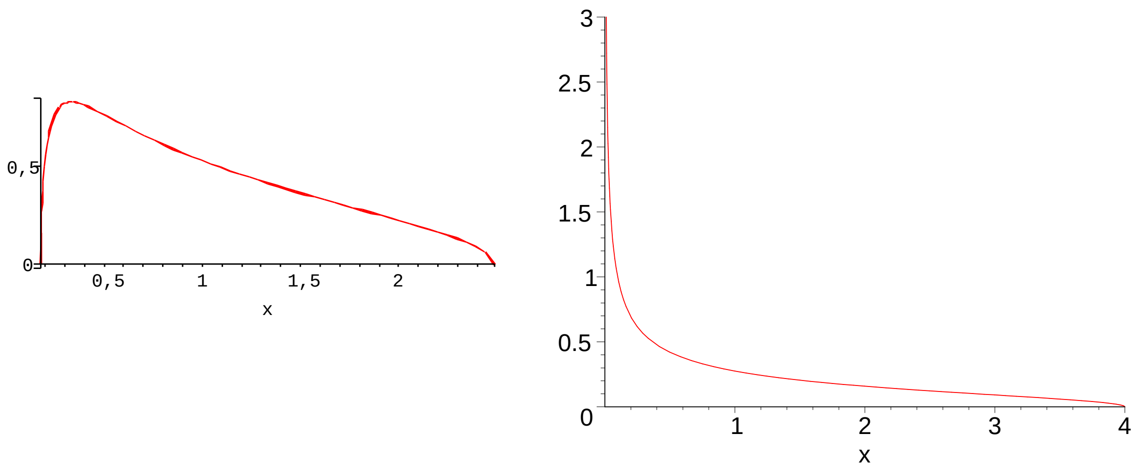
Theorem 1.1 *Denote $F_n^\lambda(x)$ the empirical eigenvalue distribution of W_p, and $F^\lambda(x)$ the so-called Marchenko-Pastur distribution, with density*

$$f^\lambda(x) := \frac{\sqrt{4\lambda - (x-\lambda-1)^2}}{2\pi\lambda x},$$

supported on the interval $\left[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2\right]$. Then

$$F_n^\lambda \overset{n\to\infty}{\longrightarrow} F^\lambda$$

weakly with probability 1.



Density of the Marchenko-Pastur distribution for $\lambda = 1/3$ and $\lambda = 1$

This theorem holds in a more general form, as we will see later.

Haagerup and Thorbjørnsen in [22] studied Wishart matrices with complex entries. They used $p \times n$ Gaussian matrices with independent complex normal entries with zero mean and variance $1/n$. In this case they proved the almost sure convergence of the empirical eigenvalue distribution of the eigenvalues by using the fact that the joint eigenvalue density is

$$\sum_{k=0}^{n-1} \varphi_k^{(m-n)}(x)^2,$$

where $\varphi_k^{(\alpha)}$ can be expressed in terms of Laguerre polynomials $L_k^{(\alpha)}$ in the following way

$$\varphi_k^{(\alpha)}(x) = \sqrt{\frac{k!}{\Gamma(k+\alpha+1)} x^\alpha \exp(-x) L_k^{(\alpha)}(x)}.$$

With this method they proved the almost sure convergence of the largest and the smallest eigenvalue, i.e.

$$\lambda_{\max} \stackrel{n\to\infty}{\longrightarrow} (1+\sqrt{\lambda})^2 \quad \text{and} \quad \lambda_{\max} \stackrel{n\to\infty}{\longrightarrow} (1-\sqrt{\lambda})^2.$$

1.4 Non selfadjoint Gaussian matrices

The simplest non selfadjoint random matrix is the $n \times n$ *standard non-selfadjoint Gaussian matrix.* As we could see, this matrix defines a standard selfadjoint Gaussian random matrix too. Similarly it gives a whole family of random matrices in the following way.

Definition 1.5 *Let $u, v \in \mathbb{R}$ such that $u^2 + v^2 = 1$, then the we call the matrix*

$$Y_n := uX_n + vX_n^* \tag{20}$$

an elliptic Gaussian matrix.

Note that in the case $u = 1/\sqrt{2}$ Y_n is the standard selfadjoint Gaussian matrix, and for $u = 1$ Y_n is the standard non-selfadjoint Gaussian matrix.

If we have A_n and A_n' independent standard $n \times n$ selfadjoint random matrices, then (5) we can construct an elliptic random matrix

$$Y_n = u\frac{A_n + \mathrm{i}A_n'}{\sqrt{2}} + v\frac{A_n - \mathrm{i}A_n'}{\sqrt{2}} = \frac{u+v}{\sqrt{2}}A_n + \frac{u-v}{\sqrt{2}}\mathrm{i}A_n', \tag{21}$$

where again

$$\left(\frac{u+v}{\sqrt{2}}\right)^2 + \left(\frac{u-v}{\sqrt{2}}\right)^2 = u^2 + v^2 = 1.$$

Since Y_n is not selfadjoint we cannot transform it into a diagonal matrix in order to get the joint eigenvalue density. Here we use the so-called Schur decomposition of the matrix Y_n.

Lemma 1.6 (Schur decomposition) *For every matrix $A \in \mathbb{C}^{n\times n}$ there exist an $n \times n$ unitary matrix U, and an upper triangular matrix Z such that*

$$A = UZU^*.$$

Proof. We are looking for an orthonormal basis $u_1, \dots, u_n$ such that the matrix A takes the upper-triangular form on this basis. We will prove the lemma by induction. If $n = 1$, then we have a trivial case. Now suppose that $n > 1$, and let u_1 be an eigenvector of A with eigenvalue λ_1, such that $\|u_1\| = 1$. If $V := u_1^\perp$, then V is invariant under $(I - u_1 u_1^*)A$, and $(I - u_1 u_1^*)Au_1 = 0$. By the induction there exists a basis $u_2, \dots u_n$, such that $(I - u_1 u_1^*)A$ takes the desired form Z_{n-1} on V. The $n \times n$ unitary matrix U with column vectors $u_1, \dots, u_n$ gives the Schur decomposition of A, since clearly

$$U^*AU = U^*(I - u_1 u_1^*)AU + U^* u_1 u_1^* AU,$$

where

$$(U^*(I - u_1 u_1^*)AU)_{ij} = u_i^*(I - u_1 u_1^*)Au_j,$$

which is zero, if either $i = 1$ or $j = 1$, and if $i, j \geq 2$ then it gives the matrix Z_{n-1}. Moreover

$$(U^* u_1 u_1^* AU)_{ij} = u_i^* u_1 u_1^* Au_j,$$

which is λ_1 if $i = j = 1$, and zero if $i \neq 1$, so we have that

$$U^*AU = \begin{pmatrix} \lambda_1 & * \\ 0 & Z_{n-1} \end{pmatrix} = Z.$$

□

We will use the Schur decomposition instead of the diagonalization in order to obtain the joint eigenvalue density of the elliptic Gaussian matrix Y_n. There exists a unitary matrix U and an upper triangular matrix Δ, such that

$$Y = U(D + \Delta)U^*,$$

where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, where $\lambda_1, \dots, \lambda_n$ are the complex eigenvalues of Y), and $\Delta_{ij} = 0$, if $i \geq j$. Again we transform the $2n^2$ variables $(\mathrm{Re}\, X_{ij})_{i,j=1}^n$, $(\mathrm{Im}\, X_{ij})_{i,j=1}^n$ into the $2n$ variables $(\mathrm{Re}\, \lambda_i)_{i=1}^n$, $(\mathrm{Im}\, \lambda_i)_{i=1}^n$ the $n(n-1)$ variables $(\mathrm{Re}\, \Delta_{ij})$, $(\mathrm{Im}\, \Delta_{ij})$, $(1 \leq i < j \leq n)$ and $n(n-1)$ variables (p_ν), $1 \leq \nu \leq n(n-1)$. U is unitary so $U^*U = I$, and therefore

$$\frac{\partial U^*}{\partial p_\nu} U + U^* \frac{\partial U}{\partial p_\nu} = 0,$$

so we use the notation

$$dS^{(\nu)} := U^* \frac{\partial U}{\partial p_\nu} = -\frac{\partial U^*}{\partial p_\nu} U.$$

U does not depend on the eigenvalues, and Δ_{ij}, so the equations, Δ and D does not depend on p_ν for $1 \leq \nu \leq n(n-1)$, so

$$dS^{(\nu)} := U^* \frac{\partial U}{\partial p_\nu} = -\frac{\partial U^*}{\partial p_\nu} U.$$

U does not depend on the eigenvalues, so

$$\frac{\partial Y}{\partial \mathrm{Re}\, \lambda_\mu} = U \frac{\partial D}{\partial \mathrm{Re}\, \lambda_\mu} U^*$$

and

$$\frac{\partial Y}{\partial \mathrm{Im}\, \lambda_\mu} = U \frac{\partial D}{\partial \mathrm{Im}\, \lambda_\mu} U^*$$

so for the entries

$$\sum_{kl} \frac{\partial Y_{kl}}{\partial \mathrm{Re}\,\lambda_\mu} \overline{U}_{ki} U_{lj} = \frac{\partial D_{ij}}{\partial \mathrm{Re}\,\lambda_\mu} = \delta_{ij}\delta_{i\mu},$$

and

$$\sum_{kl} \frac{\partial Y_{kl}}{\partial \mathrm{Im}\,\lambda_\mu} \overline{U}_{ki} U_{lj} = \frac{\partial D_{ij}}{\partial \mathrm{Im}\,\lambda_\mu} = \mathrm{i}\delta_{ij}\delta_{i\mu},$$

and if we separate the real and imaginary parts,

$$\sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial \mathrm{Re}\,\lambda_\mu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) - \sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial \mathrm{Re}\,\lambda_\mu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) = \delta_{ij}\delta_{i\mu},$$

$$\sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial \mathrm{Re}\,\lambda_\mu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) + \sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial \mathrm{Re}\,\lambda_\mu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) = 0,$$

and similarly

$$\sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial \mathrm{Im}\,\lambda_\mu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) - \sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial \mathrm{Im}\,\lambda_\mu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) = 0,$$

$$\sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial \mathrm{Im}\,\lambda_\mu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) + \sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial \mathrm{Im}\,\lambda_\mu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) = \delta_{ij}\delta_{i\mu},$$

for $1 \le \mu \le n$. Again D and Δ do not depend on p_ν for $1 \le \nu \le n(n-1)$, so

$$\frac{\partial Y}{\partial p_\nu} = \frac{\partial U}{\partial p_\nu} D U^* + U D \frac{\partial U^*}{\partial p_\nu}$$

so

$$U^* \frac{\partial Y}{\partial p_\nu} U = \, dS^{(\nu)} D - D\, dS^{(\nu)},$$

and which means for the entries

$$\sum_{k,l=1}^{n} \frac{\partial Y_{kl}}{\partial p_\nu} \overline{U}_{ki} U_{lj} = \, dS^{(\nu)}_{ij} (\lambda_i - \lambda_j),$$

so by separating the real and imaginary parts we get

$$\begin{aligned} &\sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial p_\nu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) - \sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial p_\nu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) \\ &\quad = \, d\mathrm{Re}\,S^\nu_{ij} (\mathrm{Re}\,\lambda_i - \mathrm{Re}\,\lambda_j) - \, d\mathrm{Im}\,S^\nu_{ij} (\mathrm{Im}\,\lambda_i - \mathrm{Im}\,\lambda_j), \end{aligned} \tag{22}$$

and

$$\begin{aligned} &\sum_{k,l} \frac{\partial \mathrm{Im}\,Y_{kl}}{\partial p_\nu} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) + \sum_{k,l} \frac{\partial \mathrm{Re}\,Y_{kl}}{\partial p_\nu} \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) \\ &\quad = \, d\mathrm{Im}\,S^\nu_{ij} (\mathrm{Re}\,\lambda_i - \mathrm{Re}\,\lambda_j) + \, d\mathrm{Re}\,S^\nu_{ij} (\mathrm{Im}\,\lambda_i - \mathrm{Im}\,\lambda_j). \end{aligned} \tag{23}$$

Moreover, since U and D are independent from Δ, so

$$\frac{\partial Y}{\partial \Delta_{ij}} = U \frac{\partial \Delta}{\partial \Delta_{ij}} U^*,$$

so for the entries

$$\sum_{k,l} \frac{\partial Y_{kl}}{\partial \Delta_{ij}} \overline{U}_{ki} U_{lj} = 1,$$

so

$$\sum_{k,l} \frac{\partial \mathrm{Re}\, Y_{kl}}{\partial \mathrm{Re}\, \Delta_{ij}} \mathrm{Re}\, \left(\overline{U}_{ki} U_{lj}\right) - \sum_{k,l} \frac{\partial \mathrm{Im}\, Y_{kl}}{\partial \mathrm{Re}\, \Delta_{ij}} \mathrm{Im}\, \left(\overline{U}_{ki} U_{lj}\right) = 1,$$

$$\sum_{k,l} \frac{\partial \mathrm{Re}\, Y_{kl}}{\partial \mathrm{Re}\, \Delta_{ij}} \mathrm{Im}\, \left(\overline{U}_{ki} U_{lj}\right) + \sum_{k,l} \frac{\partial \mathrm{Im}\, Y_{kl}}{\partial \mathrm{Re}\, \Delta_{ij}} \mathrm{Re}\, \left(\overline{U}_{ki} U_{lj}\right) = 0,$$

$$\sum_{k,l} \frac{\partial \mathrm{Re}\, Y_{kl}}{\partial \mathrm{Im}\, \Delta_{ij}} \mathrm{Re}\, \left(\overline{U}_{ki} U_{lj}\right) - \sum_{k,l} \frac{\partial \mathrm{Im}\, Y_{kl}}{\partial \mathrm{Im}\, \Delta_{ij}} \mathrm{Im}\, \left(\overline{U}_{ki} U_{lj}\right) = 0,$$

$$\sum_{k,l} \frac{\partial \mathrm{Re}\, Y_{kl}}{\partial \mathrm{Im}\, \Delta_{ij}} \mathrm{Im}\, \left(\overline{U}_{ki} U_{lj}\right) + \sum_{k,l} \frac{\partial \mathrm{Im}\, Y_{kl}}{\partial \mathrm{Im}\, \Delta_{ij}} \mathrm{Re}\, \left(\overline{U}_{ki} U_{lj}\right) = 1.$$

We need the determinant of the $2n^2 \times 2n^2$ matrix

$$J := \begin{pmatrix} \dfrac{\partial \mathrm{Re}\, Y_{ij}}{\partial \mathrm{Re}\, \lambda_\mu} & \dfrac{\partial \mathrm{Im}\, Y_{ij}}{\partial \mathrm{Re}\, \lambda_\mu} \\[2ex] \dfrac{\partial \mathrm{Re}\, Y_{ij}}{\partial \mathrm{Im}\, \lambda_\mu} & \dfrac{\partial \mathrm{Im}\, Y_{ij}}{\partial \mathrm{Im}\, \lambda_\mu} \\[2ex] \dfrac{\partial \mathrm{Re}\, Y_{ij}}{\partial \mathrm{Re}\, \Delta_\xi} & \dfrac{\partial \mathrm{Im}\, Y_{ij}}{\partial \mathrm{Re}\, \Delta_\xi} \\[2ex] \dfrac{\partial \mathrm{Re}\, Y_{ij}}{\partial \mathrm{Im}\, \Delta_\xi} & \dfrac{\partial \mathrm{Im}\, Y_{ij}}{\partial \mathrm{Im}\, \Delta_\xi} \\[2ex] \dfrac{\partial \mathrm{Re}\, Y_{ij}}{\partial p_\nu} & \dfrac{\partial \mathrm{Im}\, Y_{ij}}{\partial p_\nu} \end{pmatrix}.$$

Here $\partial \mathrm{Re}\, Y_{ij}/\partial \mathrm{Re}\, \lambda_\mu$, $\partial \mathrm{Im}\, Y_{ij}/\partial \mathrm{Re}\, \lambda_\mu$, $\partial \mathrm{Re}\, Y_{ij}/\partial \mathrm{Im}\, \lambda_\mu$ and $\partial \mathrm{Im}\, Y_{ij}/\partial \mathrm{Im}\, \lambda_\mu$ are an $2n \times n^2$ matrices, $\partial \mathrm{Re}\, Y_{ij}/\partial \mathrm{Re}\, \Delta_\mu$, $\partial \mathrm{Im}\, Y_{ij}/\partial \mathrm{Re}\, \Delta_\xi$, $\partial \mathrm{Im}\, Y_{ij}/\partial \mathrm{Re}\, \Delta_\xi$ and $\partial \mathrm{Im}\, Y_{ij}/\partial \mathrm{Im}\, \Delta_\mu$ are $n(n-1)/2 \times n^2$ matrices and $\partial \mathrm{Re}\, Y_{ij}/\partial p_\nu$ and $\partial \mathrm{Im}\, Y_{ij}/\partial p_\nu$ are $n(n-1) \times n^2$ matrices. Now let

$$V := \begin{pmatrix} \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) & \mathrm{Im}\,(\overline{U}_{ki} U_{lj}) \\[2ex] -\mathrm{Im}\,(\overline{U}_{ki} U_{lj}) & \mathrm{Re}\,(\overline{U}_{ki} U_{lj}) \end{pmatrix},$$

Here $\mathrm{Re}\, U^*_{ki}U_{lj}$ and $\mathrm{Im}\, U^*_{ki}U_{lj}$ are $n^2 \times n^2$ matrices, where k is fixed in a row, and the pairs (i,j) are ordered lexicographically, so V is now an $2n^2 \times 2n^2$ matrix, and by the previous equations using the notation $\lambda_{ij} := \lambda_i - \lambda_j$

$$JV = \begin{pmatrix} \delta_{ij}\delta_{i\mu} & 0 \\ 0 & \delta_{ij}\delta_{i\mu} \\ \delta_{ij}\delta_{i\xi} & 0 \\ 0 & \delta_{ij}\delta_{i\xi} \\ d\mathrm{Re}\, S^\nu_{ij}\mathrm{Re}\,(\lambda_{ij}) - \, d\mathrm{Im}\, S^\nu_{ij}\mathrm{Im}\,(\lambda_{ij}) & d\mathrm{Im}\, S^\nu_{ij}\mathrm{Re}\,(\lambda_{ij}) + \, d\mathrm{Re}\, S^\nu_{ij}\mathrm{Im}\,(\lambda_{ij}) \end{pmatrix}$$

where the (i,j) pair is fixed in one row, so if we have for the determinants of the above matrices

$$\det J \det V = \prod_{i<j} |\lambda_i - \lambda_j|^2 \det \begin{pmatrix} \delta_{ij}\delta_{i\mu} & 0 \\ 0 & \delta_{ij}\delta_{i\mu} \\ \delta_{ij}\delta_{i\xi} & 0 \\ 0 & \delta_{ij}\delta_{i\xi} \\ d\mathrm{Re}\, S^\nu_{ij} & d\mathrm{Im}\, S^\nu_{ij} \end{pmatrix},$$

since $d\mathrm{Re}\, S^\nu_{ij} = -\, d\mathrm{Re}\, S^\nu_{ji}$ and $d\mathrm{Im}\, S^\nu_{ij} = \, d\mathrm{Im}\, S^\nu_{ji}$ so we can apply

$$\det \begin{pmatrix} ax - by & ay + bx \\ ax + by & -ay + bx \end{pmatrix} = \left(a^2 + b^2\right) \det \begin{pmatrix} -x & y \\ x & y \end{pmatrix}$$

for $a = \mathrm{Re}\,(\lambda_i - \lambda_j)$, $b = \mathrm{Im}\,(\lambda_i - \lambda_j)$ and $x = \, d\mathrm{Re}\, S^\nu_{ij}$, $y = \, d\mathrm{Im}\, S^\nu_{ij}$.

Finally we have that the joint eigenvalue density of the elliptic Gaussian matrix is

$$C^{ell}_n \exp\left(-n \sum_{i=1}^n \left(\frac{(\mathrm{Re}\,\zeta_i)^2}{(u+v)^2} + \frac{(\mathrm{Im}\,\zeta_i)^2}{(u-v)^2}\right)\right) \prod_{i<j} |\zeta_i - \zeta_j|^2,$$

on the set $\mathbb{C}^n$,where C^{ell}_n is the normalizing constat depending on u and v.

Again we have results about the empirical eigenvalue distribution, which now is defined by the random measure on $\mathbb{C}$:

$$\frac{1}{n}\sum_{i=1}^n \delta(\zeta_i(Y_n)),$$

where $\zeta_1(Y_n), \dots, \zeta_n(Y_n)$ are the eigenvalues of Y_n, and $\delta(x)$ is the Dirac function concentrated at the point x. By the elliptic law of Girko in [15, 16, 17, 18], this sequence of random measures converges to the uniform distribution on the set

$$\left\{ z \in \mathbb{C} := \frac{(\mathrm{Re}\, z)^2}{(u+v)^2} + \frac{(\mathrm{Im}\, z)^2}{(u-v)^2} = 1 \right\},$$

This theorem also true in a more general form as we can see in the next section.

1.5 Random matrices with not normally distributed entries

In the case when the random matrix is not invariant under unitary conjugation, it is much more difficult to give the joint density, but we can prove similar results for the asymptotic behaviour of the empirical eigenvalue distribution.

Theorem 1.2 (Arnold) *Suppose that $A_n = (A_{ij})_{i,j=1}^n$ is an $n \times n$ random matrix, where*

- *A_{ii} are independent identically distributed random variables with $\mathbb{E}|A_{ii}|^4 < \infty$, $1 \le i \le n$;*
- *A_{ij} are independent identically distributed random variables such that $\mathbb{E}A_{ij} = 0$, $\mathbb{E}|A_{ij}|^2 = 1/n$ and $\mathbb{E}|A_{ij}|^6 < \infty$, $1 \le i < j \le n$;*
- *$A_{ij} = \overline{A}_{ji}$, if $1 \le j < i \le n$;*
- *the entries on and above the diagonal are independent.*

Then the sequence F_n of the empirical eigenvalue distribution of A_n weakly converges to the semicircle distribution with probability 1 as $n \to \infty$.

Bai and Yin in [4] proved that if the above conditions hold, then

$$\lambda_{\max}(A_n) \overset{n\to\infty}{\longrightarrow} 2 \quad \text{and} \quad \lambda_{\min}(A_n) \overset{n\to\infty}{\longrightarrow} -2.$$

The convergence of the empirical eigenvalue distribution is similar for the generalization of Wishart matrices, the so-called sample covariance matrices. The theorem of Jonsson in [27] is the following.

Theorem 1.3 (Jonsson) *Suppose that $X_p = (X_{ij})_{\substack{1\le i\le p \\ q\le j\le n}}$ is an $p \times n$ random matrix, where the entries are independent identically distributed random variables such that $\mathbb{E}X_{ij} = 0$, $\mathbb{E}|X_{ij}|^2 = 1/n$ and $E|X_{ij}|^6 < \infty$. Then the F_p sequence of the empirical eigenvalue distribution of $X_pX_p^*$ almost surely weakly converges to the Marchenko-Pastur distribution with parameter λ as $n \to \infty$ and $p/n \to \lambda \in (0,1]$. If $p/n \to \lambda > 1$ as $n \to \infty$, then the limit distribution is*

$$\left(1 - \frac{1}{\lambda}\right)\delta_0 + \frac{1}{\lambda}F_\lambda,$$

where F_λ is the Marchenko-Pastur distribution with parameter λ.

The same theorem was proven in [32]. Moreover Bai, Yin and Krishnaiah proved in [47] that if the fourth moment of the entries are finite, then the greatest and smallest eigenvalues almost surely converges to $(1+\sqrt{\lambda})^2$ and $(1-\sqrt{\lambda})^2$, respectively. The proofs of the above theorems are based on the method of moments again.

