Random Matrix Theory, Free Random Variables and Applications

Maciej A. Nowak

Fall 2019, Descartes lectures

Nowadays, it is hard to find a branch of science where random matrix theory (hereafter RMT) does not have any applications. These lectures represent an attempt to explain this omnipresence of RMT techniques and the wide scope of applications. Our guiding rule is to look at RMT as an alternative to classical probability calculus, where instead of single random variable we have to deal with a huge set of random numbers arranged in a matrix-like structure. We will show, that such construction is not only possible, but it shares amazing similarities to classical probability calculus. Moreover, this correspondence can be formalized at the mathematical level in the case where the size of the random matrix tends to infinity. By no means this is a severe restriction, since in contemporary applications the sizes of random matrices can easily reach the order of 10^4 (financial engineering), 10^7 (wireless networks) or even 10^{10} (genetics). So, almost paradoxically, the bigger the size of the random matrix, the more definite prediction we can make on its properties. The cornerstone of the this new calculus of large matrices is the study of the spectra (eigenvalues) of the matrices. Alike in classical probability calculus, powerful central limit theorems exist in random matrix theory calculus. This is the reason, why so many similar, macroscopic spectral properties are shared by diverse and unrelated to each other complex random systems, in analogy to the omnipresence of Gaussian distribution in single-valued random structures. We call this phenomenon macroscopic universality of RMT, to make a distinction from so-called microscopic universality of RMT, second powerful phenomenon responsible for wide scope of applications of RMT. Microscopic universality of RMT is the consequence of *interactions* between the eigenvalues. Since this interaction is long-ranged, certain critical spectral phenomena can emerge *locally* at the vicinity of the some points of the spectra of the matrices. The spectral behavior in the vicinity of these points (fluctuations) depends usually only on the symmetries of the system, therefore can be categorized into universality classes, shared by very different complex systems respecting the underlying symmetry of the matrix model.

These lectures are organized as follows:

- **Gaussian Orthogonal, Unitary and Symplectic Ensembles (GOE, GUE, GSpE).** We present the construction of the simplest and most famous random matrix ensembles. Then we calculate the spectral measure and we explain the notion of the Wigner's surmise. Finally we derive Wigner's semicircle, using Dyson's intuition based on so-called

Coulomb gas picture in the limiting case when the size N of the random matrix tends to infinity. This chapter exemplifies the macroscopic features (Wigner's semicircle) and microscopic features (Wigner's surmise).

- **Microscopic Universality**. Using GUE as an example, we solve the model exactly for any finite N, introducing Mehta-Gaudin technique of orthogonal polynomials. We probe various limits and we explain the concept of microscopic universality. We derive the universal kernels using insight from elementary quantum mechanics.

- **Wishart ensemble** We review the oldest example of random matrix models (Wishart, 1928), and we discuss the modern applications of this model in multivariate statistics, telecommunication and quantum chromodynamics.

- **Large N simplifications.** We explain the gist of large N simplifications, and we introduce the diagrammatic techniques for RMT. We elucidate the jargon of physicists, sometimes too carelessly used in various applications of random matrix theory, e.g. *RMT is a field theory in* $0+0$ *dimensions.*

- **Free Random Variables calculus.** Using the intuition based on diagrammatic methods, we present so-called Free Random Variables (hereafter FRV) calculus. Whenever possible, we are using the analogies to classical probability calculus. Reading of this chapter is crucial for understanding applications presented in this course.

- **Beyond stationarity.** We introduce dynamical parameter allowing to study the evolution of random matrix ensembles and we formulate corresponding stochastic matrix-valued differential equation and corresponding matrix-valued Smoluchowski-Fokker-Planck equations.

- **Beyond Gaussianity.** We abandon the domain of Gaussian random matrix models, including polynomial measures and heavy-tail measures (Lévy random matrices).

- **Beyond Hermiticity.** We extend the domain of applications of Free Random Variables to the class of matrices possessing complex (not real) spectra.

- **Extreme events.** We discuss the statistics of extreme events, explain Tracy-Widom laws and large deviations.
- **Applications in physics and mathematics.** We list and comment most prominent applications of RMT in mathematical and physical sciences.
- **Applications in quantitative finances.** We explain how to use FRV techniques for identifying signals in noisy environments, how to optimize portfolios diversification and how to seek for cross-correlations in large sets of data.
- **Applications in wireless telecommunication.** We give a justification for recent applications of RMT in the field of wireless telecommunication (mostly Multiple Input Multiple Output (MIMO) systems).
- **Applications in life sciences.** We list and comments sample applications of random matrix in genetics, neuro-informatics, demography and ecology.
- **Miscellaneous applications.** We point at applications of RMT in quantum information theory, network theory and in Machine Learning (with emphasis on recent applications to deep neural networks by Google AI).
- **Dictionary.** Random matrix theory is being developed in various domain of sciences, so often similar concepts appear simultaneously under different names and based on different notation and conventions. We provide a short dictionary facilitating navigation between publications originating from different domains (e.g. mathematics, engineering, physics).
- **Further reading.** We provide broad bibliography, organized in correspondence to the topics (chapters) listed above.

Each chapter is completed with the list of simple problems, requiring in most cases elementary operations in *Mathematica*, MATLAB or Julia. We strongly recommend solving these problems in order to get some intuition and practical skills when dealing with large sets of data.

What is Random Matrix Theory?
Some answers:

- "Random Matrix Theory is the theory which studies the matrix whose elements are random variables (equivalent to random variables which take values in the space of matrices). As such, it is a sort of the probability theory".
  from Terrence Tao, Random Matrix Theory, 2011.

- "Suppose we had a theory that could explain everything. Not just atoms and quarks but aspects of our everyday lives too. Sound impossible? Perhaps not. It's all part of the recent explosion of work in the area of physics known as a random matrix theory...
  from "Enter the Matrix", by Mark Buchanan cover story, New Scientist, 4/10/2010, Vol. 205, Issue 2755, p 28-31.

- "-Neo: What is Matrix?...
  -Morpheus: The Matrix is everywhere. It is all around us. Even now, in this very room. You can see it when you look out your window or when you turn on your television. You can feel it when you go to work... when you go to church...when you pay your taxes. "
  from movie **The Matrix** *(1999), by Wachowski siblings*

Chapter 1

GOE,GUE,GSpe

1.0.1 1 by 1 random matrix

Random matrix is a matrix-valued random variable. Before we move to matrices, let us consider the one by one "random matrix", represented by the random variable drawn from some probability density function. We consider Gaussian probability density function (pdf) $p(x)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \equiv \mathcal{N}(\mu, \sigma^2) \tag{1.1}$$

where μ and σ denote the mean and the standard deviation of the distribution. Centered, normalized Gaussian (standardized Gaussian) reads therefore $\mathcal{N}(0,1)$. It is convenient to calculate the Fourier transform of the pdf (here the standardized Gaussian), so-called characteristic function $\phi(k)$

$$\phi(k) = \int e^{ikx} p(x) dx = \frac{1}{\sqrt{2\pi}} \int e^{ikx} e^{-x^2/2} dx = e^{-k^2/2} \tag{1.2}$$

where, first, we rewrote the argument of the exponents as $(x-ik)^2/2 + k^2/2$, second, we shifted the variable $x - ik \to x$ and third, finally integrated over x. In this way we have obtained the generating function for all the moments of the Gaussian distribution

$$\phi(k) = \sum_{l=0}^{\infty} \frac{(ik)^l}{l!} m_l = e^{-k^2/2} = \sum_{l=0}^{\infty} \frac{(-k^2/2)^l}{l!} \tag{1.3}$$

where $m_l = \int x^l p(x) dx$. Comparing l.h.s. to r.h.s. of the above formula we read out the moments of the Gaussian distribution: $m_1 = 0, m_2 = 1, m_3 =$

$0, m_4 = 3, ...$ It is convenient to introduce another generating function, for the natural logarithm of the characteristic function

$$\ln \phi(k) = \sum_{l=1}^{\infty} \frac{(ik)^l}{l!} \kappa_l = -k^2/2 \tag{1.4}$$

The coefficients κ_l are called cumulants. The remarkable feature of the (centered) Gaussian distribution stems from the fact, that it has only one non-vanishing cumulant, i.e. κ_2, here normalized to 1.

Let us show the advantages of working with cumulants. Consider now the following problem. From two identical, independent [1] standard Gaussian distributions $p_i(x_i) = \mathcal{N}_i(0,1)$ we draw two random variables x_i, where $i = 1, 2$. We ask then what is the distribution of the variable $s = x_1 + x_2$. Mathematically, we have to solve the problem

$$p(s) = \int dx_1 dx_2 p_1(x_1) p_2(x_2) \delta(s - x_1 - x_2) = \int dx p_1(x) p_2(s - x) \tag{1.5}$$

i.e. we have to calculate the convolution. The Fourier transform reduces this problem to the multiplication of the corresponding characteristic functions, $\phi_{1+2}(k) = \phi_1(k) \cdot \phi_2(k)$. Then taking the logarithm transmutes the multiplication into addition,

$$\ln \phi_{1+2}(k) = \ln \phi_1(k) + \ln \phi_2(k) \tag{1.6}$$

We have obtained the "addition law": using the definition of cumulants (1.2) we see that summing of random variables generates the probability distribution, whose cumulants are the sums of corresponding cumulants of individual, independent distributions. The explicit pdf of such distribution corresponds to the inverse Fourier transform of the characteristic function $\phi_{1+2}(k)$. In the case of the Gaussian distribution we simply read $\kappa_2^{(1+2)} = \kappa_2^{(1)} + \kappa_2^{(2)} = 1 + 1 = 2$, since all other cumulants vanish. The distribution of random variable s is therefore $\mathcal{N}(0, 2)$, which we write as

$$x_1 \oplus x_2 = \mathcal{N}(0, 2) \tag{1.7}$$

A straightforward generalization gives us the result for the weighted sum of Gaussian variables $s = \sum_i^k a_i x_i$, where each x_i is drawn from independent

[1] Independence means that probability distribution factorizes, i.e. $p(x_1, x_2) = p_1(x_1) p_2(x_2)$.

$\mathcal{N}_i(\mu_i, \sigma_i^2)$:

$$a_1 x_1 \oplus a_2 x_2 \oplus ... \oplus a_k x_k = \mathcal{N}(\sum_{i=1}^{k} a_i \mu_i, \sum_{i=1}^{k} (a_i \sigma_i)^2) \tag{1.8}$$

One of the purposes of this lecture is to introduce the Reader to similar "addition laws" for the spectra of large random matrices.

1.1 Gaussian Orthogonal Ensemble

Let us fill N by N matrix M with random numbers from Gaussian distribution, e.g. from $\mathcal{N}(0, 2)$. [2] Unfortunately, such obtained matrix has a complex spectrum, which requires some special tools, to be discussed later in these notes. So let us avoid the complex spectrum by symmetrizing the matrix, i.e. by considering $X = \frac{1}{2}(M + M^T)$. The elements of X stem now from the following prescription:
– diagonal entries x_{ii} are populated from $N(0, 2)$
– off-diagonal entries x_{ij} above diagonal ($i < j$) are populated from $N(0, 1)$, as explicit from formula (1.8), since $a_1 = a_2 = \frac{1}{2}$ and $\sigma_1 = \sigma_1 = \sqrt{2}$.
– remaining off-diagonal below diagonal are identical to corresponding entries above diagonal, so the matrix is *symmetric* ($x_{ij} = x_{ji}$)

Since each drawing of random number is an independent event, the probability of formation of such random matrix is the product of individual probabilities

$$\begin{aligned} P(X) &= \frac{1}{Z_N^{GOE}} \prod_{i=1}^{N} e^{-x_{ii}^2/4} \prod_{1 \le i < j \le N} e^{-x_{ij}^2/2} \\ &= \frac{1}{Z_N^{GOE}} \prod_{i,j=1}^{N} e^{-x_{ij}/2^2} = \frac{1}{Z_N^{GOE}} e^{-\frac{1}{4}\sum_{i,j=1}^{N} x_{ij}^2} = \frac{1}{Z_N^{GOE}} e^{-\frac{1}{4}\mathrm{Tr}X^2} \end{aligned} \tag{1.9}$$

Such ensemble we call **Gaussian Orthogonal Ensemble (GOE)**. Here Z_N is the normalization constant, stemming from the normalization of the individual Gaussians, therefore $Z_N^{GOE} = (2 \cdot 2\pi)^{N/2} (2\pi)^{N(N-1)/4}$.

[2] Standard deviation equal to 2 just corresponds to my favorite convention.

In random matrix theory, we are interested rather in the eigenvalues and in the eigenvectors of random matrices, and not in the elements of the matrices. We would like therefore to express our measure in terms of the eigenvalues. In general, random N by N matrix from GOE has $N(N+1)/2$ independent elements (symmetric real matrix $X = X^T$). Taking into account that each symmetric matrix can be diagonalized by orthogonal transformation $X = O\Lambda O^T$, we can parameterize the r.h.s. of above equality as N eigenvalues λ_i forming diagonal matrix Λ and $N(N-1)/2$ "angles" parametrizing the orthogonal matrix O. Note also, that due to the cyclic properties of the trace $\mathrm{Tr}X^2 = \mathrm{Tr}\Lambda^2$. In order to see what is the price of switching from elements x_{ij} to eigenvalues, let us make an explicit calculation for the simplest case $N = 2$. First, we have to find eigenvalues of the symmetric 2 by 2 matrix X. The answer is $\lambda_\pm = \frac{1}{2}(x_{11} + x_{22}) \pm \frac{1}{2}\sqrt{(x_{11} - x_{22})^2 + 4x_{12}^2}$, so matrix $\Lambda = \mathrm{diag}(\lambda_+, \lambda_-)$. In our case the orthogonal matrix depends only on one parameter, and can be parametrized in a standard way as

$$O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.10}$$

Comparing the l.h.s to r.h.s. of $X = O\Lambda O^T$ we arrive at

$$\begin{aligned} x_{11} &= \lambda_+ \cos^2\theta + \lambda_- \sin^2\theta \\ x_{12} &= (\lambda_+ - \lambda_-)\sin\theta\cos\theta \\ x_{22} &= \lambda_+ \sin^2\theta + \lambda_- \cos^2\theta \end{aligned} \tag{1.11}$$

The Jacobian is therefore a modulus of 3 by 3 determinant

$$\begin{vmatrix} \partial_{\lambda_+} x_{11} & \partial_{\lambda_+} x_{12} & \partial_{\lambda_+} x_{22} \\ \partial_{\lambda_-} x_{11} & \partial_{\lambda_-} x_{12} & \partial_{\lambda_-} x_{22} \\ \partial_\theta x_{11} & \partial_\theta x_{12} & \partial_\theta x_{22} \end{vmatrix} = \lambda_+ - \lambda_- \tag{1.12}$$

So in our case the measure reads

$$P_2(X)dX = \frac{1}{Z_2^{GOE}} e^{-\frac{1}{4}\mathrm{Tr}X^2} dX = \mathrm{const}\ e^{-\frac{1}{4}\sum_{i=\pm}^{2}\lambda_i^2} |\lambda_- - \lambda_+| \prod_{i=\pm} d\lambda_i d\theta \tag{1.13}$$

In the next section we will argue, that the generalization of the above formula for every N reads

$$P_N(X)dX = \frac{e^{-\frac{1}{4}\mathrm{Tr}X^2} dX}{Z_N^{GOE}} = \mathrm{const}\, e^{-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2} \prod_{i<j} |\lambda_i - \lambda_j| \prod_{i=1}^{N} d\lambda_i d\Omega \tag{1.14}$$

where $d\Omega$ represents integration over $N(N-1)/2$ angles. We conclude this part with two remarks. First, note that in the case when the observable in the question does not depend on the angles, the angular integration can be done in a straightforward way, contributing only to some overall constant. Then, $P(X)$ is replaced by $P(\lambda_1, ..., \lambda_N)$, which for large N represents a considerable reduction of the dimensionality of the original problem. Second, note that the form of the Jacobian is highly non trivial, and the eigenvalues "see" each other, i.e. they *interact*. Actually, most of the properties of random matrices discussed in these lectures can be linked to this phenomenon.

1.2 Gaussian Unitary Ensemble

Instead of using real random numbers, we can use as well complex random numbers. Let us fill N by N matrix X with random numbers from Gaussian distribution according to the following prescription:
– diagonal entries x_{ii} are real and are populated from $\mathcal{N}(0,1)$
– off-diagonal entries x_{ij} above diagonal ($i < j$) are complex and are populated from $(\mathcal{N}_1(0,1) + i\mathcal{N}_2(0,1))/\sqrt{2}$, where $\mathcal{N}_1$ and $\mathcal{N}_2$ are two independent normal distributions
– remaining off-diagonal below diagonal are complex conjugated with respect to the corresponding entries above diagonal, so that the matrix is *hermitian* ($x_{ij} = \bar{x}_{ji}$)

Since each drawing of random number is an independent event, the probability of formation of such random matrix is the product of all individual N^2 entries (N diagonal and $2N(N-1)/2$ off-diagonal, where additional factor 2 comes from complex variables). The rerun of the similar argument presented in the case of GOE yields

$$P(X) = \frac{1}{Z_N^{GUE}} e^{-\frac{1}{2}\mathrm{Tr}X^2} \tag{1.15}$$

where Z_N is, as in the case of GOE, an overall constant fulfilling the normalization of the probability. Such ensemble we call **Gaussian Unitary Ensemble (GUE)**.

In order to express the measure in terms of the eigenvalues, we have to calculate the corresponding Jacobian exploiting the fact, that every hermitian matrix can be diagonalized by the unitary transformation, i.e. $X = U\Lambda U^\dagger$.

Instead of the brute force calculation presented above in the case of 2 by 2 GOE matrices, we will perform the calculation for arbitrary N by N GUE matrices exploiting the following trick:

Consider the calculation of the Jacobian corresponding to the change from three-dimensional cartesian coordinates to the spherical ones. Expressing x, y, z in terms of r, θ, ϕ and calculating 3 by 3 determinant built of corresponding derivatives alike we did in the previous example, we arrive at the known result $J = r^2 \sin\theta$. But there exists a simpler way: Let us consider an infinitesimal interval, independent on the choice coordinates, built of from small displacements along radial, polar and asimuth directions:

$$ds^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{1.16}$$

The metric tensor corresponding to this measure reads $g_{ik} = \mathrm{diag}(1, r^2, r^2 \sin^2\theta)$. Since element of volume $dV = dxdydz = \sqrt{\det g_{ik}} dr d\theta d\phi$, we immediately read out the Jacobian $J = r^2 \sin\theta$. We can repeat the above reasoning in the case of GUE matrix. Since $X = U\Lambda U^\dagger$,

$$\begin{aligned} dX &= dU\Lambda U^\dagger + Ud\Lambda U^\dagger + U\Lambda dU^\dagger \\ &= U(d\Lambda + [U^\dagger dU, \Lambda])U^\dagger \end{aligned} \tag{1.17}$$

since due to the unitarity condition $U^\dagger U = 1$ we have $dU^\dagger U + U^\dagger dU = 0$, hence $dU^\dagger = -U^\dagger dU U^\dagger$. The infinitesimal interval reads therefore

$$\mathrm{tr}\ dX dX^\dagger = \mathrm{tr}|d\Lambda + [U^\dagger dU, \Lambda]|^2 \tag{1.18}$$

Matrix Λ is diagonal, built out of real eigenvalues of hermitian matrix, i.e. $(\Lambda)_{ik} = \delta_{ik}\lambda_i$, so matrix element of the commutator reads $[U^\dagger dU, \Lambda]_{ij} = (\lambda_j - \lambda_i)[U^\dagger dU]_{ij}$. Since first term is symmetric and the second one is antisymmetric in indices (i, j), mixed terms vanish and the interval reads

$$\mathrm{tr} dX dX^\dagger = \sum_{i=1}^{N} d\lambda_i^2 + \sum_{i \neq j} (\lambda_i - \lambda_j)^2 dU_{ij} d\bar{U}_{ij} \tag{1.19}$$

From this form we can read the metric tensor corresponding to the interval: it is represented by diagonal matrix, where first N terms ("radial" coordinates) are equal to one and the subsequent $N(N-1)$ terms are of the form $(\lambda_i - \lambda_j)^2$ ("angular" coordinates), so the determinant and the Jacobian read,

respectively

$$\begin{aligned} \det g_{ij} &= \prod_{i\neq j}(\lambda_i-\lambda_j)^2 \\ J=\sqrt{\det g_{ij}} &= \prod_{i<j}(\lambda_i-\lambda_j)^2 \end{aligned} \tag{1.20}$$

The measure for GUE reads therefore

$$P_N(X)dX=\frac{e^{-\frac{1}{2}\mathrm{tr}X^2}dX}{Z_N^{GUE}}=\mathrm{const}\,e^{-\frac{1}{2}\sum_{i=1}^N\lambda_i^2}\prod_{i<j}(\lambda_i-\lambda_j)^2\prod_{i=1}^N d\lambda_i d\Omega \tag{1.21}$$

Note that the number of angles in the GOE case (equal to $N(N-1)/2$) is precisely twice smaller then corresponding number in the GUE case (equal to $N(N-1)$), which explains why the Jacobian for GOE is the square root of the Jacobian for GUE.

1.3 Gaussian Symplectic Ensemble

Till now, we were populating random matrices with random real and and random complex numbers. Let us exploit the third possibility, i.e. random quaternions. We remind that the real quaternion has the following structure

$$Q=a_0+Ia_1+Ja_2+Ka_3 \tag{1.22}$$

where I,J,K, are "square roots" of -1 and $IJK=-1$. For practical purposes, it is convenient to use the matrix representation of the quaternion. Let us identify

$$\begin{aligned} I &\leftrightarrow i\sigma_3=i\begin{pmatrix}1&0\\0&-1\end{pmatrix} \\ J &\leftrightarrow i\sigma_2=i\begin{pmatrix}0&-i\\i&0\end{pmatrix} \\ K &\leftrightarrow i\sigma_1=i\begin{pmatrix}0&1\\1&0\end{pmatrix} \end{aligned} \tag{1.23}$$

where σ_i are Pauli matrices and the unity multiplying a_0 we identify with unit two-dimensional matrix $\mathbf{1}_2$. Using the explicit form of Pauli matrices, the quaternion reads

$$Q=\begin{pmatrix}z&w\\-\bar{w}&\bar{z}\end{pmatrix} \tag{1.24}$$

where $z = a_0 + ia_1$ and $w = a_2 + ia_3$ are two complex numbers.