For the elliptic matrices, i.e the matrices

$$Y_n = uX_n + vX_n^*,$$

where X_n is a matrix with independent identically distributed entries, and $u^2+v^2=1$, in [15, 17] Girko proved the following theorem

Theorem 1.4 *Suppose that $Y_n = (Y_{ij})_{i,j=1}^n$ such that the pairs (Y_{ij}, Y_{ji}) are independent for different $i \le j$, and $\mathbb{E}Y_{ij} = 0$, $\mathbb{E}|Y_{ij}|^2 = \frac{1}{n}$, $\mathbb{E}(Y_{ij}Y_{ji}) = \frac{\tau}{n}$, and moreover there exists a $\delta > 0$ such that*

$$\sup_n \max_{1\le i,j\le n} \mathbb{E}|\sqrt{n}Y_{ij}|^{4+\delta} \le c < \infty,$$

then the empirical eigenvalue distribution converges to the elliptic distribution in probability.

In the case of non normal matrices the method of moments does not work, since we cannot check all the mixed moments. Girko used the V-transform of the empirical eigenvalue distribution μ_n of Y_n, and Girko proved the almost sure convergence as well in [16]

As we could see, the limit distribution does not depend on the distribution of the entries, we only need he finiteness of some moments.

There are some results concerning the rate of the above convergence. For example, Bai proved in [2] and [3] that the rate of convergence has the order of magnitude $O\left(n^{-\frac{1}{4}}\right)$ in the case of Wigner matrices and $O\left(n^{-\frac{5}{48}}\right)$ in the case of sample covariance matrices.

If the distribution of the entries has compact support, then the following theorem of Guionnet and Zeitouni from [20] states that the rate of this convergence is exponential.

Theorem 1.5 (Guionnet, Zeitouni) *Suppose that $A_n = (A_{ij})_{i,j=1}^n$ is an $n \times n$ selfadjoint random matrix, where the distribution of A_{ij} has a common compact support $K \subset \mathbb{C}$, and let $f : \mathbb{R}^k \to \mathbb{R}$ be a Lipschitz function, i.e.*

$$\sup_{x,y} \frac{|f(x)-f(y)|}{\|x-y\|} < \infty.$$

Then there exists a sequence δ_n and a number c depending on the function f, the diameter of the set K and the numbers $\mathbb{E}A_{ij}$ $(1 \le i,j \le n)$, such that

$$0 < \delta_n = O\left(\frac{1}{n}\right),$$

and for all $\varepsilon > \varepsilon_n$

$$\mathbb{P}\left(\left|\frac{1}{n}\,Tr\, f(A_n) - \int_{-2}^{2} f(x)w(x)\,dx\right| \geq \varepsilon\right) \leq 4e^{-cn^2(\varepsilon-\varepsilon_n)^2}.$$

Here

$$\frac{1}{n}\mathrm{Tr}\, f(A_n) = \int f(x)\, dF_n(x),$$

where F_n is the empirical eigenvalue distribution of A_n, and $f(A_n)$ is defined by the usual function calculus of selfadjoint matrices. That is , if

$$A_n = U_n^* \,\mathrm{diag}(\lambda_1, \ldots, \lambda_n) U_n$$

for an $n \times n$ unitary matrix, then

$$f(A_n) := U_n^* \,\mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n)) U_n.$$

2 Large deviation principle

2.1 The concept of the large deviation

If we have a sequence of random variables with non-random limit, the large deviation theorems state exponential rate of convergence.

The simplest example for a sequence of random variables with non-random limit is given by the law of large numbers. Let $X_1, X_2, \ldots$ a sequence of real valued independent identically distributed random variables, with mean m. Then the law of large numbers claims that the sequence of the arithmetic means of (X_n) converges to the number m as $n \to \infty$. In other words if μ_n denotes the distribution of the random variable

$$Y_n := \frac{1}{n} \sum_{i=1}^{n} X_i,$$

then

$$\mu_n \stackrel{n\to\infty}{\longrightarrow} \delta_m,$$

where δ_m is the Dirac-measure concentrated at the point m, i.e. for all $H \subset \mathbb{R}$

$$\delta_m(H) := \begin{cases} 1, & \text{if} \quad m \in H \\ 0, & \text{if} \quad m \notin H \end{cases}$$

This means that for any $G \subset \mathbb{R}$ set such that the closure of G does not contain m,

$$\mu_n(G) \stackrel{n\to\infty}{\longrightarrow} 0.$$

The large deviation principle (LDP) holds, if the rate of the above convergence is exponential. More precisely, if there exists a lower semicontinuous function $f : \mathbb{R} \to \mathbb{R}^+$, such that for all $G \subset \mathbb{R}$

$$\mu_n(G) \approx \exp\left(-L(n) \inf_{x \in G} f(x)\right)$$

then we say that the large deviation principle holds in the scale of $\frac{1}{L(n)}$. Here

$$L(n) \geq cn,$$

for some constant c. Namely, the order of magnitude of the function L is given by the degree of freedom of the random variables. The function f is called the rate function.

The first large deviation theorem was made by Cramèr in 1938 for the sample means of independent, identically distributed random variables. In the Cramèr theorem $L(n) = n$, and the rate function is the convex conjugate of the logarithmic moment generating function of the random variables. The logarithmic momentum generator function of a random variable is

$$\Lambda(\lambda) := \log\left(\mathbb{E}\left(\exp(\lambda X_i)\right)\right),$$

and for its convex conjugate

$$\Lambda^*(x) := \sup\{\lambda x - \Lambda(\lambda) : \lambda \in \mathbb{R}\}$$

and for all measurable $\Gamma \subset \mathbb{R}$

$$-\inf_{x\in\operatorname{int}\Gamma} \Lambda^*(x) \le \liminf_{n\to\infty} \frac{1}{n}\log \mu_n(\Gamma) \le \limsup_{n\to\infty} \frac{1}{n}\log\mu_n(\Gamma) \le -\inf_{x\in\operatorname{cl}\Gamma} \Lambda^*(x).$$

We can check for each independent identically distributed that Λ^* is a convex function, and it attains its minimum in the m, and $\Lambda^*(m) = 0$, because, if $m \in G \subset \mathbb{R}$, then

$$\mu_n(G) \overset{n\to\infty}{\longrightarrow} 1 = e^0 = e^{-\inf_{\{x\in G\}}\Lambda^*(x)}.$$

For example, if $X_1, X_2, \ldots$ are standard normal random variables, then

$$\begin{aligned}\Lambda(\lambda) \quad &= \log\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\lambda x - \frac{x^2}{2}}\,dx\right)\\ &= \log\left(\frac{e^{\frac{\lambda^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-\lambda)^2}{2}}\,dx\right)\\ &= \log e^{\frac{\lambda^2}{2}} = \frac{\lambda^2}{2},\end{aligned}$$

so

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\left(\lambda x - \frac{\lambda^2}{2}\right) = \frac{x^2}{2}.$$

This function attains its minimum in the point 0, which is the mean of the original random variables.

The above theorem can be proven for vector valued independent, identically distributed random variables as well.

Now recall the definition of the large deviation principle from [8].

Definition 2.1 (LDP) *Let X be a topological space, and P_n a sequence of probability measures on $\mathcal{X}$. The large deviation principle holds in the scale $L(n)^{-1}$ if there exists a lower semicontinuous function $I : \mathbb{R} \to [0, \infty]$ such that*

$$\liminf_{n\to\infty} \frac{1}{L(n)}\log P_n(G) \ge -\inf_{x\in G} I(x)$$

for all open set $G \subset X$, and

$$\limsup_{n\to\infty} \frac{1}{L(n)}\log P_n(F) \le -\inf_{x\in F} I(x)$$

for all closed set $F \subset X$. Then the function I is the so-called rate function.

Clearly, in Cramèr's theorem the topological space $\mathcal{X}$ is $\mathbb{R}$ with the usual topology on $\mathbb{R}$. The function f is lower semicontinuous, since it is given as a supremum of continuous functions.

There are still well known examples for random sequence with non-random limit. A very important theorem of statistics implies that if we have the above sequence of real independent identically distributed random variables, then if $\delta(X_i)$ denotes the random measure concentrated to the point X_i, then the random measure sequence of the so-called empirical distribution of $X_1, \dots, X_n$ defined by

$$\widehat{P}_{\underline{X}} = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i) \tag{24}$$

converges to the distribution μ_0 of X_i. It means, that if μ_n is the distribution of $\widehat{P}_{\underline{X}}$, i.e. for all $G \subset \mathcal{M}(\mathbb{R})$

$$\mu_n(G) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} \delta(X_i) \in G\right) \tag{25}$$

which is a probability measure on

$$\mathcal{M}(\mathbb{R}) := \{\text{probability measures on } \mathbb{R}\} \tag{26}$$

converges to the $\delta_\mu \in \mathcal{M}(\mathbb{R})$. The corresponding large deviation theorem was made by Sanov. In his theorem the scale $L(n) = n$ again, since again we have n independent random variables, so the degree of freedom is again n. The topological space $\mathcal{X}$ is $\mathcal{M}(\mathbb{R})$, and the topology is given in the following way. Let

$$G_{f,\varepsilon,\mu} : \left\{\nu \in \mathcal{M}(\mathbb{R}) : \left|\int_{\mathbb{R}} f(x)\, d\mu - \int_{\mathbb{R}} f(x)\, d\nu\right| < \varepsilon\right\}, \tag{27}$$

where f is an element of the set $C_b(\mathbb{R})$ of all bounded continuous functions, $\mu \in \mathcal{M}(\mathbb{R})$, and $\varepsilon > 0$. These sets form the basis of the topology on $\mathcal{M}(\mathbb{R})$, which is the topology of the weak convergence. This space is metrizable with the Lévy metric

$$\mathcal{L}(\mu,\nu) := \inf\left\{\varepsilon > 0 : \mu(F) \le \nu(F_\varepsilon), \nu(F) \le \mu(F_\varepsilon), \text{ for every closed } F \subset \mathbb{R}\right\}, \tag{28}$$

where

$$F_\varepsilon := \left\{x \in \mathbb{R} := \inf_{y\in F} |x-y| < \varepsilon\right\}.$$

Let $D(\,.\,\|\mu_0) : \mathcal{M}(\mathbb{R}) \to [0,\infty]$ is

$$D(\mu\|\mu_0) := \begin{cases} \displaystyle\int_{\mathbb{R}} f(x)\log f(x)\, d\mu_0(x), & \text{if} \quad \mu \ll \mu_0 \quad \text{and} \quad f = \dfrac{d\mu}{d\mu_0} \\ +\infty, & \text{if} \quad \mu \not\ll \mu_0 \end{cases} \tag{29}$$

for $\mu \in \mathcal{M}(\mathbb{R})$. This function is the so-called *relative entropy* of μ with respect to the measure μ_0. The relative entropy is not a metric on $\mathcal{M}(\mathbb{R})$, because the symmetry does not hold, but

- $D(\nu\|\mu) \geq 0$
- $D(\nu\|\mu) = 0$ if and only if $\nu = \mu$.

The relative entropy is a convex function because of the convexity of the function

$$x \to x \log x,$$

and it is lower semicontinuous. Then the following large deviation theorem holds

Theorem 2.1 (Sanov) *For the sequence μ_n given by (25) the large deviation theorem holds on the scale n, and with the rate function*

$$I(\nu) := D(\nu\|\mu).$$

The properties of the relative entropy imply, that I attains its minimum 0 at the point μ.

2.2 Large deviations for random matrices

When we talk about a large deviation theorem for random matrices, it concerns the empirical eigenvalue density. It will be similar to the Sanov theorem, since the empirical eigenvalue distribution of an $n \times n$ random matrix is the sample mean of the Dirac measures concentrated in n random variables, which are the eigenvalues of the matrix. In the simplest case, if we have diagonal matrix with independent, identically distributed entries, then the Sanov theorem implies the large deviation theorem. But in most cases random matrix consists of n^2 random variables, and the eigenvalues are not independent.

Assume that $T_n(\omega)$ is a random $n \times n$ matrix with complex eigenvalues $\zeta_1(\omega), \dots, \zeta_n(\omega)$. (If we want, we can fix an ordering of the eigenvalues, for example, regarding their absolute values and phases, but that is not necessary.) The *empirical eigenvalue distribution* of $T_n(\omega)$ is the random atomic measure

$$\widehat{P}_n(\omega) := \frac{\delta(\zeta_1(\omega)) + \dots + \delta(\zeta_n(\omega))}{n}.$$

Therefore $\widehat{P}_n$ is a random measure, or in other words a measure-valued random variable. Now denote P_n the distribution of $\widehat{P}_n$, which means P_n is a probability measure on $\mathcal{M}(\mathbb{C})$.

The degree of freedom is n^2, since a random matrix consists of n^2 random variables, so $L(n) = n^2$. The limit measure of the eigenvalue distribution is the unique minimizer of the rate function.

For the matrices mentioned in the Section 1 we know, that the limit of this random measure sequence is a non-random measure so there is a chance to prove the large deviation theorem for the rate of convergence of these sequences of random variables. First consider the simplest example for random matrix.

Suppose that D_n is an $n \times n$ diagonal random matrix with independent identically distributed real entries. Suppose moreover that the continuous density function f is supported on the interval $[a, b]$, and $f(x) > 0$ if $a < x < b$. Then the joint density of the eigenvalues is

$$f_{\lambda_1,\dots,\lambda_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i)$$

on $[a, b]^n$. This gives a measure ν_n on $\mathbb{R}^n$. If X is compact and $\mathcal{A}$ is a base for the topology, then the large deviation principle is equivalent to the following conditions (Theorem 4.1.11. and 4.1.18 in [8]):

$$-I(x) = \inf_{x\in G, G\in\mathcal{A}} \left\{ \limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\} = \inf_{x\in G, G\in\mathcal{A}} \left\{ \liminf_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\} \tag{30}$$

for all $x \in X$. Now suppose that $G \subset \mathcal{M}([a, b])$ is a neighbourhood of $\mu \in \mathcal{M}([a, b])$, and denote P_n the distribution of the empirical eigenvalue distribution of D_n. Then if

$$G_0 := \left\{ (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n \middle| \mu_\lambda := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i} \in G \right\},$$

then

$$\begin{aligned} P_n(G) = \nu_n(G_0) &= \int \dots \int_{G_0} \exp\left(\sum_{i=1}^{n} \log f(\lambda_i) \right) d\lambda_1 \dots d\lambda_n \\ &= \int \dots \int_{G_0} \exp\left(n \int \log f(x)\, d\mu_\lambda(x) \right) \le \exp\left(n \sup_{\mu'\in G} \int \log f(x)\, d\mu'(x) \right), \end{aligned}$$

so

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(G) \le \sup_{\mu'\in G} \int \log f(x)\, d\mu_\lambda(x),$$

so by the weak* continuity of $\mu' \mapsto \int \log f(x)\, d\mu'(x)$ we have

$$\inf_{G:\mu\in G} \left(\limsup_{n\to\infty} \frac{1}{n} \log P_n(G) \right) \le \int \log f(x)\, d\mu(x).$$

For the other equality of (30) we suppose that the measure μ has continuous density g, since we can approximate with the measure with density function

$$\int_a^b \varphi_\varepsilon(x - y)\, d\mu(y),$$

where φ is a C^∞ function supported on the interval $[-\varepsilon, \varepsilon]$, such that

$$\int_{-\varepsilon}^{\varepsilon} \varphi(x) = 1.$$

Moreover we can assume that $\delta < g(x) < \delta^{-1}$ for $x \in [a,b]$ and for some $\delta > 0$, if we have a convex combination of the normalized Lebesgue measure on $[a,b]$ and μ, and we take the limit as the coefficient of μ tends to 1. Then for all n consider the partition of the interval

$$a = y_0^{(n)} < y_1^{(n)} < \cdots < y_{n-1}^{(n)} < y_n^{(n)} = b,$$

such that

$$\int_{y_{i-1}^{(n)}}^{y_i^{(n)}} g(x)\,dx = \frac{1}{n},$$

for $1 \le i \le n$, then

$$\frac{\delta}{n} \le x_i^{(n)} - x_{i-1}^{(n)} \le \frac{1}{n\delta}.$$

Set

$$\Delta_n = \left\{ (\lambda_1, \dots, \lambda_n) \middle| x_{i-1}^{(n)} < \lambda_i < x_i^{(n)} \right\},$$

then for n large enough $\Delta_n \subset G_0$, so we have

$$\begin{aligned}
P_n(G) &= \nu_n(G_0) \ge \nu_n(\Delta_n) \\
&\ge \int \dots \int_{\Delta_n} \exp\left(\sum_{i=1}^n \log f(x_i) \right) dx_1 \dots dx_n \\
&\ge \left(\frac{\delta}{n} \right)^n \exp\left(\sum_{i=1}^n \min_{x \in [x_{i-1}^{(n)}, x_i^{(n)}]} \log f(x) \right),
\end{aligned}$$

so since

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n \min_{x \in [x_{i-1}^{(n)}, x_i^{(n)}]} \log f(x) = \int g(x) \log f(x)\,dx = \int \log f(x)\,d\mu(x)$$

thus

$$\inf_{G:\mu\in G} \left(\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \right) \ge \int \log f(x)\,d\mu(x).$$

In this way we proved the Sanov theorem for the random variables with density f.

In Section 1 we could see, that for the convergence of the empirical eigenvalue distribution there is no need to know the density of the entries. Again we will use the exact form of the joint density of the eigenvalues as above, which is known only in the case of random matrices which are invariant under unitary conjugation. So in this section we will study only Gaussian random matrices.

The first large deviation theorem for random matrices was proven by Ben Arous and Guionnet in [5], and it concerns the standard selfadjoint Gaussian matrices.

Theorem 2.2 (Ben Arous, Guionnet) *Let $\widehat{P}_n$ is the empirical eigenvalue distribution of the standard selfadjoint Gaussian matrix A_n, i.e. a random measure on $\mathbb{R}$.*

Then the large deviation principle holds on the scale n^{-2} with rate function

$$I^{sa}(\mu) := -\iint_{\mathbb{R}^2} \log|x-y|\, d\mu(x)\, d\mu(y) + \frac{1}{2}\int_{\mathbb{R}} x^2\, d\mu(x) - B, \tag{31}$$

Where

$$B_{sa} = -\lim_{n\to\infty}\frac{1}{n^2}\log C_n^{sa} = \frac{1}{4}\log 2 + \frac{3}{8},$$

and C_n^{sa} is the normalization constant defines in (14).

In their paper they proved the large deviation theorem for real case as well. Moreover they proved the large deviation for the sequence of matrices $p(A_n)$, where $p : \mathbb{R} \to \mathbb{R}$ is a bounded, positive diffeomorphism, and $p(A_n)$ is again defined by the usual function calculus of selfadjoint matrices. In this case the topological space is again $\mathcal{M}(\mathbb{R})$ with the topology of the weak convergence.

The next theorem was made by Hiai and Petz in [23] about the Wishart type random matrices, when $p/n \stackrel{n\to\infty}{\longrightarrow} \lambda < 1$.

Theorem 2.3 (Hiai, Petz) *Let $\widehat{P}_n$ is the empirical eigenvalue distribution of the $p\times p$ Wishart matrix, i.e. a random measure on $\mathbb{R}^+$. Then the large deviation principle holds on the scale p^{-2} with rate function*

$$I^{wish}(\mu) := -\frac{1}{2}\iint_{(\mathbb{R}^+)^2} \log|x-y|\, d\mu(x)\, d\mu(y) + \frac{1}{2}\int_{\mathbb{R}}(x-(\lambda-1)\log x)\, d\mu(x) - B_{wish}, \tag{32}$$

Where

$$\begin{aligned} B_{wish} &= -\lim_{n\to\infty}\frac{1}{p^2}\log C_{n,p}^{wish} \\ &= \frac{1}{4}\left(3\lambda - \lambda^2\log\lambda + (1-\lambda)^2\log(1-\lambda)\right) \end{aligned} \tag{33}$$

In this paper Hiai and Petz proved more. They considered $p\times p$ positive matrices with the joint eigenvalue density function

$$\frac{1}{Z_n}\exp\left(-n\sum_{i=1}^{p} Q(\lambda_i)\right)\prod_{i=1}^{n}\lambda_i^{\gamma(n)}\prod_{1\le i<j\le p}|\lambda_i-\lambda_j|^{2\beta},$$

where $\beta > 0$ fixed, and Q is a real continuous function such that for all $\varepsilon > 0$

$$\lim_{x\to\infty} x\exp\left(-\varepsilon Q(x)\right) = 0. \tag{34}$$

Then the large deviation principle hold if $p/n \stackrel{n\to\infty}{\longrightarrow} \lambda > 1$ and $\gamma(n)/n \stackrel{n\to\infty}{\longrightarrow} \gamma > 0$.

We know the convergence for the case $p/n \ge 1$, and by the following lemma 2.3 proves the large deviation principle as well.

Lemma 2.2 *For $n \in \mathbb{N}$ let $\widetilde{P}_n$ be a random probability measure on a complete separable metric space $\mathcal{X}$. Let μ_0 be a fixed probability measure on $\mathcal{X}$ and $0 < \alpha_n < 1$ such that $\alpha_n \stackrel{n\to\infty}{\longrightarrow} \alpha \in (0,1)$. Suppose that $(\widetilde{P}_n)$ is exponentially tight, i.e. for all $L \geq 0$ there exists a $K_L \subset \mathcal{X}$ compact set, such that*

$$\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(K_L^c) \leq -L, \tag{35}$$

where K_L^c denotes the complement of K_L. If $(\widetilde{P}_n)$ satisfies the large deviation principle at the scale $L(n)$ with rate function $\widetilde{I}$ on $\mathcal{M}(\mathcal{X})$, then the sequence of random measures

$$(1-\alpha_n)\mu_0 + \alpha_n \widetilde{P}_n$$

satisfies the same with good rate function

$$I(\mu) := \begin{cases} \widetilde{I}(\widetilde{\mu}), & \text{if} \quad \mu = (1-\alpha_n)\mu_0 + \alpha_n\widetilde{\mu} \\ \infty, & \text{otherwise.} \end{cases}$$

If we apply the above lemma for $\alpha_n = \frac{n}{p}$ and $\mu_0 = \delta_0$, we have that the large deviation principle hold for the singular Wishart matrices as well, i.e. in the case when $n < p$.

Finally Hiai and Petz proved the following theorem in [33].

Theorem 2.4 (Hiai, Petz) *Let $\widehat{P}_n$ is the empirical eigenvalue distribution of the $n\times n$ Gaussian elliptic random matrix*

$$Y_n := uX_n + vX_n^*,$$

where $u^2 + v^2 = 1$. Then $\widehat{P}_n$ is a random measure on $\mathbb{C}$. Then the large deviation principle holds on the scale n^{-2} with rate function

$$I^{ell}(\mu) := -\iint_{\mathbb{C}^2} \log|z-w|\, d\mu(z)\, d\mu(w) + \int_{\mathbb{C}} \left(\frac{Re\, z^2}{(u+v)^2} + \frac{Im\, z^2}{(u-v)^2} \right) d\mu(z) - B_{ell}, \tag{36}$$

Where

$$B_{ell} = -\lim_{n\to\infty} \frac{1}{n^2} \log C_n^{ell} = \frac{3}{4}. \tag{37}$$

By the following theorem large deviations of the empirical eigenvalue distribution of random matrices imply other large deviation theorems. (See Theorem 4.2.1 in [8])

Theorem 2.5 (Contraction principle) *If the sequence $\mu_n \in \mathcal{M}(\mathcal{X})$ satisfies the large deviation principle with rate function I and $f : \mathcal{X} \to \mathcal{Y}$ is a continuous function, then the sequence ν_n defined by*

$$\nu_n(B) := \mu_n(f^{-1}(B))$$

satisfies the large deviation principle with rate function

$$J(y) := \inf\{I(x) \big| f(x) = y\}.$$

For example for a continuous $\varphi : \mathbb{C} \to \mathbb{C}$ function consider $f_\varphi : \mathcal{M}(\mathcal{C}) \to \mathbb{C}$,

$$f_\varphi(\mu) := \int \varphi(x)\, d\mu(x).$$

This function is continuous in the weak* topology, so if the large deviation theorem holds for the distribution P_n of the empirical eigenvalue distribution of the $n \times n$ random matrix X_n , then the distribution of

$$\int \varphi(x)\, d\mu_n(x) = \frac{1}{n} \sum_{i=1}^{n} \varphi(\lambda_i(X_n))$$

satisfies the large deviation theorem too. On the other hand the exact form of the rate function

$$J(y) := \inf \left\{ \iint \log |z - w|\, d\mu(z)\, d\mu(w) \middle| \int \varphi(z)\, d\mu(z) = y \right\}$$

is rather difficult.

2.3 Potential theory and large deviations

Next we recall some definitions and theorems of potential theory [36].

Definition 2.3 *For a signed measure ν on a K compact subset of $\mathbb{C}$*

$$I(\nu) := \iint_{K^2} \log \frac{1}{|z - w|}\, d\nu(z)\, d\nu(w) \tag{38}$$

is the so-called logarithmic energy of ν.

Definition 2.4 *For a signed measure ν on a K compact subset of $\mathbb{C}$*

$$\Sigma(\nu) := \iint_{K^2} \log |z - w|\, d\nu(z)\, d\nu(w) \tag{39}$$

is the so-called free entropy of ν.

Since

$$\Sigma(\nu) = \inf_{\alpha<0} \iint_{K^2} \max(\log |z - w|, \alpha)\, d\nu(z)\, d\nu(w),$$

this functional is upper semi-continuous. We want to show its concavity. The following lemma is strongly related to the properties of the logarithmic kernel $K(z, w) = \log |z - w|$ (cf. Theorem 1.16 in [29]).

Lemma 2.5 *Let ν be a compactly supported signed measure on $\mathbb{C}$ such that $\nu(\mathbb{C}) = 0$. Then $\Sigma(\nu) \leq 0$, and $\Sigma(\nu) = 0$ if and only if $\nu = 0$.*

From this lemma we can deduce strictly concavity of the functional Σ. First we prove that

$$\Sigma\left(\frac{\mu_1 + \mu_2}{2}\right) \geq \frac{\Sigma(\mu_1) + \Sigma(\mu_2)}{2}, \tag{40}$$

for all $\mu_1, \mu_2 \in \mathcal{M}(K)$, moreover the equality holds if and only if $\mu_1 = \mu_2$. For this, apply Lemma 2.5 for the signed measure $\nu = \mu_1 - \mu_2$. We get in the case of $\mu_1 \neq \mu_2$

$$0 > \Sigma(\mu_1 - \mu_2) = \Sigma(\mu_1) + \Sigma(\mu_2) - 2 \iint_{K^2} \log|z - w| \, d\mu_1(z) \, d\mu_2(w),$$

thus

$$\frac{\Sigma(\mu_1) + \Sigma(\mu_2)}{2} < \iint_{K^2} \log|z - w| \, d\mu_1(z) \, d\mu_2(w),$$

and

$$\Sigma\left(\frac{\mu_1 + \mu_2}{2}\right) = \frac{\Sigma(\mu_1) + \Sigma(\mu_2)}{4} + \frac{1}{2} \iint_{K^2} \log|z - w| \, d\mu_1(z) \, d\mu_2(w) > \frac{\Sigma(\mu_1) + \Sigma(\mu_2)}{2}.$$

The concavity is the property

$$\Sigma(\lambda\mu_1 + (1 - \lambda)\mu_2) \geq \lambda\Sigma(\mu_1) + (1 - \lambda)\Sigma(\mu_2) \tag{41}$$

for an arbitrary $\lambda \in [0, 1]$. If $\Sigma(\mu_1) = -\infty$ or $\Sigma(\mu_2) = -\infty$, then this holds trivially. Next assume that $\Sigma(\mu_1) > -\infty$ and $\Sigma(\mu_2) > -\infty$. Then we have (41) for dyadic rational λ from the midpoint concavity (40). For an arbitrary $\lambda \in [0, 1]$ we proceed by approximation. For a fixed sequence $\varepsilon_n > 0$, $\varepsilon_n \to 0$, there exist $i(n), k(n) \in \mathbb{N}$ such that

$$\left|\left(\frac{i(n)}{2^{k(n)}} - \lambda\right)\Sigma(\mu_1) + \left(\lambda - \frac{i(n)}{2^{k(n)}}\right)\Sigma(\mu_2)\right| < \varepsilon.$$

By the midpoint concavity

$$\begin{aligned} \lambda\Sigma(\mu_1) + (1 - \lambda)\Sigma(\mu_2) - \varepsilon_n &< \frac{i(n)}{2^{k(n)}}\Sigma(\mu_1) + \left(1 - \frac{i(n)}{2^{k(n)}}\right)\Sigma(\mu_2) \\ &\leq \Sigma\left(\frac{i(n)}{2^{k(n)}}\mu_1 + \left(1 - \frac{i(n)}{2^{k(n)}}\right)\mu_2\right). \end{aligned}$$

Here

$$\frac{i(n)}{2^{k(n)}}\mu_1 + \left(1 - \frac{i(n)}{2^{k(n)}}\right)\mu_2 \stackrel{n\to\infty}{\longrightarrow} \lambda\mu_1 + (1 - \lambda)\mu_2,$$

and the upper semi-continuity of Σ implies

$$\limsup_{n\to\infty} \Sigma\left(\frac{i(n)}{2^{k(n)}}\mu_1 + \left(1 - \frac{i(n)}{2^{k(n)}}\right)\mu_2\right) \le \Sigma(\lambda\mu_1 + (1-\lambda)\mu_2),$$

which gives the concavity (41) and the equality can hold only in the trivial case.

Since for all $\mu \in \mathcal{M}(K)$

$$I(\mu) = -\Sigma(\mu),$$

the above properties of Σ imply that the logarithmic energy is a convex, lower semi-continuous function.

Definition 2.6 *The quantity*

$$\mathrm{cap}(K) := e^{-V}$$

is called the logarithmic capacity of K*, where*

$$V := \inf\{I(\mu) : \mu \in \mathcal{M}(K)\}.$$

The logarithmic potential of $\mu \in \mathcal{M}(K)$ *is the function*

$$U^{\mu}(z) := \int_K \log\frac{1}{|z-w|}\, d\mu(w) \tag{42}$$

defined on K*.*

Definition 2.7 *Let* $F \subset \mathbb{C}$ *be a closed set, and* $Q : F \to (-\infty, \infty]$ *be a lower semi-continuous function. The integral*

$$I_Q(\mu) := \iint_{F^2} \log\frac{1}{|z-w|}\, d\mu(z)\, d\mu(w) + 2\int_F Q(z)\, d\mu(z) \tag{43}$$

is called weighted energy.

The weight function is

$$w(z) := \exp(-Q(z)) \tag{44}$$

is admissible if it satisfies the following conditions

- w is upper semicontinuous;
- $F_0 := \{z \in F : w(z) > 0\}$ has positive capacity;
- if F is unbounded then $|z|w(z) \to 0$ as $|z| \to \infty$, $z \in F$.

We can recognize, that the rate functions in the large deviation theorems are weighted energy functionals with different rate functions. For example, in the case of selfadjoint Gaussian matrices the weight function

$$w^{sa}(x) = \exp\left(-\frac{x^2}{4}\right)$$

which is clearly an admissible weight function.

Now consider a theorem (cf. Theorem I.1.3 in [36]) about the minimizer of the weighted energy.

Theorem 2.6 *Let $w = exp(-Q)$ be an admissible weight on a closed set Σ, and let*

$$V_Q := \inf\{I_Q(\mu) : \mu \in \mathcal{M}(F)\}.$$

Then the following properties hold.

- *V_Q is finite.*
- *There exists a unique element $\mu_Q \in \mathcal{M}(F)$ such that*

  $$I_Q(\mu_Q) = V_Q.$$

  Moreover μ_Q has finite logarithmic energy.
- *$S_Q := \operatorname{supp}(\mu_Q)$ is compact, $S_Q \subset F_0$, and has positive capacity.*

Definition 2.8 *The measure μ_Q is called the equilibrium or extremal measure associated with w.*

The following result tells about the minimizer of the weighted potential (cf. Theorem I.3.3 in [36]).

Proposition 2.9 *Let Q as above. Assume that $\sigma \in \mathcal{M}(K)$ has compact support, $E(\sigma) < \infty$ and there exists a constant F such that*

$$U^\sigma(z) + Q(z) = F$$

if $z \in \operatorname{supp}\sigma$, and

$$U^\sigma(z) + Q(z) \geq F$$

if $z \in K$. Then σ is the measure in $\mathcal{M}(K)$ such that

$$I_Q(\sigma) = \inf_{\mu \in \mathcal{M}(K)} I_Q(\mu),$$

i.e., σ is the so-called equilibrium measure associated with Q.

The above proposition gives a very useful hint to find the equilibrium measure of a weighted energy. For example its corollary is the following theorem, which helps us to prove that the rate function of the large deviation principle for the selfadjoint Gaussian matrices has the Wigner semicircle distribution as the unique minimizer, since it can be written in the form

$$\frac{1}{2\pi}\sqrt{4-t^2}=\frac{1}{2\pi}\int_{|t|}^{2}\frac{u}{\sqrt{u^2-t^2}}\,du,$$

on $[-2,2]$, so

$$-U(x)=\frac{1}{2\pi}\int_{-2}^{2}\log|x-t|\int_{|t|}^{2}\frac{u}{\sqrt{u^2-t^2}}\,du\,dt=\int_{0}^{2}\frac{u}{2}\cdot\frac{1}{\pi}\int_{-u}^{u}\frac{\log|x-t|}{\sqrt{u^2-t^2}}\,dt\,du.$$

Here by $t=u\cos\vartheta$ we have

$$\frac{1}{\pi}\int_{-u}^{u}\frac{\log|x-t|}{\sqrt{u^2-t^2}}\,dt=-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{1}{|x-u\cos\vartheta|}\,d\vartheta.$$

If we apply the so-called Joukowski transformation (See [36] Example 3.5)

$$x=\frac{u}{2}\left(\zeta+\frac{1}{\zeta}\right),$$

then

$$\zeta=\begin{cases}\operatorname{sgn}(x)\dfrac{x+\sqrt{x^2-u^2}}{u} & \text{if} \quad |x|>u\\[2ex] \operatorname{sgn}(x)\mathrm{i}\sqrt{u^2-x^2} & \text{if} \quad 0\le|x|\le u\end{cases}$$

Then since

$$|x-u\cos\vartheta|=\left|\frac{u}{2}(\zeta+\zeta^{-1})-(e^{\mathrm{i}\vartheta}+e^{-\mathrm{i}\vartheta})\right|=\frac{u}{2}|\zeta-e^{\mathrm{i}\vartheta}||\zeta^{-1}-e^{\mathrm{i}\vartheta}|,$$

and by

$$\frac{1}{2\pi}\int_{0}^{2\pi}\log\frac{1}{|z-re^{\mathrm{i}\varphi}|}\,d\varphi=\begin{cases}-\log r, & \text{if } |z|\le r\\ -\log|z|, & \text{if } |z|>r,\end{cases},$$

thus

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{1}{|x-u\cos\vartheta|}\,d\vartheta\\ &=\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{2}{|\zeta-t||\zeta^{-1}-t|}=\begin{cases}\log 2-\log\left|x+\sqrt{x^2-u^2}\right|, & \text{if} \quad |x|>u\\ \log 2-\log u, & \text{if} \quad |x|\le u.\end{cases}\end{aligned}$$

Then if $-2\le x\le 2$

$$\begin{aligned}-U(x)\quad&=-\log 2+\frac{1}{2}\int_{|x|}^{2}u\log u\,du+\frac{1}{2}\int_{0}^{|x|}u\log|x+\sqrt{x^2-u^2}|\,du\\ &=-\log 2+\frac{1}{2}\left[\frac{u^2\log u}{2}-\frac{u^2}{4}\right]_{|x|}^{2}+\int_{0}^{|x|}\frac{u}{2}\log|x|\,du\\ &\qquad+\frac{|x|^2}{2}\int_{0}^{1}v\log|1+\sqrt{1-v^2}|\,dv=-\frac{1}{2}+\frac{|x|^2}{4},\end{aligned}$$

since

$$\int_0^1 v \log|1+\sqrt{1-v^2}|\,dv = \int_0^1 \frac{v}{1+\sqrt{1-v^2}} \cdot \frac{v^2}{\sqrt{1-v^2}}\,dv$$
$$= \int_0^1 \frac{v(1-\sqrt{1-v^2})(1+\sqrt{1-v^2})}{(1+\sqrt{1-v^2})\sqrt{1-v^2}}\,dv = \int_0^1 \left(\frac{v}{\sqrt{1-v^2}} - v\right)\,dv = \frac{1}{2}.$$

If $|x| > 2$, then by the symmetry we can suppose that $x > 2$, and with similar calculations

$$U(x) = \log 2 - \frac{1}{2}\int_0^2 u \log|x+\sqrt{x^2-u^2}|\,du$$
$$= -\log 2 - \frac{x^2}{4} - \log(x+\sqrt{x^2-4}) + \frac{x}{4}\sqrt{x^2-4} + \frac{1}{2},$$

and since here the weight function

$$Q(x) := \frac{x^2}{4},$$

so

$$U(x) + Q(x) = \frac{1}{2} \quad \text{if} \quad |x| \leq 2,$$

and

$$U(x) + Q(x) \geq \frac{1}{2} \quad \text{if} \quad |x| > 2,$$

so the semicircular distribution is equilibrium measure of the weighted energy, i.e. the unique minimizer of rate function I_{wig}.

Proposition 2.9 can be used to prove that the unique minimizer of I_{wish} is the Marchenko-Pastur distribution, and the minimizer of I_{ell} is the uniform distribution on the corresponding ellipse. Later we will use this Proposition to find the equilibrium measure of a weighted energy.

We could see, that the rate function of the large deviation theorem for random matrices is a weighted logarithmic energy, which has a unique equilibrium measure μ_0, so we can write the rate function in the following form

$$I(\mu) = \mathcal{I}_Q(\mu) - \mathcal{I}_Q(\mu_0),$$

so we can consider the rate function I again as a relative entropy with respect to the minimizer μ_0.

3 Haar unitaries

Apart from the selfadjoint random matrices there is an other important set of normal random matrices, the unitary random matrices. We already used non-random unitary matrices in the previous sections, but now we recall the definition, since in the sequel we will study random unitary matrices.

A unitary matrix $U = (U_{ij})$ is a matrix with complex entries and $UU^* = U^*U = I$. In terms the entries these relations mean that

$$\sum_{j=1}^{n} |U_{ij}|^2 = \sum_{i=1}^{n} |U_{ij}|^2 = 1, \text{for all } 1 \leq i, j \leq n, \tag{45}$$

$$\sum_{l=1}^{n} U_{il}\overline{U}_{lj} = 0, \text{for all } 1 \leq i, j \leq n, \quad i \neq j. \tag{46}$$

In other words an $n \times n$ matrix is unitary if the columns (or rows) are pairwise orthogonal unit vectors.

The set $\mathcal{U}(n)$ of $n \times n$ unitary matrices forms a compact topological group with respect to the matrix multiplication and the usual topology, therefore there exists a unique (up to the scalar multiplication) translation invariant measure on $\mathcal{U}(n)$, the so-called Haar measure. We will consider a random variable U_n which maps from a probability space to $\mathcal{U}(n)$, and take its values uniformly from $\mathcal{U}(n)$, i.e. if $H \subset \mathcal{U}(n)$, then

$$\mathbb{P}(U_n \in H) = \gamma(H),$$

where γ is the normalized Haar measure on $\mathcal{U}(n)$. We call this random variable a Haar unitary random variable, or shortly Haar unitary.

Although the distribution of the entries cannot be normal, since the absolute values must lie on the interval $[0, 1]$, some properties of the normal variables play important role in the construction of the Haar unitary random matrices.

3.1 Construction of a Haar unitary

Next we recall how to get a Haar unitary from a Gaussian matrix with independent entries by the Gram-Schmidt orthogonalization procedure on the column vectors. Suppose that we have a complex random matrix Z whose entries Z_{ij} are mutually independent standard complex normal random variables. We perform the Gram-Schmidt orthogonalization procedure on the column vectors Z_i $(i = 1, 2, \ldots, n)$, i.e.

$$U_1 = \frac{Z_1}{\|Z_1\|},$$

and

$$U_i = \frac{Z_i - \sum_{l=1}^{i-1} \langle Z_i, U_l \rangle U_l}{\left\| Z_i - \sum_{l=1}^{i-1} \langle Z_i, U_l \rangle U_l \right\|}, \tag{47}$$

where

$$\|(X_1, X_2, \dots, X_n)\| = \sqrt{\sum_{k=1}^{n} |X_k|^2}\,.$$

Lemma 3.1 *The above column vectors U_i constitute a unitary matrix $U = (U_i)_{i=1,\dots,n}$. Moreover, for all $V \in \mathcal{U}(n)$ the distributions of U and VU are the same.*

Proof. From the proof of Lemma 1.4, we know, that the distributions of Z and VZ are the same. The ith column of VU is exactly VU_i and we have

$$VU_i = \frac{VZ_i - \sum_{l=1}^{i-1} \langle Z_i, U_l \rangle VU_l}{\left\| Z_i - \sum_{l=1}^{i-1} \langle Z_i, U_l \rangle U_l \right\|} = \frac{VZ_i - \sum_{l=1}^{i-1} \langle VZ_i, VU_l \rangle VU_l}{\left\| VZ_i - \sum_{l=1}^{i-1} \langle VZ_i, VU_l \rangle VU_l \right\|} \tag{48}$$

which is the Gram-Schmidt orthogonalization of the vectors VZ_i. Since we showed above that Z and VZ are identically distributed, we conclude that U and VU are identically distributed as well. Since the left invariance characterizes the Haar measure on a compact group, the above constructed U is Haar distributed and its distribution is right invariant as well. □

The column vectors of a unitary matrix are pairwise orthogonal unit vectors. On the bases of this fact we can determine a Haar unitary in a slightly different way. The complex unit vectors form a compact space on which the unitary group acts transitively. Therefore, there exist a unique probability measure invariant under the action. Let us call this measure uniform. To determine a Haar unitary, we choose the first column vector U_1 uniformly from the space of n-vectors. U_2 should be taken from the $n-1$ dimensional subspace orthogonal to U_1 and choose it uniformly again. In general, if already $U_1, U_2, \dots, U_j$ is chosen, we take U_{j+1} from the $n-j$ dimensional subspace orthogonal to $U_1, U_2, \dots, U_j$, again uniformly. The column vectors constitute a unitary matrix and we check that its distribution is left invariant. Let V be a fixed unitary. We show that the vectors $VU_1, VU_2, \dots, VU_n$ are produced by the above described procedure. They are obviously pairwise orthogonal unit vectors. VU_1 is uniformly distributed by the invariance property of the distribution of U_1. Let $V(1)$ be such a unitary that $V(1)VU_1 = VU_1$. Then $V^{-1}V(1)VU_1 = U_1$ and the choice of U_2 gives that $V^{-1}V(1)VU_2 \sim U_2$. It follows that $V(1)VU_2 \sim VU_2$. Since $V(1)$ was arbitrary VU_2

is uniformly distributed in the subspace orthogonal to VU_1. Similar argument works for $VU_3, \dots, VU_n$. The Gram-Schmidt orthogonalization of the columns of a Gaussian matrix gives a concrete realization of this procedure. Now suppose that A is a random matrix with independent identically distributed entries, where the distribution of the entries has finite mean. Then if the distribution of the entries is absolutely continuous with respect to the Lebesgue measure, then we can construct a random unitary matrix with the above methods. This unitary random matrix is not translation invariant, because the only unitary invariant distribution according to Theorem 1.3 is the normal distribution. If the distribution is not continuous, then A can be singular with positive probability, so the Gram-Schmidt orthogonalization does not work almost surely.

3.2 General properties of Haar unitaries

The entries of a Haar unitary random matrix are clearly not independent, since for example the sum of the square of the absolute values of the entries in the same row or column must be 1. It is difficult to find the joint density of the entries, but now from the translation invariance of the Haar measure and from the construction we can state several facts about the entries.

For example since permutation matrices are in $\mathcal{U}(n)$, and by multiplying with an appropriate permutation matrix every row and column can be transformed to any other row or column, so the translation invariance of a Haar unitary U implies that all the entries have the same distribution.