Let us fill N by N matrix X with random quaternions from Gaussian distribution according to the following prescription:
– diagonal entries x_{ii} are real ($Q = a_0\mathbf{1}_2$) and a_0 are populated from $\mathcal{N}(0, 1/2)$
– off-diagonal entries x_{ij} above diagonal ($i < j$) are quaternions of the form $(\mathcal{N}_0(0,1) + I\mathcal{N}_1(0,1) + J\mathcal{N}_2(0,1) + K\mathcal{N}_3(0,1))/2$, where $\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ are four independent standard Gaussian distributions, i.e. $z = (\mathcal{N}_0 + i\mathcal{N}_1)/2, w = (\mathcal{N}_2 + i\mathcal{N}_3)/2$.
– remaining off-diagonal below diagonal are complex conjugated quaternions with respect to the corresponding entries above diagonal
Corresponding measure reads

$$P(X) = \frac{1}{Z_N^{GSE}} e^{-\mathrm{tr}X^2} \tag{1.25}$$

where Z_N is an overall constant fulfilling the normalization of the probability. Such ensemble we call **Gaussian Symplectic Ensemble (GSE)**.

Comparing the number of the "angles" for GSE we notice that this number is twice larger comparing to GUE, since now we have a pair of complex numbers for the off-diagonal elements. It is therefore natural to expect, that the corresponding Jacobian for GSE is the square of the Jacobian for GUE.

1.4 Threefold way

We can unify all three above ensembles by introducing parameter β, equal to 1 for GOE (real numbers), equal to 2 for GUE (complex numbers) and equal to 4 for GSE (quaternions). *It is also convenient to rescale the variance with N, which corresponds to augment the above presciptions for drawing GOE, GUE and GSE ensembles with additional division of the random matrix by* $\sqrt{N}$. Then

$$P_N^{(\beta)} d_\beta X = \frac{1}{Z_N^{(\beta)}} e^{-\frac{N\beta}{2}\mathrm{Tr}V(H)} d_\beta X \tag{1.26}$$

For the Gaussian case, $V(X) = \frac{1}{2\sigma^2}X^2$, where σ is arbitrary parameter, which we have put equal to 1 in the previous sections, but in general, for non-Gaussian ensembles, $V(X)$ could be a polynomial in X. Switching to eigenvalues and integrating over the "angles" we arrive at the important

formula

$$P_N^{(\beta)}(\lambda_1,...,\lambda_N)\prod_{i=1}^{N}d\lambda_i=\frac{1}{Q_N^{(\beta)}}e^{-\frac{N\beta}{2}\sum_{i=1}^{N}V(\lambda_i)}\prod_{i<j}|\lambda_i-\lambda_j|^{\beta}\prod_{i=1}^{N}d\lambda_i \quad (1.27)$$

where $V(\lambda_i)=\frac{\lambda_i^2}{2\sigma^2}$ and the normalization constant reads:

$$Q_N^{(\beta)}=(2\pi)^{N/2}\left(\frac{2\sigma^2}{\beta N}\right)^{-(2N+\beta N(N-1))/4}\prod_{j=1}^{N}\frac{\Gamma(\beta j/2)}{\Gamma(\beta/2)} \quad (1.28)$$

This non-trivial normalization, taking into account the volume of the integrated "angles", can be derived from the **Selberg integral**. This is the first of few "famous" integrals in Random Matrix Theory

$$\begin{aligned} S_N(a,b,c) &\equiv \int_0^1...\int_0^1 t_i^{a-1}(1-t_i)^{b-1}\prod_{i<j}|t_i-t_j|^{2c}dt_1...dt_N \\ &= \prod_{j=0}^{N-1}\frac{\Gamma(a+jc)\Gamma(b+jc)\Gamma(1+(j+1)c)}{\Gamma(a+b+(N+j-1)c)\Gamma(1+c)} \end{aligned} \quad (1.29)$$

1.5 Poisson distribution versus Wigner surmise

We consider the sequence of eigenvalues (energies) E_i. Lets the probability per unit energy that we will observe the peak is given by

$$dp=\rho dE \quad (1.30)$$

where ρ is the *constant* rate and dE is infinitesimal. We try to estimate the probability of finding no level till the value E, (probability of the gap), which we denote as $p(0,E)$. To do it, let us consider probability of the gap between 0 and $E+dE$, which is the product of probability of the gap till E and the probability of the gap in dE, which from (1.30) is $1-\rho dE$, i.e.

$$p(0,E+dE)=p(0,E)(1-\rho dE) \quad (1.31)$$

Above is equivalent (note that dE is infinitesimal) to the differential equation

$$\frac{dp(0,E)}{dE}=-\rho p(0,E) \quad (1.32)$$

with obvious solution $p(0,E) = Ce^{-\rho E}$. Since $p(0,0) = 1$, $C = 1$. Probability of observing the next level in the interval $[E, E+dE]$ is therefore $p_P(E) = e^{-\rho E}\rho dE$. Introducing the average energy

$$< E >= \int_0^\infty E\rho e^{-\rho E} dE = \frac{1}{\rho} \equiv D \tag{1.33}$$

and rescaling the variables $E/D \equiv s$ we finally arrive at the Poisson law

$$p_P(s) = e^{-s} \tag{1.34}$$

We contrast this law against the so-called Wigner surmise - the probability of finding the gap in the statistical ensemble of 2 by 2 random matrices belonging to GOE. Up to irrelevant here normalization constant the probability of finding the gap S in the spectrum reads

$$p_W(S) \sim \int e^{-\frac{\lambda_+^2}{4}} e^{-\frac{\lambda_-^2}{4}} |\lambda_+ - \lambda_-| \delta(S - (\lambda_+ - \lambda_-)) d\lambda_+ d\lambda_- \tag{1.35}$$

where we used the result (1.13). We factorize the integrals introducing sum and difference of eigenvalues ($p = \lambda_+ + \lambda_-, r = \lambda_+ - \lambda_-,$) and elementary integrations lead to $p(S) \sim Se^{\frac{-S^2}{8}}$. Imposing the conditions $\int p(S)dS = 1$ and $\int Sp(S)dS = 1$ (normalization condition and the requirement that the average spacing is one), we arrive at **the Wigner's surmise**

$$p_W^{(GOE)}(s) = \frac{\pi}{2} s e^{-\frac{\pi}{4}s^2} \tag{1.36}$$

Similarly we get two other surmises, for GUE and GSE. The answer reads

$$\begin{aligned} p_W^{(GUE)}(s) &= \frac{32}{\pi^2} s^2 e^{-\frac{4}{\pi}s^2} \\ p_W^{(GSE)}(s) &= \frac{2^{18}}{3^6\pi^3} s^4 e^{-\frac{64}{9\pi}s^2} \end{aligned} \tag{1.37}$$

The spectacular agreement of the GOE Wigner surmise with the nuclear data for the spectra of the excited resonances was the beginning of the triumphal march of RMT in complex systems. In his seminal paper in 1967, Wigner has challenged the paradigm of the statistical physics. He proposed the scheme, in which the Hamiltonian itself was a statistical object, restricted only by some class of the global symmetry. Originally, in his construction he used

the symmetric matrices whose elements were just drawn from binomial distribution (0 and 1). Due to the universality of the prediction, the same class is also represented by the Gaussians. Taking into account that $\int p(s)ds = 1$ and restriction $< s >= 1$, Wigner prediction was parameter-free. The surmise was based on a calculation of 2 by 2 matrices. The spectacular agreement of the surmise comparing to the exact result for the probability of the gap, obtained only few years after, shows the main effects comes from the repulsion of the adjacent eigenvalues, and the effect of the repulsion for the triples, quadruples etc is only marginal. Figure 1 shows the ensemble of ca1000 resonances, compared to the predictions of GOE Wigner surmise. Similar successful predictions were soon obtained for systems in atomic and molecular physics, elasto-mechanic modes in irregular crystals and chaotic billiards. Part of the mystery of the success of Wigner's surmise for GOE was resolved by so-called Bohigas-Giannoni-Schmidt hypothesis (still unproven), stating that the spectra of time reversal invariant systems whose classical analogs were chaotic show the same fluctuations as Gaussian Orthogonal Ensembles. The hypothesis was then generalized for two remaining surmises of the threefold way, for rotationally invariant systems with lack of time-reversal symmetry (GUE) and with time reversal symmetry but lack of rotational symmetry (GSE). Still, the BGS hypothesis does not resolve the mystery of the microscopic universality of the gaps in all complex systems. One of the most stunning empirical results is perhaps the agreement with GUE surmise based on 70 millions spacing between the adjacent zeros of Riemann ζ function starting at 10^{20} zero. Other examples include similarity of hadrons spectra for GUE or predictions for the spacing of the spectra of Dirac operators on lattice. Wigner surmise is also observe in stock market data , patterns of waiting times for some Mexican bus companies and in various models of percolation, e.g. in ice pack patterns in Antarctics. We, humans, are not free from surmise restrictions as well - we have it in our bones, in our genes, and in our brain. Perhaps the Reader may come with even more exotic examples of the omnipresence of the Wigner surmises.

Exercise: Generate 10^3 matrices of various sizes (2 by 2, 4 by 4, etc) for GOE. Plot the difference of the adjacent eigenvalues (staying far from the edges of the whole spectrum), and compare the results to the Wigner surmise.

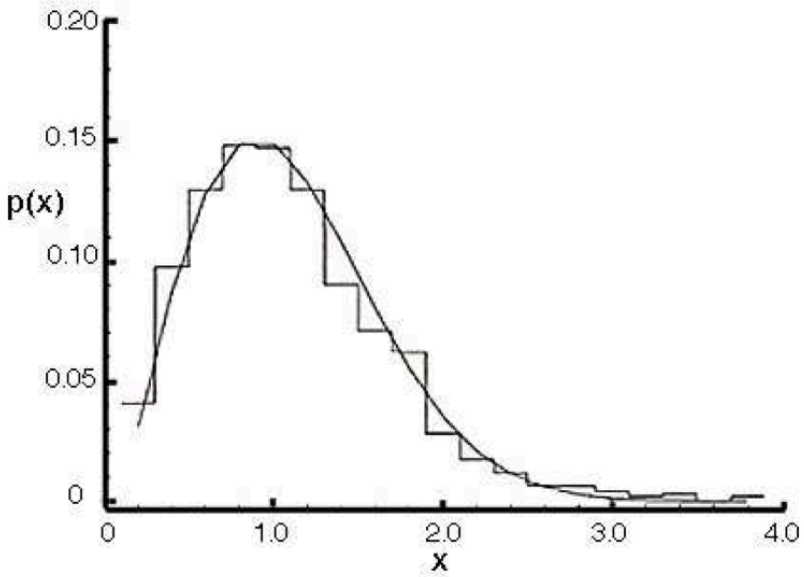


Figure 1.1: Wigner' surmise fitted to spacing distribution of 932 s-wave resonances in the interaction ^{238}U$+n$ at energies up to 20 KeV. From F.WJ.Firk, A.J. Miller; arXiv:0909.4914v1.

1.6 Wigner's semicircle

We will now study the large N limit of all three Gaussian ensembles of the threefold way. We explain the shape of the average spectral density for the Gaussian threefold way as follows: We integrate (22)

$$\begin{aligned}\int P_N^\beta(\lambda_1,...\lambda_N)\prod d\lambda_i &= \text{const}\int d\lambda_1...d\lambda_N\prod_i e^{-\frac{N\beta}{2}V(\lambda_i)}\prod_{i<j}|\lambda_i-\lambda_j|^\beta \\ &\sim \int d\lambda_1...d\lambda_N e^{-N^2S(\lambda_1,...,\lambda_N)} \qquad (1.38)\end{aligned}$$

where

$$S=\frac{\beta}{2}\left[\frac{1}{N}\sum V(\lambda_i)-\frac{2}{N^2}\sum_{i<j}\ln|\lambda_i-\lambda_j|\right] \qquad (1.39)$$

We will now calculate the above integral in the limit when $N\to\infty$, using the saddle point method. The condition of the extremum of S, i.e. $\frac{\partial S}{\partial\lambda_i}=0$, reads

$$V'(\lambda_i^*)=\frac{1}{N}\sum_{i\neq j}\frac{2}{\lambda_i^*-\lambda_j^*} \qquad (1.40)$$

where the asterisk denotes the saddle point solution. Note that the parameter β has disappeared from the saddle point condition, implying the similar shape for all three ensembles. We define the auxiliary complex-valued function $G^*(z)=\frac{1}{N}\sum_{i=1}^N\frac{1}{z-\lambda_i^*}$. Squaring this function, and using in the last line the saddle point condition (1.40) we arrive at

$$\begin{aligned}[G(z)^*]^2 &= \frac{1}{N^2}\sum_i\frac{1}{(z-\lambda_i^*)^2}+\frac{1}{N^2}\sum_{i\neq j}\frac{1}{z-\lambda_i^*}\frac{1}{z-\lambda_j^*} \\ &= -\frac{1}{N}\frac{dG^*(z)}{dz}+\frac{1}{N^2}\sum_{i\neq j}\frac{1}{\lambda_i^*-\lambda_j^*}\left(\frac{1}{z-\lambda_j^*}-\frac{1}{z-\lambda_i^*}\right) \\ &= -\frac{1}{N}\frac{dG^*(z)}{dz}+\frac{2}{N^2}\sum_{i<j}\frac{1}{\lambda_i^*-\lambda_j^*}\frac{1}{z-\lambda_j^*} \\ &= -\frac{1}{N}\frac{dG^*(z)}{dz}+\frac{1}{N}\sum_j\frac{V'(\lambda_j^*)}{z-\lambda_j^*} \qquad (1.41)\end{aligned}$$

In the case of the simplest, Gaussian potential, $V^{'}(\lambda_j) = \lambda_j/\sigma^2$, the above equation takes the form

$$[G^*(z)]^2 = -\frac{1}{N}\frac{dG^*(z)}{dz} + (zG^*(z) - 1)/\sigma^2 \tag{1.42}$$

In the limit of large N (and assuming that the derivative is finite - important point to which we will come back soon) we arrive at the algebraic (quadratic) equation

$$\sigma^2[G^*(z)]^2 - zG^*(z) + 1 = 0 \tag{1.43}$$

1.6.1 Green's function alias resolvent alias Stieltjes alias Cauchy transform

We define a complex-valued function Green's function (resolvent)

$$G(z) \equiv \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z\mathbf{1}_N - X}\right\rangle = \frac{1}{N}\sum_{k=0}^{\infty}\frac{\left\langle \mathrm{Tr}X^k\right\rangle}{z^{k+1}} \tag{1.44}$$

where by $< ... >$ we mean averaging over the measure $P_N^{\beta}(\lambda_1, ...\lambda_N)\prod d\lambda_i$. Let us say that are interested in averaged spectral density of our ensemble, i.e. in the observable

$$\rho(\lambda) = \frac{1}{N}\left\langle \sum_i \delta(\lambda - \lambda_i)\right\rangle = \frac{1}{N}\left\langle \mathrm{Tr}(\lambda\mathbf{1}_N - X)\right\rangle \tag{1.45}$$

Combining two above equations we arrive at the Stieltjes (Cauchy) transformation

$$G(z) = \int \rho(\lambda)\frac{d\lambda}{z-\lambda} \tag{1.46}$$

where the eigenvalue density $\rho(\lambda)$ is defined on the real interval I and the complex number z is defined outside this interval. Expanding both equations around $z = \infty$ we arrive at the spectral moments

$$M_k = \frac{1}{N}\left\langle \mathrm{Tr}X^k\right\rangle = \int \rho(\lambda)\lambda^k d\lambda \tag{1.47}$$

The Green's function is therefore the generating function of the spectral moments. In random matrix theory it plays exactly the role of the characteristic function (Fourier transform of the pdf), generating moments of pdf's

in classical theory of probability. The calculation of the Green's function is very often simpler comparing to the calculation of the spectral distribution. Knowing Green's function, one can reconstruct the spectral distribution. The easiest way is to consider the Green's function in the limit $z = \lambda^{'} + i\epsilon$, where $\epsilon \to 0$. Then

$$\Im G(z)|_{z=\lambda'+i\epsilon} = \Im \int \frac{\rho(\lambda)d\lambda}{\lambda' - \lambda + i\epsilon} = -\int \rho(\lambda)\frac{\epsilon}{|\lambda - \lambda'|^2 + \epsilon^2}d\lambda \qquad (1.48)$$

Noting that in the limit $\epsilon \to 0$ the expression $\frac{1}{\pi}\frac{\epsilon}{|\lambda-\lambda^{'}|^{+}\epsilon^2}$ represents the Dirac delta function $\delta(\lambda - \lambda^{'})$, we arrive at

$$-\frac{1}{\pi}\lim_{\epsilon\to 0}\Im G(z)|_{z=\lambda'+i\epsilon} = \int \rho(\lambda)\delta(\lambda - \lambda^{'})d\lambda = \rho(\lambda^{'}) \qquad (1.49)$$

so the spectral function can be read our from the discontinuities (jumps) of the resolvent along the real axis.

We can move now back to our saddle point analysis. We expect, that the true Green's function can be approximated in the large N limit by $G(z) = G^*(z) + O(1/N)$. In the large N limit, the solution of the quadratic equation is

$$G(z) \approx G^*(z) = \frac{1}{2\sigma^2}(z \mp \sqrt{z^2 - 4\sigma^2}) \qquad (1.50)$$

Since we expect that our spectral moments are finite, we know that in the limit $z \to \infty$ asymptotic behavior of the Green's function is $G(z) \sim 1/z$. This condition chooses the upper sign. We can read out now the imaginary part of the Green's function, coming from the condition $z^2 - 4\sigma^2 \leq 0$. The answer reads

$$\rho(\lambda) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - \lambda^2} \qquad (1.51)$$

This is the famous Wigner's semicircle law (actually, this is a semi-ellipse, and not a semicircle).

1.6.2 Real part of the Green's function

Since imaginary part of the Green's function turned out to be so useful, we may wonder if the real part of the Green's function is of any use as well. In

the large N expansion, we approximate $G^*(z) \approx G(z)$, which allows us to rewrite the saddle point equation as

$$2 \lim_{\epsilon \to 0} \Re G(z)|_{z=\lambda+i\epsilon} = V'(\lambda) \tag{1.52}$$

so the real part of the Green's encodes the shape of the random matrix eigenvalue potential. For the Gaussian $V(\lambda) = \frac{\lambda^2}{2\sigma^2}$ the explicit form of the saddle point equation reads

$$2\mathrm{PV} \int \frac{\rho(\lambda')d\lambda'}{\lambda - \lambda'} = \frac{\lambda}{\sigma^2} \tag{1.53}$$

where PV denotes the principal value of the integral. We can now view the above equation as a singular integral equation for the unknown spectral density $\rho(\lambda)$. One can easily check, that the solution gives the Wigner semicircle. The easiest way is to expand unknown spectral function in terms of Chebyshev polynomials and use the relation

$$\int_{-1}^{1} \frac{\sqrt{1-y^2}U_{n-1}(y)dy}{x-y} = \pi T_n(x), \tag{1.54}$$

where T_n and U_n are the Chebyshev polynomials of the first and second kind, respectively. Inspection of the formula (1.53) shows, that the r.h.s. is just, modulo rescaling, the first Chebyshev polynomial $T_1(x)$, where $x = \frac{\lambda}{2\sigma}$, so, changing the variables $y = \frac{\lambda'}{2\sigma}$ in (1.54) we immediately recover (1.51). The fact that from the knowledge of real part we can infer the imaginary part, or vice versa, is one of the consequences of the complex analysis, known as dispersive relations (or Kramers-Kronig relations).

1.6.3 Warning

In the process of above derivation, we have assumed that the $1/N$ expansion is everywhere well-defined. Well, let us calculate the derivative $\frac{dG(z)}{dz}$ for Wigner's semicircle, which we have neglected in equation (1.42). The derivative explodes at the endpoints of the spectrum, which strongly suggests the breakdown of the $1/N$ expansion at some isolated points of the spectrum. Indeed, this is the case. In the next section we will solve exactly the GUE for any N, which will allow us later to trace the subtleties of the $1/N$ expansion. **Exercise** Calculate first 10 moments M_k from your numerics, and compare

the result with analytics, i.e. expand $G(z)$ at $z = \infty$ up to 10th term. Then check if the obtained sequence of integers $1, 2, ...$ appear in Sloane's Online Encyclopedia of Integer Sequences (www.oeis.org).