Theorem 3.1 *From the construction of a Haar unitary one can deduce easily the distribution of the entries:*

$$\frac{n-1}{\pi}(1-r^2)^{n-2} r\, dr\, d\vartheta,$$

Proof. We know from the construction and from Lemma 1.2, that

$$U_{11} = \frac{Z_{11}}{\sqrt{\sum_{i=1}^n Z_{i1}^2}} = \frac{R_1 e^{\mathrm{i}\vartheta_1}}{\sqrt{\sum_{i=1}^n R_i^2}}, \tag{49}$$

where $Z_{i1} = R_i e^{\mathrm{i}\vartheta_i}$, $R_1^2, \dots, R_n^2$ are independent exponentially distributed random variables with parameter 1, and $\vartheta_1, \dots, \vartheta_n$ are independent uniformly distributed random variables on the interval $[0, 2\pi]$. Clearly the phase of U_{11} depends only on ϑ_1, and it is independent from the absolute value of the entry, and uniform on the interval $[0, 2\pi]$. For the absolute value, we know, that the density function of the sum k independent identically distributed exponential random variables with parameter λ is

$$f_k(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \tag{50}$$

on $x \in \mathbb{R}^+$, so

$$\begin{aligned}
\mathbb{P}(|U_{11}| \leq r) &= \mathbb{P}\left(\frac{R_1}{\sqrt{\sum_{i=1}^n R_i^2}} < r\right) \\
&= \mathbb{P}\left(R_1^2 < \frac{r^2 \sum_{i=2}^n R_i^2}{1-r^2}\right) \\
&= \int_0^\infty \int_0^{\frac{r^2}{1-r^2}y} e^{-x} \frac{y^{n-2}e^{-y}}{(n-2)!}\, dx\, dy \\
&= \frac{1}{(n-1)!} \int_0^\infty \left(1 - e^{\frac{r^2}{1-r^2}y}\right) y^{n-2} e^{-y}\, dy \\
&= \frac{1}{(n-2)!} \left(\int_0^\infty y^{n-2} e^{-y}\, dy - \int_0^\infty y^{n-2} e^{-\frac{y}{1-r^2}}\, dy\right) \\
&= 1 - (1-r^2)^{n-1} \\
&= 2(n-1) \int_0^r \rho(1-\rho)^{n-2}\, d\rho,
\end{aligned}$$

since from (1) we know the kth moment the exponential random variable. $\square$

Lemma 3.2 *The joint distribution of $U_{11}, \ldots, U_{n-1,1}$ is uniform on the set*

$$\left\{(x_1, \ldots, x_{n-1}) : \sum_{i=1}^n x_i^2 \leq 1\right\}.$$

Proof. Suppose that $X_1, \ldots, X_n$ are independent exponentially distributed random variables with parameter 1, then

$$|U_{j1}|^2 = \frac{X_j}{\sum_{i=1}^n X_i},$$

so the joint distribution of $|U_{11}|^2, \ldots, |U_{n-1,1}|^2$ is same as the joint distribution of

$$\frac{X_1}{\sum_{i=1}^n X_i}, \ldots, \frac{X_{n-1}}{\sum_{i=1}^n X_i}.$$

The joint density of $X_1, \ldots X_n$ is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) := e^{-(x_1+\cdots+x_n)} \tag{51}$$

on $(\mathbb{R}^+)^n$, so if we use the transformation

$$(x_1, \ldots, x_n) \mapsto \left(\frac{x_1}{\sum_{i=1}^n x_i}, \ldots, \frac{x_{n-1}}{\sum_{i=1}^n x_i}, \sum_{i=1}^n x_i\right),$$

and we integrate with respect of the last variable, then we have the density. The Jacobian of the transformation has the determinant

$$\det\begin{pmatrix} \frac{\sum_{i=1}^n x_i - x_1}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{-x_1}{\left(\sum_{i=1}^n x_i\right)^2} & \dots & \frac{-x_1}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{-x_1}{\left(\sum_{i=1}^n x_i\right)^2} \\ \frac{-x_2}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{\sum_{i=1}^n x_i - x_2}{\left(\sum_{i=1}^n x_i\right)^2} & \dots & \frac{-x_2}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{-x_2}{\left(\sum_{i=1}^n x_i\right)^2} \\ \vdots & & \ddots & & \vdots \\ \frac{-x_{n-1}}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{-x_{n-1}}{\left(\sum_{i=1}^n x_i\right)^2} & \dots & \frac{\sum_{i=1}^n x_i - x_{n-1}}{\left(\sum_{i=1}^n x_i\right)^2} & \frac{-x_{n-1}}{\left(\sum_{i=1}^n x_i\right)^2} \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}$$

$$= \det\begin{pmatrix} \frac{\sum_{i=1}^n x_i}{\left(\sum_{i=1}^n x_i\right)^2} & 0 & \dots & 0 & 0 \\ 0 & \frac{\sum_{i=1}^n x_i}{\left(\sum_{i=1}^n x_i\right)^2} & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & \frac{\sum_{i=1}^n x_i}{\left(\sum_{i=1}^n x_i\right)^2} & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} = \left(\sum_{i=1}^n x_i\right)^{-(n-1)}.$$

With (51) we have that the joint density function of the new random variables depends only $\sum_{i=1}^n x_i$. If we integrate with respect to this variable, then we get, that the joint density of the other $n-1$ random variables is constant. We obtained that the joint density of $|U_1 1|^2, \dots, |U_{n-1,1}|^2$ is uniform on the set $\{(x_1, \dots, x_{n-1}) : \sum_{i=1}^n x_i \leq 1\}$, so since the phase of U_{1i} are independent uniformly distributed on $[0, 2\pi]$ we proved the lemma. $\square$

Since we know the density of the entries we can compute the even moments of their absolute value. For every $k \in \mathbb{Z}^+$,

$$\mathbb{E}\left(|U_{ij}|^{2k}\right) = \binom{n+k-1}{n-1}^{-1} \tag{52}$$

for all $1 \leq i, j \leq n$. This can be easily computed from the density function as follows

$$
\begin{aligned}
\mathbb{E}\left(|U_{ij}|^{2k}\right) &= (n-1)\int_0^1 r^{2k+1}(1-r^2)^{n-2}\,dr \\
&= (n-1)\left(-\left[r^{2k}\frac{(1-r^2)^{n-1}}{n-1}\right]_0^1 + \frac{k}{n-1}\int_0^1 r^{2k-1}(1-r^2)^{n-1}\,dr\right) \\
&= k\cdot\frac{k-1}{n}\int_0^1 r^{2k-3}(1-r^2)^n\,dr \\
&= \frac{k!}{n\ldots(n+k-2)}\int_0^1 r(1-r^2)^{n+k-2} = \\
&= \binom{n+k-1}{n-1}^{-1}
\end{aligned}
$$

Clearly the entries are not independent, and the entries in the same row or column are more correlated then the others. The correlation coefficients can be computed as follows. Since

$$
\mathbb{E}|U_{11}|^2 = \mathbb{E}\left(\sum_{j=1}^n |U_{11}|^2|U_{1j}|^2\right) = (n-1)\mathbb{E}\left(|U_{11}|^2|U_{12}|^2\right) + \mathbb{E}\left(|U_{11}|^4\right),
$$

so

$$
\mathbb{E}\left(|U_{11}|^2|U_{12}|^2\right) = \frac{1}{n-1}\left(\frac{1}{n} - \frac{2}{(n+1)n}\right) = \frac{1}{n(n+1)},
$$

so the correlation coefficient is

$$
\frac{\mathbb{E}\left(|U_{11}|^2|U_{12}|^2\right) - \mathbb{E}|U_{11}|^2\mathbb{E}|U_{12}|^2}{\mathbb{E}\left(|U_{11}|^4\right) - \left(\mathbb{E}\left(|U_{11}|^2\right)\right)^2} = -\frac{1}{n-1}.
$$

For the entries in different row and column, we can use the fact

$$
\sum_{i=1}^n |U_{11}|^2|U_{2i}|^2 = |U_{11}|^2
$$

to calculate

$$
\mathbb{E}\left(|U_{11}|^2U_{22}|^2\right) = \frac{1}{n-1}\left(\frac{1}{n} - \frac{1}{n(n+1)}\right) = \frac{1}{n^2-1},
$$

therefore the correlation coefficient here is

$$
\frac{\mathbb{E}\left(|U_{11}|^2|U_{22}|^2\right) - \mathbb{E}|U_{11}|^2\mathbb{E}|U_{22}|^2}{\mathbb{E}\left(|U_{11}|^4\right) - \left(\mathbb{E}\left(|U_{11}|^2\right)\right)^2} = \frac{1}{(n-1)^2}.
$$

(see p. 139 in [26]).

Theorem 3.2 *Since*

$$
\mathbb{P}(|\sqrt{n}U_{ij}|^2 \geq x) = \left(1 - \frac{x}{n}\right)^{n-1} \to e^{-x}
$$

$\sqrt{n}U_{ij}$ *converges to a standard complex normal variable.*

□

3.3 Joint eigenvalue density

Let U be a Haar distributed $n \times n$ unitary matrix with eigenvalues $\lambda_1, \lambda_1, \ldots, \lambda_n$. The eigenvalues are random variables with values in $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$.

The joint density of the eigenvalues was obtained by Weyl [41],

$$\frac{1}{(2\pi)^n n!} \prod_{i<j} |e^{\mathrm{i}\vartheta_i} - e^{\mathrm{i}\vartheta_j}|^2 \tag{53}$$

with respect to $\vartheta_1 \ldots\ d\vartheta_n$. Now we write down a shortened form of the proof (see p. 135 in [26]).

At any point of $U \in \mathcal{U}(n)$ the matrix

$$dU^* U + U^*\, dU = \ d(U^* U) = 0,$$

so

$$dL := -\mathrm{i} U^*\, dU$$

is an infinitesimal Hermitian matrix. Since the Haar measure γ_n on $\mathcal{U}(n)$ is invariant under multiplication by a unitary matrix we have

$$\gamma_n(\,dU) = C \prod_{i=1}^{n} dL_{ii} \prod_{i<j} dL_{ij}\, dL_{ij}^*.$$

For every $U \in \mathcal{U}(n)$ there exist $V \in \mathcal{U}(n)$ and a D diagonal matrix, such that

$$U = VDV^*,$$

where the non-zero entries of D are the eigenvalues of U, so D can be written in the form $D := \operatorname{diag}\left(e^{\mathrm{i}\vartheta_1}, \ldots e^{\mathrm{i}\vartheta_n}\right)$, since the eigenvalues are on the unit circle. The matrices V and D are not unique, so we can assume, that for the infinitesimal Hermitian matrix $dM := -\mathrm{i}V^*\, dV$ the entries in the diagonal are zero, so $dM_{ii} = 0$ for $1 \le i \le n$. Since

$$\begin{aligned} dL \quad &= -\mathrm{i}VD^*V^*\, d\,(VDV^*) \\ &= -\mathrm{i}VD^*V^*\,(\,dVDV^* + V\,dDV^* + VD\,dV^*) \\ &= V\,(D^*\,dMD - \mathrm{i}D^*\,dD - \,dM)\,V^* \end{aligned}$$

since $D^*D = I$. For the element of the matrix V^*DV we get

$$(V^*\,dLV)_{ii} = -\mathrm{i}e^{-\mathrm{i}\vartheta_i}\, de^{\mathrm{i}\vartheta_i} = \ d\vartheta_i,$$

and for $i < j$

$$(V^*\,dLV)_{ij} = e^{\mathrm{i}(\vartheta_j - \vartheta_i)}\, dM_{ij} - \ dM_{ij} = e^{-\mathrm{i}\vartheta_i}\left(e^{\mathrm{i}\vartheta_j} - e^{\mathrm{i}\vartheta_j}\right) dM_{ij}.$$

Finally we have

$$\prod_{i=1}^{n} dL_{ii} \prod_{i<j} dL_{ij}\, dL_{ij}^{*} = \prod_{i<j} \left| e^{\mathrm{i}\vartheta_i} - e^{\mathrm{i}\vartheta_j} \right| \prod_{i=1}^{n} d\vartheta_i \prod_{i<j} dM_{ij}\, dM_{ij}^{*}.$$

The normalization constant can be computed in several ways. We use here the properties of complex contour integral as follows

$$\begin{aligned}
&\int_0^{2\pi} \dots \int_0^{2\pi} \prod_{i<j} |e^{\mathrm{i}\vartheta_i} - e^{\mathrm{i}\vartheta_j}|^2 d\vartheta_1 \dots d\vartheta_n \\
&\quad = (-\mathrm{i})^n \oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} \prod_{i<j} (z_i - z_j)(\overline{z}_i - \overline{z}_j) dz_1 \dots dz_n \\
&\quad = (-\mathrm{i})^n \oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} \prod_{i<j} (z_i - z_j)(z_i^{-1} - z_j^{-1}) dz_1 \dots dz_n \\
&\quad = (-\mathrm{i})^n \oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} \det \left[z_i^{j-1} \right]_{i,j=1}^n \det \left[z_i^{-(j-1)} \right]_{i,j=1}^n dz_1 \dots dz_n = \\
&\quad = (-\mathrm{i})^n \oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} \sum_{\pi \in S_n} (-1)^{\sigma(\pi)} \times \\
&\qquad\qquad \times \prod_{i=1}^n z_i^{\pi(i)-1} \sum_{\rho \in S_n} (-1)^{\sigma(\rho)} \prod_{i=1}^n z_i^{-(\rho(i)-1)} dz_1 \dots dz_n \\
&\quad = n!(-\mathrm{i})^n \oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} dz_1 \dots dz_n
\end{aligned}$$

Since by the theorem of residue those terms of the above sum vanish, where there exists a z_i on the power different from -1. So in the above sum it is enough to consider the case when

$$\pi(i) = \rho(i)$$

for all $1 \leq i \leq n$. Therefore we take the summation over the $n!$ elements of S_n. Again by the theorem of residue

$$\oint_{\{|z|=1\}^n} z_1^{-1} \dots z_n^{-1} dz_1 \dots dz_n = (2\pi \mathrm{i})^n,$$

which gives the normalization constant.

From this we have the joint eigenvalue density function of any powers of Haar unitary random matrices. In [34] we used the above method of complex contour integral in order to prove the following theorem.

Theorem 3.3 *For $m \geq n$ the random variables $\lambda_0^m, \lambda_1^m, \dots, \lambda_{n-1}^m$ are independent and uniformly distributed on $\mathbb{T}$.*

Proof. Since the Fourier transform determines the joint distribution measure of $\lambda_0^m, \lambda_1^m, \dots, \lambda_{n-1}^m$ uniquely, it suffices to show that

$$\int_{[0,2\pi]^n} z_0^{k_0 m} z_1^{k_1 m} \dots z_{n-1}^{k_{n-1} m} \prod_{i<j} |z_i - z_j|^2 \, dz_0 dz_1 \dots dz_{n-1} = 0 \tag{54}$$

if at least one $k_j \in \mathbb{Z}$ is different from 0, where $dz_i = d\varphi_i / 2\pi$ for $z_i = e^{\mathrm{i}\varphi_i}$. We use the following notation for the above Vandermonde determinant.

$$\Delta(z_0, z_1, \dots, z_{n-1}) := \prod_{i<j} (z_i - z_j) = \det \left[z_i^j \right]_{0 \le i \le n-1,\, 0 \le k \le n-1}. \tag{55}$$

(What we have here is the so-called Vandermonde determinant.) Then one can write (54) as the complex contour integral on the unit circle as follows

$$\begin{aligned}
&\int_{[0,2\pi]^n} z_0^{k_0 m} z_1^{k_1 m} \dots z_{n-1}^{k_{n-1} m} \Delta(z_0, \dots, z_{n-1}) \Delta(z_0^{-1}, \dots, z_{n-1}^{-1}) \, dz_0 dz_1 \dots dz_{n-1} \\
&= \oint_{\{|z|=1\}^n} z_0^{k_0 m} z_1^{k_1 m} \dots z_{n-1}^{k_{n-1} m} \Delta(z_0, \dots, z_{n-1}) \Delta(z_0^{-1}, \dots, z_{n-1}^{-1}) z_0^{-1} \dots z_{n-1}^{-1} \, dz_0 \dots dz_{n-1} \\
&= \oint_{\{|z|=1\}^n} z_0^{k_0 m - 1} z_1^{k_1 m - 1} \dots z_{n-1}^{k_{n-1} m - 1} \sum_{\pi \in S_n} (-1)^{\sigma(\pi)} z_0^{\pi(0)} \dots z_{n-1}^{\pi(n-1)} \\
&\qquad \times \sum_{\rho \in S_n} (-1)^{\sigma(\rho)} z_0^{-\rho(0)} \dots z_{n-1}^{-\rho(n-1)} \, dz_0 \dots dz_{n-1}.
\end{aligned}$$

By the theorem of residue, we get nonzero terms only in the case, where the exponent of z_i is -1 for all $0 \le i \le n-1$. This means, that we need the permutations where

$$k_j m + \pi(j) - \rho(j) - 1 = -1 \qquad (0 \le j \le n-1),$$

so

$$k_j m = \rho(j) - \pi(j).$$

Here $|\rho(j) - \pi(j)| \le n-1$, and $|k_j m| \ge m \ge n$, if $k_j \ne 0$, so if at least one $k_j \in \mathbb{Z}$ is different from 0, then there exists no solution. This proves the theorem. $\square$

3.4 Asymptotics of the trace of polynomials of the Haar unitary

In this section we give a more elementary proof of the theorem of Diaconis and Shahshahani in [10]. We used the method of moments in [34] in order to obtain the same theorem.

Let $U_n = (U_{ij})_{1 \le i,j \le n}$ be a Haar distributed unitary random matrix. In this section we are interested in the convergence of $\operatorname{Tr} U_n$ as $n \to \infty$. Since the correlation between the diagonal entries decreases with n, one expects on the basis of the central limit theorem, that the limit of the trace has complex normal distribution. In the proof we need the following technical lemma which tells us that the expectation of most of the product of the entries are vanishing.

Lemma 3.3 ([26]) *Let $i_1, \dots, i_h, j_1, \dots j_h \in \{1, \dots, n\}$ and $k_1, \dots, k_h, m_1, \dots, m_h$ be positive integers for some $h \in \mathbb{N}$. If*

$$\sum_{i_r=u} (k_r - m_r) \neq 0 \qquad \text{for some} \qquad 1 \leq u \leq n$$

or

$$\sum_{j_r=v} (k_r - m_r) \neq 0 \qquad \text{for some} \qquad 1 \leq v \leq n,$$

then

$$\mathbb{E}\left((U_{i_1 j_1}^{k_1} \overline{U}_{i_1 j_1}^{m_1}) \dots (U_{i_h j_h}^{k_h} \overline{U}_{i_h j_h}^{m_h}) \right) = 0.$$

Proof. Suppose that $t := \sum_{i_r=u} (k_r - m_r) \neq 0$. The translation invariance of U implies that multiplying this matrix by $V = \mathrm{Diag}(1, \dots, 1, e^{i\vartheta}, 1, \dots, 1) \in \mathcal{U}(n)$ from the left we get

$$\mathbb{E}\left((U_{i_1 j_1}^{k_1} \overline{U}_{i_1 j_1}^{m_1}) \dots (U_{i_h j_h}^{k_h} \overline{U}_{i_h j_h}^{m_h}) \right) = e^{it\vartheta} \mathbb{E}\left((U_{i_1 j_1}^{k_1} \overline{U}_{i_1 j_1}^{m_1}) \dots (U_{i_h j_h}^{k_h} \overline{U}_{i_h j_h}^{m_h}) \right),$$

for all $\vartheta \in \mathbb{R}$. □

Theorem 3.4 *Let U_n be a sequence of $n \times n$ Haar unitary random matrices. Then $Tr U_n$ converges in distribution to a standard complex normal random variable as $n \to \infty$.*

Proof. For the sake of simplicity we write U instead of U_n. First we study the asymptotic of the moments

$$\begin{aligned} \mathbb{E}\left((\operatorname{Tr} U)^k (\overline{\operatorname{Tr} U})^k \right) &= \mathbb{E}\Big(\Big(\sum_{i=1}^n U_{ii} \Big)^k \Big(\sum_{j=1}^n \overline{U}_{jj} \Big)^k \Big) \\ &= \sum_{i_1, \dots, i_k=1}^n \sum_{j_1, \dots, j_k=1}^n \mathbb{E}(U_{i_1 i_1} \dots U_{i_k i_k} \overline{U}_{j_1 j_1} \dots \overline{U}_{j_k j_k}), \end{aligned}$$

$k \in \mathbb{Z}^+$. By Lemma 3.3 parts of the above sum are zero, we need to consider only those sets of indices $\{i_1, \dots, i_k\}$ and $\{j_1, \dots, j_k\}$ which coincide (with multiplicities). Consider a summand $\mathbb{E}(|U_{i_1 i_1}|^{2k_1} \dots |U_{i_r i_r}|^{2k_r})$, where $\sum_{l=1}^r k_l = k$. From the Hölder inequality

$$\mathbb{E}(|U_{i_1 j_1}|^{2k_1} \dots |U_{i_r j_r}|^{2k_r}) \leq \prod_{l=1}^r \sqrt[2^l]{\mathbb{E}(|U_{i_l j_l}|^{2 \cdot 2^l k_l})} = \prod_{l=1}^r \binom{n + 2^l k_l - 1}{2^l k_l - 1}^{-1/2^l} = O\left(n^{-k}\right). \tag{56}$$

The number of those sets of indices, where among the numbers $i_1, \dots, i_k$ there are at least two equal is at most

$$k!\binom{k}{2}n^{k-1} = O(n^{k-1}).$$

By (56) the order of magnitude of these factors is $O(n^{-k})$, so this part of the sum tends to zero as $n \to \infty$. Next we assume that $i_1, \dots, i_k$ are different. Since by translation invariance any row or column can be replaced by any other, we have

$$\mathbb{E}(|U_{i_1 i_1}|^2 \dots |U_{i_k i_k}|^2) = \mathbb{E}(|U_{11}|^2 \dots |U_{kk}|^2)) =: M_k^n. \tag{57}$$

It is enough to determine this quantity and to count how many of these terms are in the trace. The length of the row vectors of the unitary matrix is 1, hence

$$\sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \mathbb{E}\left(|U_{i_1 1}|^2 \dots |U_{i_k k}|^2\right) = 1. \tag{58}$$

We divide the sum into two parts: the number of terms with different indices is $n!/(n-k)!$, and again the translation invariance implies that each of them equals to M_k^n, and we denote by ε_k^n the sum of the other terms. Therefore

$$\varepsilon_k^n = 1 - \frac{n!}{(n-k)!}M_k^n \leq k!\binom{k}{2}O(n^{-k}) \to 0,$$

and

$$M_k^n = \frac{(1-\varepsilon_k^n)(n-k)!}{n!}.$$

Now we can count how many expectations of value M_k^n are there in the sum (56). We can fix the indices $i_1, \dots, i_k$ in $n!/(n-k)!$ ways, and we can permute them in $k!$ ways to get the indices $j_1, \dots, j_k$. The obtained equation

$$\lim_{n\to\infty} \mathbb{E}\left((\operatorname{Tr} U_n)^k (\overline{\operatorname{Tr} U_n})^k\right) = \lim_{n\to\infty} \frac{n!}{(n-k)!}k!\frac{(1-\varepsilon_k^n)(n-k)!}{n!} = k!$$

finishes the proof. For the mixed moments we have by Lemma 3.3

$$\mathbb{E}\left((\operatorname{Tr} U_n)^k (\overline{\operatorname{Tr} U_n})^m\right) = 0 \qquad (k \neq m),$$

and we have proven the convergence of all moments. The only thing is left to conclude the convergence in distribution is to show that the moments determine uniquely the limiting distribution (VIII. 6 in [14]). Although we have complex random variables, the distribution of the phase is uniform, and we can consider them as real valued random variables. The Stirling formula implies that

$$\sum_{k\in\mathbb{N}} (k!)^{-\frac{1}{k}} \geq \sum_{k\geq M} \left(\left(\frac{2k}{e}\right)^k\right)^{-\frac{1}{k}} = \frac{e}{2}\sum_{k\geq M}\frac{1}{k} = \infty.$$

for a large $M \in \mathbb{N}$, since $\sqrt{2k\pi} \leq 2^k$, if $k \geq 2$. $\square$

The convergence for the higher powers was done also by Diaconis and Shashahani in [10]. Here we use elementary methods.

Theorem 3.5 *Let Z be standard complex normal distributed random variable, then for the sequence of U_n $n \times n$ Haar unitary random matrices $Tr U_n^l \to \sqrt{l} Z$ in distribution.*

Proof. We use the method of moments again. Lemma 3.3 implies that we only have to take into consideration $\mathbb{E}\left(\left(\operatorname{Tr} U_n^l\right)^k \left(\overline{\operatorname{Tr} U_n^l}\right)^k\right)$, for all $k \in \mathbb{Z}^+$.

$$
\begin{aligned}
&\mathbb{E}\left(\left(\operatorname{Tr} U_n^l\right)^k \left(\overline{\operatorname{Tr} U_n^l}\right)^k\right) \\
&= \mathbb{E}\Bigg(\Big(\sum_{i_1,\dots,i_l} U_{i_1 i_2} U_{i_2 i_3} \dots U_{i_{l-1} i_l} U_{i_l i_1}\Big)^k \Big(\sum_{j_1,\dots,j_l} \overline{U}_{j_1 j_2} \overline{U}_{j_2 j_3} \dots \overline{U}_{j_{l-1} j_l} \overline{U}_{j_l j_1}\Big)^k\Bigg) \\
&= \sum \mathbb{E}\Big(U_{i_1 i_2} \dots U_{i_l i_1} U_{i_{l+1} i_{l+2}} \dots U_{i_{2l} i_{l+1}} \dots U_{i_{l(k-1)+1} i_{l(k-1)+2}} \dots U_{i_{kl} i_{l(k-1)+1}} \\
&\quad \times \overline{U}_{j_1 j_2} \dots \overline{U}_{j_l j_1} \overline{U}_{j_{l+1} j_{l+2}} \dots \overline{U}_{j_{2l} j_{l+1}} \dots \overline{U}_{j_{l(k-1)+1} j_{l(k-1)+2}} \dots \overline{U}_{j_{kl} j_{l(k-1)+1}}\Big),
\end{aligned}
$$

where the indices $i_1, \dots, i_{kl}, j_1, \dots, j_{kl}$ run from 1 to n, and by Lemma 3.3 if the sets $\{i_1, \dots, i_{kl}\}$ and $\{j_1, \dots, j_{kl}\}$ are different, then the expectation of the product is zero. It follows from the Cauchy and Hölder inequalities, and (56), that

$$
\begin{aligned}
&\left|\mathbb{E}\left(U_{i_1 i_2} \dots U_{i_{kl} i_{l(k-1)+1}} \overline{U}_{j_1 j_2} \dots \overline{U}_{j_{kl} j_{l(k-1)+1}}\right)\right| \\
&\quad \le \mathbb{E}\left|U_{i_1 i_2} \dots U_{i_{kl} i_{l(k-1)+1}} \overline{U}_{j_1 j_2} \dots \overline{U}_{j_{kl} j_{l(k-1)+1}}\right| \\
&\quad \le \sqrt{\mathbb{E}\left(|U_{i_1 i_2}|^2 \dots |U_{i_{kl} i_{l(k-1)+1}}|^2 |\overline{U}_{j_1 j_2}|^2 \dots |\overline{U}_{j_{kl} j_{l(k-1)+1}}|^2\right)} \le O\left(n^{-kl}\right).
\end{aligned}
\tag{59}
$$

Again the number of the set of indices, where there exist at least two equal indices is at most $O(n^{kl-1})$, so the sum of the corresponding expectations tends to zero as $n \to \infty$. Suppose that all the indices are different. There exist $\frac{n!}{(n-kl)!}(kl)! = O(n^{kl})$ of these kinds of index sets, and now we will prove, that most of the corresponding products have order of magnitude less than n^{-kl-1}. Consider for any $0 \le r \le kl$

$$
N_k^n(r) := \mathbb{E}\left(|U_{12}|^2 |U_{23}|^2 \dots |U_{r1}|^2 U_{r+1,r+2} \dots U_{kl-1,kl} U_{kl,r+1} \overline{U}_{r+2,r+1} \dots \overline{U}_{r+1,kl}\right).
$$

Note that $N_k^n(kl) = N_k^n(kl-1) = M_{kl}^n$, and if $\{i_1, \dots i_{kl}\} = \{j_1, \dots, j_{kl}\}$, and all the indices are different, then the corresponding term equals to $N_k^n(r)$ for some $0 \le r \le kl$. Using the orthogonality of the rows for $0 \le r \le kl-2$

$$
\mathbb{E}\left(\sum_{j=1}^n |U_{12}|^2 |U_{23}|^2 \dots |U_{r1}|^2 U_{r+1,r+2} \dots U_{kl-1,j} U_{kl,r+1} \overline{U}_{r+2,r+1} \dots \overline{U}_{r+1,j}\right) = 0. \tag{60}
$$

If $j \ge kl$, then the permutation invariance implies, that

$$
\mathbb{E}\left(|U_{12}|^2 |U_{23}|^2 \dots |U_{r1}|^2 U_{r+1,r+2} \dots U_{kl-1,j} U_{kl,r+1} \overline{U}_{r+2,r+1} \dots \overline{U}_{r+1,j}\right) = N_k^n(r),
$$

so we can write from (60)

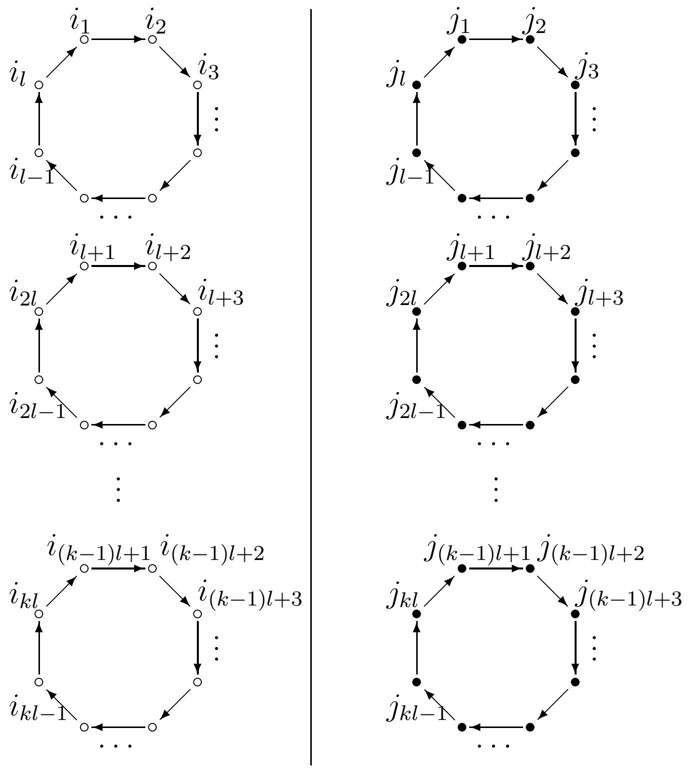
$$(n-kl)N_k^n(r)$$
$$= -\mathbb{E}\left(\sum_{j=1}^{kl}|U_{12}|^2|U_{23}|^2\dots|U_{r1}|^2U_{r+1,r+2}\dots U_{kl-1,j}U_{kl,r+1}\overline{U}_{r+2,r+1}\dots\overline{U}_{r+1,j}\right).$$

On the right side there is a sum of kl numbers which are less than $O(n^{-kl})$ because of (59), so this equation holds only if $N_k^n(r) \le O(n^{-kl-1})$.

We have to compute the sum of the expectations

$$\mathbb{E}\left(|U_{i_1i_2}|^2\dots|U_{i_li_1}|^2\dots|U_{i_{(k-1)l+1}i_{(k-1)l+2}}|^2\dots|U_{i_{kl}i_{(k-1)l+1}}|^2\right) = M_{kl}^n.$$

Now we count the number of these summands, so first we fix the set of sequences of length l $I_{l,k} = \{(i_{(u-1)l+1},\dots,i_{ul}), 1 \le u \le k\}$, and we try to find the set $J_{l,k} = \{(j_{(u-1)l+1},\dots,j_{ul}), 1 \le u \le k\}$, which gives M_{kl}^n. If the product contains $U_{i_ri_{r+1}}$, then it has to contain $\overline{U}_{i_ri_{r+1}}$, so if i_r and i_{r+1} are in the same sequence of $I_{l,k}$, then $j_s = i_r$ and $j_t = i_{r+1}$ have to be in the same sequence of $J_{l,k}$, and $t = s+1$ modulo l.



On the picture we have two directed graphs corresponding the indices in one term of the sum. The white vertices are the I indices, with directed edges $\overrightarrow{(i_u,i_v)}$, if there is $U_{i_ui_v}$ occurs in the product, and the black vertices denotes the J indices with directed edges $\overrightarrow{(j_u,j_v)}$, if there is $\overline{U}_{j_uj_v}$ occurs in the product. The calculations above showed, that the two graph has the same vertices and the same edges, so the permutation of the I indices holds the components and the order of the vertices in a component.

This means, that for all $1 \le u \le k$ there exists a sequence $(i_{(v-1)l+1},\dots,i_{vl}) \in I_{k,l}$ and a cyclic permutation π of the numbers $\{(v-1)l+1,\dots,vl\}$ such that $(j_{(u-1)l+1},\dots,j_{ul}) = (i_{\pi((v-1)l+1)},\dots,i_{\pi(vl)})$. We conclude, that for each $I_{l,k}$ there are $k!l^k$ sets $J_{l,k}$,since we can permute the sets of $I_{l,k}$ in $k!$ ways, and in all sets there are l cyclic permutations.

Clearly there are $\frac{n!}{(n-kl)!}$ sets $I_{l,k}$, so

$$\lim_{n\to\infty} \mathbb{E}\left(\left(\operatorname{Tr} U_n^l\right)^k \left(\overline{\operatorname{Tr} U_n^l}\right)^k \right) = \lim_{n\to\infty} \frac{n!}{(n-kl)!} k! l^k \frac{(1-\varepsilon_{kl}^n)(n-kl)!}{n!} = k! l^k,$$

and as in the proof of Theorem 3.4 this is the kth moment of $(\sqrt{l}Z)\overline{(\sqrt{l}Z)}$. □

Finally we prove that the limits of the trace of different powers are independent. The method of computation is the same as in the previous sections.

Theorem 3.6 *Let U_n be a sequence of Haar unitary random matrices as above. Then $Tr U_n$, $Tr U_n^2, \dots, Tr U_n^l$ are asymptotically independent.*

Proof. We will show, that the joint moments of $\operatorname{Tr} U_n, \operatorname{Tr} U_n^2, \dots \operatorname{Tr} U_n^l$ converge to the joint moments of $Z_1, \sqrt{2}Z_2, \dots, \sqrt{l}Z_l$, where $Z_1, Z_2, \dots Z_l$ are independent standard complex normal random variables. The latter joint moments are

$$\mathbb{E}\left(\prod_{i=1}^{l} i^{\frac{a_i+b_i}{2}} Z_i^{a_i} \overline{Z}_i^{b_i} \right) = \prod_{i=1}^{l} i^{\frac{a_i+b_i}{2}} \mathbb{E}\left(Z_i^{a_i} \overline{Z}_i^{b_i} \right) = \prod_{i=1}^{l} \delta_{a_i b_i} a_i! i^{a_i},$$

so we will prove that

$$\mathbb{E}\left(\prod_{i=1}^{l} \left(\operatorname{Tr} U_n^i\right)^{a_i} \left(\overline{\operatorname{Tr} U_n^i}\right)^{b_i} \right) = \prod_{i=1}^{l} \delta_{a_i b_i} a_i! i^{a_i}.$$

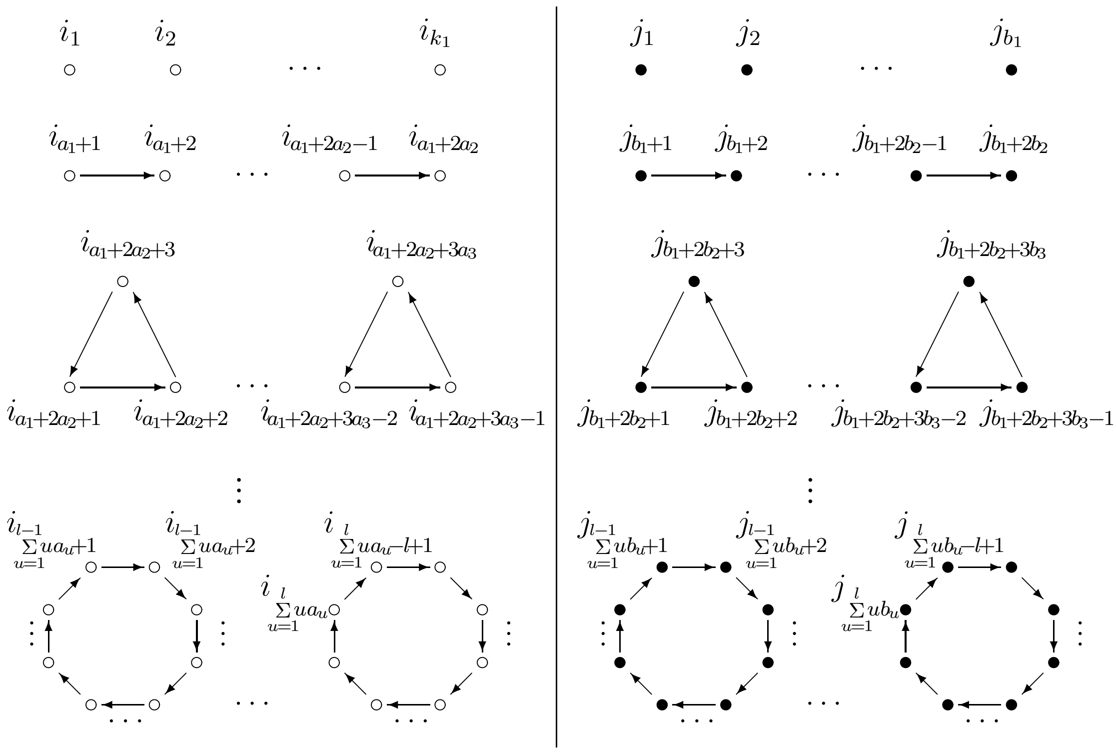
From Lemma 3.3, if $\sum_{i=1}^{l} i a_i \neq \sum_{i=1}^{l} i b_i$, then the moment

$$\mathbb{E}\left(\prod_{i=1}^{l} \left(\operatorname{Tr} U_n^i\right)^{a_i} \prod_{i=1}^{l} \left(\overline{\operatorname{Tr} U_n^i}\right)^{b_i} \right) = 0.$$

This implies, that it is enough to consider the case, when $\sum i a_i = \sum i b_i$. We have to take the summation over $n^{\sum i a_i}$ set of indices, since again if the indices in the first product does dot coincides with the ones from the second product (with multiplicity), then the expectation is zero according to Lemma 3.3. The order of magnitude of each summand is at most

$$O\left(n^{-\sum i a_i}\right),$$

as above, so if not all the indices are different, then the sum of these expectations tends to zero, as $n \to \infty$. The same way as in the proof of the previous theorem, those summands where there is a $U_{i_r i_{r+1}} \overline{U}_{i_r i_s}$, $i_{r+1} \neq i_s$ in the product are small. So now we have to sum the expectations $M^n_{\Sigma i a_i}$.



If we fix the set of first indices I, then again the sequences of the appropriate J, have to be cyclic permutations of the sequences of I. So again if we consider the graphs corresponding to the two sets of indices, we can permute the vertices by components. This means that the number of the sequences of length i in I is the same as in J, which means $a_i = b_i$ for all $1 \leq i \leq l$. The number of the I sets is $\frac{n!}{(n-\sum ia_i)!}$, so we have arrived to

$$\lim_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^{l} \left(\operatorname{Tr} U_n^i\right)^{a_i} \left(\overline{\operatorname{Tr} U_n^i}\right)^{b_i}\right)$$
$$= \lim_{n\to\infty} \frac{n!}{(n-\sum ia_i)!} \prod_{i=1}^{l} \delta_{a_i,b_i} i^{a_i} a_i! \frac{\left(1-\varepsilon^n_{\sum ia_i}\right)(n-\sum ia_i)!}{n!} = \prod_{i=1}^{l} \delta_{a_i,b_i} a_i! i^{a_i}.$$

□

Diaconis and Evans in [9] generalized the result for infinite series of Haar unitary random matrices. Their result is the following.

Theorem 3.7 *Consider an array of complex numbers a_{nj}, where $n, j \in \mathbb{N}$. Suppose there exists σ^2 such that*

$$\lim_{n\to\infty} \sum_{j=1}^{\infty} |a_{nj}|^2 \min(j,n) = \sigma^2.$$

Suppose also that there exist a sequence of positive integers $\{m_n : n \in \mathbb{N}\}$ *such that*

$$\lim_{n\to\infty} \frac{m_n}{n} = 0,$$

and

$$\lim_{n\to\infty} \sum_{j=m_n+1}^{\infty} |a_{nj}|^2 \min(j,n) = 0.$$

Then

$$\sum_{j=1}^{n} a_{nj}\, Tr\left(U_n^j\right) \xrightarrow{n\to\infty} \sigma Z$$

in distribution, where Z *is a standard complex normal random variable.*

For the polynomials of random matrices the theorem can be proven by the same methods as before. The proof of Diaconis and Evans based on the fact that for any $j, k \in \mathbb{N}$

$$\mathbb{E}\left(\operatorname{Tr} U_n^j \overline{\operatorname{Tr} U_n^k}\right) = \delta_{jk} \min(j,k).$$

Diaconis and Shahshahani mentioned a very important consequence of their theorem, namely that it implies the convergence of the empirical eigenvalue distribution to the uniform distribution on the circle, since the Fourier transform of a $\mu \in \mathcal{M}(\mathbb{T})$ is given by the sequence

$$\int_{\mathbb{T}} z^k \, d\mu(z), \qquad k \in \mathbb{Z}.$$

Now if γ is the uniform distribution on $\mathbb{T}$, then

$$\int_{\mathbb{T}} z^k \, d\gamma(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{\mathrm{i}k\varphi} \, d\varphi = \begin{cases} 1, & \text{if} \quad k = 0 \\ 0, & \text{if} \quad k \neq 0 \end{cases}$$

If the eigenvalues of the $n \times n$ Haar unitary U_n are $\zeta_1, \dots, \zeta_n$, then

$$\int_{\mathbb{T}} z^k \, d\left(\frac{1}{n}\sum_{i=1}^{n} \delta(\zeta_i)\right)(z) = \frac{1}{n}\sum_{i=1}^{n} \zeta_i^k = \frac{1}{n}\operatorname{Tr} U_n^k.$$

By the Chebyshev inequality for $k \neq 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\operatorname{Tr} U_n^k\right| > \varepsilon\right) = \mathbb{P}\left(\left|\operatorname{Tr} U_n^k\right| > n\varepsilon\right) \leq \frac{\mathbb{E}\left(\operatorname{Tr} U_n^k \operatorname{Tr}\left(U_n^*\right)^k\right)}{n^2\varepsilon^2} = O\left(\frac{1}{n^2}\right),$$

so

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{1}{n}\operatorname{Tr} U_n^l\right| > \varepsilon\right) < \infty,$$

which means by the Borel-Cantelli lemma, that

$$\frac{1}{n}\operatorname{Tr} U_n^k \xrightarrow{n\to\infty} 0,$$

with probability 1. If $k = 0$, then

$$\frac{1}{n}\mathrm{Tr}\, U_n^k = \frac{1}{n}\mathrm{Tr}\, I_n = 1,$$

where I_n is the $n \times n$ identity matrix. Thus the limit Fourier transform coincides with the Fourier transform of the uniform distribution, therefore by the unicity of the Fourier transform, the limit of the empirical eigenvalue distribution is the uniform distribution on $\mathbb{T}$.

3.5 Orthogonal random matrices

The set of $n \times n$ orthogonal random matrices is again a compact topological group, so we can define a Haar distributed orthogonal random matrix. The construction is similar, but we start from a matrix with real valued standard normal random variables. Applying the Gram-Schmidt orthogonalization gives the random matrix O_n.

The permutation invariance of the matrix implies that the entries of O_n have the same distribution, and by the construction, the square of the entries has beta distribution with parameters $\left(\frac{1}{2}, \frac{n-1}{2}\right)$, so it has the density

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} x^{-\frac{1}{2}}(1-x)^{\frac{n-3}{2}},$$

on the interval $[0, 1]$. Now using the symmetry of O_{ij} we have that

$$\begin{aligned}
\mathbb{P}(O_{ij} < x) \quad &= \frac{1}{2} + \frac{1}{2}\mathbb{P}\left(O_{ij}^2 < x^2\right) \\
&= \frac{1}{2} + \frac{1}{2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^{x^2} t^{-\frac{1}{2}}(1-t)^{\frac{n-3}{2}}\, dt \\
&= \frac{1}{2} + \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_0^{x} \left(1-y^2\right)^{\frac{n-3}{2}} \\
&= \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \int_{-1}^{x} \left(1-y^2\right)^{\frac{n-3}{2}}
\end{aligned}$$

Similarly to Theorem 3.2 we have the limit distribution of the normalized entries.

Theorem 3.8 *The density of $\sqrt{n}O_{ij}$ is on the interval $[-\sqrt{n}, \sqrt{n}]$*

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} \left(1 - \frac{y^2}{n}\right)^{\frac{n-3}{2}} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

so it converges to a standard normal variable in distribution.

We need only the convergence of the constant. Since

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi},$$

$\Gamma(n)=(n-1)!$ and

$$\Gamma\left(n+\frac{1}{2}\right)=\Gamma\left(\frac{1}{2}\right)\prod_{i=1}^{n}\left(n-i+\frac{1}{2}\right)=\frac{\sqrt{\pi}(2n)!}{2^{2n}n!}.$$

by the Stirling formula we have that for $n=2k$

$$\begin{aligned}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}&=\frac{\Gamma\left(k\right)}{\Gamma\left(k-\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}=\frac{((k-1)!)^2 2^{2(k-1)}}{\pi(2(k-1))!}\\&\approx\frac{\left(\frac{k-1}{e}\right)^{2(k-1)}2^{2(k-1)}2\pi(k-1)}{\pi\left(\frac{2(k-1)}{e}\right)^{2(k-1)}\sqrt{4\pi(k-1)}}=\frac{\sqrt{k-1}}{\sqrt{\pi}}=\frac{\sqrt{n-2}}{\sqrt{2\pi}},\end{aligned}$$

and for $n=2k+1$

$$\begin{aligned}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}&=\frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(k\right)\Gamma\left(\frac{1}{2}\right)}=\frac{(2k)!}{(k-1)!2^{2k}k!}\\&\approx\frac{k\left(\frac{2k}{e}\right)^{2k}\sqrt{4\pi k}}{\left(\frac{k}{e}\right)^{2k}2\pi k 2^{2k}}=\frac{\sqrt{k}}{\sqrt{\pi}}=\frac{\sqrt{n-1}}{\sqrt{2\pi}},\end{aligned}$$

so we arrived to

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n}\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\xrightarrow{n\to\infty}\frac{1}{\sqrt{2\pi}}.$$

$\square$

The moments of O_{ij} can be computed from the density, which are important, if we want to prove a theorem which is similar to the Theorem 3.6. The proof of that theorem showed, that it is enough to know the second moment of the entries, and the order of magnitude of the other ones. The odd moments are clearly 0. The $2k$th even moment $M_{k,n}$ can be computed by partial integration, i.e.

$$\begin{aligned}M_{k,n}&:=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{-1}^{1}x^{2k}\left(1-x^2\right)^{\frac{n-3}{2}}\\&=\frac{2k-1}{n-1}\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_{-1}^{1}x^{2(k-1)}\left(1-x^2\right)^{\frac{n-1}{2}}=\frac{2k-1}{n}M_{k-1,n+1},\end{aligned}$$

because

$$\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n+2}{2}\right)}=\frac{n-1}{n}.$$

By induction

$$M_{k,n} = \prod_{i=1}^{k} \frac{2k-2i+1}{n+i-1} = O\left(n^{-k}\right),$$

and

$$M_{2,n} = \frac{1}{n}.$$

Clearly the limit distribution of the trace cannot be complex valued, since the entries are real. We use the method of moments again, so we need the moments of the standard normal variable. It is well known, that for an $\eta \sim N(0,1)$

$$\mathbb{E}\eta^n = \begin{cases} \frac{(2k)!}{2^k k!}, & \text{if} \quad n = 2k \\ 0, & \text{if} \quad n = 2k+1 \end{cases}$$

We need the analogue of Lemma 3.3 for orthogonal matrices.

Lemma 3.4 *Let $i_1, \ldots, i_h, j_1, \ldots j_h \in \{1, \ldots, n\}$ and $k_1, \ldots, k_h$ be positive integers for some $h \in \mathbb{N}$. If $\sum_{i_r = u} k_r$ is odd for some $1 \le u \le n$, or $\sum_{j_r = v} k_r$ is odd for some $1 \le v \le n$, then*

$$E\left(O_{i_1 j_1}^{k_1} \ldots O_{i_h j_h}^{k_h}\right) = 0.$$

The proof goes similarly to the proof of Lemma 3.3, but we can use it only in the $\vartheta = \pi$ case, since the entries are real.

From this, the following theorem holds.

Theorem 3.9 *Let O_n be a sequence of Haar unitary random matrices as above. Then $Tr O_n \stackrel{n\to\infty}{\longrightarrow} N(0,1)$.*

Proof. The proof of this convergence is similar to the Theorem 3.4, so we use the method of moments, and we consider for $k \in N$

$$\mathbb{E}\left(\operatorname{Tr} O_n\right)^k = \sum_{i_1, \ldots i_k} \mathbb{E}\left(O_{i_1 i_1} O_{i_2 i_2} \ldots O_{i_k i_k}\right).$$

Now we can use Lemma 3.4 to show that it is enough to sum the terms, where in the corresponding sequence of indices contains each index with even multiplicity. This implies, that if k is odd, then the kth moment of the trace vanishes as $n \to \infty$. If $k = 2m$, then from Cauchy inequality we have that each term has the order of magnitude $O\left(n^m\right)$, so it is enough to consider the sum of the terms where each index occurs exactly twice. We can choose the m indices in $\binom{n}{m}$ ways, and then we choose the places where we put the same indices in $\frac{(2m)!}{2^m m!}$ ways, and then we order the indices in $m!$ ways. So

$$\lim_{n\to\infty} \mathbb{E}\left(\operatorname{Tr} O_n\right)^{2m} = \frac{(2m)!}{2^m m!},$$

which is exactly the $2m$th moment of the standard normal variable. $\square$

The above theorem is not true for higher powers of O_n. For example with combinatorial methods we get that

$$\mathbb{E}\left(\operatorname{Tr} O_n^{2l}\right) \xrightarrow{n\to\infty} 1.$$

Using the Fourier transform one can easily check, that the limit of the empirical eigenvalue distribution of O_n as $n \to \infty$ is again the uniform distribution on the unit circle.