1.7 Dyson's electrostatic analogy (Coulomb gas picture)

From the high school we remember the Coulomb law, stating that the electric field created by the point charge q dies out isotropically with the distance as $1/r^2$, i.e. $E \sim \frac{q}{r^2}$. Alternatively, we may use the concept of electrostatic potential $V(r) \sim -\frac{q}{r}$, whose gradient (radial derivative) gives the electric field. The high school textbooks usually do not specify, that the Coulomb law is the consequence of more general Gauss law, stating that the electric field from the point charge q scales in any dimension d like one over the surface of the sphere in d dimensions. Indeed, for $d = 3$ we recover high-school imprinted Coulomb law, since the surface of the sphere is $4\pi r^2$. But in $d = 2$, the surface of the sphere is just the circumference of the circle, $2\pi r$, so the Coulomb law in 2 dimensions reads $E_2 \sim \frac{q}{r}$, and electrostatic potential in two dimensions reads therefore as $V_2 \sim \ln r$. Looking at (1.39) , we notice that eigenvalues λ_i behave like electric charges of the same sign, confined to the line (real axis), but subjected to the 2-d Coloumbic repulsion. Additionally, each charge is confined by the potential V, in case of the Gaussian ensembles by harmonic potential $\lambda^2/2\sigma^2$. In the large N limit, attractive potential term behaves like $N \times 1/N \sim O(1)$, and is counterbalanced by the repulsive electrostatic potential scaling like $N^2 \times 1/N^2 \sim O(1)$. The saddle point solution is therefore realizing the equilibrium between the minimization of the energy terms (potential) and the maximization of the disorder by the repulsive electrostatic terms. For each particular form of the potential, the resulting equilibrium density of charges will take the different form. In the case of the Gaussian potential, the density of charges gives the Wigner semicircle. For more complicated potential, e.g. for $V(\lambda) = \frac{\lambda^2}{2\sigma^2} - \frac{g_4}{4}\lambda^2$ the resulting equilibrium can take other shapes (including the spectra located on two disconnected intervals), depending on the values of the parameters of the potential. Dyson electrostatic picture is a powerful analytic tool, with several application for so-called statistics of large deviations It allows also to formulate spectral law for the random matrices with complex eigenvalues.

We will address both these topics in the later parts of these lectures.

Chapter 2

Microscopic universality in the GUE case

2.1 Orthogonal polynomial method

Previous chapters gave us the gist of the spectral questions in random matrix theory. We will try now to systematize the issue of spectral relations, introducing general observable, *symmetric* function $F(\lambda_1, ..., \lambda_N)$, and its expectation value, with they respect of jpdf function $p_N(\lambda_1, ..., \lambda_N)$, i.e.

$$< F >\equiv \int ... \int F(\lambda_1, ...\lambda_N) p_N(\lambda_1, ..., \lambda_N) d\lambda_1 ... d\lambda_N \tag{2.1}$$

Here $p(\lambda_1, \lambda_2, ..., \lambda_N)d\lambda_1 ... d\lambda_N$ is the probability of finding one eigenvalue at interval $[\lambda_1, \lambda_1 + d\lambda_1]$, another at interval $[\lambda_2, \lambda_2 + d\lambda_2]$ etc, regardless of the labeling used for the ordered set. Note that the Green's function was already an example of such object as $< F >$, belonging to so-called *linear statistics* case. Three classes of linear statistics observables are of primary importance:

1. *Correlation functions:*
   Introduced by Dyson, correlation functions R_k are defined as

   $$R_k(\lambda_1, ...\lambda_k) = \frac{N!}{(N-k)!} \int ... \int p_N(\lambda_1, ..., \lambda_k, ...\lambda_N) d\lambda_{k+1} ... d\lambda_N \tag{2.2}$$

   Modulo the normalization, they provide the probability of finding K eigenvalues, irrespectively of their order, in the infinitesimal vicinity of the set of eigenvalues $\lambda_1, ...\lambda_k$.

2. *Clusters:*
One may ask the question about the probability of finding the fixed k-tuple of eigenvalues, i.e. n-point correlation, from which we want to eliminate the probability of finding $1, ...k-1$-tuples of eigenvalues. Clusters T_k are defined as follows

$$T_k(\lambda_1, ...\lambda_k) = \sum (-1)^{k-m}(m-1)! R_{S_1}...R_{S_M} \tag{2.3}$$

where the sum runs over all partitions of $(1, ...k)$, and S_m denote non-empty subsets corresponding to these partitions. Explicitly,

$$\begin{aligned} &T_1(\lambda_1) = R_1(\lambda_1) \\ &T_2(\lambda_1, \lambda_2) = -R_2(\lambda_1, \lambda_2) + R_1(\lambda_1)R_1(\lambda_2) \\ &T_3(\lambda_1, \lambda_2, \lambda_3) = R_3(\lambda_1, \lambda_2, \lambda_3) - R_2(\lambda_1, \lambda_2)R_1(\lambda_3) - R_2(\lambda_2, \lambda_3)R_1(\lambda_1) - \\ &R_2(\lambda_1, \lambda_3)R_1(\lambda_2) + 2R_1(\lambda_1)R_1(\lambda_2)R_1(\lambda_3) \end{aligned} \tag{2.4}$$

3. *Gaps:*
The third class answers to the question, what is the probability of finding precisely n_i eigenvalues in corresponding I_i intervals of the spectrum, denoted as $P(n_1, ..., n_m; I_1, ...I_m)$. In particular, $P(0, I)$ gives the probability of having no eigenvalues in the gap (interval) I. Wigner's surmise was an answer to this question in the case of Gaussian ensembles with $N = 2$, but now we hope to find the answer for any N.

2.1.1 Case $\beta = 2$

In the case $\beta = 2$, one can, (at least formally), particularly easily provide the general answer for correlations, clusters and gap functions. The procedure is based on ingenious observation, that the Jacobian $\prod_{i<j}(\lambda_i - \lambda_j)^2$ is a square of the Van der Monde determinant, i.e. modulo the irrelevant sign, it reads:

$$\begin{aligned} \prod_{i<j}(\lambda_i - \lambda_j) &= (-1)^{\frac{N(N-1)}{2}} \det \begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ \lambda_1 & \lambda_2 & ... & \lambda_{N-1} & \lambda_N \\ \lambda_1^2 & \lambda_2^2 & ... & \lambda_{N-1}^2 & \lambda_N^2 \\ ... & ... & ... & ... & ... \\ \lambda_1^{N-1} & \lambda_2^{N-1} & ... & \lambda_{N-1}^{N-1} & \lambda_N^{N-1} \end{pmatrix} \\ &\equiv \Delta_N(\lambda_1, ...\Lambda_N) \end{aligned} \tag{2.5}$$

Let us now introduce the generic polynomial $P_k(\lambda) = a_k\lambda^k + a_{k-1}\lambda^{k-1} + ...a_0$. Note that (2.6) can be rewritten as

$$\Delta(\lambda_1, ..., \lambda_N) = \frac{(-1)^{N(N-1)/2}}{a_0 a_1 .... a_{N-1}} \det \begin{pmatrix} P_0(\lambda_1) & ... & P_0(\lambda_N) \\ ... & ... & ... \\ P_{N-1}(\lambda_1) & ... & P_{N-1}(\lambda_N) \end{pmatrix} \tag{2.6}$$

Indeed, let us first multiply and divide first row by a_0, second row by a_1, etc, and finally the last row by a_{N-1}. Now we can add first row to the second one, than first and second to the third one, etc, and finally all rows from the first till the before last to the last one. Collecting remaining from division factors $a_0 a_1 ... a_{N-1}$ in front of the determinant yields the final formula. Note that till know we have not specified the type of polynomials $P_k(\lambda)$. Note, that in the previous chapter we have obtained, using Selberg integral for $\beta = 2$, the generic form

$$p_N(\lambda_1, ..., \lambda_N) = \frac{1}{N! Q_N^{(2)}} e^{-N\sum_i V(\lambda_i)} \Delta_N^2 \tag{2.7}$$

Note additional $N!$ from symmetrization, so let us introduce $q_N^{(2)} \equiv Q_N^{(2)} N!$. Let us absorb now prefactors $e^{-NV_i(\lambda_i)}$ into determinants, in a following way: we multiply first column of each determinant by $e^{-NV(\lambda_1)/2}$, second by $e^{-NV(\lambda_2)/2}$, etc., finally the last one by $e^{-NV(\lambda_N)/2}$. Introducing the notion $\phi_{i-1}(\lambda_j) \equiv e^{-NV(\lambda_j)/2} P_{i-1}(\lambda_j)$ we may rewrite the final formula for the jpd as

$$p_N(\lambda_1, ...\lambda_N) = \frac{1}{q_N^{(2)}} \frac{1}{[\prod_{i=0}^N a_i]^2} [\det \phi_{i-1}(\lambda_k)|_{i,j=1,..N}]^2 \tag{2.8}$$

It is obvious now, that the optimal choice of the polynomials depends of the shape of the potential $V(\lambda)$. The largest simplifications occur if we choose functions $\phi_{i-1}(\lambda_j)$ to be orthonormal, i.e., equivalently, polynomials P_k orthonormal with respect to the weight function $w(\lambda) = e^{-NV(\lambda)}$. At the last step of our construction we fix the normalization prefactors in (2.8). We use another famous integral formula, known as **Andreief identity**:

$$\int \det[f_j(\lambda_k)]_{j,k=1,N} \det[g_j(\lambda_k)]|_{j,k=1,N} d\lambda_1 ... d\lambda_N = N! \det[\int d\lambda f_j(\lambda) g_k(\lambda)]_{j,k=1,..,N} \tag{2.9}$$

Substituting $f_j(\lambda_k) = g_j(\lambda_k) = \phi = \phi_{j-1}(\lambda_k)$ and integrating over all eigenvalues and using the orthonormality condition allows us to fix $q_N^{(2)}[\prod a_i]^2 = N!$, so we get

$$p_N(\lambda_1, ..., \lambda_N) = \frac{1}{N!}\det[A_{ij}]^2 \tag{2.10}$$

where $\phi_{i-1}(\lambda_j) \equiv A_{ij}$. Now, since

$$[\det A]^2 = \det A \, \det A = \det A^T \, \det A = \det A^T A = \det(\sum_{j=1}^{N} A_{ji} A_{jk}) \tag{2.11}$$

In such a way we have obtained the final formula for the jpdf in the form of universal two-point function named *kernel*

$$\begin{aligned} p_N(\lambda_1, ...\lambda_N) &= \frac{1}{N!}\det[K_N(\lambda_i, \lambda_k)]|_{i,k=1,...,N} \\ K_N(\lambda_i, \lambda_j) &= e^{-\frac{N}{2}(V(\lambda_i)+V(\lambda_j))} \sum_{k=0}^{N-1} P_k(\lambda_i) P_k(\lambda_j) \end{aligned} \tag{2.12}$$

where we have explicitly rewritten the kernel in terms of the weights and corresponding orthogonal polynomials.

2.1.2 Properties of the kernel

The kernel function has remarkable properties:

- Normalization condition

$$\int K_N(\lambda, \lambda) d\lambda = N \tag{2.13}$$

- Basic chain property

$$\int K_N(\lambda, \mu) K_N(\mu, \lambda') = K_N(\lambda, \lambda') \tag{2.14}$$

- Self-reproducing property

$$\begin{aligned} &\int ... \int \det K_N(\lambda_i, \lambda_j)|_{i,j=1,...N} d\lambda_{k+1} ... d\lambda_N = \\ &(N-k)! \det K_N(\lambda_i, \lambda_k)|_{i,j=1,...k} \end{aligned} \tag{2.15}$$

The first two relations are obvious, due to the orthonormality of the polynomials. The last one also comes from orthonormality of the polynomials, the prefactor counting non-vanishing permutations from the definition of the original determinant.

The last equation allows to write the generic Dyson correlation function in a particularly simple form

$$R_k(\lambda_1, ..., \lambda_k) = \det[K_N(\lambda_i, \lambda_j)]|_{i,j=1,...,k} \tag{2.16}$$

In particular, the average spectral distribution reads

$$\rho(\lambda) = K_N(\lambda, \lambda) = e^{-NV(\lambda)} \sum_{j=0}^{N-1} [P_j(\lambda)]^2 \tag{2.17}$$

Also all cluster correlation functions are, by definition, expressible in terms of correlation functions R, therefore, in terms of the kernels, e.g.

$$T_2(\lambda_1, \lambda_2) = [K_N(\lambda_1, \lambda_2)]^2 \tag{2.18}$$

The same feature holds for the gap probability

$$P(0, I) = \int_{-I/2}^{I/2} ... \int_{-I/2}^{I/2} \prod_{i=1} (1 - \theta_I(\lambda_k)) P_N(\lambda_1, ..., \lambda_N) d\lambda_1 ... d\lambda_N =$$
$$\sum_{j=0}^{N} \frac{(-1)^j}{j!} \int ... \int \det \begin{pmatrix} K_N(\lambda_1, \lambda_1) & ... & K_N(\lambda_1, \lambda_j) \\ \vdots & & \vdots \\ K_N(\lambda_j, \lambda_1) & ... & K_N(\lambda_j, \lambda_j) \end{pmatrix} d\lambda_1 ... d\lambda_j \tag{2.19}$$

where θ_I is equal to 1 if the eigenvalue belongs to the interval I or is equal to zero otherwise, and the r.h.s. of the above formula is the result of simple combinatorics resulting from collecting all terms in the product of $(1 - \theta_i)$ terms.

2.1.3 Christoffel-Darboux formula - further simplification of the kernel

In the case when orthogonal polynomials fulfill three step recurrence relation

$$\lambda P_l(\lambda) = a_{l+1} P_{l+1}(\lambda) + b_l P_l(\lambda) - a_l P_{l-1}(\lambda) \tag{2.20}$$

one can considerably simplify the kernel. Let us multiply (2.20) by $P_l(\mu)$, obtaining

$$\lambda P_l(\lambda)P_l(\mu) = a_{l+1}P_{l+1}(\lambda)P_l(\mu) + b_l P_l(\lambda)P_l(\mu) - a_l P_{l-1}(\lambda)P_l(\mu) \quad (2.21)$$

Now, we copy the above equation renaming $\lambda \leftrightarrow \mu$, and we subtract the renamed copy from (2.21). We obtain

$$(\lambda - \mu)P_l(\lambda)P_l(\mu) = B_{l+1} - B_l \quad (2.22)$$

where $B_l \equiv a_l[P_{l-1}(\lambda)P_l(\mu) - P_{l-1}(\mu)P_l(\lambda)]$. Now, we sum (2.22) from $l = 1$ to $N - 1$, obtaining

$$(\lambda - \mu)\sum_{l=1}^{N-1} P_l(\lambda)P_l(\mu) = B_N - B_1 \quad (2.23)$$

since, due to the alternating signs in the subsequent terms only B_n and B_1 survive the cancellations in the sum. Observing that $P_0 = 1$ and $P_1(x) = x$, we can rewrite the sum as

$$\sum_{l=0}^{N-1} P_l(\lambda)P_l(\mu) = \frac{B_N}{\lambda - \mu} \quad (2.24)$$

Above equation, known as **Christoffel-Darboux relation**, allows to simplify the kernel

$$K_N(\lambda_i, \lambda_j) = a_N e^{-\frac{N}{2}(V(\lambda_i)+V(\lambda_j))}\frac{P_{N-1}(\lambda_i)P_N(\lambda_j) - P_{N-1}(\lambda_j)P_N(\lambda_i)}{\lambda_i - \lambda_j} \quad (2.25)$$

2.1.4 Example: Gaussian Unitary Ensemble

In the case of GUE, we seek the polynomials which are defined on the real axis and are orthogonal with respect to the weight $w(\lambda) = e^{-NV(\lambda)} = e^{-\frac{N\lambda^2}{2\sigma^2}}$. The answer is given by the well known in physics and statistics Hermite polynomials [1]. In our case, taking into account the proper normalization, explicit formulae read

$$\begin{aligned} P_k^{(N)}(\lambda) &= \left(\frac{N}{2\sigma^2}\right)^{1/4} h_l\left[\sqrt{\frac{N}{2\sigma^2}}\lambda\right] \\ a_k &= \sigma\sqrt{\frac{k}{N}} \end{aligned} \quad (2.26)$$

[1] Note that the physicists and the statisticians use usually different conventions for Hermite polynomials.

with the convention $\int_{\infty}^{\infty} e^{-x^2} h_i(x) h_j(x) dx = \delta_{ij}$.
Exercise: Using the above listed explicit representations for the weight and for orthogonal polynomials, plot the average spectral function $\rho(\lambda)$ represented by (2.17) for $N = 1, 2, 5, 10$ and 20. Observe how the subsequent curves approximate better and better the Wigner's semicircle.

2.1.5 GOE ($\beta = 1$) and GSpE ($\beta = 4$)

In the case of the Gaussian Orthogonal Ensemble the Jacobian reads Δ_N and not Δ_N^2, so the trick based on orthogonality of polynomials does not work any longer. In this case, one uses another trick, introducing the Pfaffian, which can be viewed as a "square root" of determinant. Related trick works also for $\beta = 4$ case, i.e. the Gaussian Symplectic Ensemble. The final formulae for the kernels are much more involved comparing to the presented here $\beta = 2$ case. We refer the interested Reader to the rich literature on this subject.

2.1.6 Subtleties of the large N expansion

We would like to stress, that the construction of the kernel allowed us to express all the observables of the random matrix model in terms of the single two-point function (kernel), where the dependence on the dimension of the matrix appears only as an *index*. Nevertheless, in the next section, we will see that large N limit has its subtleties and has to be performed with particular care.

2.1.7 Explicit form of the kernel

Correlations of eigenvalues probed on the scale of the typical separation between them are independent on the probability density function of matrix elements. They fall into several classes, depending on the point x_0 of the spectrum at which their behavior is probed. The shape of the spectral density, in turn, determines the microscopic scale s by demanding that in the interval $[x_0, x_0 + s]$ one expects one eigenvalue to occur. Looking at the form of the Wigner semicircle, we immediately identify two distinct regions corresponding to microscopic scalings.

1. **Bulk.** We ask how many out of original N eigenvalues will appear in a narrow bin of size s around some x_0, lying between the endpoints of

the spectrum. Since the interval s is very narrow, one can approximate the density as locally constant $\rho(x_0)$. This leads to

$$n_{bulk} \sim N \int_{x_0-s/2}^{x_0+s/2} \rho(x)dx \sim N \int_{x_0-s/2}^{x_0+s/2} \rho(x_0)dx \sim Ns\rho(x_0), \quad (2.27)$$

which implies that the bulk microscopic scale is $s \sim \frac{1}{N\rho(x_0)}$.

2. **Soft edge.** At any of the two endpoints, the scaling with N is different, since the macroscopic spectral density around both endpoints $x = \pm 2$ vanishes like $\sqrt{x}$. Counting the eigenvalues close to the edge (let us say, near $x = 2$), leads to

$$n_{soft} \sim N \int_0^s \sqrt{x}dx \sim Ns^{3/2}, \quad (2.28)$$

thus the edge microscopic scale is set to $s \sim N^{-2/3}$.

In order to infer the asymptotic form of the kernel $K_N(\lambda, \mu)$ we have to perform a rather subtle large N limit, since not only the order of the polynomial depends on N, but also its argument. This is so-called Plancherel-Rotach limit - the explicit fomulae, obtained by saddle point approximation, are elegantly explained e.g. in Ian Fyodorov lectures.

In our lectures, we will follow, however, the different scenario, borrowing form the generic arguments by Borodin and Olshanski [**?**] and inspired by Tao [**?**] presentation for the Gaussian Unitary Ensemble.

First, we recall that the joint probability distribution for GUE we interpreted as a square of the Slater determinant, built from first N one-particle "wavefunctions" $\phi_l(\lambda)$. The fact that the resulting orthogonal polynomials in these wavefunctions are of Hermite type, immediately brings to every physicist a connotation with quantum harmonic oscillator. Let us recall the quantum harmonic oscillator, given by the Hamiltonian $\hat{H} = \frac{p^2}{2m} + \frac{1}{2}kx^2$, where the momentum operator $p = -i\hbar\frac{d}{dx}$. In order to simplify our analysis, let us put $m = k = \frac{1}{2}$ and the Planck constant $\hbar = 1$. Then, the resulting Schroedinger equation reads (as every textbook in quantum mechanics shows)

$$[-\frac{d^2}{dx^2} + \frac{1}{4}x^2]\phi_l(x) = E_l\phi_l(x) \quad (2.29)$$

with well known solutions

$$E_l = l + 1/2$$
$$\phi(l) = \gamma_l \left(\frac{1}{2}\right)^{1/4} e^{-x^2/4} h_l(x/\sqrt{2}) \tag{2.30}$$

where $\gamma_l = (\sqrt{\pi} 2^l l)^{-1/2}$, and $l = 0, 1, 2, 3, ...$ Note that modulo additional rescaling of the argument by $x = \lambda/\sqrt{N}$, this is exactly the form or orthogonal "wavefunctions" in GUE.