3.6 Large deviation theorem for unitary random matrix

We know that the limit of the empirical eigenvalue distribution of the Haar unitary random matrix is the uniform distribution on the unit circle $\mathbb{T} := \{z : |z| = 1\}$. For the rate of the convergence the large deviation theorem was proven by Hiai and Petz. The theorem concerns not only the Haar unitary random matrices but the unitary random matrices whose distribution is exponential with respect to the Haar measure. So suppose that γ_n is the Haar measure on the $\mathcal{U}(n)$ set of $n \times n$ unitary matrices, and $Q : \mathbb{T} \to \mathbb{R}$ is a continuous function. Now for each $n \in \mathbb{N}$ take the measure $\nu_n \in \mathcal{M}(\mathcal{U}(n))$ as

$$\nu_n := \frac{1}{Z_n} \exp(-n\operatorname{Tr} Q(U))\, d\gamma_n(U),$$

where Z_n is the normalizing constant. Then the joint eigenvalue density is

$$\frac{1}{Z_n} \exp\left(-n \sum_{i=1}^{n} Q(\zeta_i)\right) \prod_{i<j} |\zeta_i - \zeta_j|^2.$$

Now consider a sequence of $n \times n$ unitary matrices with distribution ν_n, and denote P_n the sequence of the distribution of empirical eigenvalue distribution of the matrices. Then each P_n is a measure on $\mathcal{M}(\mathbb{T})$, and the following theorem holds.

Theorem 3.10 (Hiai, Petz) *There exists the finite limit*

$$B := \lim_{n\to\infty} \log Z_n,$$

and the sequence (P_n) satisfies the large deviation principle in the scale n^{-2} with rate function

$$I(\mu) := \iint_{\mathbb{T}^2} \log \frac{1}{|\zeta - \eta|}\, d\mu(\zeta)\, d\mu(\eta) + \int_{\mathbb{T}} Q(\zeta)\, d\mu(\zeta) + B.$$

Furthermore there exists a unique $\mu_0 \in \mathcal{M}(\mathbb{T})$ such that $I(\mu_0) = 0$.

The case $Q \equiv 0$ gives the large deviation for the sequence of Haar unitary random matrices, and in this case the minimizing measure is the uniform distribution on $\mathbb{T}$, but generally it is difficult to find the limit of the empirical eigenvalue distribution.

4 Truncations of Haar unitaries

Let U be an $n \times n$ Haar distributed unitary matrix. By truncating $n-m$ bottom rows and $n-m$ last columns, we get an $m \times m$ matrix $U_{[n,m]}$. The distribution of the entries is clearly the same as in the case of Haar unitaries. By the construction, the distribution of $U_{n,m}$ is invariant under conjugation, and multiplying by any $V \in \mathcal{U}(m)$.

4.1 Joint eigenvalue density

The truncated matrix is not unitary but it is a contraction, because suppose, that there exists an $x = (x_1, \dots, x_m) \in \mathbb{C}^m$, $\|x\| = 1$ such that

$$\|U_{[n,m]}x\|^2 = x^* U^*_{[n,m]} U_{[n,m]} x > 1,$$

then for $x' = (x_1, \dots, x_m, 0 \dots 0) \in \mathbb{C}^n$ and for the matrix $C = (U_{ij})_{\substack{n-m+1 \le i \le n \\ 1 \le j \le m}}$

$$\|Ux'\|^2 = \|U_{[n,m]}x\|^2 + \|Cx\|^2 \ge \|U_{[n,m]}x\|^2 > 1.$$

So we proved, that $U_{[n,m]}$ is a contraction, so $\|U_{[n,m]}\| \le 1$, and therefore the eigenvalues $z_1, z_2, \dots, z_m \in D^m$, where $D = \{z \in \mathbb{C} : |z| \le 1\}$ is the unit disc. According to [48] the joint probability density of the eigenvalues is

$$C_{[n,m]} \prod_{i<j} |\zeta_i - \zeta_j|^2 \prod_{i=1}^m (1 - |\zeta_i|^2)^{n-m-1}$$

on D^m. Now we sketch the proof of this result. Let U_m be an $m \times m$ Haar unitary matrix and write it in the block-matrix form

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A is an $n \times n$, B is $n \times (m-n)$, C is $(m-n) \times n$ and D is an $(m-n) \times (m-n)$ matrix. The space of $n \times n$ (complex) matrices is easily identified with $\mathbb{R}^{2n^2}$ and the push forward of the usual Lebesgue measure is denoted by λ_n. It was obtained in [7] that for $m \ge 2n$, the distribution measure of the $n \times n$ matrix A is absolute continuous with respect to λ_n and the density is

$$C(n,m) \det(1 - A^*A)^{m-2n} \mathbf{1}_{\|A\| \le 1} d\lambda_n(A)\,. \tag{61}$$

To determine the joint distribution of the eigenvalues $\zeta_1, \zeta_2, \dots, \zeta_n$ of A, we need only the matrices A and C, and by a unitary transformation we transform A to an upper triangular form

$$\begin{pmatrix} \zeta_1 & \Delta_{1,2} & \Delta_{1,3} & \dots & \Delta_{1,n} \\ 0 & \zeta_2 & \Delta_{2,3} & \dots & \Delta_{2,n} \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & \zeta_n \\ C_1 & C_2 & C_3 & \dots & C_n \end{pmatrix}, \tag{62}$$

where $C_1, C_2, \dots, C_n$ are the column vectors of the matrix C. First we consider the case $m = 1$. In this case the eigenvalue of the 1×1 matrix is the first entry of the first row, so it has the density $(1 - |z|^2)^{n-1}$.

For $m \geq 2$ we get by the Schur decomposition that

$$A = T(z + \Delta)T^{-1},$$

where T is an appropriate unitary matrix, $Z = \mathrm{diag}(z_1, \dots, z_m)$, and $\Delta = (\Delta_{ij})_{1 \leq i < j \leq m-1}$ is a strictly upper triangular matrix. The matrix $dL = -\mathrm{i}T^{-1}dT$ is Hermitian and the we can assume, that $dL_{ii} = 0$ for $1 \leq i \leq m$. Then from Mehta

$$dA = \prod_{i<j} |z_i - z_j|^2 \prod_{i=1}^{m} dz_i \prod_{i<j} d\Delta_{ij} dT_{ij}.$$

By the orthogonality of the rows for $i < j$

$$\overline{z}_i \Delta_{ij} + C_i^* C_j + \sum_{k<i} \overline{\Delta}_{ki} \Delta_{kj} = 0,$$

so

$$\Delta_{ij} = -\frac{1}{\overline{z}_i} \left(C_i^* C_j + \sum_{k<i} \overline{\Delta}_{ki} \Delta_{kj} \right), \tag{63}$$

and the columns are unit vectors so

$$C_i^* C_j + \sum_{k<i} |\Delta_{ki}|^2 + |z_i|^2 = 1. \tag{64}$$

So since the entries of the matrix Δ are determined by the matrices C and Z, we get the joint density if we integrate the joint density of $Z + \Delta$ and C with respect the elements of C. First we integrate with respect to the last column, because all the other columns can be constructed without the last one.

From (63) we get, that since

$$d\left(-\frac{1}{\overline{z}}\right) = \frac{1}{|z|^2} dz$$

thus any modification of z_i modify Δ_{im} by $1/|z_i|^2$, which gives a $\prod_{i<m} 1/|z_i|^2$ in the density function.

There exists $(n-m) \times (n-m)$ matrices $X^{(i)}$ such that

$$\Delta_{ij} = \frac{1}{\overline{z}_i} C_i^* X^{(i)} C_j.$$

Since $\Delta_{1j} = -\frac{1}{\overline{z}_1} C_1^* C_j$, $X^{(1)} = I$. If we know $X^{(1)}, \dots, X^{(i-1)}$

$$\Delta_{ij} = -\frac{1}{\overline{z}_i} C_i^* C_j + \sum_{k<i} C_i^* X^{(k)} \frac{C_k C_k^*}{|z_k|^2} X^{(k)} C_j,$$

so

$$X^{(i)} = I + \sum_{k<i} X^{(k)} \frac{C_k C_k^*}{|z_k|^2} X^{(k)}.$$

Then

$$C_i^* C_i + \sum_{k<i} \overline{\Delta}_{ki} \Delta_{ki} = C_i^* X^{(i)} C_i,$$

so the vectors C_i must satisfy the equations

$$C_i^* X^{(i)} C_i = 1 - |z_i|^2, \tag{65}$$

so $C_{1i}, \ldots, C_{n-m-1,i}$ lies inside the ellipsoid given by $X^{(i)}$. By Lemma 3.2 we need the integral of the uniform density on this ellipsoid, i.e. the volume of this set defined in (65). In order to obtain the volume it is enough to know the determinant of $X^{(i)}$.

$$X^{(i)} = I + \sum_{k<i} X^{(k)} \frac{C_k C_k^*}{|z_k|^2} X^{(k)} = X^{(i-1)} + X^{(i-1)} \frac{C_{i-1} C_{i-1}^*}{|z_{i-1}|^2} X^{(i-1)},$$

so

$$\det X^{(i)} = \det X^{(i-1)} \det \left(I + \frac{C_{i-1} C_{i-1}^*}{|z_{i-1}|^2} X^{(i-1)} \right).$$

Here

$$\frac{C_{i-1} C_{i-1}^*}{|z_{i-1}|^2} X^{(i-1)} C_{i-1} = \frac{C_{i-1}^* C_{i-1} + \sum_{k<i-1} |\Delta_{k,i-1}|^2}{|z_{i-1}|^2} C_{i-1} = \left(\frac{1}{|z_{i-1}|^2} - 1 \right) C_{i-1},$$

so the matrix

$$I + \frac{C_{i-1} C_{i-1}^*}{|z_{i-1}|^2} X^{(i-1)}$$

has the eigenvalue $1/|z_{i-1}|^2$ with multiplicity 1, and all the other eigenvalues are 1, so

$$\det X^{(i)} = \frac{\det X^{(i-1)}}{|z_{i-1}|^2} = \prod_{j<i} \frac{1}{|z_j|^2}.$$

Now we integrate with respect to the first column. For fixed $\Delta_{1,m} \ldots \Delta_{m-1,m}$ the distribution of $C_{1,m}, \ldots, C_{n-m-1,m}$ is uniform on the set

$$|C_{1,m}|^2 + \cdots + |C_{n-m-1,m}|^2 \leq 1 - |z_m|^2 - |\Delta_{1,m}|^2 \ldots |\Delta_{m-1,m}|^2,$$

i.e. inside the ellipsoid defined by (65). The volume of this $n - m - 1$ dimensional complex ellipsoid is

$$\frac{(1 - |z_m|^2)^{n-m-1}}{\det X^{(m)}} = (1 - |z_m|^2)^{n-m-1} \prod_{i<m} |z_i|^2,$$

so the form the last column we get $(1-|z_m|^2)^{n-m-1}$. Since only the last column depends on z_m, and the joint density function of the eigenvalues must be symmetric in $z_1,\dots,z_m$, so the joint density function of the eigenvalues is given by

$$\prod_{1\le i<j\le m} |z_i-z_j|^2 \prod_{i=1}^m \left(1-|z_i|^2\right)^{n-m-1}.$$

Since the normalizing constant $C_{[n,m]}$ was not given in [48], we computed it by integration in [34]. To do this, we write $\zeta_i = r_i e^{\mathrm{i}\varphi_i}$ and $d\zeta_i = r_i\,dr_i\,d\varphi_i$. Then

$$\begin{aligned}
C_{[n,m]}^{-1} &= \int_{D^m} \prod_{1\le i<j\le m} |z_i-z_j|^2 \prod_{i=1}^m (1-|z_i|^2)^{n-m-1}\,dz \\
&= \int_{[0,1]^m}\int_{[0,2\pi]^m} \prod_{1\le i<j\le m} |r_i e^{i\varphi_i} - r_j e^{i\varphi_j}|^2 \prod_{i=1}^m (1-r_i^2)^{n-m-1} \prod_{i=1}^m r_i\,d\varphi\,dr.
\end{aligned}$$

Next we integrate with respect to $d\varphi = d\varphi_1\,d\varphi_2\dots d\varphi_m$ by transformation into complex contour integral what we evaluate by means of the residue theorem.

$$\begin{aligned}
&\int_{[0,2\pi]^n} \prod_{1\le i<j\le m} |r_i e^{i\varphi_1} - r_j e^{i\varphi_j}|^2\,d\varphi \\
&\qquad = (-i)^m \int_{\mathbb{T}^n} \prod_{1\le i<j\le m} |r_i z_i - r_j z_j|^2 \prod_{i=1}^m z_i^{-1}\,dz \\
&\qquad = (-i)^m \int_{\mathbb{T}^n} \prod_{1\le i<j\le m} (r_i z_i - r_j z_j)(r_i z_i^{-1} - r_j z_j^{-1}) \prod_{i=1}^m z_i^{-1}\,dz \\
&\qquad = (-i)^m \int_{\mathbb{T}^n} \prod_{i=1}^m z_i^{-1} \det\begin{bmatrix} 1 & 1 & \dots & 1 \\ r_1 z_1 & r_2 z_2 & \dots & r_m z_m \\ \vdots & & \ddots & \vdots \\ r_1^{m-1} z_1^{m-1} & r_2^{m-1} z_2^{m-1} & \dots & r_m^{m-1} z_m^{m-1} \end{bmatrix} \times \\
&\qquad\qquad \times \det\begin{bmatrix} 1 & 1 & \dots & 1 \\ r_1 z_1^{-1} & r_2 z_2^{-1} & \dots & r_m z_m^{-1} \\ \vdots & & \ddots & \vdots \\ r_1^{m-1} z_1^{-(m-1)} & r_2^{m-1} z_2^{-(m-1)} & \dots & r_m^{m-1} z_m^{-(m-1)} \end{bmatrix} dz \\
&\qquad = (-i)^m \int_{\mathbb{T}^n} \prod_{i=1}^m z_i^{-1} \sum_{\pi\in S_m} (-1)^{\sigma(\pi)} \prod_{i=1}^m (r_i z_i)^{\pi(i)-1} \sum_{\rho\in S_m} (-1)^{\sigma(\rho)} \prod_{i=1}^m (r_i z_i^{-1})^{\rho(i)-1}\,dz\,.
\end{aligned}$$

We have to find the coefficient of $\prod_{i=1}^m z_i^{-1}$, this gives that only $\rho=\pi$ contribute and the integral is

$$(2\pi)^m \sum_{\rho\in S_m} \prod_{i=1}^m (r_i)^{2(\rho(i)-1)}.$$

So we have

$$\begin{aligned}
C_{[n,m]}^{-1} &= (2\pi)^m \int_{[0,1]^m} \sum_{\rho\in S_m} \prod_{i=1}^{m} (r_i)^{2(\rho(i)-1)} \prod_{i=1}^{m} (1-r_i^2)^{n-m-1} \prod_{i=1}^{m} r_i \, dr \\
&= (2\pi)^m m! \prod_{i=1}^{m} \int_0^1 r_i^{2i-1}(1-r_i^2)^{n-m-1}\, dr_i
\end{aligned}$$

and the rest is done by integration by parts:

$$\begin{aligned}
\int_0^1 r^{2k+1}(1-r^2)^{n-m-1}\, dr &= \frac{k}{n-m}\int_0^1 r^{2k-1}(1-r^2)^{n-m}\, dr \\
&= \frac{k!}{(n-m)\dots(n-m+k-1)}\int_0^1 r(1-r^2)^{n-m+k-1}\, dr \\
&= \binom{n-m+k-1}{k}^{-1} \frac{1}{2(n-m+k)}.
\end{aligned}$$

Therefore

$$C_{[n,m]}^{-1} = \pi^m m! \prod_{k=0}^{m-1} \binom{n-m+k-1}{k}^{-1} \frac{1}{n-m+k}. \tag{66}$$

4.2 Limit distribution of the truncation

In this section we study the limit of $U_{[n,m]}$ when $n\to\infty$ and m is fixed. Clearly here we need some normalization, otherwise the entries and the eigenvalues vanish as the matrix size goes to infinity.

Now we consider $\sqrt{n/m}\,U_{[n,m]}$. Its joint probability density of the eigenvalues is simply derived from the above density of $U_{[n,m]}$ by the transformation

$$(\zeta_1,\dots,\zeta_m) \mapsto \left(\sqrt{\frac{m}{n}}\zeta_1,\dots,\sqrt{\frac{m}{n}}\zeta_m\right),$$

and it is given as

$$\begin{aligned}
C_{[n,m]}&\left(\frac{m}{n}\right)^m \prod_{i<j}\left|\sqrt{\frac{m}{n}}\zeta_i - \sqrt{\frac{m}{n}}\zeta_j\right|^2 \prod_{i=1}^{m}\left(1-\frac{m|\zeta_i|^2}{n}\right)^{n-m-1} \\
&= \frac{1}{\pi^m m!}\prod_{k=0}^{m-1}\binom{n-m+k-1}{k}(n-m+k)\left(\frac{m}{n}\right)^{m(m+1)/2} \\
&\qquad \times \prod_{i<j}|\zeta_i-\zeta_j|^2 \prod_{i=1}^{m}\left(1-\frac{m|\zeta_i|^2}{n}\right)^{n-m-1}
\end{aligned}$$

Now consider the asymptotic behaviour of the density.

$$
\begin{aligned}
&C_{[n,m]}\left(\frac{m}{n}\right)^m \prod_{i<j}\left|\sqrt{\frac{m}{n}}\zeta_i-\sqrt{\frac{m}{n}}\zeta_j\right|^2\prod_{i=1}^m\left(1-\frac{m|\zeta_i|^2}{n}\right)^{n-m-1}\\
&=\frac{1}{\pi^m m!}\prod_{k=0}^{m-1}\frac{n^{k+1}(1+o(1))}{k!}\left(\frac{m}{n}\right)^{m(m+1)/2}\prod_{i<j}|\zeta_i-\zeta_j|^2\prod_{i=1}^m\left(1-\frac{m|\zeta_i|^2}{n}\right)^{n-m-1}\\
&=\frac{m^{m(m+1)/2}}{\pi^m\prod_{k=1}^m k!}(1+o(1))\prod_{i<j}|\zeta_i-\zeta_j|^2\prod_{i=1}^m\left(1-\frac{m|\zeta_i|^2}{n}\right)^{n-m-1}.
\end{aligned}
$$

The limit of the above as $n\to\infty$ is

$$
\frac{m^{m(m+1)/2}}{\pi^m\prod_{k=1}^m k!}\exp\left(-m\sum_{i=1}^m|\zeta_i|^2\right)\prod_{i<j}|\zeta_i-\zeta_j|^2, \tag{67}
$$

which is exactly the joint eigenvalue density of the standard $m\times m$ non-selfadjoint Gaussian matrix.

4.3 Large deviation theorem for truncations

In the case of selfadjoint Gaussian random matrices, Wishart matrices and elliptic Gaussian random matrices the limit of the empirical eigenvalue distribution was known, and from the joint eigenvalue density we could get the rate function, and we found that the unique minimizer of the rate function is the limit of the empirical eigenvalue distribution. Now we have different kind of random matrices, and we don't know the limit of the empirical eigenvalue distribution, but we have the joint eigenvalue density. So now we will prove the large deviation theorem with the rate function which we get from the joint eigenvalue density, and then we try to find the unique minimizer of the rate function with the tools of potential theory mentioned in the Section 2 in order to get the limit distribution.

The following theorem, which is the main result of the dissertation was published in [35].

Theorem 4.1 *[Petz, Réffy] Let $U_{[m,n]}$ be the $n\times n$ truncation of an $m\times m$ Haar unitary random matrix and let $1<\lambda<\infty$. If $m/n\to\lambda$ as $n\to\infty$, then the sequence of empirical eigenvalue distributions $P_n=P_{[m,n]}$ satisfies the large deviation principle in the scale $1/n^2$ with rate function*

$$
I(\mu):=-\iint_{\mathcal{D}^2}\log|z-w|\,d\mu(z)\,d\mu(w)-(\lambda-1)\int_{\mathcal{D}}\log(1-|z|^2)\,d\mu(z)+B,
$$

for $\mu\in\mathcal{M}(\mathcal{D})$, where

$$
B:=-\frac{\lambda^2\log\lambda}{2}+\frac{\lambda^2\log(\lambda-1)}{2}-\frac{\log(\lambda-1)}{2}+\frac{\lambda-1}{2}.
$$

Furthermore, there exists a unique $\mu_0 \in \mathcal{M}(\mathcal{D})$ given by the density

$$d\mu_0(z) = \frac{(\lambda-1)r}{\pi\left(1-r^2\right)^2}\, dr\, d\varphi, \quad z = re^{\mathrm{i}\varphi}$$

on $\{z : |z| \leq 1/\sqrt{\lambda}\}$ such that $I(\mu_0) = 0$.

Set

$$F(z,w) := -\log|z-w| - \frac{\lambda-1}{2}\left(\log(1-|z|^2) + \log(1-|w|^2)\right),$$

and

$$F_\alpha(z,w) := \min(F(z,w),\alpha),$$

for $\alpha > 0$. Since $F_\alpha(z,w)$ is bounded and continuous

$$\mu \in \mathcal{M}(\mathcal{D}) \mapsto \iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu(z)\, d\mu(w).$$

is continuous in the weak* topology, when the support of μ is restricted to a compact set. The functional I is written as

$$\begin{aligned} I(\mu) \quad &= \iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w) + B \\ &= \sup_{\alpha>0} \iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu(z)\, d\mu(w) + B\,, \end{aligned}$$

hence I is lower semi-continuous.

We can write I in the form

$$I(\mu) = -\Sigma(\mu) - (\lambda-1)\int_{\mathcal{D}} \log(1-|z|^2)\, d\mu(z) + B.$$

Here the first part $-\Sigma(\mu)$ is strictly convex (as it was established in the previous section) and the second part is affine in μ. Therefore I is a strictly convex functional. If X is compact and $\mathcal{A}$ is a base for the topology, then the large deviation principle is equivalent to the following conditions (Theorem 4.1.11 and 4.1.18 in [8]):

$$-I(x) = \inf_{x\in G, G\in\mathcal{A}} \left\{\limsup_{n\to\infty} \frac{1}{n^2}\log P_n(G)\right\} = \inf_{x\in G, G\in\mathcal{A}} \left\{\liminf_{n\to\infty} \frac{1}{n^2}\log P_n(G)\right\}$$

for all $x \in X$. We apply this result in the case $X = \mathcal{M}(\mathcal{D})$, and we choose

$$\left\{\mu' \in \mathcal{M}(\mathcal{D}) : \left|\int_{\mathcal{D}} z^{k_1}\overline{z}^{k_2}\, d\mu'(z) - \int_{\mathcal{D}} z^{k_1}\overline{z}^{k_2}\, d\mu(z)\right| < \varepsilon \text{ for } k_1 + k_2 \leq m\right\}.$$

to be $G(\mu; m, \varepsilon)$. For $\mu \in \mathcal{M}(\mathcal{D})$ the sets $G(\mu; m, \varepsilon)$ form a neighbourhood base of μ for the weak* topology of $\mathcal{M}(\mathcal{D})$, where $m \in \mathbb{N}$ and $\varepsilon > 0$. To obtain the theorem, we have to prove that

$$-I(\mu) \geq \inf_G \left\{ \limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\},$$

and

$$-I(\mu) \leq \inf_G \left\{ \liminf_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\},$$

where G runs over neighbourhoods of μ. The large deviation theorem implies the almost sure weak convergence.

Theorem 4.2 *Let $U_{[m,n]}$, P_n and μ_0 as in Theorem 4.1. Then*

$$P_n(\omega) \stackrel{n\to\infty}{\longrightarrow} \mu_0$$

weakly with probability 1.

Proof. For fixed $f : \mathcal{D} \to \mathbb{C}$ bounded and continuous function and $\varepsilon > 0$ we define the sets

$$\Omega_n := \left\{ \left| \int_{\mathcal{D}} f(z)\, dP_n(\omega, z) - \int_{\mathcal{D}} f(z)\, d\mu_0(z) \right| \geq \varepsilon \right\}$$

for all $n \in \mathbb{N}$. Then

$$\text{Prob}\ (\omega \in \Omega_n) = P_n \left(\mu \in \mathcal{M}(\mathcal{D}) : \left| \int_{\mathcal{D}} f(z)\, d\mu(z) - \int_{\mathcal{D}} f(z)\, d\mu_0(z) \right| \geq \varepsilon \right).$$

The set

$$F := \left\{ \mu \in \mathcal{M}(\mathcal{D}) : \left| \int_{\mathcal{D}} f(z)\, d\mu(z) - \int_{\mathcal{D}} f(z)\, d\mu_0(z) \right| \geq \varepsilon \right\}$$

is closed, so Theorem 4.1 implies that

$$\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(F) \leq - \inf_{\mu \in F} I(\mu).$$

Because of lower semi-continuity of I, the sets $\{\mu : I(\mu) > c\}$ are open in $\mathcal{M}(\mathcal{D})$ for all $c \in \mathbb{R}$. Since F is compact, and

$$F \subset \bigcup_{c>0} \{\mu : I(\mu) > c\},$$

there exists a $\gamma > 0$, such that $I(\mu) \geq \gamma$ for all $\mu \in F$. The large deviation theorem above implies, that

$$\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(F) \leq -\gamma,$$

so for all $0 < \delta < \gamma$, there exists $N \in \mathbb{N}$, such that if $n \geq N$, then

$$P_n(F) \leq e^{-n^2(\gamma-\delta)}.$$

We get for n large enough, that

$$\text{Prob } (\omega \in \Omega_n) = P_n(F) \leq e^{-n^2(\gamma-\delta)},$$

thus

$$\sum_{n=1}^{\infty} \text{Prob } (\omega \in \Omega_n) < \infty,$$

for all $\varepsilon > 0$, so the Borel-Cantelli lemma implies that

$$\int_{\mathcal{D}} f(z)\, dP_n(\omega, z) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{D}} f(z)\, d\mu_0(z) \qquad \text{a.s.}$$

Since this is true for all bounded and continuous function on $\mathcal{D}$, the weak convergence follows. $\square$

Now we prove Theorem 4.1. Our method is again based on the explicit form of the joint eigenvalue density. First we compute the limit of the normalizing constant (66). Compute as follows.

$$\begin{aligned}
B =&: \lim_{n\to\infty} \frac{1}{n^2} \log C_{[m,n]} \\
&= \lim_{n\to\infty} \frac{1}{n^2} \log \left(\pi^n n! \prod_{j=0}^{n-1} \binom{m-n+j-1}{j}^{-1} \frac{1}{m-n+j} \right) \\
&= -\lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n-1} \log \binom{m-n+j-1}{j} \\
&= -\lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n-1} \sum_{i=1}^{j} \log \frac{m-n-1+i}{i} \\
&= -\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n-1} (n-1-i) \log \frac{m-n-1+i}{i} \\
&= -\lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{n-1-i}{n-1} \log \frac{m-n-1+i}{i}.
\end{aligned}$$

Here the limit of a Riemannian sum can be recognized and this gives an integral:

$$\begin{aligned}
B \quad &= -\int_0^1 (1-x) \log \left(\frac{\lambda - 1 + x}{x} \right) dx \\
&= -\frac{\lambda^2 \log \lambda}{2} + \frac{\lambda^2 \log(\lambda-1)}{2} - \frac{\log(\lambda-1)}{2} + \frac{\lambda-1}{2}.
\end{aligned}$$

The lower and upper estimates are stated in the form of lemmas.

Lemma 4.1 *For every $\mu \in \mathcal{M}(\mathcal{D})$,*

$$\inf_G \left\{ \limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\} \le -\iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w) - B$$

where G runs over a neighbourhood base of μ.

Proof. For $\zeta = (\zeta_1, \ldots, \zeta_n) \in \mathcal{D}^n$ set a measure

$$\mu_\zeta = \frac{1}{n} \sum_{i=1}^n \delta(\zeta_i).$$

Moreover for any neighbourhood G of $\mu \in \mathcal{M}(\mathcal{D})$ put

$$G_0 = \{\zeta \in \mathcal{D}^n : \mu_\zeta \in G\} \subset \mathcal{D}^n.$$

Then we get

$$\begin{aligned}
&P_n(G) = \overline{\nu}_n(G_0) \\
&= \frac{1}{Z_n} \int \cdots \int_{G_0} \exp\left((n-1) \sum_{i=1}^n \log\left(1 - |\zeta_i|^2\right) \right) \prod_{1 \le i < j \le n} |\zeta_i - \zeta_j|^2 \, d\zeta_1 \ldots d\zeta_n \\
&= \frac{1}{Z_n} \int \cdots \int_{G_0} \exp\left(-2 \sum_{1 \le i < j \le n} F(\zeta_i, \zeta_j) \right) d\zeta_1 \ldots d\zeta_n \\
&\le \frac{1}{Z_n} \int \cdots \int_{G_0} \exp\left(-n^2 \iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu_\zeta(z)\, d\mu_\zeta(w) + n\alpha \right) d\zeta_1 \ldots d\zeta_n \\
&= \frac{1}{Z_n} \exp\left(-n^2 \inf_{\mu' \in G} \iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu'(z)\, d\mu'(w) + n\alpha \right).
\end{aligned}$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G) \le -\inf_{\mu' \in G} \iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu'(z)\, d\mu'(w) - \lim_{n\to\infty} \frac{1}{n^2} \log C_{[m,n]}.$$

Thanks to the weak* continuity of

$$\mu' \mapsto \iint F_\alpha(z,w)\, d\mu'(z)\, d\mu'(w)$$

we obtain

$$\inf_G \left\{ \limsup_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\} \le -\iint_{\mathcal{D}^2} F_\alpha(z,w)\, d\mu(z)\, d\mu(w) + B.$$

Finally, letting $\alpha \to \infty$ yields inequality. □

Lemma 4.2 *For every $\mu \in \mathcal{M}(\mathcal{D})$,*

$$\inf_G \left\{ \liminf_{n\to\infty} \frac{1}{n^2} \log P_n(G) \right\} \geq - \iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w) - B,$$

where G runs over a neighbourhood base of μ.

Proof. If

$$\iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w)$$

is infinite, then we have a trivial case. Therefore we may assume that this double integral is finite. Since $F(z,w) = \infty$ on the boundary of the unit circle, we assume, that the support of the measure μ is distinct from the boundary, since

$$\iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w) = \infty$$

in this case. Since $F(z,w)$ is bounded from below, we have

$$\iint_{\mathcal{D}^2} F(z,w)\, d\mu(z)\, d\mu(w) = \lim_{k\to\infty} \iint_{\mathcal{D}^2} F(z,w)\, d\mu_k(z)\, d\mu_k(w)$$

with the conditional measure

$$\mu_k(B) = \frac{\mu(B \cap \mathcal{D}_k)}{\mu(\mathcal{D}_k)},$$

for all Borel set B, where

$$\mathcal{D}_k := \left\{ z : |z| \leq 1 - \frac{1}{k} \right\}.$$

So it suffices to assume, that the support of μ is contained in $\mathcal{D}_k$ for some $k \in \mathbb{N}$. Next we regularize the measure μ. For any $1/k(k+1) > \varepsilon > 0$, let φ_ε be a nonnegative C^∞-function supported in the disc $\{z : |z| < \varepsilon\}$ such that

$$\int_{\mathcal{D}} \varphi_\varepsilon(z)\, dz = 1,$$

and $\varphi_\varepsilon * \mu$ be the convolution of μ with φ_ε. This means that $\varphi_\varepsilon * \mu$ has the density

$$\int_{\mathcal{D}} \varphi_\varepsilon(z-w)\, d\mu(w)$$

on $\mathcal{D}_{k+1}$. Thanks to concavity and upper semi-continuity of Σ restricted on probability measures with uniformly bounded supports, it is easy to see that

$$\Sigma(\varphi_\varepsilon * \mu) \geq \Sigma(\mu).$$

Also

$$\lim_{\varepsilon\to 0}\int_{\mathcal{D}} \log(1-|z|)^2\, d(\varphi_\varepsilon * \mu)(z) = \int_{\mathcal{D}} \log\left(1-|z|^2\right)\, d\mu(z),$$

since $\log(1-|z|^2)$ is bounded on $\mathcal{D}_{k+1}$. Hence we may assume that μ has a continuous density on the unit disc. Now let γ be the uniform distribution on the unit disc. Then it suffices to show the required inequality for $(1-\delta)\mu+\delta\gamma$ $(0<\delta<1)$, since again by the concavity of Σ we have

$$\Sigma((1-\delta)\mu+\delta\gamma) \geq (1-\delta)\Sigma(\mu).$$

After all we may assume that μ has a continuous density f on the unit disc $\mathcal{D}$, and $\delta \leq f(z)$ for some $\delta > 0$. Next let $k = [\sqrt{n}]$, and choose

$$0 = r_0^{(n)} \leq r_1^{(n)} \leq \cdots \leq r_{k-1}^{(n)} \leq r_k^{(n)} = 1,$$

such that

$$\mu\left(\left\{z = re^{\mathrm{i}\varphi} : r \in [r_{i-1}^{(n)}, r_i^{(n)}]\right\}\right) = \frac{1}{k} \quad \text{for} \quad 1 \leq i \leq k.$$

(We have partitioned the disc into annuli of equal measure.) Note that

$$k^2 \leq n \leq k(k+2),$$

and there exists a sequence $l_1, \ldots, l_k$ such that $k \leq l_i \leq k+2$, for $1 \leq i \leq k$, and $\sum_{i=1}^k l_i = n$. For fixed i let

$$0 = \varphi_0^{(n)} \leq \varphi_1^{(n)} \leq \cdots \leq \varphi_{l_i-1}^{(n)} \leq \varphi_{l_i}^{(n)} = 2\pi,$$

such that

$$\mu\left(\left\{z = re^{\mathrm{i}\varphi} : r \in [r_{i-1}^{(n)}, r_i^{(n)}], \varphi \in [\varphi_{j-1}^{(n)}, \varphi_j^{(n)}]\right\}\right) = \frac{1}{kl_i} \quad \text{for} \quad 1 \leq j \leq l_i.$$

In this way we divided $\mathcal{D}$ into n pieces, $S_1^{(n)}, \ldots, S_n^{(n)}$. Here

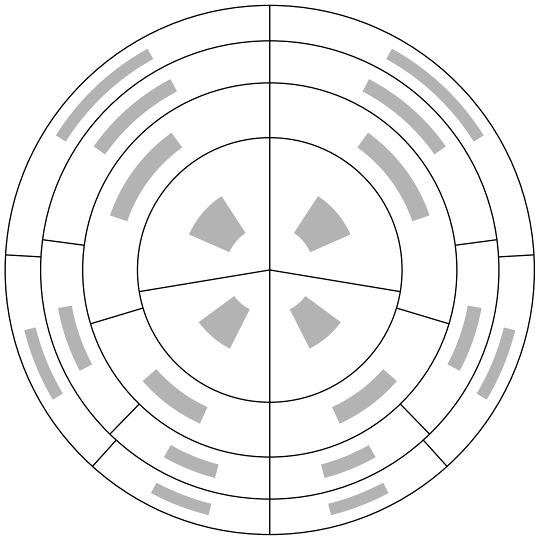
$$\frac{\delta(1-\varepsilon_n)}{n} \leq \frac{\delta}{kl_i} = \int_{S_i^{(n)}} dz \leq \frac{1}{k^2\delta} \leq \frac{1+\varepsilon_n'}{n\delta}, \tag{68}$$

where $\varepsilon_n = 2/(\sqrt{n}+2) \to 0$ and $\varepsilon_n' = 1/(\sqrt{n}-1) \to 0$ as $n \to \infty$. We can suppose, that

$$\lim_{n\to\infty}\left(\max_{1\leq i\leq n} \operatorname{diam}\left(S_i^{(n)}\right)\right) = 0. \tag{69}$$

In each part $S_i^{(n)}$ we take a smaller one $D_i^{(n)}$, similarly to $S_i^{(n)}$ by dividing the radial and phase intervals above into three equal parts, and selecting the middle ones, so that

$$\frac{\delta(1-\varepsilon_n)}{9n} \leq \int_{D_i^{(n)}} dz \leq \frac{1+\varepsilon_n'}{9n\delta}. \tag{70}$$



A division for μ with density $\dfrac{1}{2\pi}r(2+r\cos\vartheta)$ in case of $n=20$.
The white parts denote the sets $S_i^{(n)}$ the grey ones the set $D_i^{(n)}$.

We set

$$\Delta_n := \left\{(\zeta_1,\ldots,\zeta_n) : \zeta_i \in D_i^{(n)}, 1\le i\le n\right\}.$$

For any neighbourhood G of μ

$$\Delta_n \subset \{\zeta \in \mathcal{D}^n : \mu_\zeta \in G\}$$

for every n large enough. Then

$$\begin{aligned}
P_n(G) &\ge \overline{\nu}_n(\Delta_n)\\
&= \frac{1}{Z_n}\int\cdots\int_{\Delta_n}\exp\left((n-1)\sum_{i=1}^{n}(\lambda-1)\log\left(1-|\zeta_i|^2\right)\right)\\
&\qquad\qquad\qquad\qquad \times\prod_{1\le i<j\le n}|\zeta_i-\zeta_j|^2\,d\zeta_1\ldots d\zeta_n\\
&\ge \frac{1}{Z_n}\exp\left((n-1)(\lambda-1)\sum_{i=1}^{n}\min_{\zeta\in D_i^{(n)}}\log\left(1-|\zeta|^2\right)\right)\\
&\qquad\qquad\qquad\qquad \times\prod_{1\le i<j\le n}\left(\min_{\zeta\in D_i^{(n)},\eta\in D_j^{(n)}}|\zeta-\eta|^2\right)\gamma(\Delta_n)\\
&\ge \frac{1}{Z_n}\left(\frac{\delta(1-\varepsilon_n)}{9n}\right)^{n^2}\exp\left((n-1)(\lambda-1)\sum_{i=1}^{n}\min_{\zeta\in D_i^{(n)}}\log\left(1-|\zeta|^2\right)\right)\\
&\qquad\qquad\qquad\qquad \times\prod_{1\le i<j\le n}\left(\min_{\zeta\in D_i^{(n)},\eta\in D_j^{(n)}}|\zeta-\eta|^2\right).
\end{aligned}$$

Here for the first part of the sum

$$\begin{aligned}\lim_{n\to\infty} \frac{(n-1)(\lambda-1)}{n^2} \sum_{i=1}^{n} \min_{\zeta\in D_i^{(n)}} \log\left(1-|\zeta|^2\right) \\ = \lim_{n\to\infty} \frac{\lambda-1}{n} \sum_{i=1}^{n} \min_{\zeta\in D_i^{(n)}} \log\left(1-|\zeta|^2\right) \\ = (\lambda-1)\int_{\mathcal{D}} \log\left(1-|\zeta|^2\right) f(\zeta)\, d\zeta,\end{aligned}$$

because (69) implies, that the sum is the Riemannian sum of the above integral. So it remains to prove that

$$\begin{aligned}\liminf_{n\to\infty} \frac{2}{n^2} \sum_{1\le i<j\le n} \log \left(\min_{\zeta\in D_i^{(n)}, \eta\in D_j^{(n)}} |\zeta-\eta| \right) \\ \geq \iint_{\mathcal{D}^2} f(\zeta) f(\eta) \log|\zeta-\eta|\, d\zeta\, d\eta. \end{aligned} \tag{71}$$

We have

$$\iint_{\mathcal{D}^2} f(\zeta)f(\eta)\log|\zeta-\eta|\,d\zeta\,d\eta \le 2\sum_{1\le i<j\le n} \int_{S_i^{(n)}}\int_{S_j^{(n)}} f(\zeta)f(\eta)\log|\zeta-\eta|\,d\zeta\,d\eta, \tag{72}$$

since in the sum we left the terms where we integrate on the $S_i^{(n)}$, which are negative if n is large enough, since then $\operatorname{diam} S_i^{(n)} < 1$, so

$$\log|\zeta-\eta| < 0, \text{if} \quad \zeta,\eta \in S_i^{(n)}.$$

For the rest of the summands we have

$$\begin{aligned} 2\sum_{1\le i<j\le n} \int_{S_i^{(n)}}\int_{S_j^{(n)}} f(\zeta)f(\eta)\log|\zeta-\eta|\,d\zeta\,d\eta \\ \le 2\sum_{1\le i<j\le n} \log\left(\max_{\zeta\in S_i^{(n)},\eta\in S_j^{(n)}}|\zeta-\eta|\right)\int_{S_i^{(n)}} f(\zeta)\,d\zeta \int_{S_j^{(n)}} f(\eta)\,d\eta \\ \le \frac{2(1+\varepsilon_n)^2}{n^2}\sum_{i<j}\log\left(\max_{\zeta\in S_i^{(n)},\eta\in S_j^{(n)}}|\zeta-\eta|\right). \end{aligned}$$

Since the construction of $S_i^{(n)}$ and $D_i^{(n)}$ yields

$$\lim_{n\to\infty}\frac{2(1+\varepsilon_n)^2}{n^2}\sum_{1\le i<j\le n}\log\left(\frac{\max_{\zeta\in S_i^{(n)},\eta\in S_j^{(n)}}|\zeta-\eta|}{\min_{\zeta\in D_i^{(n)},\eta\in D_j^{(n)}}|\zeta-\eta|}\right) = 0,$$

we obtain (71). Here the equality does not hold because of (72). $\square$

4.4 The limit of the empirical eigenvalue distribution

The following lemma is the specialization of Proposition 2.9 to a radially symmetric function $Q : \mathcal{D} \to (-\infty, \infty]$, i. e., $Q(z) = Q(|z|)$. We assume that Q is differentiable on $(0, 1)$ with absolute continuous derivative bounded below, moreover $rQ'(r)$ increasing on $(0, 1)$ and

$$\lim_{r\to 1} rQ'(r) = \infty.$$

Let $r_0 \geq 0$ be the smallest number for which $Q'(r) > 0$ for all $r > r_0$, and we set R_0 be the smallest solution of $R_0 Q'(R_0) = 1$. Clearly $0 \leq r_0 < R_0 < 1$.

Lemma 4.3 *If the above conditions hold, then the functional I_Q attains its minimum at a measure μ_Q supported on the annulus*

$$S_Q = \{z : r_0 \leq |z| \leq R_0\},$$

and the density of μ_Q is given by

$$d\mu_Q(z) = \frac{1}{2\pi}(rQ'(r))'\, dr\, d\varphi, \qquad z = re^{\mathrm{i}\varphi}.$$

Proof. The proof is similar to the one of Theorem IV. 6. 1 in [36]. Using the formula

$$\frac{1}{2\pi}\int_0^{2\pi} \log\frac{1}{|z - re^{\mathrm{i}\varphi}|}\, d\varphi = \begin{cases} -\log r, & \text{if } |z| \leq r \\ -\log|z|, & \text{if } |z| > r, \end{cases}$$

we get that

$$\begin{aligned} U^\mu(z) &= \frac{1}{2\pi}\int_{r_0}^{R_0} (rQ'(r))' \int_0^{2\pi} \log\frac{1}{|z - re^{\mathrm{i}\varphi}|}\, d\varphi\, dr \\ &= -\log|z| \int_{r_0}^{|z|} (rQ'(r))'\, dr - \int_{|z|}^{R_0} (r(Q'(r))' \log r\, dr \\ &= -\log|z|(|z|Q'(|z|) - r_0 Q'(r_0)) \\ &\qquad -R_0 Q'(R_0)\log R_0 + |z|Q'(|z|)\log|z| + Q(R_0) - Q(z) \\ &= Q(R_0) - \log R_0 - Q(z), \end{aligned}$$

for $z \in S_Q$, since $r_0 = 0$ or $Q'(r_0) = 0$. We have

$$U^\mu(z) + Q(z) = Q(R_0) - \log R_0,$$

which is clearly a constant. Let $|z| < r_0$. Then

$$\begin{aligned} U^\mu(z) &= -\int_{r_0}^{R_0} (r(Q'(r))' \log r\, dr \\ &= -R_0 Q'(R_0)\log R_0 + \lim_{r\to r_0} rQ'(r)\log r + Q(R_0) - Q(r_0) \\ &= -\log R_0 + Q(R_0) - Q(r_0), \end{aligned}$$

since $\lim_{r\to 0} r\log r = 0$, and $Q(r_0) = 0$ if $r_0 \neq 0$. So

$$U^\mu(z) + Q(z) = Q(R_0) - \log R_0 - Q(r_0) + Q(z) \geq Q(R_0) - \log R_0,$$

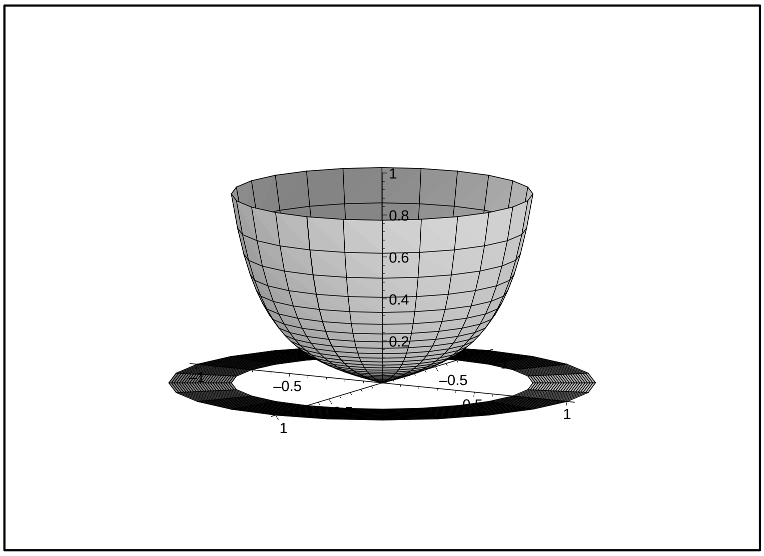
due to definition of r_0 and the monotonicity of $rQ'(r)$ implies $Q(z) \geq Q(r_0)$ for $|z| \leq r_0$. Let $|z| > R_0$ Then

$$U^\mu(z) = -\log|z| \int_{r_0}^{R_0} (r(Q'(r))'\, dr = -\log|z|.$$

So

$$U^\mu(z) + Q(z) = Q(z) - \log|z| \leq Q(R_0) - \log R_0,$$

since for $|z| > 1/\sqrt{y}$, $|z|Q'(|z|) \geq 1$, so $Q(z) = \log|z|$ is increasing. Therefore μ_Q satisfies conditions of Theorem 2.9 and it must be the minimizer.