Second, we make another important link to quantum mechanics. Note, that from the point of view of quantum mechanics, the elements kernel operator $\hat{K}_N$ are just $< x|\hat{K}_N|y >= \sum_{l=0}^{N-1} | < x|\phi_l >< \phi_l|y >$, so the kernel operator is a projection operator on first N states of harmonic oscillator. Indeed, due to the orthogonality of the wave functions, $\hat{K_N}^2 = \hat{K}_N$, or, equivalently, we recover chain relation $\int K_N(x,y) K_N(y,z) dy = K_N(x,z)$. If N goes to infinity, kernel operator approaches the unit operator, due to the completeness of the states of the harmonic oscillator. However, if simultaneously we will scale with N the arguments of the kernel, $< x(N)|\phi_l >= \phi_l(x(N))$, our unit operator will get deformed. This is precisely this deformation which yields the explicit form of the kernel, as we show now. The complete set of eigenfunctions provides a resolution of identity $\mathbf{1} = \sum_{l=0}^{\infty} |\phi_l >< \phi_l|$. The range of the projection onto the first N states can be formally written as $\hat{H} \leq E_N$ or, using the explicit form of the Schroedinger equation with identification

$$-\frac{d^2}{dx^2} + \frac{1}{4}x^2 \leq N + 1/2 \tag{2.31}$$

- **Sine kernel.** In the bulk, $x/\sqrt{N} = x_0 + \frac{s}{N\rho(x_0)}$ (note that we made the rescaling of the l.h.s. in order to absorb the N factor in the exponent in the GUE measure), so $dx = \frac{ds}{\sqrt{N}\rho(x,0)}$. In the large N limit, under the change of the variables $x \to s$, the operator bound on the Schroedinger equation 3.17 reads simply

$$-N\rho(x_0)^2 \frac{d^2}{ds^2} + 1/4(Nx_0^2 + O(1/N)) \leq N \tag{2.32}$$

Using the explicit form of the Wigner semicircle density, the above bound is simplified to

$$-\frac{d^2}{ds^2} \leq \pi^2. \tag{2.33}$$

On the l.h.s. we recognize the Schrödinger operator for a free particle, therefore the natural procedure for resolving this bound is to use plane waves, i.e. move to the momentum space via the Fourier transformation:

$$\begin{aligned} F(q) &= \int_{-\infty}^{\infty} e^{2\pi i t q} f(t) dt, \\ f(t) &= \int_{-\infty}^{\infty} e^{-2\pi i t q} F(q) dq. \end{aligned} \tag{2.34}$$

The spectral deformation in the momentum space reads therefore

$$q^2 \leq \frac{\pi^2}{(2\pi)^2} = \frac{1}{4}. \tag{2.35}$$

Combination of Fourier transforms provides a representation of an identity operator

$$f(t') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2\pi i t' q} e^{2\pi i t q} f(t) dt dq. \tag{2.36}$$

The deformation (2.35) projects the above identity operator onto

$$\mathbf{P}[f(t')] = \int_{-\infty}^{\infty} \left[\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i t' q} e^{2\pi i t q} dq \right] f(t) dt, \tag{2.37}$$

Microscopic scaling in the bulk restricts the range of momenta to the interval $-\frac{1}{2} \leq q \leq \frac{1}{2}$. Calculation of the integral in square brackets yields the projection in the position basis, which is the sine kernel

$$K_{Sine}(t, t') = \frac{\sin(\pi(t' - t))}{\pi(t' - t)}. \tag{2.38}$$

- **Airy kernel.** At the soft edge (e.g. at the right one) we introduce the scaling variable s as $x/\sqrt{N} = 2 + sN^{2/3}$. Now, $dx = ds/N^{1/6}$, and in the large N limit generic bound (3.17) on the Schoredinger equation of harmonic oscillator is transformed into

$$-\frac{d^2}{ds^2} + s \leq 0. \tag{2.39}$$

On the l.h.s. we recognize the Schrödinger operator with the linear potential. Knowing, that the quantum mechanical solution of this problem is given in by the Airy function, we identify the differential Airy operator in 3.21, and we directly resort to *the Airy transform* [**?**]

$$F(z) = A[f(t)] = \int_{-\infty}^{\infty} Ai(z-t)f(t)dt \tag{2.40}$$

and its inverse

$$f(t) = \int_{\infty}^{\infty} F(z)Ai(z-t)dz. \tag{2.41}$$

Using the Airy transform for the operator bound (3.21), and the fact that Airy function fullfils $Ai''(x) = xAi(x)$ we express the spectral deformation in dual variable t simply as

$$z \leq 0. \tag{2.42}$$

Combining both Airy transforms we obtain the identity operator

$$f(t') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Ai(t'-z)Ai(t-z)f(t)dtdz \tag{2.43}$$

The deformation condition (2.42) turns the above identity operator into a projection

$$\mathbf{P}[f(t')] = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{0} Ai(t'-z)Ai(t-z)dz\right]f(t)dt \tag{2.44}$$

so the kernel reads

$$\begin{aligned} K_{Airy}(t,t') &= \int_{-\infty}^{0} Ai(t'-z)Ai(t-z)dz \\ &= \frac{Ai(t')Ai'(t) - Ai'(t')Ai(t)}{t'-t}, \end{aligned} \tag{2.45}$$

where on the r.h.s. we presented the more familiar form of the Airy kernel based on relation

$$\left[\frac{Ai(s-z)Ai'(t-z) - Ai'(s-z)Ai(t-z)}{s-t}\right]' = Ai(s-z)Ai(t-z) \tag{2.46}$$

where prime denotes differentiation with respect to z.

2.1.8 Semicircle from Bohr-Sommerfeld quantization condition

Before the formulation of Schroedinger equation, so-called "old quantum theory" made use of semi-classical picture, based on Bohr-Sommerfeld quantization condition. In its original form, Bohr-Sommerfeld conditiion reads

$$\oint p dx = 2\pi\hbar N \tag{2.47}$$

where the integration is over the cycle between the classical turning points (x_-, x_+). Upon the identification of $1/\hbar \leftrightarrow N$, and using the fact that $E \sim N\hbar = 1$ and classical $p = \sqrt{E - x^2/4}$, so the turning points corresponding to zero momentum read $x = \pm 2$, we see

$$\frac{1}{2\pi}\oint p dx = \frac{1}{2\pi}\int_{-2}^{2}\sqrt{4-x^2} = 1 \tag{2.48}$$

so Wigner's semicircle (actually semi-ellipse) stems from the quantization condition of classical elliptical orbits $p^2 + \frac{1}{4}x^2 = 1$. It is not so surprising, if we recall that originally, we got Wigner's semicircle as a result of saddle point approximation for 2-dimensional Coulomb gas in the large N limit, and Bohr-Sommerfeld quantization condition comes as well from the saddle point approximation, where the role of $N \to \infty$ is played by the inverse of the Planck constant $\hbar$ in the semiclasical limit $\hbar \to 0$.

Chapter 3

Wishart ensemble

Wishart ensembles represent a particularly important domain of random matrices, with multiple fundamental applications ranging from multivariate statistics, through wireless telecommunication to computer simulations of Quantum Chromodynamics - contemporary theory of strong interactions. Wishart ensemble is also, historically, the first direct construction of the random matrix ensemble.

In the case of classical probability, the distribution of the sum $Q = \sum_{t=1}^{T} x_t^2$, i.e. the sum of squares of the standard normal distributions, plays a crucial role in testing the hypothesis. ..... This distribution bears the name of χ_T^2 with T degrees of freedom and its pdf reads

$$p_T(x) = \frac{1}{2^{T/2}\Gamma(T/2)} x^{T/2-1} e^{-x/2} \quad \text{where } x > 0 \tag{3.1}$$

In 1928, John Wishart posed the question, what is the analog of the χ^2 distribution, if instead of single measurement x_t we perform a multivariate measurement $\vec{x}_t$, where $\vec{x}$ is an N dimensional vector. The single measurement can be therefore represented as an element of the matrix X_{it}, where $i = 1, ..., N$ and $t = 1, ..., T$. The (real) Wishart distribution is therefore the distribution of the $N \times N$ matrix $M = XX^T$, where each element is drawn from the standard Gaussian distribution. In the case when $N \leq T$, after elaborate calculations, Wishart obtained the probability density

$$P^{(1)}(M)dM = Q_T^{(1)} e^{-\mathrm{Tr}M/2} \mathrm{det} M^{(T-N-1)/2} dM \tag{3.2}$$

where $Q_T^{(1)}$ is the normalization constant. Comparing above formulae we see, that indeed, Wishart distribution is the generalization of the χ^2 distribution

for the case $N > 1$. Similar formula can be obtained in the case, when elements of X are complex Gaussians, with $< |X_{it}|^2 >= 1$. Then $M = XX^\dagger$ is distributed accordingly

$$P^{(2)}(M)dM = Q_T^{(2)} e^{-\mathrm{Tr}M} \mathrm{det} M^{T-N} dM \tag{3.3}$$

Today, the simplest way to obtain Wishart distributions is to perform the integration $\int \delta(M - XX^\dagger)P(X)dX$, where

$$P_\beta(X)d_\beta X = \frac{1}{Z_{T,N}^{(\beta)}} e^{-\beta \mathrm{tr} XX^\dagger/2} d_\beta X \tag{3.4}$$

where $\beta = 1, 2$ and $d_1 X = \prod_{i,t}^{N,T} dX_{it}$ and $d_2 X = \prod_{i,t}^{N,T} d\Re X_{it} d\Im X_{it}$. Let us show how this works in the case when $\beta = 2$, since we will be studying this case in more detail in the following chapters. First, we "open" the matrix-valued Dirac's delta with the help of the integral representation

$$\delta(M - XX^\dagger) = \frac{1}{(2\pi)^{N^2}} \int d\mathcal{F} e^{i\mathrm{tr}\mathcal{F}(M - XX^\dagger)} \tag{3.5}$$

where $\mathcal{F}$ is an $N \times N$ hermitian matrix. Gaussian integration over X yields, up to a constant,

$$P^{(2)}(M) \sim \int d\mathcal{F} \mathrm{det}(1 + i\mathcal{F})^{-T} e^{i\mathrm{tr}\mathcal{F}M} = const\ (\mathrm{det}M)^{T-N} e^{-\mathrm{tr}M} \tag{3.6}$$

This last integration was done with the help of powerful Ingham-Siegel integral (see the appendix), the subsequent example from the list of our fundamental integrals in random matrix theories.

The case when $N > T$ is more difficult. In such case, the elements of matrix M are not independent, there is a large redundancy, due to the repetition of rows and columns during the multiplication process $XX^\dagger$. The corresponding density has to include this redundancy in terms of constraints. The analogue of Wishart formulae in this case can be written as well for real and complex cases, but are more complicated. Sometimes this case is named as anti-Wishart distribution.

Let us move now towards the spectral distribution for the Wishart case. Conventionally, one considers the matrix $C \equiv \frac{1}{T} XX^\dagger$, which represents correlations between entries i and j averaged over the time series of length T. The jpdf for matrix C reads therefore

$$P^{(\beta)}(\lambda_1, ...\lambda_N) = C_{N,T}^{(\beta)} \prod \lambda_i^{(T-N+1)\beta/2+1} e^{-T\sum_i \lambda_i/2} |\Delta_N|^\beta \tag{3.7}$$

Ignoring the constant, the first term and second term come from the fact, that the determinant and the trace operations are invariant under diagonalization, the last one is the price (Jacobian) for switching from elements to eigenvalues. The Jacobian is identical to the presented before construction for GOE/GUE ensembles, i.e. is given by appropriate power of the Van der Monde determinate. The only difference is that now the eigenvalues cannot be negative, since C is a positive operator. This observation allows us to write down immediately the jpdf for complex Wishart as a square of the Slater determinant, alike we did in the case of the GUE.

$$P_N(\lambda_1, ..., \lambda_N) = \frac{1}{N!} \left(\det \left[\psi_{j-1}^{(N)}(\lambda_k) \right]_{j,k=1}^{N} \right)^2 = \frac{1}{N!} \left[\det K_N(\lambda_i, \lambda_j) \right] \quad (3.8)$$

with the correlation *kernel*

$$K_N(\lambda, \mu) = \sum_{l=0}^{N-1} \psi_l^{(N)}(\lambda) \psi_l^{(N)}(\mu) \quad (3.9)$$

where $\psi_l^{(N)}(\lambda) = e^{-T\lambda/2} \lambda^{\alpha/2} P_l^{(N)}(\lambda)$, where $\alpha = T - N$. This form again suggests links to Quantum Mechanics. Moreover, we see that the most natural choice of polynomials is dictated by the weight $w_{N,T}(\lambda) = \lambda^{\alpha} e^{-T\lambda}$. Such polynomials, orthonormal on the positive part of the real axis, are the associated Laguerre polynomials. This observation suggests the link to hydrogen atom problem, where very similar associated Laguerre functions appear in the radial part of the Schroedinger equation. Indeed, the radial part of the Schroedinger equation reads

$$\frac{d^2 y(r)}{dr^2} + \left[\frac{2\mu e^2}{r\hbar^2} - \frac{l(l+1)}{r^2} \right] y(r) = -\frac{2\mu E}{\hbar^2} \quad (3.10)$$

where $y(r) = rR(r)$ in the standard notation. Switching to dimensionless variable $x = r\epsilon$, where $(\epsilon/2)^2 = -2\mu E/\hbar^2$ and, finally, putting all the physical constants to 1 with the convention $2\mu = 1$, we recover [**?**]

$$\frac{d^2 y(x)}{dx^2} + \left[-\frac{1}{4} + \frac{1}{\epsilon x} - \frac{l(l+1)}{x^2} \right] y(x) = 0 \quad (3.11)$$

where $y = y_n^l = e^{-x/2} x^{(k+1)/2} L_j^k(x)$, where $l(l+1) = \frac{k^2-1}{4}$, hence $k = 2l + 1$ and the principal quantum number reads $n = j + l + 1$. Note, that $\epsilon = 1/N$, or, equivalently, $E = -1/4N^2$, since in our units Bohr 's radius equals to

2. To map random matrix problem to the hydrogen atom, we notice, that $\psi_l(\lambda) = \sqrt{x}y(x)$. This completes the dictionary between hydrogen atom problem and the Wishart kernel.

The final formula for ψ expressed in terms of the parameters of the Wishart ensemble reads

$$\frac{d^2\psi_k}{dx^2} + \frac{1}{x}\frac{d\psi_k}{dx} + \frac{1+2k+\alpha}{2x}\psi_k - \frac{\alpha^2}{4x^2}\psi_k = \frac{1}{4}\psi_k \tag{3.12}$$

Next, we would like to see the analog of Wigner semicircle for the complex Wishart ensemble. There are several ways to get this result. First, we may repeat the Coulomb gas calculation based on saddle point expansion. We can also use the shortcut, using the insight from quantum mechanics and apply semiclassical methods. In this case we will exploit again the semiclassical limit, alike we did in the GUE case.

To have the finite support of the spectral density in the large N limit, we rescale the x variable as xT, We notice that in the limit $N, T \to \infty$, the kernel is equivalent to the semiclassical projection of the region of the phase space (p, x) in the limit when radial momentum operator p tends to zero (in analogy to $\hbar \leftrightarrow 1/T$). The quantization rule is then expressed as the Bohr-Sommerfeld quantization condition,

$$T \oint p(xT)dx = (N + \frac{1}{2})2\pi \tag{3.13}$$

or equivalently

$$\int_{r_-}^{r_+} \rho(x)dx = 1 \tag{3.14}$$

with the resulting spectral deformation, obtained from the WKB condition $p(x) = \sqrt{2\mu}\sqrt{E - V_{eff}}$

$$\rho(x) = \frac{1}{2\pi c x}\sqrt{(r_+ - x)(x - r_-)} \tag{3.15}$$

Here $c \equiv \lim_{N,T\to\infty} \frac{N}{T}$ and $r_\pm = (1 \pm \sqrt{c})^2$ are classical returns points in WKB approximation and $V_{eff} = -\frac{1}{r} + \frac{l(l+1)}{r^2}$ We have therefore obtained Marchenko-Pastur distribution (3.15) as an *exact, semiclassical limit of the quantum mechanical hydrogen atom problem.* In the sections devoted to free variables, we will re-obtained Marcenko-Pastur formula in a rigorous way, interpreting it as a free Poisson distribution.

3.1 Microscopic scaling as a spectral deformation

Alike in the GUE case, we can now look at the microscopic scaling. In the case of the bulk, and the edges, which develop square root behavior, repetition of the GUE procedure yields again sine an Airy kernels, respectively. However, when $r_- \to 0$, we encounter new scaling in the Marcenko-Pastur distribution, so- called hard edge, in distinction to the soft edge, described by the Airy kernel. Indeed, In the limit when $c \to 1$, return point r_- approaches the zero, and the MP distribution behaves like $1/\sqrt{x}$. If we now ask, how many out of original N eigenvalues will appear in a narrow bin s around zero, we get

$$n_{hard} \sim N \int_0^s \frac{dx}{\sqrt{x}} \sim N\sqrt{s} \tag{3.16}$$

which implies that the mean level spacing of order 1 in this region implies the scaling $s \sim 1/N^2$.

We can now approach the problem of spectral deformations. For finite N, the spectral bound of the Schroedinger operator can be viewed formally as $\hat{H} \leq E_N$, or, using the explicit form of the equation (3.12), as

$$\frac{d^2}{dx^2} + \frac{1}{x}\frac{d}{dx} + \frac{1+2k+\alpha}{2x} - \frac{\alpha^2}{4x^2} \geq \frac{1}{4} \tag{3.17}$$

The bulk and soft edge scalings recover the sine and Airy kernel, as expected. For completeness, we repeat beneath the reasoning. However, the hard edge, yields a new scaling law.

1. **Sine kernel.** Using the bulk scaling $x/T = x_0 + \frac{s}{N\rho(x_0)}$, and performing the large N, T limit we obtain from (3.17) the relation

$$\frac{d^2}{ds^2} \geq \frac{(x-r_+)(x-r_-)}{4c^2x^2\rho^2(x_0)} \tag{3.18}$$

Using the explicit form of $\rho(x_0)$, i.e. the Marcenko-Pastur spectral density at x_0, the above bound reads simply

$$-\frac{d^2}{dz^2} \leq \pi^2 \tag{3.19}$$

so the spectral bound identical to the GUE case.

2. **Airy kernel.** Using the soft edge scaling $x/T = r_\pm + sN^{-2/3}$ (rescaling by $1/T$ absorbs the T in the exponent of the weight $w(\lambda)$) and performing the large N and T limit we obtain the equation from generic bound (3.17)

$$c^{3/2} r_\pm^2 \frac{d^2}{ds^2} - s \geq 0 \tag{3.20}$$

Changing variables $z = [c^{3/2} r_\pm^2]^{-1/3}$ converts the above bound onto more familiar form

$$-\frac{d^2}{dz^2} + z \leq 0 \tag{3.21}$$

i.e. the case again identical to the soft edge of the GUE.

3. **Bessel kernel.** Using the hard edge scaling $x/T \to sN^2$, and performing the large N limit (note that $k \sim N$), we obtain the equation

$$\frac{d^2}{ds^2} + \frac{1}{s}\frac{d}{ds} + \frac{1}{s} - \frac{\alpha^2}{4s^2} \geq 0 \tag{3.22}$$

Changing variables $z = 2\sqrt{s}$ converts the above bound onto more familiar form

$$\Delta_\alpha \equiv -\frac{d^2}{dz^2} - \frac{1}{z}\frac{d}{dz} - \frac{\alpha^2}{z^2} \leq 1 \tag{3.23}$$

where on the l.h.s. we recognize Bessel operator, appearing in quantum mechanical problems with polar angle symmetry. To see the deformation caused by hard edge scaling in the above equation we define *Hankel transform*

$$F_\alpha(t) = H_\alpha[f(z)] = \int_0^\infty z f(z) J_\alpha(z) dz \tag{3.24}$$

and the inverse Hankel transform is given as

$$f(z) = \int_0^\infty t F_\alpha(t) J_\alpha(tz) \tag{3.25}$$

Since the Hankel transform of the Bessel operator reads $H_\alpha[\Delta_\alpha f(z)] = t^2 F_\alpha(t)$ [**?**], the spectral deformation in dual variable t (note that t cannot be negative) reads simply

$$t \leq 1 \tag{3.26}$$

Combining both Hankel transforms we obtain the identity operator

$$F_\alpha(t^{'}) = \int_0^\infty \int_0^\infty ztJ_\alpha(t^{'}z)J_\alpha(tz)F_\alpha(t)dtdz \tag{3.27}$$

The deformation condition (3.26) projects the above identity operator onto

$$\mathbf{P}[F_\alpha(t^{'})] = \int_0^\infty \left[\int_0^1 ztJ_\alpha(t^{'}z)J_\alpha(tz)F_\alpha(t)dz\right] dz \tag{3.28}$$

so the kernel, understood as a projection, reads

$$K_{Bessel}(t,t^{'}) = \int_0^1 ztJ_\alpha(t^{'}z)J_\alpha(tz)dz \tag{3.29}$$

Coming back to original variables $z = \sqrt{s}$ and introducing $t = \sqrt{x}$ and $t^{'} = \sqrt{y}$ we write the kernel as

$$K_{Bessel}(x,y) = \frac{1}{4}\int_0^1 J_\alpha(\sqrt{xs})J_\alpha(\sqrt{ys})ds = \frac{J_\alpha(\sqrt{x})J_\alpha^{'}(\sqrt{y})\sqrt{y} - \sqrt{x}J_\alpha^{'}(\sqrt{x})J_\alpha(\sqrt{y})}{2(x-y)} \tag{3.30}$$

where on the r.h.s. we presented the more familiar form of the kernel based on Lommel integral and the prime denotes the differentiation with respect to the argument. Hard edge scaling deforms the upper half plane in s variable onto the strip between the parallel lines $s = 0$ and $s = 1$.

3.2 Related ensembles

3.2.1 Anti-Wishart

We have already mentioned the case of anti-Wishart. The spectral analysis of anti-Wishart is easy, if one realizes, what is the difference between the spectra of $XX^\dagger$ and $X^\dagger X$ matrices. In the case $N \leq T$, the dimension of the matrix $XX^\dagger$ is $N \times N$, whereas the dimension of the matrix $X^\dagger X$ is $T \times T$. The rank of the Wishart matrix is N, but the rank of the Anti-Wishart matrix is also N, so $T - N$ eigenvalues of anti-Wishart matrix have to be zero. Due to cyclic properties of the trace, all moments for Wishart and anti-Wishart are identical, so all non-zero eigenvalues are the same. The spectral functions are therefore related, due to the cyclic properties of the

trace. Explicitly, in the case of Anti-Wishart, Marcenko-Pastur (hereafter MP) law reads

$$\rho_{a-W} = (1-c)\delta(\lambda) + c\rho_{MP}(\lambda) \tag{3.31}$$

3.2.2 Quatercircle distribution

The case $c = 1$ of MP distribution yields the probability distribution $\rho(\lambda)d\lambda$, which is called quater-cricle distribution. Indeed, a change of variables $\lambda \to \lambda^2$, converts the quater-circle law onto Wigner-semicricle.

3.2.3 F and MANOVA

In multivariate statistics, two Wishart-related distributions play important role. So-called F-distribution is simply the ratio of two independent Wisharts, $W_1 W_2^{-1}$. So-called MANOVA (Multivariate ANalysis Of VAriance) is related to F, since this is a distribution of $W_1(W_1+W_2)^{-1}$. Using the powerful tools of free random variable calculus we will soon obtain painlessly the corresponding spectral distributions for both of them.

3.2.4 Chiral ensemble

A particularly important class of random matrices, related to Wishart, are chiral random matrix models. They have the form of $(N+T) \times (N+T)$ block matrix

$$\begin{pmatrix} 0_{N\times N} & X \\ X^\dagger & 0_{T\times T} \end{pmatrix} \tag{3.32}$$

For $N > T$, such matrix has $N - T$ zero eigenvalues, whereas non-zero eigenvalues come in pairs of opposite sign, being the pairs of square roots of the positive eigenvalues $X^\dagger X$ matrix. Such matrix mimics the properties of the Euclidean massless Dirac operators in non-Abelian Yang-Mills theory (Quantum Chromodynamics), whose non-zero eigenvalues also come in pairs, due the chiral property $[D_\mu(A)\gamma_\mu, \gamma_5]_+$, where γ's are Dirac matrices. In the case of chiral matrix models, the role of γ_5 matrix, anti-commuting with (3.32), is played by the "chirality" operator $\mathrm{diag}(\mathbf{1}_N, -\mathbf{1}_M)$. Due to the universality properties, the smallest eigenvalues of the Euclidean QCD should therefore obey the strictures of the hard edge (Bessel kernel). Indeed, the

lattice calculations have verified this conjecture in an impressive way (see figure XX).

Figure 3.1: Microscopic spectral density of the Euclidean Dirac operator measured in latticed simulations at hard edge, versus analytical prediction. From Wettig et al.

3.2.5 Any β

The case $\beta = 1$ is important in applications, since the measured multivariate data are usually real. The complex case $\beta = 2$ is also important, due to the applications in wireless telecommunication, when elements represents the scattered waves. (X) The case $\beta = 4$ (quaternionic Wishart) is also not academic, since the chiral $\beta = 4$ is important in applications of Quantum Chromodynamics. In all three cases, the macroscopic spectral density is given by Marchenko-Pastur distribution, modulo the rescaling of the variance of the Gaussians. The microscopic features, which are governed by the power of the Van der Monde determinate, are however different. Alike in the case of Gaussian ensembles, the $\beta = 2$ is the easiest, due to the determinantal structure, exploited as a Slater determinant in the previous chapters. Cases $\beta = 1, 4$ are more difficult, due to the need of the use of Pfaffians, and the resulting microscopic kernels are more complicated, and are given as a single and double integrals of the expressions involving Bessel functions, respectively. For the original formulae, we refer to XXXX. Last, but not least, we comment the case of an arbitrary β, in light of the extension of Dumitru and Edelman for the Gaussian ensembles, discussed in the previous chapter. Also in the case of the Wishart ensemble (named also as Laguerre ensemble), similar construction is possible and serves as a tremendous simplification in numerical simulations.

3.3 Appendix - Ingham-Siegel type of integrals

We will calculate the Ingham-Siegel type of the integral

$$P_{T,N}(M) = \int d\mathcal{F} e^{-\mathcal{F}M} \det(1+i\mathcal{F})^{-T} \tag{3.33}$$

by iteration method. Let us first decompose the matrices $\mathcal{F} \equiv \mathcal{F}_N$ and $M \equiv M_N$ as

$$\mathcal{F}_N = \left(\begin{array}{c|c} f_{11} & f^\dagger \\ \hline f & \mathcal{F}_{N-1} \end{array}\right) \qquad M_N = \left(\begin{array}{c|c} m_{11} & m^\dagger \\ \hline m & M_{N-1} \end{array}\right) \tag{3.34}$$

We will derive a recurrence relation by first integrating over f_{11} and then over the vector f. To this end we use the identity

$$\det(1+i\mathcal{F}_N) = \left(1 + if_{11} + f^\dagger (1+i\mathcal{F}_{N-1})^{-1} f\right) \cdot \det(1+i\mathcal{F}_{N-1}) \tag{3.35}$$

which is a special case ($n=1, m=N-1$) of a block $(n+m)\times(n+m)$ matrix determinant identify $\det M = \det(A - BD^{-1}C)\det D$, where

$$M = \left(\begin{array}{cc} A & B \\ C & D \end{array}\right) \tag{3.36}$$

and matrices A, B, C, D are $(n\times n), (n\times m), (m\times n), (m\times m)$, respectively. The integral over f_{11} can be done by residues giving

$$\begin{aligned} &\int d\mathcal{F}_{N-1} e^{i\mathrm{tr}\mathcal{F}_{N-1}M_{N-1}} (\det(1+i\mathcal{F}_{N-1}))^{-T} m_{11}^{T-1} e^{-m_{11}} \\ &\cdot \int df df^\dagger e^{-m_{11} f^\dagger (1+i\mathcal{F}_{N-1})^{-1} f} e^{i(f^\dagger m + m^\dagger f)} \end{aligned} \tag{3.37}$$

The last integral (second line) is Gaussian giving

$$\frac{1}{m_{11}^{N-1}} \cdot \det(1+i\mathcal{F}_{N-1}) \cdot e^{-\frac{1}{m_{11}} m^\dagger (1+i\mathcal{F}_{N-1}) m} \tag{3.38}$$

Substituting this back into (3.37) leads to

$$\int d\mathcal{F}_{N-1} (\det(1+i\mathcal{F}_{N-1}))^{-(T-1)} e^{i\mathrm{tr}\mathcal{F}_{N-1}\left(M_{N-1} - \frac{1}{m_{11}} m m^\dagger\right)} m_{11}^{T-N} e^{-m_{11} - \frac{m^\dagger m}{m_{11}}} \tag{3.39}$$

Hence we are led to the recurrence relation

$$P_{T,N}(M_N) = C''_{T,N}\, m_{11}^{T-N} e^{-m_{11} - \frac{m^\dagger m}{m_{11}}} P_{T-1,N-1}\left(M_{N-1} - \frac{1}{m_{11}} m m^\dagger\right) \quad (3.40)$$

We will now use the recurrence relation (3.40) to determine $P_{T,N}(M)$. The repeated use of (3.40) reduces the problem to calculating $P_{T,1}(M)$ which is just

$$P_{T,1}(m) = m^{T-1} e^{-m} \quad (3.41)$$

Since in the Wishart case the eigenvalues are generically distinct we may diagonalize the matrix M, and then the recurrence relation can be solved immediately [**?**] to get

$$P_{T,N}^{WISHART}(M) = C'''_{T,N}\, (\det M)^{T-N} e^{-\mathrm{tr}\, M} \quad (3.42)$$

Chapter 4

Large N simplifications - RMT "Diagrammar"

4.0.1 Preliminaries

We are interested in finding the distribution of eigenvalues λ_i, in the limit when N (the size of the matrix H) is infinite. The average spectral distribution reads

$$\rho(\lambda) = \lim_{N\to\infty} \frac{1}{N} \left\langle \sum_{i=1}^{N} \delta(\lambda - \lambda_i) \right\rangle \tag{4.1}$$

where λ_i are eigenvalues of a random hermitian matrix H and brackets $\langle \ldots \rangle$ denote averaging over a given ensemble of $N \times N$ random hermitian matrices generated with the probability

$$P(H)dH \propto e^{-N\mathrm{Tr}V(H)}dH. \tag{4.2}$$

For hermitian matrices eigenvalues λ_i's lie on the real axis. Such spectral density is an analogue of one dimensional probability density function (pdf) in the classical theory of probability. It is convenient to introduce the generating function for the spectral moments of the pdf. In classical probability such generating function is called characteristic function, and is simply the Fourier transform of the pdf, $F(k)$, as we have pointed in the first chapter. It's matrix analogue is the complex-valued resolvent (Green's function)

$$G_H(z) = \lim_{N\to\infty} \frac{1}{N} \left\langle \mathrm{Tr}\, \frac{1}{z\mathbf{1} - H} \right\rangle . \tag{4.3}$$

from which one can reconstruct the spectral density function (4.1)

$$\rho(\lambda) = \frac{1}{2\pi i} \lim_{\epsilon \to 0^+} \left(G(\lambda - i\epsilon) - G(\lambda + i\epsilon) \right). \tag{4.4}$$

using the well-known formula $\frac{1}{\lambda \pm i0^+} = \text{P.V.}\frac{1}{\lambda} \mp i\pi\delta(\lambda)$. The symbol $\mathbf{1}$ will be used throughout the paper to denote identity matrices of different size, here. Here it is a N-by-N identity matrix. Expanding the Greens function around $z = \infty$ (or, equivalently, around $x = 0$ after changing variables $z = \frac{1}{x}$), we see that indeed the Green's function is a generating function for spectral moments $\mu_n = \lim_{N\to\infty} \frac{1}{N} \langle \text{Tr} H^n \rangle$

$$G(z) = \sum_{n=0}^{\infty} \frac{\mu_n}{z^{n+1}} \tag{4.5}$$

with $\mu_0 = 1$, as follows from the $1/z$-expansion of (4.3). In classical probability it is convenient to introduce the function $r(k) = \log F(k)$, generating the cumulants. Cumulants are very useful concept in classical probability theory, since they are additive under the convolution of variables coming from independent pdfs. An analogue of such generating function in the probability theory of infinite random matrices is so-called Voiculescu R-transform.

$$R(w) = \sum_{n=1}^{\infty} \kappa_n w^{n-1} \tag{4.6}$$

We will see soon that so called free cumulants $\kappa_n = \lim_{N\to\infty} \frac{1}{N} \langle\langle \text{Tr} H^n \rangle\rangle$ correspond to planar connected moments denoted here by double brackets.

We will also see that the relation between the generating function for spectral moments $G(z)$ and the generating function for connected moments $R(w)$ is given as

$$G(z) = \frac{1}{z - R(G(z))} \tag{4.7}$$

which is equivalent to

$$G\left(R(w) + \frac{1}{w} \right) = w \,. \tag{4.8}$$

We see, that $R(w)$ is modulo the shift $1/w$, the functional inverse of the Green's function. It is convenient to introduce $B(w) = R(w) + \frac{1}{w}$, sometimes

called anecdotally as a Blue's function. Now relation between these two generating functions is particularly simple

$$B[G(z)] = z \quad \text{or} \quad G[B(w)] = w \tag{4.9}$$

since when $z = B(w)$ then $w = G(z)$. The only missing element in building the complete analogy between two probability calculi is to find the counterpart of *independence* in the case of ensembles of infinite, non-commuting random matrices. This analogue is called *freeness* and will be explained in detail using diagrammatic tools, so common in physics. In order to introduce gently the Reader to diagrammatic notation, we consider first the already known case of large N limit of the Gaussian Unitary Ensemble (GUE).

4.0.2 Gaussian Case

In the GUE case, $V(H) = \frac{H^2}{2}$. To calculate any moment $\mu_m = \frac{1}{N} < \mathrm{tr} H^m >$ we introduce the "partition function" with the source J ($N \times N$ matrix), normalized in such a way that $Z(0)$=1

$$Z(J) = \int dH e^{-\frac{N}{2}\mathrm{trH}^2 + \mathrm{tr}HJ} \tag{4.10}$$

Note that

$$\begin{aligned} \frac{\partial Z(J)}{\partial J_a^b} &= \int dH H_b^a e^{-\frac{N}{2}\mathrm{trH}^2 + \mathrm{tr}HJ} \\ \frac{\partial^{(2)} Z(J)}{\partial J_a^b \partial J_c^d} &= \int dH H_b^a H_d^c e^{-\frac{N}{2}\mathrm{trH}^2 + \mathrm{tr}HJ} \end{aligned} \tag{4.11}$$

so any expectation value of the string of m matrices H can be obtained as

$$< H_b^a H_d^c H_f^e ... >= \frac{\partial^{(m)}}{\partial J_a^b \partial J_c^d \partial J_e^f ...} Z(J)|_{J=0} \tag{4.12}$$

In the case of the Gaussian measure we can calculate $Z(J)$ exactly. Indeed, by completing the square and integrating the Gaussian after the change of the variables $H \to H + \frac{J}{N}$ we arrive at

$$Z(J) = \left(\int dH e^{-\frac{N}{2}[\mathrm{tr}H^2 + \frac{2}{N}\mathrm{tr}HJ + \frac{1}{N^2}\mathrm{tr}J^2]} \right) e^{\frac{1}{2N}\mathrm{tr}J^2} = e^{\frac{1}{2N}\mathrm{tr}J^2} \tag{4.13}$$

Now we can calculate moments explicitly

$$\begin{aligned}
\frac{\partial^{(2)} Z(J)}{\partial J_a^b \partial J_c^d}|_{J=0} &= < H_b^a H_d^c >= \frac{1}{N}\delta_b^a \delta_d^c \\
\frac{\partial^{(4)} Z(J)}{\partial J_a^b \partial J_c^d \partial J_e^f \partial J_g^h}|_{J=0} &= < H_b^a H_d^c H_f^e H_h^g > \\
&= \frac{1}{N^2}(\delta_d^a \delta_b^c \delta_h^e \delta_f^g + \delta_h^a \delta_b^g \delta_f^c \delta_d^e + \delta_f^a \delta_b^e \delta_g^d \delta_h^c) \quad (4.14)
\end{aligned}$$

Obviously, next non-vanishing term (average of the string of 6 matrices, since odd strings vanish due to the parity of the measure) involves 15 terms, each composed of 6 Kronecker deltas. To simplify the rapidly increasing combinatorics and proliferation of indices, we introduce subsequently two tricks - first, so-called *Feynman diagrams*, second, the large N expansion introduced by 't Hooft's *planar diagrams* dominance in the context of non-Abelian field theories. The first trick is based on identification of few building blocks in the above complicated expressions and representing them as pictograms. Then we "draw" as pictures the graphical representations of complicated expressions, being able at each step to decode the picture back into the mathematical formula. As an example we consider the Green's function matrix G_{ab}, whose normalized trace is just the Green's function $G(z)$. Explicitly (and using the fact that odd averages vanish)

$$\begin{aligned}
G(z)_{ab} &= \left\langle \left(\frac{1}{z\mathbf{1}_N - H} \right)_{ab} \right\rangle \\
&= \left\langle \left(\frac{1}{z}\mathbf{1}_N + \mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} H \mathbf{1_N} \frac{1}{z} + \mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} + . \right)_{ab} \right\rangle \quad (4.15)
\end{aligned}$$

where we have expanded formally the above expression around $z = \infty$. We can easily identify two basic building blocks - first, $(\mathbf{1}_N \frac{1}{z})_{ab}$, which we depict as a single, dotted line, symbolizing Kronecker delta $[\mathbf{1}_N]_{ab} = \delta_b^a$ multiplied by $1/z$, and $< H_b^a H_d^c >= \frac{1}{N}\delta_b^a \delta_d^c$, which we depict as a double line, as shown at the figure 4.0.2. Let us look explicitly at the second term in explicit notation (see figure 4.2)

$$\begin{aligned}
&\left\langle (\mathbf{1}_N \frac{1}{z} H \mathbf{1}_N \frac{1}{z} H \mathbf{1_N} \frac{1}{z})_{ab} \right\rangle = \\
&\frac{1}{z^3} \left\langle \delta_{ac} H_{cd} \delta_{de} H_{ef} \delta_{fb} \right\rangle =
\end{aligned}$$

$$\frac{1}{z^3}\delta_{ac}\delta_{de}\delta_{fb}\left\langle H_{cd}H_{ef}\right\rangle = \\ \frac{1}{z^3}\delta_{ac}\delta_{de}\delta_{fb}\frac{1}{N}\delta_{cf}\delta_{de} = \\ \frac{1}{z^3}\frac{1}{N}\delta_{dd}\delta_{ab} = \frac{1}{z^3}\delta_{ab} \tag{4.16}$$

In the third line we used the double line notation for the expectation value of two matricial elements of H, then the fact the $\delta_{dd} = N$. Note that in above expression we use known summation convention, that repeated indices $(c, d, e, f,)$ are implicitly summed over. Graphically, summation is just depicted as gluing two lines. Then, using both rules and expressions (4.14), Green's function is depicted as the first line on figure 4.0.2. Note that $\sum_a G_{aa} = NG(z)$, since G_{ab} is proportional to δ_{ab}, due to the invariance of the trace operation under diagonalization of the matrix H.

Next, we perform large N limit. Every double line brings factor $1/N$, every loop brings factor N, since $\mathrm{Tr}\mathbf{1} = N$. Inspection of diagrams for G of figure 4.0.2 shows, that all depicted terms behave as $O(1)$, except the last one, which vanishes a $1/N^2$. Indeed, the last depicted diagram for G has two double lines ("propagators" in quantum field theory slang), but no loops.

Next, we introduce one-line irreducible diagrams, i.e. such diagrams, which do not fall apart if we cut one internal line. Third diagram on the r.h.s. in first picture of the fig. 4.0.2 on the r.h.s. is one line reducible, but the fourth is not. We collect all 1LIR diagrams into one object, which we call $\Sigma(z)$. By definition,

$$[G(z)]_{ab} = [\mathbf{1}_N\frac{1}{z} + \mathbf{1}_N\frac{1}{z}\Sigma\mathbf{1}_N\frac{1}{z} + \mathbf{1}_N\frac{1}{z}\Sigma\mathbf{1}_N\frac{1}{z}\Sigma\mathbf{1}_N\frac{1}{z} + ...]_{ab} = \left[\frac{1}{z-\Sigma(z)}\right]_{ab} \tag{4.17}$$

or simply $G(z) = \frac{1}{z-\Sigma(z)}$ for the traced object. In physicist slang, Σ is often called self-energy. As a last step, we close the construction, by expressing self-energy in terms of the Green's function, i.e we seek equation of the form $\Sigma(G) = f(G)$. Due to the planarity (large N expansion all diagrams are "nested", and we do depict above relation (known as Schwinger-Dyson equation in quantum field theory) as last line on figure 4.0.2, which reads

$$\Sigma(z) = G(z) \tag{4.18}$$

Finally, combining equations $G = 1/(z-\Sigma)$ and $\Sigma = G$, we obtain the known, quadratic equation for $G(z)$. We choose the solution which is convergent at $z = \infty$, and its imaginary part reproduces the Wigner semicircle.

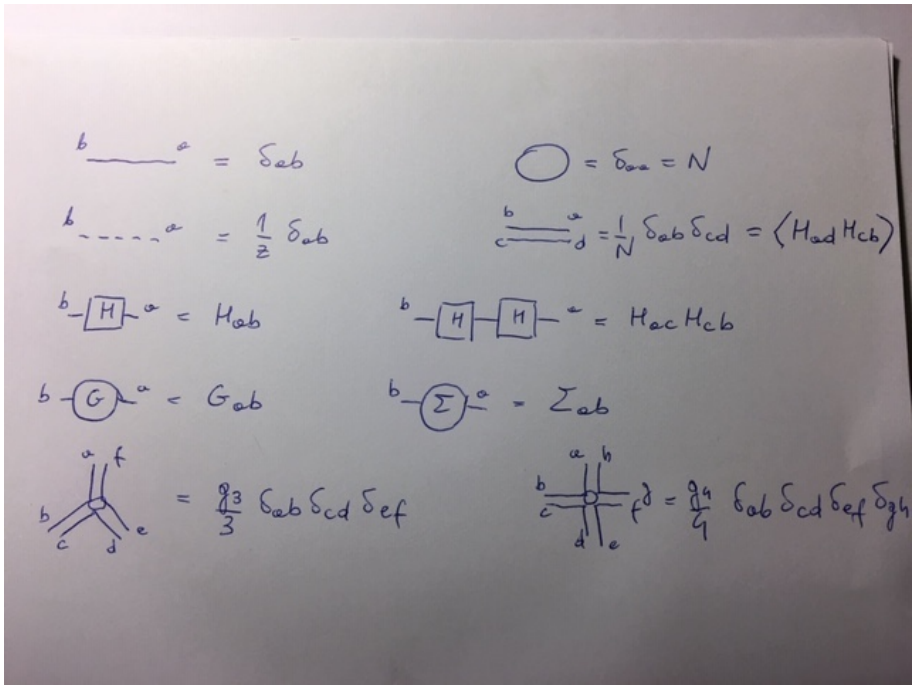


Figure 4.1: Feynman rules

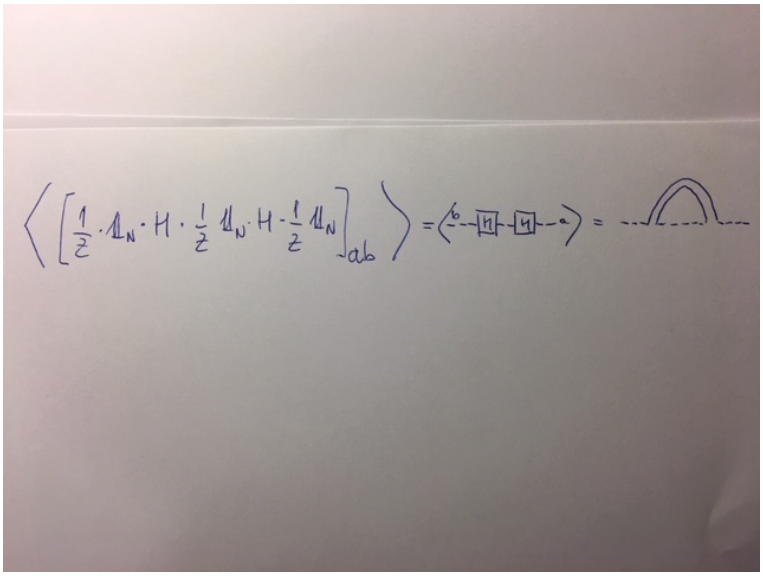


Figure 4.2: Example

4.0.3 Planar diagrams

In case of more complicated potentials, one can calculate (4.5) by Gaussian perturbation theory. One does it by splitting the integration measure (4.2) into a Gaussian part and a residual part

$$P(H) = \mathcal{N}^{-1} e^{-N\frac{g_2}{2}\mathrm{Tr}H^2} e^{-N\sum_{n\neq 2}\frac{g_n}{n}\mathrm{Tr}H^n} = P_0(H) e^{-N\sum_{n\neq 2}\frac{g_n}{n}\mathrm{Tr}H^n} \quad (4.19)$$