Density of μ_0 in case of $\lambda = 1/2$

The last step is to minimize I. Now we apply Lemma 4.3 for

$$Q(z) := -\frac{\lambda - 1}{2} \log\left(1 - |z|^2\right)$$

on $\mathcal{D}$. This function satisfies the conditions of the lemma. Hence the support of the limit measure μ_0 is the disc

$$S_\lambda = \left\{ z : |z| \leq \frac{1}{\sqrt{\lambda}} \right\},$$

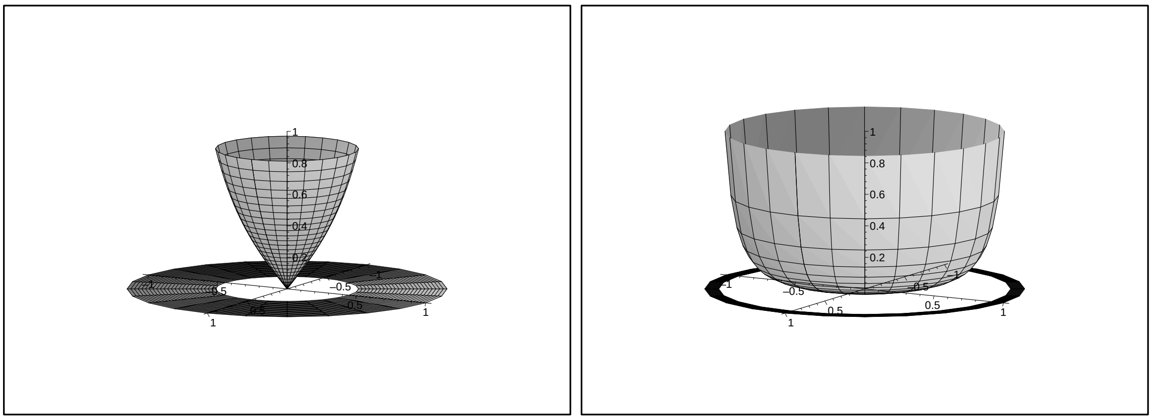
and the density is given by

$$d\mu_0 = \frac{1}{\pi}(rQ'(r))'\, dr\, d\varphi = \frac{1}{\pi} \frac{(\lambda - 1)r}{\left(1 - r^2\right)^2}\, dr\, d\varphi, \quad z = re^{\mathrm{i}\varphi}.$$

For this μ_0 again

$$\begin{aligned} I(\mu_0) &= \frac{1}{2} Q\left(\frac{1}{\sqrt{\lambda}}\right) + \frac{1}{2}\log\lambda + \frac{1}{2}\int_{S_\lambda} Q(z)\,d\mu_0(z) + B \\ &= -\frac{\lambda-1}{2}\log(\lambda-1) + \frac{1}{2\lambda}\log\lambda - \frac{(\lambda-1)^2}{2\pi}\int_0^{2\pi}\int_0^{\frac{1}{\sqrt{\lambda}}} \frac{r\log(1-r^2)}{(1-r^2)^2}\,dr\,d\varphi \\ &= -\frac{\lambda-1}{2}\log(\lambda-1) + \frac{1}{2\lambda}\log\lambda - \frac{\lambda-1}{2}\left(\lambda\log\left(\frac{\lambda-1}{\lambda}\right)+1\right) + B = 0. \end{aligned}$$

The uniqueness of μ_0 satisfying $I(\mu_0) = 0$ follows from the strict convexity of I. So we have the limit of the empirical eigenvalue distribution.



Density of μ_0 for $\lambda = 5$ and $\lambda = 1/5$

If $\lambda = 1$, then the proof goes on the same line, until the point of the upper limit. In that case we cannot assume, that the support of μ is distinct from the boundary of $\mathcal{D}$, since $F(z, w)$ in finite on the boundary.

□

Let Q_m be an $m \times m$ projection matrix of rank n, and let U_m be an $m \times m$ Haar unitary. Then the matrix $Q_m U_m Q_m$ has the same non-zero eigenvalues as $U_{[m,n]}$, but it has $m - n$ zero eigenvalues, similarly to the case of the Wishart matrices. There for we can use the 2.2 for the sequence of empirical eigenvalue distributions, and the large deviation result for $U_{[m,n]}$ is easily modified to have the following.

Theorem 4.3 *Let $1 < \lambda < \infty$ and Q_m, U_m as above. If $m/n \to \lambda$ as $n \to \infty$, then the sequence of empirical eigenvalue distributions $Q_m U_m Q_m$ satisfies the large deviation principle in the scale $1/n^2$ with rate function*

$$\tilde{I}(\widetilde{\mu}) := \begin{cases} I(\mu), & \text{if } \widetilde{\mu} = (1-\lambda^{-1})\delta(0) + \lambda^{-1}\mu, \\ +\infty, & \text{otherwise} \end{cases}$$

Furthermore, the measure

$$\widetilde{\mu}_0 = (1 - \lambda^{-1})\delta(0) + \lambda^{-1}\mu_0$$

is the unique minimizer of $\tilde{I}$, *and* $\tilde{I}(\widetilde{\mu}_0) = 0$.

5 Some connection to free probability

Let $\mathcal{A} \subset B(\mathcal{H})$. $\mathcal{A}$ is called a *unital C^* algebra*, if $\mathcal{A}$ is a $*$-algebra, (with the adjoint as the involution $*$), $\mathcal{A}$ is unital (i.e. $I_{\mathcal{H}} \in \mathcal{A}$), and $\mathcal{A}$ is closed with respect to the norm topology.

A linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ is called *state*, if $\varphi(I_{\mathcal{H}}) = 1$, and $\varphi(a^*a) \geq 0$ for every $a \in \mathcal{A}$.

Definition 5.1 *If $\mathcal{A}$ is a unital C^* algebra, and $\varphi : \mathcal{A} \to \mathbb{C}$ is a state, then we call the pair $(\mathcal{A}, \varphi)$ a non-commutative probability space, and an element of $\mathcal{A}$ is a non-commutative random variable.*

For example, if $\mathcal{H} := \mathbb{C}^n$, then $B(\mathcal{H})$ is the set $M_n(\mathbb{C})$ of $n \times n$ matrices with complex entries endowed with the state

$$\varphi(A) = \frac{1}{n} \operatorname{Tr} A = \frac{1}{n} \sum_{i=1}^{n} A_{ii}$$

is a noncommutative probability space. This is a unital algebra with the $n \times n$ identity matrix as the unit, and the involution maps the matrix into its adjoint. The normalized trace is a linear, unit preserving map, since the trace of the $n \times n$ identity matrix is n.

The state φ is *tracial*, if

$$\varphi(ab) = \varphi(ba) \tag{73}$$

for all $a, b \in \mathcal{A}$. The state φ is *faithful*, if

$$\varphi(a^*a) > 0 \tag{74}$$

for all $0 \neq a \in \mathcal{A}$.

It is easy to check, that the normalized trace on the noncommutative probability space of matrices is tracial and faithful. In the following we will assume that we have a noncommutative probability space $(\mathcal{A}, \varphi)$ with a faithful tracial state φ. The following definition is from Voiculescu ([40]).

Definition 5.2 *Let $(\mathcal{A}, \varphi)$ a noncommutative probability space, and let $\mathcal{A}_i$ be subalgebras of $\mathcal{A}$. We say that the family $(\mathcal{A}_i)_{i \in I}$ is in free relation if for every $n \in \mathbb{N}$, and $i_1, \dots, i_n \in I$, where*

$$i_1 \neq i_2 \neq \dots \neq i_{n-1} \neq i_n \neq i_1,$$

if $a_k \in \mathcal{A}_{i_k}$, and $\varphi(a_k) = 0$ for $1 \leq k \leq n$, then

$$\varphi(a_1 a_2 \dots a_n) = 0.$$

Definition 5.3 *The set $a_1, \dots, a_k$ of non-commutative random variables are free, if the generated subalgebras are free, i.e. for any set of polynomials with two non commuting variables $p_1, \dots, p_n$ such that*

$$\varphi(p_j(a_{i_j}, a_{i_j}^*)) = 0$$

for all $1 \leq j \leq n$, the

$$\varphi\left(p_1(a_{i_1}, a_{i_1}^*) \dots p_n(a_{i_n}, a_{i_n}^*)\right) = 0,$$

where

$$i_1 \neq i_2 \neq \dots \neq i_{n-1} \neq i_n \neq i_1.$$

The following definition gives other important quantities of noncommutative random variables (see [38]).

Definition 5.4 *The Fuglede-Kadison determinant of a noncommutative random variable a is defined by*

$$\Delta(a) := \exp\left(\varphi\left(\log|a|\right)\right).$$

The Brown measure of a noncommutative random variable a is

$$\mu_a = \frac{1}{2\pi}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \log\Delta\left(a - (x + y\mathrm{i})\right)$$

in distribution sense.

Consider $M_n(\mathbb{C})$ with the normalized trace. If we have an $n \times n$ matrix B_n, such that $\lambda_i(B_n) > 0$ $(1 \leq i \leq n)$ are the eigenvalues of the B_n, then

$$\exp\left(\mathrm{Tr}\,\log B_n\right) = \exp\left(\sum_{i=1}^{n} \lambda_i(B_n)\right) = \prod_{i=1}^{n} \lambda_i(B_n) = \det B_n.$$

Then for any $n \times n$ matrix A_n

$$\Delta(A_n) = \exp\left(\frac{1}{n}\mathrm{Tr}\left(\log(A_n A_n^*)^{\frac{1}{2}}\right)\right) = \sqrt[n]{\det(A_n A_n^*)^{\frac{1}{2}}} = \sqrt[n]{|\det A_n|}.$$

Now in order to obtain the Brown measure of A_n, we use that the solution of the Laplacian equation

$$\frac{1}{2\pi}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) E(x + y\mathrm{i}) = \delta_0,$$

where δ_0 is the Dirac delta distribution, is the function

$$E(x + y\mathrm{i}) := \log|x + y\mathrm{i}|.$$

This means that

$$\frac{1}{2\pi}\int_{\mathbb{C}} f(x + \mathrm{i}y)\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \log|\lambda - (x + y\mathrm{i})|\; d(x + y\mathrm{i}) = f(\lambda),$$

so

$$\begin{aligned}
&\int_{\mathbb{C}} f(x+\mathrm{i}y)\,d\mu_{A_n} \\
&\quad= \frac{1}{2n\pi}\sum_{i=1}^{n}\int_{\mathbb{C}} f(x+\mathrm{i}y)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\log|\lambda_i(A_n)-(x+y\mathrm{i})|\,d(x+y\mathrm{i}) \\
&\quad= \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i(A_n)),
\end{aligned}$$

where $\lambda_1(A_n),\ldots,\lambda_n(A_n)$ are the eigenvalues of A_n, so

$$\mu_{A_n} = \frac{1}{n}\sum_{i=1}^{n}\delta(\lambda_i(A_n)).$$

As we could see, the space of $n\times n$ matrices with the normalized trace is a noncommutative probability space in which the above definitions can be treated easily. This is why we use matrices to approximate the noncommutative random variables by a sequence of matrices as the matrix size increases. This approximation can be useful, if we know some ,,continuity" of the above properties. Unfortunately, the Fuglede-Kadison determinant is not continuous, since it is not bounded if the eigenvalues are small. If we have random matrix approximation then the probability of the small eigenvalues vanishes, so we will use random matrices.

Definition 5.5 *Let a be a noncommutative random variable, and A_n is a sequence of $n\times n$ random matrices, such that*

$$\frac{1}{n}\mathbb{E}\left(Tr(P(A_n,A_n^*))\right)\overset{n\to\infty}{\longrightarrow}\varphi\left(P(a,a^*)\right) \tag{75}$$

for all noncommutative polynomial P with two variables, the we say that A_n is a random matrix model of a. ([24].) In this case we say that (a,a^*) is the limit in distribution of (A_n,A_n^*). Let $a_1,\ldots,a_k$ be noncommutative random variables and $A_n^{(1)},\ldots,A_n^{(k)}$ be $n\times n$ random matrices. The latter form a random matrix model for $a_1,\ldots,a_k$ if

$$\frac{1}{n}\mathbb{E}\,Tr\,P\left(A_n^{(1)},\ldots,A_n^{(k)},A_n^{(1)*},\ldots,A_n^{(k)*}\right)\overset{n\to\infty}{\longrightarrow}\varphi(P(a_1,\ldots,ak,a_1^*,\ldots,a_k^*))$$

for all polynomials P of $2k$ non commuting variables.

We can define the random matrix model of k noncommutative random variables in the following way.

For example we call a a semicircular element, if $a=a^*$, and

$$\varphi(a^k)=\begin{cases}\dfrac{1}{m+1}\dbinom{2m}{m}, & \text{if} \quad k=2m\\[2ex] 0, & \text{if} \quad k \text{ odd.}\end{cases}$$

The random matrix model of the semicircular element is the sequence of $n \times n$ Wigner matrices. It is easy to check all the mixed moments, since the Wigner matrices are selfadjoint.

Like in (21) if we have two semicircular element in free relation, then

$$y := ua + vb,$$

where

$$u^2 + v^2 = 1$$

is the so-called elliptic element. It is more difficult to prove that the random matrix model of the elliptic element is the sequence of elliptic random matrices, since we need all the joint moments.

We call a $u \in \mathcal{A}$ a Haar unitary, if it is unitary, i.e.

$$uu^* = u^*u = I_{\mathcal{H}},$$

and its moments

$$\varphi(u^k) = \begin{cases} 0, & \text{if} \quad k = 0 \\ 1, & \text{if} \quad k \neq 0. \end{cases}$$

These two properties gives that

$$\varphi\left(P(u, u^*)\right) = \alpha_0,$$

where α_0 is the coefficient of the constant term in P. For a U_n sequence of $n \times n$ Haar distributed unitary random matrices it is we have that from Theorem 3.6 that

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\, U_n^k \xrightarrow{n\to\infty} 0, \tag{76}$$

if $k \neq 0$, so this sequence can be a random matrix model of u.

The Brown measures of the above mentioned noncommutative random variables (i.e. the semicircular, elliptic and Haar unitary elements) are the limit distribution of the empirical eigenvalue distributions of the corresponding random matrix models (see [21]). It is reasonable since the Brown measure can be considered as the density function of the noncommutative random variables. Since the convergence of the empirical eigenvalue distribution is fast (the large deviation principle holds in each cases), therefore the derivatives, that is the ,,densities” converge to the corresponding density function.

We proved the large deviation theorem for the truncations of the Haar unitary random matrices in Section 4, and it implied the large deviation theorem for the random matrices $Q_nU_nQ_n$, where Q_n is an $n \times n$ non-random projection ($Q_n^* = Q_n$, and $Q_n^2 = Q_n$) with rank m, and

$$\frac{m}{n} \xrightarrow{n\to\infty} \frac{1}{\lambda}.$$

Now we try to find a noncommutative random variable for this random matrix model, and check if the Brown measure of this random variable coincides with the obtained limit distribution.

We now that a random matrix model for a Haar unitary element is the sequence of Haar unitary random matrices. It is easy to see, that Q_n is a random matrix model for a projection $q \in \mathcal{A}$, (i.e. $q^2 = q$ and $q^* = q$), such that

$$\varphi(q) = \frac{1}{\lambda}.$$

Since Q_n and q are selfadjoint, so it is enough to check that

$$\frac{1}{n}\mathbb{E}\mathrm{Tr}\, Q_n^k = \frac{1}{n}\mathbb{E}\mathrm{Tr}\, Q_n = \frac{m}{n} \stackrel{n\to\infty}{\longrightarrow} \frac{1}{\lambda} = \varphi(q) = \varphi(q^k).$$

So we have the q and u limit of Q_n and U_n, we want to know their relationship. For this we have the following definition from Voiculescu.

Definition 5.6 *Let $a_1(n), \dots a_k(n)$ be noncommutative random variables in the probability space $(\mathcal{A}_n, \varphi_n)$. They are asymptotically free if there are free noncommutative random variables $a_1, \dots, a_k$ in the noncommutative probability space $(\mathcal{A}, \varphi)$ such that*

$$\varphi_n\left(P\left(a_1(n), \dots, a_k(n), a_1(n)^*, \dots, a_k(n)^*\right)\right) \stackrel{n\to\infty}{\longrightarrow} \varphi_n\left(P\left(a_1, \dots, a_k, a_1^*, \dots, a_k^*\right)\right)$$

for every polynomial P of $2k$ non-commuting variables.

We will use the following theorem in order to have that the limits u and q are in free relation. The following theorem was again proven by Voiculescu (see Theorem 4.3.1 of [26]).

Theorem 5.1 *Let S, T be sets of indices, and $(U_n(s))_{s\in S}$ an independent family of $n \times n$ Haar unitary random matrices. Let $(D_n(t))_{t\in\mathbb{T}}$ be a family of $n \times n$ non-random matrices such that*

$$\sup_n \|D_n(t)\| < \infty$$

for all $t \in T$ (here $\|\,.\,\|$ denotes the operator norm), and $(D_n(t), D_n^*(t))_{t\in T}$ has the limit. Then the family

$$\left\{(U_n(s), U_n(s)^*)_{s\in S}, (D_n(t), D_n(t)^*)_{t\in T}\right\}$$

is asymptotically free.

Now we will apply the theorem only for index sets with one element, and the non-random matrices $D_n := Q_n$. As we proved above, if $m/n \stackrel{n\to\infty}{\longrightarrow} 1/\lambda$, then the sequence Q_n has the limit q. Then we get that the matrices

$$\left\{(U_n, U_n^*), (Q_n, Q_n^*)\right\},$$

are asymptotically free, so the limits, q and u are in free relation.

So now we have t that $Q_n U_n Q_n$ is the random matrix model for noncommutative random variable quq, where u is a Haar unitary, q is a projection with rank $1/\lambda$, and they are in free relation.

In [21] Haagerup and Larsen found that the radial density of the Brown measure of this noncommutative random variable is

$$f_{qu}(s) = \frac{1 - \frac{1}{\lambda}}{\pi(1 - s^2)^2} = \frac{\lambda - 1}{\lambda\pi(1 - s^2)^2}$$

on the interval $\left[0, \frac{1}{\sqrt{\lambda}}\right]$, and

$$\mu_{qu}(\{0\}) = 1 - \frac{1}{\lambda}.$$

If $a, b \in \mathcal{A}$ are noncommutative random variables, then the Brown measure of ab and ba is the same, so

$$\mu_{quq} = \mu_{q^2u} = \mu_{qu}.$$

Again we got that the limit of the empirical eigenvalue distribution of the random matrix model is the Brown measure μ_{qu} of the noncommutative random variable.

References

[1] L. Arnold. On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20:262–268, 1967.

[2] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. I. Wigner matrices. *Ann. Probab.*, 21(2):625–648, 1993.

[3] Z. D. Bai. Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices. *Ann. Probab.*, 21(2):649–672, 1993.

[4] Z. D. Bai and Y. Q. Yin. Convergence to the semicircle law. *Ann. Probab.*, 16(2):863–875, 1988.

[5] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields*, 108(4):517–542, 1997.

[6] G. Ben Arous and O. Zeitouni. Large deviations from the circular law. *ESAIM Probab. Statist.*, 2:123–134 (electronic), 1998.

[7] B. Collins. *Intégrales matricielles et Probabilités Non-Commutatives*. Ph.D. thesis. University of Paris 6, 2003.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*. Jones and Bartlett Publishers, Boston, MA, 1993.

[9] P. Diaconis and S. N. Evans. Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.*, 353(7):2615–2633 (electronic), 2001.

[10] P. Diaconis and M. Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994. Studies in applied probability.

[11] F. J. Dyson. Statistical theory of the energy levels of complex systems. I. *J. Mathematical Phys.*, 3:140–156, 1962.

[12] F. J. Dyson. Statistical theory of the energy levels of complex systems. II. *J. Mathematical Phys.*, 3:157–165, 1962.

[13] F. J. Dyson. Statistical theory of the energy levels of complex systems. III. *J. Mathematical Phys.*, 3:166–175, 1962.

[14] W. Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

[15] V. L. Girko. The elliptic law. *Teor. Veroyatnost. i Primenen.*, 30(4):640–651, 1985.

[16] V. L. Girko. The elliptic law: ten years later. I. *Random Oper. Stochastic Equations*, 3(3):257–302, 1995.

[17] V. L. Girko. The elliptic law: ten years later. II. *Random Oper. Stochastic Equations*, 3(4):377–398, 1995.

[18] V. L. Girko. Strong elliptic law. *Random Oper. Stochastic Equations*, 5(3):269–306, 1997.

[19] A. Guionnet. Large deviations and stochastic calculus for large random matrices. *Probab. Surv.*, 1:72–172 (electronic), 2004.

[20] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000.

[21] U. Haagerup and F. Larsen. Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. *J. Funct. Anal.*, 176(2):331–367, 2000.

[22] U. Haagerup and S. Thorbjørnsen. Random matrices with complex Gaussian entries. *Expo. Math.*, 21(4):293–337, 2003.

[23] F. Hiai and D. Petz. Eigenvalue density of the Wishart matrix and large deviations. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, 1(4):633–646, 1998.

[24] F. Hiai and D. Petz. Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math. (Szeged)*, 66(3-4):809–834, 2000.

[25] F. Hiai and D. Petz. A large deviation theorem for the empirical eigenvalue distribution of random unitary matrices. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(1):71–85, 2000.

[26] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.

[27] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.*, 12(1):1–38, 1982.

[28] J. P. Keating and N. C. Snaith. Random matrices and L-functions. *J. Phys. A*, 36(12):2859–2881, 2003. Random matrix theory.

[29] N. S. Landkof. *Foundations of modern potential theory*. Springer-Verlag, New York, 1972. Translated from the Russian by A. P. Doohovskoy, Die Grundlehren der mathematischen Wissenschaften, Band 180.

[30] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[31] M. L. Mehta. *Random matrices*. Academic Press Inc., Boston, MA, second edition, 1991.

[32] F. Oravecz and D. Petz. On the eigenvalue distribution of some symmetric random matrices. *Acta Sci. Math. (Szeged)*, 63(3-4):383–395, 1997.

[33] D. Petz and F. Hiai. Logarithmic energy as an entropy functional. In *Advances in differential equations and mathematical physics (Atlanta, GA, 1997)*, volume 217 of *Contemp. Math.*, pages 205–221. Amer. Math. Soc., Providence, RI, 1998.

[34] D. Petz and J. Réffy. On asymptotics of large Haar distributed unitary matrices. *Period. Math. Hungar.*, 49(1):103–117, 2004.

[35] D. Petz and J. Réffy. Large deviation for the empirical eigenvalue density of truncated haar unitary matrices. *Probab. Theory Related Fields*, to appear.

[36] E. B. Saff and V. Totik. *Logarithmic potentials with external fields*, volume 316 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1997. Appendix B by Thomas Bloom.

[37] P. Śniady. Gaussian random matrix models for q-deformed Gaussian variables. *Comm. Math. Phys.*, 216(3):515–537, 2001.

[38] P. Śniady. Random regularization of Brown spectral measure. *J. Funct. Anal.*, 193(2):291–313, 2002.

[39] R. Speicher. Free probability theory and random matrices. In *Asymptotic combinatorics with applications to mathematical physics (St. Petersburg, 2001)*, volume 1815 of *Lecture Notes in Math.*, pages 53–73. Springer, Berlin, 2003.

[40] D. Voiculescu. Limit laws for random matrices and free products. *Invent. Math.*, 104(1):201–220, 1991.

[41] H. Weyl. *The Classical Groups. Their Invariants and Representations.* Princeton University Press, Princeton, N.J., 1939.

[42] K. Wieand. Eigenvalue distributions of random unitary matrices. *Probab. Theory Related Fields*, 123(2):202–224, 2002.

[43] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. of Math. (2)*, 62:548–564, 1955.

[44] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

[45] E. P. Wigner. Random matrices in physics. *SIAM Rev.*, 9:1–23, 1967.

[46] J. Wishart. Genaralized product moment distribution in samples. *Biometrika*, 20 A:32–52, 1928.

[47] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.

[48] K. Życzkowski and H.-J. Sommers. Truncations of random unitary matrices. *J. Phys. A*, 33(10):2045–2057, 2000.

Nyilatkozat

Alulírott Réffy Júlia kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Budapest, 2005. május 10.

....

Réffy Júlia