The Gaussian part $P_0(H)$ is then used to calculate averages $\langle \ldots \rangle_0$ while the remaining expression is left inside the brackets and is averaged with respect to P_0. The constant $\mathcal{N}$ is an overall normalization. This non-Gaussian part is perturbatively expanded in g_n, so effectively one has to calculate averages of various powers of H with respect to the Gaussian measure. The novel element comparing to the GUE case is the presence of the new type of strings of H matrices originating from terms with g_n when $n > 2$. Let us consider an example of the cubic term $\frac{g_3}{3}TrH^3$ in the measure. In the lowest non-trivial expansion of the exponent, it can be written as

$$\frac{g_3}{3} trH^3 = \frac{g_3}{3}\delta^a_f\delta^c_b\delta^e_d H^a_b H^c_d H^e_f \equiv V^{ace}_{fbd} H^a_b H^c_d H^e_f \quad (4.20)$$

For such vertex V we introduce a new pictogram, as shown on figure 4.0.2. Each term in the expansion for G function has a graphical representation,

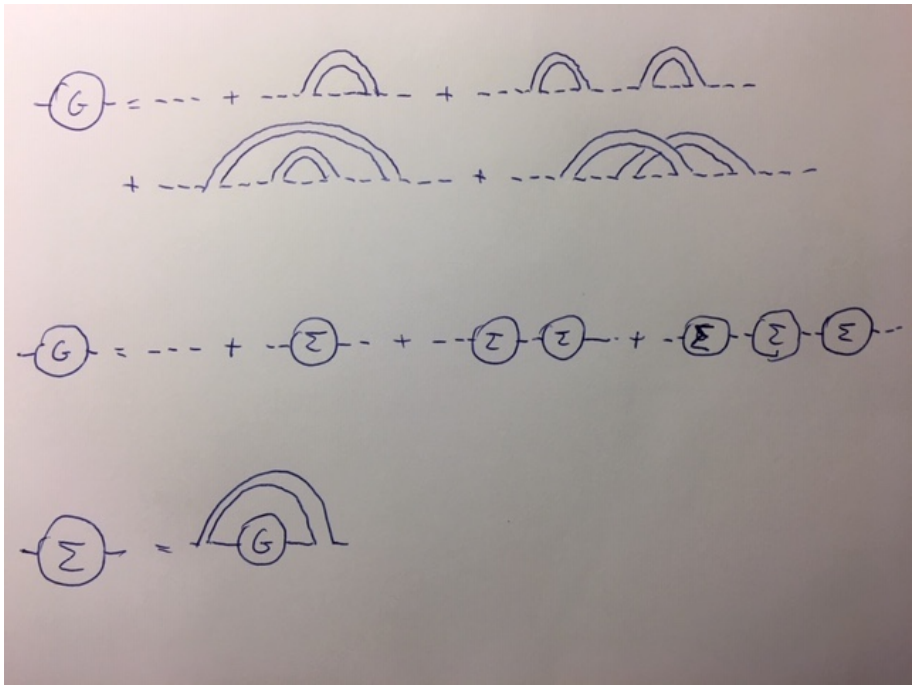


Figure 4.3: Green's function, self energy and Schwinger-Dyson equation

similar to Feynman diagrams known from quantum field theory (see figure 4.5) . Since all lines in the diagram are proportional to the delta function this equation reduces to a scalar equation for $G(z)$. Our basic set of pictograms is now enlarged, on top of dashed lines terms and double lines we have triple vertices. Note that the averaging we do perturbatively with respect to the GUE measure. Similarly one can do for other vertices. Sticking to the cubic one only, we simply draw all the pictures contributing to the Greens function, alike on figure 4.5. Note, that since each vertex brings a factor N, alike the loop, only planar diagrams survive the large N limit. As an example, let us consider sample diagram shown on figure sevendiag contributing to the to the seventh moment $\frac{1}{N}\langle \mathrm{Tr} H^7 \rangle$ which of order $\frac{1}{z^8}$ in the series expansion (4.5) since it has eight horizontal lines. The diagram contains seven cubic vertices g_3^7 and one quartic vertex g_4 that are generated by the perturbative expansion of the residual part of (4.19).

Thus the calculation of $G(z)$ amounts to summing all (infinitely many) contributions from planar diagrams with two endpoints as shown in figure (4.5).

Actually in the most general case one should rather consider a matrix form of the Green's function $G_{ij}(z)$) where i and j are indices of two endpoints $i = 1, \ldots, N$, $j = 1, \ldots, N$ (see figure 4.5) and calculate the scalar function (4.3) afterwards as the normalized trace $G(z) = \frac{1}{N}\mathrm{Tr} G(z)$. Also the self-energy equation (**??**) should formally be written in a matrix form. However in our case all generating matrices are proportional to Kronecker delta functions $G_{ij}(z) = G(z)\delta_{ij}$, $\Sigma_{ij}(z) = \Sigma(z)\delta_{ij}$, $\langle H_{ij} H_{kl} \rangle_0 \sim \delta_{il}\delta_{jk}$ so all equations like (**??**) and (**??**) reduce to scalar equations for the coefficients multiplying the delta functions.

A graphical interpretation of equation (**??**) becomes clear if one rewrites it as an infinite geometric series

$$G(z) = \frac{1}{z} + \frac{1}{z}\Sigma(z)\frac{1}{z} + \frac{1}{z}\Sigma(z)\frac{1}{z}\Sigma(z)\frac{1}{z} + \ldots \ . \tag{4.21}$$

which can be seen in figure 4.6. This figure tells us that all diagrams in $G(z)$ can be constructed by lining up one-line-irreducible diagrams one after another. An example of such a one-line-irreducible diagram contributing to $\Sigma(z)$ is shown in figure 4.7. Such diagrams are characterized by the property that they cannot be disconnected by cutting one line as opposed to diagrams generated by $G(z)$. Indeed, as one can see in figure 4.6 a diagram in $G(z)$ can be disconnected by cutting any horizontal line like that between two con-

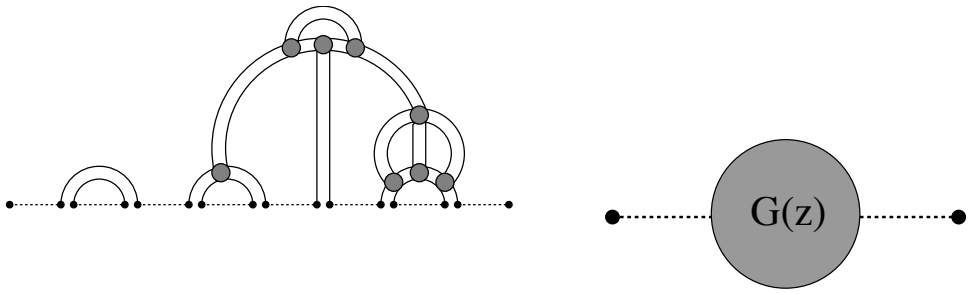


Figure 4.4: (Left) An example of a diagram contributing to the generating function $G_{ij}(z)$. Two end-points should be labeled by indices ij. Each horizontal dashed line corresponds to $\frac{1}{z}\delta_{ab}$ while a double line represents the expectation value (the propagator) $\langle H_{ab}H_{cd}\rangle_0 = \frac{1}{Ng_2}\delta_{ad}\delta_{bc}$. Each pair of dots on the horizontal line corresponds to a factor H_{ab} inside the average $\langle \mathrm{Tr}H^7\rangle_0$. (Right) The graphical notation for the generating function $G(z)$. It generates diagrams having two end-points which include for example the one shown on the left.

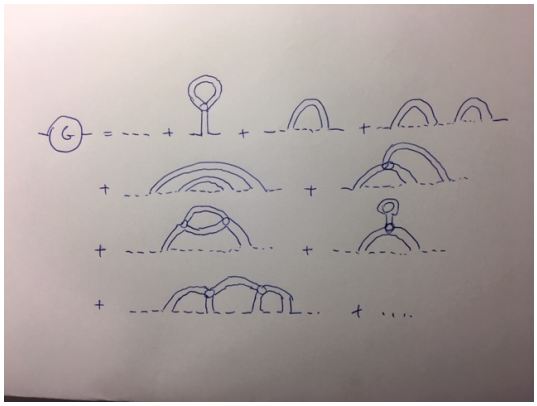


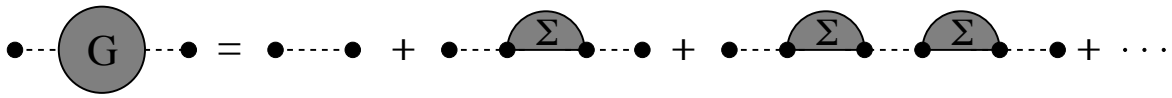
Figure 4.5: Green's function with cubic vertices.

secutive Σ's. The diagrammatic equation in figure 4.6 can be interpreted as a definition of the generating function $\Sigma(z)$ of one-line-irreducible diagrams.

It turns out that one can again write down an independent equation relating $\Sigma(z)$ to $G(z)$, i.e. the Schwinger Dyson equation. One can namely observe that any one-line-irreducible diagram can be obtained from diagrams generated by $G(z)$ as shown in figure 4.8 by adding a spider structure making them one-line-irreducible. Each bubble κ_n of the spider with n double legs corresponds to a connected moment (free cumulant) of order n. This equation tells us that

$$\Sigma(z) = \frac{1}{N}\langle\langle \mathrm{Tr} H\rangle\rangle + \frac{1}{N}\langle\langle \mathrm{Tr} H^2\rangle\rangle G(z) + \frac{1}{N}\langle\langle \mathrm{Tr} H^3\rangle\rangle G^2(z) + \ldots = R(G(z)) \tag{4.22}$$

The diagrammatic equations in figures 4.6 and 4.8 belong to the category of Dyson-Schwinger equations known from quantum field theory. They are equivalent to the equations (**??**) and (**??**) discussed in the previous section.


G
=
+
Σ
+
Σ
Σ
+ · · ·

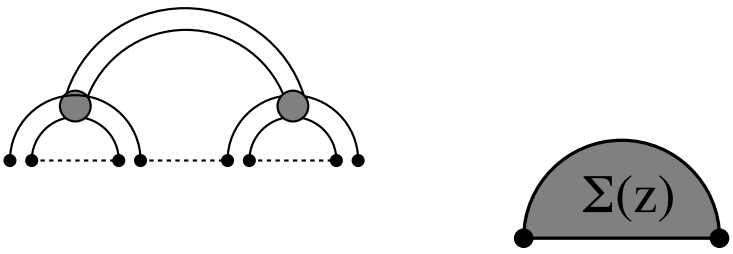


Figure 4.7: (Right) An example of an one-line-irreducible diagram. (Left) The graphical notation for the generating function $\Sigma(z)$ of one-line-irreducible diagrams.

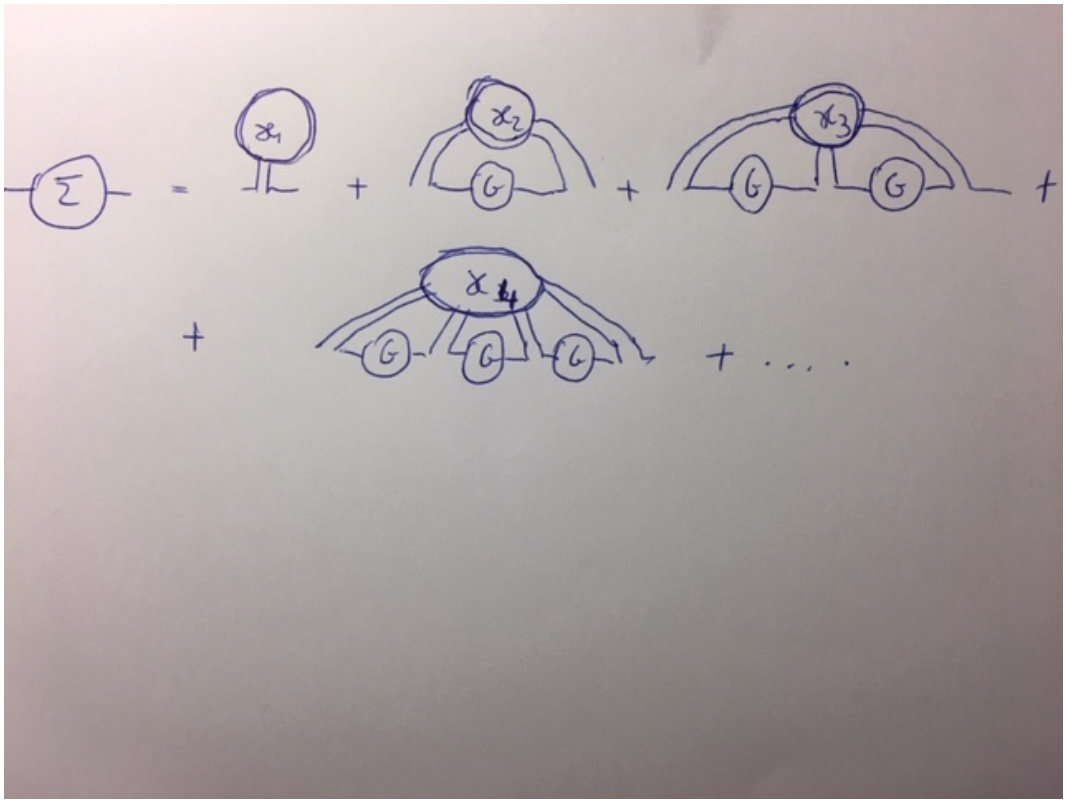


Figure 4.8: A one-line-irreducible diagram can be obtained from a one-line-reducible diagram by adding to it a minimal diagrammatic structure complying with the measure (4.19) which makes it one-line-irreducible. Such a minimal structure is provided by diagrams corresponding to planar connected moments κ_k (free cumulants) (see figure 4.9) which we indicated by bubbles surrounded by double circles in the figure. This double ring around the bubble is chosen to make it similar to double brackets used in our notation for connected averages. Diagrams in such a bubble are connected. The difference between diagrams corresponding to planar connected moments (cumulants) and spectral moments is explained in figure 4.9.

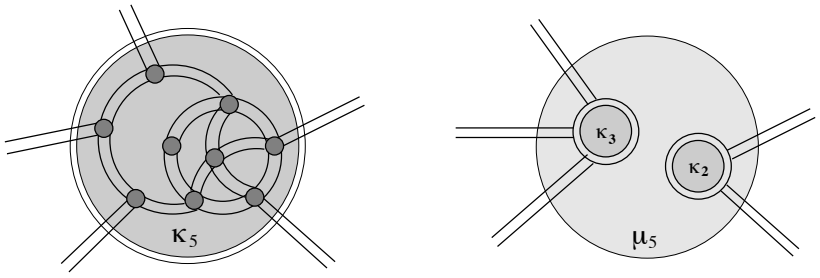


Figure 4.9: (Left) An example of a diagram generated by fifth free cumulant $\kappa_5 = \frac{1}{N}\langle\langle \mathrm{Tr}H^5\rangle\rangle$. All diagrams in the bubble must be connected in contrast to the diagrams generated by spectral moments. (Right) An example of the decomposition of some diagrams generated by the fifth spectral moment $\mu_5 = \frac{1}{N}\langle \mathrm{Tr}H^5\rangle$ into two connected moments $\kappa_2\kappa_3$. Some other diagrams in μ_5 can be decomposed into $\kappa_1\kappa_2\kappa_2$ or any other combination of cumulants as long as the number of external legs is five. Only a small subset of diagrams in μ_5 corresponds to those of κ_5.

Chapter 5

Free Random Variables -Addition law

5.1 Addition law intutively: R transform

Before we introduce formally the concept of freeness, let us try to gain some intuition. The R transform [**?**] is important because it allows one to concisely write down a law of addition of (free) independent large matrices. Consider first a factorized measure for two large matrices H_1, H_2 in the limit $N \to \infty$

$$P(H_1, H_2) = P_{H_1}(H_1) P_{H_2}(H_2) \tag{5.1}$$

where $P_{H_1}(H_1) \sim \exp[-N\mathrm{Tr}V_{H_1}(H_1)]$ and $P_{H_2}(H_2) \sim \exp[-N\mathrm{Tr}V_{H_2}(H_2)]$. Then consider a matrix $H = H_1 + H_2$. The law of addition tells us how to calculate the spectral density of H for given spectral densities of H_1 and H_2.

The idea is based on the observation that connected planar moments (free cumulants) of the sum $H = H_1 + H_2$ split into two independent parts

$$\frac{1}{N}\langle\langle \mathrm{Tr}(H_1 + H_2)^k \rangle\rangle = \frac{1}{N}\langle\langle \mathrm{Tr}H_1^k \rangle\rangle + \frac{1}{N}\langle\langle \mathrm{Tr}H_2^k \rangle\rangle \ . \tag{5.2}$$

The reason for this separation of connected moments can be easily understood in terms of Feynman diagrams. All mixed connected moments $\langle\langle \mathrm{Tr}H_1^a H_2^b H_1^c H_2^d ... \rangle\rangle$ disappear just because there is no direct line in a connected diagram between a vertex of type H_1 and H_2 since the H_1H_2-propagator is zero $\langle (H_1)_{ij}(H_2)_{kl} \rangle_0 = 0$. The crossed pairs of double lines corresponding to H_1 and H_2 vanish in the large N limit, since they represent non-planar

contribution. So all external lines of a bubble generated by k-th cumulant correspond either to H_1 or to H_2. In other words free cumulants fulfil a simple equation

$$\kappa_{H_1+H_2,n} = \kappa_{H_1,n} + \kappa_{H_2,n} \tag{5.3}$$

and thus

$$R_{H_1+H_2}(w) = R_{H_1}(w) + R_{H_2}(w) \ . \tag{5.4}$$

The argument given above is equivalent to a reasoning based on non-crossing partitions used to prove this law in [**?**]. The law of free addition (5.4) is also sufficient to calculate spectral moments of the free sum $\mu_n = \frac{1}{N}\langle \mathrm{Tr} H^n \rangle = \frac{1}{N}\langle \mathrm{Tr}(H_1 + H_2)^n \rangle$ if one knows the spectral moments of H_1 and H_2. The recipe follows from the relations (4.7) and (4.8):

1. Using (4.8) calculate $R_{H_1}(w)$ for given $G_{H_1}(z)$ and $R_{H_2}(w)$ for $G_{H_2}(z)$.
2. Construct the R transform $R_{H_1+H_2}(w)$ for the sum using the addition law (5.4).
3. Calculate $G_{H_1+H_2}(z)$ for $R_{H_1+H_2}(w)$ using (4.8) and calculate spectral moments $\langle \mathrm{Tr}(H_1 + H_2)^n \rangle$ and the spectral density of $H = H_1 + H_2$ using (4.4).

5.1.1 Addition law - mathematically rigid construction for R transform

Let us define the linear operation $\frac{1}{N}\langle \mathrm{tr} A \rangle \equiv \phi(A)$. Then in particular,

$$G_H(z) = \phi(\frac{1}{z\mathbf{1} - H}) \tag{5.5}$$

where we have informally written the inverse of the operator $(z\mathbf{1} - H)$ as $\frac{1}{z\mathbf{1}-H}$.

Infinite matrices $H_1, ...H_k$ are mutually free if

$$\phi\left([P_1(H_{i_1}) - \phi(P_1(H_{i_1})]...[P_n(H_{i_n}) - \phi(P_n(H_{i_n})]\right) = 0 \tag{5.6}$$

whenever $P_1, ..., P_n$ are polynomials and indices $i_1, ..., i_n$ from the set $\{1....k\}$ are such, that no two adjacent i_k are equal.

This definition is a bit cryptic, so let us consider a series of examples. Obviously, for any fixed H_i, we have $\phi(H_i - \phi(H_i)) = \phi(H_i) - \phi(H_i) = 0$. Let us consider now a pair H_1, H_2, in the case of lowest order polynomial, i.e.

$$\phi\left([H_1 - \phi(H_1)][H_2 - \phi(H_2]\right) = \phi(H_1H_2) - \phi(H_1)\phi(H_2) - \phi(H_1)\phi(H_2) + \phi(H_1)\phi(H_2) \quad (5.7)$$

Since the l.h.s. is zero, we just get $\phi(H_1H_2) = \phi(H_1)\phi(H_2)$, which is identical to the factorization property of the moments of independent variables in classical probability, i.e. $< xy >=< x >< y >$. The first surprise comes when we consider

$$\phi\left([H_1) - \phi(H_1)][H_2 - \phi(H_2][H_1) - \phi(H_1)][H_2 - \phi(H_2]\right) \quad (5.8)$$

Since this quantity is equal to zero for free H_1, H_2, little algebra of multiplication and linearity shows that

$$\phi(H_1H_2H_1H_2) = \phi(H_1^2)\phi^2(H_2) + \phi(H_2^2)\phi^2(H_1) - \phi^2(H_1)\phi^2(H_2) \quad (5.9)$$

On the other hand, when we consider the product

$$\phi\left([H_1) - \phi(H_1)][H_1) - \phi(H_1)][H_2 - \phi(H_2][H_2 - \phi(H_2]\right) \quad (5.10)$$

similar reasoning gives different result, i.e.

$$\phi(H_1H_1H_2H_2) = \phi(H_1^2)\phi(H_2^2) \quad (5.11)$$

This is expected, since free H_1 and H_2 do not commute. Freeness is also quite different from classical independence - here we rather get the sums of certain factorized moment, whereas in the classical case moments for mixed independent variables always factorize. Note also that in free world condition $\phi(H_1) = \phi(H_2) = 0$ forces $\phi(H_1H_2H_1H_2)$ to vanish, whereas in the case of classical probability, from the condition that independent means vanish $< x >=< y >= 0$ we cannot infer similar property for $< xyxy >=< x^2y^2 >=< x^2 >< y^2 >$, i.e. the product of variances of independent variables does not vanish, if the means are zero!

Let us now take Green's function $G_{H_1}(z) = \phi(\frac{1}{z\mathbf{1}-H_1})$, and substitute $z \to B_{H_1}(w)$, where B_{H_1} is the Blue's function. By definition of the Blue's function we read $w = \phi(\frac{1}{B_{H_1}(w)\mathbf{1}-H_1})$. Let us parametrize the inverse of matrix $(B_{H_1}(w)\mathbf{1} - H_1)$ as

$$\frac{1}{B_{H_1}(w)\mathbf{1} - H_1} \equiv w(\mathbf{1} - Q_{H_1}(w)) \quad (5.12)$$

where Q_{H_1} is some complicated function of matrix H_1, with only restriction that $\phi(Q_{H_1}(w)) = 0$, in order to recover the formula for the Green's function. Let us rewrite now the above relation for two supposed to be free matrices H_1 and H_2. We get

$$\begin{aligned} B_{H_1}(w)\mathbf{1} - H_1 &= \frac{1}{w}(\mathbf{1} - Q_{H_1}(w))^{-1} \\ B_{H_2}(w)\mathbf{1} - H_2 &= \frac{1}{w}(\mathbf{1} - Q_{H_2}(w))^{-1} \end{aligned} \tag{5.13}$$

Adding above equations we arrive at

$$H_1 + H_2 = B_{H_1}(w)\mathbf{1} + B_{H_2}(w)\mathbf{1} - \frac{1}{w}[(\mathbf{1} - Q_{H_1}(w))^{-1} + (\mathbf{1} - Q_{H_2}(w))^{-1}] \tag{5.14}$$

The expression in square brackets we rewrite as

$$[...] = [(\mathbf{1} - Q_{H_1}(w))^{-1}(\mathbf{1} - Q_{H_2}(w) + \mathbf{1} - Q_{H_1}(w))(\mathbf{1} - Q_{H_2}(w))^{-1}] \tag{5.15}$$

Using the trivial (matricial) identity

$$\begin{aligned} \mathbf{1} &= (\mathbf{1} - Q_{H_1}(w))^{-1}(\mathbf{1} - Q_{H_1}(w))(\mathbf{1} - Q_{H_2}(w))(\mathbf{1} - Q_{H_2}(w))^{-1} \\ &= (\mathbf{1} - Q_{H_1}(w))^{-1}(\mathbf{1} - Q_{H_1}(w) - Q_{H_2}(w) + Q_{H_1}(w)Q_{H_2}(w))(\mathbf{1} - Q_{H_2}(w))^{-1} \end{aligned}$$

we can now rewrite (5.14) as

$$\begin{aligned} H_1 + H_2 = &\left[B_{H_1}(w) + B_{H_2}(w) - \frac{1}{w}\right]\mathbf{1} + \\ &-\frac{1}{w}[(\mathbf{1} - Q_{H_1}(w))^{-1}(\mathbf{1} - Q_{H_1}(w)Q_{H_2}(w))(\mathbf{1} - Q_{H_2}(w))^{-1}] \end{aligned} \tag{5.16}$$

Inverting both sides we arrive at

$$\begin{aligned} &(H_1 + H_2 - \left[B_{H_1}(w) - B_{H_2}(w) + \frac{1}{w}\right]\mathbf{1})^{-1} = \\ &-w(\mathbf{1} - Q_{H_2}(w))(1 - Q_{H_1}(w)Q_{H_2}(w))^{-1}(1 - Q_{H_2}(w)) \end{aligned} \tag{5.17}$$

Let us apply now the operation ϕ on both sides of the equation. It is crucial to observe, that on the right hand side only the unit matrix survives, since $\phi(\mathbf{1}) = 1$. By construction, $\phi(Q_{H_1}(w)) = \phi(Q_{H_2})(w) = 0$, and representing the middle factor of the r.h.s. as a geometric series, we notice that the ϕ

operating on all mixed terms with alternating products of Q_{H_1} and Q_{H_2} have to vanish by the freeness condition. We arrive then at

$$\phi\left(\left[B_{H_1}(w)+B_{H_2}(w)-\frac{1}{w}\right]\mathbf{1}-(H_1+H_2)\right)=w \tag{5.18}$$

Finally, let us denote

$$B_{H_1}(w)+B_{H_2}(w)-\frac{1}{w}\equiv B_{H_1+H_2}(w) \tag{5.19}$$

Substituting now $w\to G_{H_1+H_2}(z)$ we recover the definition of the Green's function for the sum of free matrices H_1 and H_2, i.e.

$$G_{H_1+H_2}(z)=\phi(z\mathbf{1}-(H_1+H_2))=\frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z\mathbf{1}-(H_1+H_2)}\right\rangle \tag{5.20}$$

Remembering that Blues function is related to R-transform as $B(w)=R(w)+\frac{1}{w}$ we recover from (5.19) the additivity law for the R-transform, i.e.

$$R_{H_1}(w)+R_{H_2}(w)=R_{H_1+H_2}(w) \tag{5.21}$$

i.e. the addition law for the free cumulants.

5.1.2 Important examples

- Convolution of two discrete measures

  Let us consider discrete probability distribution, composed of two values $\{-1/2,+1/2\}$, with equal probability of drawing them, i.e. equal to $1/2$. Let us first consider the convolution of two such pdf's in classical probability, i.e. we consider pdfs as *independent*. As a result of "summation", we get the new pdf, also discrete, with values $\{-1,0,+1\}$ and corresponding probabilities $\{1/4,1/2,1/4\}$, since value zero we get get in two ways $(+1/2+(-1/2))$ or $((-1/2)+1/2)$.

  Let us now solve the same problem in free probability. The spectral distribution for each free distribution reads

$$\rho(\lambda)=\frac{1}{2}\left(\delta(\lambda-\frac{1}{2})+\delta(\lambda+\frac{1}{2})\right) \tag{5.22}$$

which implies that the Green's function reads

$$G(z) = \int \frac{\rho(\lambda)}{z-\lambda} = \frac{1}{2}\left[\frac{1}{z-\frac{1}{2}} + \frac{1}{z+\frac{1}{2}}\right] \tag{5.23}$$

It is easy to find the corresponding Blue's function, by substitution $z \to B(w)$, so we arrive at equation

$$w = \frac{4B(w)}{4B^2(w)-1} \tag{5.24}$$

with obvious solution

$$B(w) = \frac{1}{2}(\frac{1}{w} + \frac{\sqrt{1+w^2}}{w}) \tag{5.25}$$

where we kept the sign which is consistent with small w behavior of the Blue's function. Let us now consider an addition of two identical, **free** such distributions, which we index as H_1 and H_2. An addition law gives us

$$B_{H_1}(w) + B_{H_2}(w) - \frac{1}{w} \equiv B_{H_1+H_2}(w) = \frac{\sqrt{1+w^2}}{w} \tag{5.26}$$

Now, the play the trick $w \to G_{H_1+H_2}(z)$, so by the definition of Blue's function allows the above equation reads

$$z = \frac{\sqrt{1+G^2_{H_1+H_2}(z)}}{G_{H_1+H_2}(z)} \tag{5.27}$$

with obvious from asymptotics solution

$$G_{H_1+H_2}(z) = +\frac{1}{\sqrt{z^2-1}} \tag{5.28}$$

The corresponding spectral function we recover from the analysis of the cuts, i.e

$$\rho_{H_1+H_2}(\lambda) = -\frac{1}{\pi}\lim_{\epsilon\to 0} \Im G_{H_1+H_2}(z)|_{z=\lambda+i\epsilon} = \frac{1}{\pi\sqrt{1-\lambda^2}} \tag{5.29}$$

The result is surprising, since the distribution is continuous, which would never happen as a result of discrete convolutions in classical

probability. To understand this result, let us try to form two large, N by N matrices H_1 and H_2, which would be free in the limit when $N \to \infty$. Let us choose them as $H_1 = \mathrm{diag}(+1/2, +1/2, -1/2, ..., 1/2)$ and $H_2 = \mathrm{diag}(+1/2, -1/2, -1/2, ..., -1/2$, so we put randomly equal number of eigenvalues $\pm 1/2$ on the diagonals. If we simply sum such both diagonal; matrices, we will recover the discrete pdf known from the classical case. However, if we would consider sum $H_1 + UH_2U^\dagger$, where U is a random unitary matrix (Haar measure), we would get the expected continuous distribution, see the exercise. Freenees of two ensembles means therefore, that their corresponding eigenvectors are maximally decorrelated.

Exercise: Generate several sums $M = H_1 + UH_2U$, where $H_{1,2}$ are diagonal matrices with randomly distributed equal number of values $\pm 1/2$, and U is a Haar measure. Then, diagonalize an ensemble of M's, and plot the resulting spectral distribution, and compare it to analytic result (5.29), known as a *free arcsine distribution*. Hint. There are many ways to generate U. Most laborious comes from definition $UU^\dagger = 1$, i.e. generate random vector u_1 and normalize it. Then generate second vector u_2, make it orthogonal to the first one (e.g. by Gram-Schmidt procedure), and normalize, then draw the third one, made it orthogonal to both u_1, u_2, normalize etc. An easier way is to use the random eigenvectors corresponding to eigenvalues of the GUE ensemble, one only has to normalize them, since they are random and orthogonal by construction.

- Pastur equation

Let us consider now the problem of addition of (1) the measure corresponding to the discrete set of eigenvalues λ_i, $(i = 1, 2, ..., K)$ appearing with the corresponding probabilities p_i and (2) the measure of the Gaussian Unitary Ensemble. The Green's function for the discrete set reads

$$G_D(z) = \sum_i^K \frac{p_i}{z - \lambda_i} \tag{5.30}$$

hence, substitution $z \to B_D(w)$ yields equation

$$w = \sum_{i=1}^{K} \frac{p_i}{B_D(w) - \lambda_i} \tag{5.31}$$

The Blue's function for random GUE is simply $B_R(w) = R_R(w) + \frac{1}{w} = w + \frac{1}{w}$, since only second free cumulant does not vanish for GUE (here, for simplicfity, we put the value of this cumulant to one). Now, the addition law tell us

$$B_{D+R}(w) = B_D(w) + B_R(w) - \frac{1}{w} = B_D(w) + w \tag{5.32}$$

hence $B_D(w) = B_{D+R}(w) - w$. Plugging in this relation into (5.31) we arrive at w

$$w = \sum_{i}^{K} \frac{p_i}{B_{D+R}(w) - w - \lambda_i} = G_D[B_{D+R}(w) - w] \tag{5.33}$$

Finally, replacing in the above equation $w \to G_{D+R}(z)$ and using the definition of the Blue's function, we arrive at the Pastur equation

$$G_{D+R}(z) = G_D(z - G_{D+R}(z)) \tag{5.34}$$

In the case when discrete set is composed on one eigenvalue, we get the GUE measure, shifted by the value of this eigenvalue. In the case of two eigenvalues, let us say $\pm a$, and equal probabilities of finding them we arrive at cubic equation

$$G(z) = \frac{1}{2}\left(\frac{1}{z - a - G(z)} + \frac{1}{z + a - G(z)}\right) \tag{5.35}$$

and so forth.

Exercise. (a) Check numerically Pastur equation in the above case of $\lambda_i = \pm a$. First perform the simulation. Note, that the eigenvectors of both ensembles are already maximally decorrelated, due to the Gaussian random character of GUE, so simply add diagonal matrix with randomly distributed values $\pm a$ to the GUE matrix, and perform averaging over such example for $a = 1/2, 1, 2, 10$. (b) Compare these results with analytic solution, i.e. solve analytically the resulting cubic

(Cardano) equation, identify the right solution for $G(z)$ from the three possible, and plot the spectral function $\rho(\lambda, a)$ as a 3d plot ρ, λ, a. At which value of a the support of the spectrum splits from single interval onto two intervals?

- Free central limit theorem.

  Let us consider K free ensembles H_i, and let us ask what is the spectral distribution of the sum $\sum_{i=1}^{K} \frac{H_i}{\sqrt{K}}$ in the limit when K tends to infinity. We only assume, that second cumulant of each H_i is finite.

  Before we tackle this problem, let us show that if we know the Green's function of H, we know it also for aH, where constant a is real. Indeed

$$G_{aH} = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z - aH}\right\rangle = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{az/a - aH}\right\rangle = \frac{1}{a}G_H(\frac{z}{a}) \quad (5.36)$$

  Then, by definition

$$B_{aH}(w) = aB_H(aw) \quad (5.37)$$

  and similarly $R_{aH}(w) = aR_H(aw)$. Let us define the R transform for each H_i as $R_i(w) \equiv wr_0^{(i)}(w)$. Then, the sum of rescaled by a matrices H_i fulfills an addition law

$$R(w) = \sum_{i=1}^{K} a^2 w r_0^{(i)}(aw) \quad (5.38)$$

  therefore

$$B(w) = \sum_{i=1}^{K} a^2 w r_0^{(i)}(aw) \quad (5.39)$$

  Substitution $w \to G(z)$ and $a = 1/\sqrt{K}$ yields

$$z = \sum_{i=1}^{k} \frac{G(z)}{K} r_0^{(i)}\left(\frac{G(z)}{\sqrt{K}}\right) + \frac{1}{G(z)} \quad (5.40)$$

  In the large K limit the argument of $r_0^{(i)}$ tends to zero. Denoting $\lim_{K\to\infty} \frac{1}{K}\sum_{i=1}^{K} r_0^{(i)}(0) \equiv \kappa_2$ we arrive at the quadratic equation

$$G(z) = \kappa_2 G(z) + \frac{1}{G(z)} \tag{5.41}$$

therefore at the Wigner's semicircle $\rho(\lambda) = \frac{1}{2\pi\kappa_2}\sqrt{\lambda^2 - 4\kappa_2}$. We have therefore confirm, that Wigner semicircle is the result of free central limit theorem, alike the Gaussian distribution is the result of central limit theorem.

Chapter 6

Free random variables - Multiplication laws

In previous chapter, we have seen how the concept of addition of independent random variables in classical probability can evolve into the idea of addition of large, non-commuting free matrices, i.e. the the additive R-transform. In this chapter we formulate similar analogy, but for the multiplication of random variables.

6.0.1 Multiplication law in classical probability

As previously, let us start from the classical probability theory. We consider two random variables x_1, x_2 from corresponding, independent, known pdfs $p_i(x_i)$, where $i = 1, 2$, and we ask, how we can infer from this knowledge the form of the pdf of the product $y = x_1 x_2$. Mathematically, we formulate this question as

$$p(y) = \int dx_1 dx_2 p_1(x_1) p_2(x_2) \delta(y - x_1 x_2) = \int \frac{dx}{x} p_1(\frac{y}{x}) p_2(x) \tag{6.1}$$

The way to unravel the distribution is to apply a Mellin transform, defined as

$$m(t) = \int_0^\infty dx x^{t-1} p(x) \tag{6.2}$$

Indeed, calculating the Mellin transform of $p(y)$, and using the first equality in (6.1), i.e. the one with Dirac delta, we arrive at

$$\int dy y^{t-1} p(y) \equiv m(t) = m_1(t) m_2(t) \tag{6.3}$$

where $m_i(t) = \int_0^\infty dx_i x_i^{t-1} p_i(x_i)$, for $i = 1, 2$. To get the desired $p(y)$ in the last step we have to use the inverse Mellin transform, defined as

$$p(y) = \frac{1}{2\pi i} \int_\Gamma m(t) t^{-y} dt \tag{6.4}$$

where Γ denotes the integration from minus infinity to plus infinity, along the imaginary axes, for any real $t = c$ belonging to the analyticity strip of $m(t)$,

Mellin transform is defined for positive $p(x)$, which seems to be a problem for several classical pdf's, including the Gaussian distribution. One can circumvent this problem using the following trick (Epstein). Let us take as an example two identical, independent Gaussian distributions, $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Let us rewrite $p_1(x_1) = p_{1+}(x_1) + p_{1-}(x_1)$, where p_{1+} is the positive "half" of the Gaussian distribution, i.e. the distribution, which is, by definition, zero for negative arguments. Similarly, p_{1-} is the negative "half" of the Gaussian, i.e. the distribution, which is, by definition, zero for positive argument. We perform similar decomposition for the second Gaussian, arriving in this way to four contributions to $p(y)$, two of them defined on the positive part of the y axis, two of them defined on the negative part. For example,

$$m_{1+}(t) = \int_0^\infty x^{t-1} p_{1+}(x) = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{t-1} e^{-\frac{t^2}{2}} dt = \frac{2^{(t-3)/2}}{\sqrt{\pi}} \Gamma(t/2), \quad t > 0 \tag{6.5}$$

so the product of Mellin transforms reads $m_{++}(t) = m_{1+}(t) m_{2+}(t) = \frac{2^{t-3}}{\pi} \Gamma^2(t/2)$ for $t > 0$. The contribution $p_{++}(y)$, where $y > 0$ is given by the inverse Mellin transform,

$$p_{++}(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-t} m_{++}(t) = \frac{1}{2\pi} K_0(y) \quad y > 0 \tag{6.6}$$

where $K_0(y)$ is a modified Bessel function, and c is any positive real number, since the Euler's Gamma function is analytic for positive arguments. Similarly we consider three other combinations m_{+-}, m_{-+}, m_{--}, actually, yielding to identical result like m_{++}. The final answer for the distribution of the product of independent Gaussian variables reads therefore

$$p(y) = p_{++}(y) + p_{+-}(y) + p_{-+}(y) + p_{--}(y) = \frac{1}{\pi} K_0(y) \quad y \in (-\infty, +\infty) \tag{6.7}$$

This distribution has a integrable singularity at $y = 0$, since diverges at zero as $-\ln|y|$, and dies at $\pm\infty$ alike $\sqrt{\frac{\pi}{2|y|}}e^{-|y|}$.

Similarly, we may ask the question, what is the distribution of the quotient (ratio) of two, independent random variables. Since the ratio $q \equiv x_1/x_2$ we may rewrite as multiplication $x_1 \cdot \frac{1}{x_2}$, the Mellin transform for $\int p(x_2)(1/x_2)^{t-1}dx_2$ reads, by definition, $m_2(2-t)$. The distribution of the ratio is therefore given by

$$p(q) = \frac{1}{2\pi i}\int_\Gamma q^{-t} m_1(t) m_2(2-t) dt \tag{6.8}$$

Let us consider an example: the ratio of two independent Gaussian distributions. Repeating similar decomposition into positive and negative parts alike above, we arrive at

$$\begin{aligned} m_{++}(t) &= m_{1+}(t)\cdot m_{2+}(2-t) = \\ &= \frac{2^{(t-3)/2}}{\sqrt{\pi}}\Gamma(t/2)\cdot\frac{2^{(-t-1)/2}}{\sqrt{\pi}}\Gamma(1-t/2) = \frac{1}{4}\frac{1}{\sin t\pi/2} \end{aligned} \tag{6.9}$$

where in the last line we have used the property of Euler's Gamma function $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}$. The inverse Mellin transform reads therefore

$$p_{++}(q) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} q^{-t} m_{++}(t) = \frac{1}{2\pi}\frac{1}{1+q^2}, \quad q > 0 \tag{6.10}$$

where $c \in (0,2)$. Adding similar contributions from decomposition (actually, all equal to each other via symmetries) we arrive at

$$p(q) = p_{++}(q) + p_{+-}(q) + p_{-+}(q) + p_{--}(q) = \frac{1}{\pi}\frac{1}{1+q^2} \quad q \in (-\infty, +\infty) \tag{6.11}$$

The distribution of the ratio of two Gaussian variables is given by the Cauchy distribution. Note, that all moments of Cauchy distribution are divergent.

These two examples show, that multiplication of pdfs is less intuitive comparing to the addition of pdfs. Now, we would like to paralel similar construction in the case of free random matrices.

6.0.2 Multiplication in free probability

The main problem for determining the spectral distribution for the product of random matrices comes from non-commutativity: even if we multiply two

hermitian matrices A and B, the resulting matrix $H = AB$ is not hermitian, since $H^\dagger = (AB)^\dagger = B^\dagger A^\dagger = BA \neq AB$, so the spectrum of H is usually complex. In the next chapter we will see how to extend the free random variable calculus also in such case. Luckily, if at least one of the matrices A and/or B is positive, we may play the following trick. To calculate the Greens function $G_H(z) = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z-AB}\right\rangle$ we have to know all the moments $m_n = \frac{1}{N}\langle \mathrm{tr}(AB)^n\rangle$. If A is positive, due to the cyclic properties of the trace we have $m_n = \frac{1}{N}\left\langle \mathrm{tr}(A^{1/2}BA^{1/2})^n\right\rangle$. But matrix $H' = A^{1/2}BA^{1/2}$ is hermitian, contrary to $H = AB$, which allows us to formulate the following multiplication law expressed in terms of so-called Voiculescu S-transform, provided we know the spectral moments of A and B.

The multiplication law, expressed in terms of the S transform, reads [**?**]

$$S_{A\cdot B}(z) = S_A(z)S_B(z) \tag{6.12}$$

and the S transform is defined by

$$S(z) = \frac{1+z}{z}\chi(z), \quad \text{where} \quad \chi\left(zG(z)-1\right) = \frac{1}{z}\ . \tag{6.13}$$

or equivalently, $\frac{1}{\chi(w)}G[\frac{1}{\chi(w)}] - 1 = w$.
The algorithm for "multiplication" is similar to that for "addition":
(i) Knowing $G_A(z)$ and $G_B(z)$, calculate $S_A(z)$ and $S_B(z)$ using (6.13).
(ii) Use the multiplication law (6.12).
(iii) Use again (6.13) to derive $G_{AB}(z)$ for the product of AB. We see, that the S-transform in free probability is the analogue of the Mellin transform in classical probability calculus.

Let us first derive some useful relations between the R and S transforms. Changing variables $z = yG(y) - 1$ in (6.13) we get

$$S(yG(y)-1) = \frac{1}{y - \frac{1}{G(y)}}\ . \tag{6.14}$$

Using $G(y) = (y - \Sigma(y))^{-1}$, we can rewrite the last equation as

$$S\left(G(y)\Sigma(y)\right) = \frac{1}{\Sigma(y)}\ . \tag{6.15}$$

Setting $\Sigma(z) = R(G(z))$ and taking the reciprocals of both sides we arrive at

$$\frac{1}{S\left(G(y)R(G(y))\right)} = R(G(y))\ . \tag{6.16}$$

Changing variables once again to $z = G(y)$ we obtain the equation

$$R(z) = \frac{1}{S\left(zR(z)\right)} \tag{6.17}$$

which gives an explicit relation between the R and S transforms. The S transform can be defined only if the R transform does not vanish at the origin: $R(0) \neq 0$. This corresponds to random matrices with a non-vanishing first moment (cumulant) $\frac{1}{N}\langle \mathrm{Tr} H\rangle = \frac{1}{N}\langle\langle \mathrm{Tr} H\rangle\rangle \neq 0$. Otherwise the S transform cannot be defined as a power series and all the manipulations presented above break down. The last equation can be inverted. Let us introduce a new variable $y = zR(z)$. Now (6.17) reads

$$S(y) = \frac{1}{R\left(\frac{y}{R(z)}\right)} = \frac{1}{R\left(\frac{y}{R\left(\frac{y}{R(z)}\right)}\right)} = \frac{1}{R\left(\frac{y}{R\left(\frac{y}{R(\ldots)}\right)}\right)} \tag{6.18}$$

where z can be recursively eliminated by repeating the substitution $z = \frac{y}{R(z)}$ ad infinitum. This leads to a function which is nested infinitely many times forming a sort of continued fraction. The last equation can alternatively be written as

$$S(z) = \frac{1}{R\left(zS(z)\right)} \tag{6.19}$$

which is an inverse formula to (6.17). The two equations can written in a symmetric way as mutually inverse maps

$$z = yS(y) \quad \text{and} \quad y = zR(z)\ . \tag{6.20}$$

As an example we consider a Wishart random matrix $W = \frac{1}{T}XX^\dagger$. In this case, $R_(z) = \frac{1}{1-rz}$, where r is the rectangularity (ratio of the number of rows N of X to the number of columns T of X. Using (6.19) we obtain

$$S(z) = 1 - rzS(z) \tag{6.21}$$

and hence $S(z) = \frac{1}{1+rz}$.

6.0.3 Diagrammatic derivation of the multiplication law

We are now ready to diagrammatically derive the S transform and the corresponding multiplication law. The initial point of the construction is to

consider a $2N \times 2N$ block matrix $\mathcal{H}$ and its even powers

$$\mathcal{H} = \left(\begin{array}{c|c} 0 & A \\ \hline B & 0 \end{array} \right) , \quad \mathcal{H}^{2k} = \left(\begin{array}{c|c} (AB)^k & 0 \\ \hline 0 & (BA)^k \end{array} \right) . \tag{6.22}$$

The upper-left corner of $\mathcal{H}^{2k}$ involves solely the powers of AB, which we are interested in. In order to have an access to the traces of individual blocks in the matrix we introduce the block trace operation defined as follows.

$$\mathrm{Tr}_{\mathrm{b}2} \left(\begin{array}{c|c} X & Y \\ \hline Z & V \end{array} \right)_{2N \times 2N} \equiv \left(\begin{array}{c|c} \mathrm{Tr}\ X & \mathrm{Tr}\ Y \\ \hline \mathrm{Tr}\ Z & \mathrm{Tr}\ V \end{array} \right)_{2 \times 2} . \tag{6.23}$$

which reduces $2N \times 2N$ matrices to 2×2 ones. The upper-left corner of the reduced matrix $\mathrm{Tr}_{\mathrm{b}2}\mathcal{H}^{2k}$ is equal $\mathrm{Tr}(AB)^k$ while of $\mathrm{Tr}_{\mathrm{b}2}\mathcal{H}^{2k+1}$ is equal zero. So now the idea is to reformulate the problem of calculating the Green's function for the product

$$G_{AB}(z) = \frac{1}{N} \left\langle \mathrm{Tr} \frac{1}{z1 - AB} \right\rangle \tag{6.24}$$

as a problem of calculating the upper-left corner of the Green's function $\mathcal{G}(w)$ for the matrx $\mathcal{H}$:

$$\mathcal{G}(w) = \left(\begin{array}{cc} \mathcal{G}_{11}(w) & \mathcal{G}_{12}(w) \\ \mathcal{G}_{21}(w) & \mathcal{G}_{22}(w) \end{array} \right) = \frac{1}{N} \left\langle \mathrm{Tr}_{\mathrm{b}2} \frac{1}{w1 - \mathcal{H}} \right\rangle . \tag{6.25}$$

where w is a complex number and 1 is a unity matrix of dimensions $2N \times 2N$. One can easily check that

$$G_{AB}(z = w^2) = \frac{\mathcal{G}_{11}(w)}{w} \tag{6.26}$$

since only every second (even) power of $\mathcal{G}(w)$ contributes to the power expansion of $\mathcal{G}_{11}(w)$, which is thus a power expansion in $z = w^2$.

The next step is to define self-energy $\Sigma(w)$. It is a 2×2 matrix

$$\Sigma(w) \equiv \left(\begin{array}{cc} \Sigma_{11}(w) & \Sigma_{12}(w) \\ \Sigma_{21}(w) & \Sigma_{22}(w) \end{array} \right) \tag{6.27}$$

that is related to the Green's function as

$$\mathcal{G}(w) = (w1 - \Sigma(w))^{-1} \tag{6.28}$$

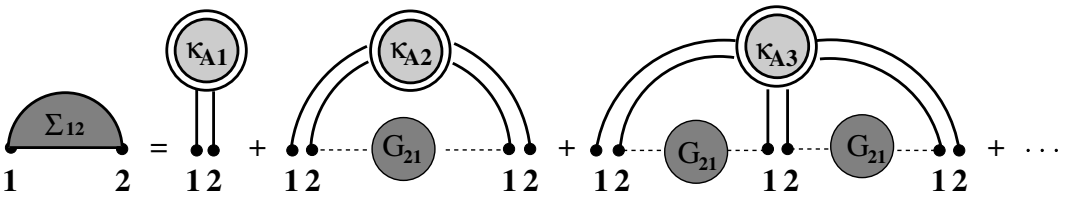


Figure 6.1: The spider diagrams correspond to free cumulants generated by the matrix $A = \mathcal{H}_{12}$ and therefore the double dots on the horizontal line are indexed by 12. So on the horizontal line we have alternating indices $1212\ldots12$ and this enforces all the $\mathcal{G}$-bubbles to have indices 21 as one can see in the figure. Therefore there are only $\mathcal{G}_{21}$ bubbles in the diagram and the corresponding equation is $\Sigma_{12}(w) = R_A(\mathcal{G}_{21})$. The analogous equation for B-cumulants is $\Sigma_{21}(w) = R_B(\mathcal{G}_{12})$. Similarly, one can see that $\Sigma_{11}(w) = \Sigma_{22}(w) = 0$ since one of the double dots on the horizontal line would need to have identical indices, for which as we know from (6.22) the double line is equal zero.

in analogy to the case from previous chapter, $G(w) = 1/(w - \Sigma(w))$. All the matrices in the last equation are of 2×2 dimensions. This is the first Dyson-Schwinger equation. To write down the second one – a counterpart of $\Sigma(z) = R[G(z)]$ from the previous chapter, it is convenient to use its diagrammatic representation as that in figure 4.8. Instead of a scalar equation $\Sigma(z) = R[G(z)]$ we will have a matrix equation for 2×2 matrices $\mathcal{G}$ and Σ (6.25) and (6.27). Since we have now have 2×2 matrices it is crucial to work out the index structure of the corresponding equation. This structure stems from the correspondence $A \leftrightarrow \mathcal{H}_{12}$ and $B \leftrightarrow \mathcal{H}_{21}$ that follows from the position of the blocks in $\mathcal{H}$ (6.22).

The only non-vanishing cumulants are $\frac{1}{N}\langle\langle \mathrm{Tr}\mathcal{H}_{12}^n\rangle\rangle = \frac{1}{N}\langle\langle \mathrm{Tr}A^n\rangle\rangle \equiv \kappa_{A,n}$ or $\frac{1}{N}\langle\langle \mathrm{Tr}\mathcal{H}_{21}^n\rangle\rangle = \frac{1}{N}\langle\langle \mathrm{Tr}B^n\rangle\rangle \equiv \kappa_{B,n}$ while all mixed ones vanish due to freeness, as we discussed in the previous chapter. Due to this, the index structure of non-vanishing one-line-irreducible diagrams is restricted to that shown in figure 6.1 and its counterpart obtained by exchanging $1 \leftrightarrow 2$ and $A \leftrightarrow B$. The diagrammatic equations discussed in figure 6.1 can be sumarized as

$$\Sigma(w) = \begin{pmatrix} 0 & R_A(\mathcal{G}_{21}(w)) \\ R_B(\mathcal{G}_{12}(w)) & 0 \end{pmatrix} . \tag{6.29}$$

Inserting this into (6.28) yields

$$\begin{pmatrix} \mathcal{G}_{11}(w) & \mathcal{G}_{12}(w) \\ \mathcal{G}_{21}(w) & \mathcal{G}_{22}(w) \end{pmatrix} = \begin{pmatrix} w & -R_A\left(\mathcal{G}_{21}(w)\right) \\ -R_B\left(\mathcal{G}_{12}(w)\right) & w \end{pmatrix}^{-1} \tag{6.30}$$

which gives a direct relation between the Green's function $\mathcal{G}(w)$ and the R transform.

We shall now rewrite this equation in a way which explicitly exhibits multiplicative structure. Note that in the following manipulations we do not need to assume anything about the first moment i.e. whether the ensemble is centred or not. Inverting the matrix on the right hand side we obtain

$$\mathcal{G}_{12}(w) = \frac{1}{\mathrm{Det}} R_A(\mathcal{G}_{21}(w)) \ , \quad \mathcal{G}_{21}(w) = \frac{1}{\mathrm{Det}} R_B(\mathcal{G}_{12}(w)) \tag{6.31}$$

$$\mathcal{G}_{11}(w) = \mathcal{G}_{22}(w) = \frac{w}{\mathrm{Det}} \ . \tag{6.32}$$

where Det is the determinant of the matrix, $w1 - \Sigma(w)$, on the right hand side of (6.30):

$$\mathrm{Det} = w^2 - R_A\left(\mathcal{G}_{21}(w)\right) R_B\left(\mathcal{G}_{12}(w)\right) \ . \tag{6.33}$$

Inserting two last equations to (6.26) we obtain

$$G_{AB}(z) = \frac{\mathcal{G}_{11}(w)}{w} = \frac{1}{\mathrm{Det}} = \frac{1}{z - R_A\left(\mathcal{G}_{21}(w)\right) R_B\left(\mathcal{G}_{12}(w)\right)} \tag{6.34}$$

where $z = w^2$. Comparing the denominator in this equation to that of the standard equation $G_{AB}(z) = 1/(z - R_{AB}(G_{AB}(z))$ we get

$$R_{AB}\left(G_{AB}(z)\right) = R_A\left(\mathcal{G}_{21}(w)\right) R_B\left(\mathcal{G}_{12}(w)\right) \ . \tag{6.35}$$

At this stage we see the first hint of a multiplicative structure emergence. In order to complete this equation we also need (6.31). Let us set $g = G_{AB}(z)$, $g_A = \mathcal{G}_{12}(w)$ and $g_B = \mathcal{G}_{21}(w)$ to simplify arguments in the R transforms in the last equation. Using this substitution we can write (6.35) and (6.31) in a compact form as a closed set of equations for the R transform of the product

$$R_{AB}(g) = R_A(g_B) R_B(g_A) \tag{6.36}$$

and

$$g_A = gR_A(g_B) \ , \quad g_B = gR_B(g_A) \ . \tag{6.37}$$

This is the multiplication law formulated in terms of the R transform. Its main advantage in comparison to the S transform is that it can be applied even to centred ensembles (i.e. having vanishing mean) including the case when both are centred (see the examples in next chapter).

The difference with respect to the conventional multiplication law $S_{AB}(z) = S_A(z)S_B(z)$ is that the individual factors appearing in (6.36) are not expressed uniquely in terms of the properties of a *single* random matrix ensemble e.g. the factor $R_A(\cdot)$ is evaluated on g_B which is related to the ensemble B. However it is straightforward to obtain from (6.36)-(6.37) the conventional multiplication law as we shall illustrate below.

Let us introduce a new variable $y = gR_{AB}(g)$. We can now express g_B – the argument of R_A purely in terms of the properties of ensemble A:

$$g_B = gR_B(g_A) = g\frac{R_{AB}(g)}{R_A(g_B)} = \frac{y}{R_A(g_B)} = \frac{y}{R_A\left(\frac{y}{R_A(\ldots)}\right)} \tag{6.38}$$

Now each of the factors in (6.36) depends on a single ensemble. We may do the same for the left hand side, which becomes (6.18)

$$R_{AB}\left(\frac{y}{R_{AB}(g)}\right) = R_{AB}\left(\frac{y}{R_{AB}\left(\frac{y}{R_{AB}(\ldots)}\right)}\right) = \frac{1}{S_{AB}(y)} \tag{6.39}$$

Putting these formulas together, we can finally write (6.36) using only the variable y [**?**]

$$R_{AB}\left(\frac{y}{R_{AB}\left(\frac{y}{R_{AB}(\ldots)}\right)}\right) = R_A\left(\frac{y}{R_A\left(\frac{y}{R_A(\ldots)}\right)}\right) R_B\left(\frac{y}{R_B\left(\frac{y}{R_B(\ldots)}\right)}\right) \tag{6.40}$$

which amounts to the standard formulation for the multiplication law [**?**] as follows from (6.18)

$$S_{AB}(y) = S_A(y)S_B(y) \ . \tag{6.41}$$

The necessity of assuming noncentred distributions comes from the fact that the implicit continued fractions appearing in (6.18) make sense only for $R(z) \sim const + \mathcal{O}(z)$ with nonzero constant term [**?**].

6.0.4 Division in free probability

Since division of two free matrices X_1/X_2 is equivalent to multiplication of X_1 and X_2^{-1}, we may use the concept of S-transform, provided that for positive X_2 we know $S_{X_2^{-1}}(z)$. First, note that it is easy to relate Green's functions for X and X^{-1}. Indeed

$$\begin{aligned} G_{X^{-1}}(z) &= \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z-X^{-1}}\right\rangle = \frac{1}{N}\left\langle \mathrm{tr}\frac{X}{zX-1}\right\rangle = \\ &= \frac{1}{zN}\left\langle \mathrm{tr}\frac{ZX-1+1}{zX-1}\right\rangle = \frac{1}{z}-\frac{1}{zN}\left\langle \mathrm{tr}\frac{1}{1-zX}\right\rangle = \\ &= \frac{1}{z}-\frac{1}{z^2}G_X(\frac{1}{z}) \end{aligned} \tag{6.42}$$

Let us define generic $M_A(z) = zG_A(z) - 1 = \sum_{k=1}^{\infty} m_k z^{-k}$ where moments $m_k = \frac{1}{N}\left\langle \mathrm{tr}A^k\right\rangle$. Then, equation (6.42) reads

$$M_{X^{-1}}(z) + M_X(1/z) = -1 \tag{6.43}$$

Subsequently, we define the generic functional inverse of M_A as N_A, i.e. $M_A[N_A(w)] = w$ or $N_A[M_A(z)] = z$. Note that in standard notation of free random variables $N_A(z) \equiv 1/\chi_A(z)$. First, setting $1/z = N_X(w)$ we get $w - M_{X^{-1}}[1/N_X(w)] = -1$, or

$$M_{X^{-1}}[1/N_X(w)] = -1 - w. \tag{6.44}$$

Second, we apply $N_{X^{-1}}$ to both sides of the above equation, which yields

$$1/N_X(w) = N_{X^{-1}}(-1-w) \tag{6.45}$$

Finally, using the fact that $S(w) = \frac{1+w}{w}\frac{1}{N(w)}$, we rewrite the last formula as our final relation

$$S_X(w)\cdot S_{X^{-1}}(-1-w) = 1 \tag{6.46}$$

or, renaming $w \equiv -1 - z$,

$$S_X(-1-z)\cdot S_{X^{-1}}(z) = 1 \tag{6.47}$$

As an example, we consider Marchenko-Pastur distribution with rectangularity r. Since $S_{MP}(w) = \frac{1}{1+rw}$ we read out from (6.47) that $S_{MP^{-1}}(w) =$

$1 - r - rw$.

Exercise (optional) The Reader may easily calculate now the spectral distribution of so-called free F-distribution, i.e. the spectral distribution of the quotient X_1/X_2, where X_1, X_2 come from two free Wishart distributions, with rectangularities r_1 and r_2, respectively.

Chapter 7

Beyond Stationarity

7.1 Time, Spectral Viscosity and Microscopic Universality

In 1962, Dyson suggested a galvanizing way to understand the jpdf of the eigenvalues of Gaussian random matrices [1]. In order to find jpdf, he introduced an auxiliary dynamics in some fictitious "time", which, in the large "time" limit, led to the stationary state (Gibbs state) representing the desired jpdf. In his own words, he pointed [**?**], *after considerable and fruitless efforts to develop a Newtonian theory of ensembles, we discovered that the correct procedure is quite different and much simpler. The* x_i[eigenvalues] *should be interpreted as positions of particles in Brownian motion.* The resulting stationary distributions (originally for hermitian or for unitary random matrices) were obtained as a result of Ornstein-Uhlenbeck diffusion with a drift force coming from electrostatic-like repulsion of eigenvalues. The success of this description has contributed to multiple applications of random matrix models in practically all branches of science. The notion of "time" has evolved as well, so nowadays it can be a physical parameter, representing either the real time or, e.g., the length of a mesoscopic wire, the area of a string or an external temperature. In this chapter, we will follow this strategy, applied to so-called **characteristic polynomial** $\pi(z)$, defined as

$$\pi_N(z) = \langle \det(z - X) \rangle = \left\langle \prod_i^N (z - \lambda_i) \right\rangle \tag{7.1}$$

[1] Note that has happened before the orthogonal polynomial method was invented.

First, we will see, that that resulting dynamical equation for $\pi(z)$ is given by the diffusion equation, with diffusive constant behaving like $1/N$. We will then link this equation to so-called viscid Burgers equation, known from the toy-models of turbulence. We will point at the singular limit of $N \to \infty$ of this equation, and we will identify spectral shock waves corresponding to the breakdown of the $1/N$ expansion. The same procedure will allow us also to study large N limit of orthogonal polynomials, and to write down the form of universal kernels, both in the vicinity of the shocks waves, and elsewhere (in "the bulk"). This surprising link stems from the fact, that characteristic and orthogonal polynomials are identical!

$$\pi_N(z) = P_N(z) \tag{7.2}$$

Below, we sketch the proof of this result. By definition,

$$P_N(z) = \frac{1}{Q_N} \int \prod d\lambda_i e^{-\frac{N}{2}\sum_i V(\lambda_i)} \prod_i (z-\lambda_i) \prod_{i<j} (\lambda_i - \lambda_j)^2. \tag{7.3}$$

Let us substitute $z \to \lambda_{N+1}$ in $P_N(z)$, then multiply the $P_N(\lambda_{N+1})$ by $Q_N \lambda_{N+1}^k$, and integrate over the λ_{N+1} with the weight $w(\lambda_{N+1}) = e^{-NV(\lambda_{N+1})/2}$. The considered integral reads

$$I = \int \prod_{i=1}^{N+1} w(\lambda_i) d\lambda_i \Delta_{N+1}(\lambda_1, .., \lambda_{N+1}) \Delta_N(\lambda_1, ..., \lambda_N) \lambda_{N+1}^k \tag{7.4}$$

where we have included the factor $\prod_i (\lambda_{N+1} - \lambda_i)$ into the definition of the first Δ_N from the squared term. We expand now $\Delta_N \lambda_N^k$ using the definition of the determinant. Note that, due to the symmetries of the second determinant, we can focus only on the single permutation, let us say on $\lambda_1^0 \lambda_2^1 .... \lambda_N^{N-1} \lambda_{N+1}^k$. For such a choice, we absorb the product in such a way, that the j-th column is multiplied by a factor λ_j^{j-1}, and the last column is multiplied by the λ_{N+1}^k. The resulting expression is equivalent, modulo rescaling $(N+1)!/N! = N+1$ to the expansion of the beneath written determinant D with the respect of the last column

$$D = \det \begin{pmatrix} 1 & \lambda_1 & ... & \lambda_1^{N-1} & \lambda_1^k \\ 1 & \lambda_2 & ... & \lambda_2^{N-1} & \lambda_2^k \\ ... & ... & ... & ... & ... \\ ... & ... & ... & ... & ... \\ 1 & \lambda_{N+1} & ... & \lambda_{N+1}^{N-1} & \lambda_{N+1}^k \end{pmatrix} \tag{7.5}$$

We see that for each $k = 0, ..., N-1$ two columns of the determinant are identical. For $k = N$, we arrive at

$$Q_N \int d\lambda w(\lambda) P_N(\lambda) \lambda^N = \frac{Q_{N+1}}{N+1} \tag{7.6}$$

hence

$$\int d\lambda w(\lambda) P_N(\lambda) P_k(\lambda) = \delta_{N,k} \frac{Q_{N+1}}{(N+1)Q_N} \tag{7.7}$$

which, after normalization, reproduces gives orthogonal polynomials condition.

7.1.1 Dynamics

We introduce an N by N hermitian matrix H, where $H_{ij} = x_{ij} + iy_{ij}$ for $i \neq j$ and $H_{ii} = x_{ii}$. To ensure hermiticity, $x_{ij} = x_{ji}$ and $y_{ij} = -y_{ji}$. Let each x_{ij} and each y_{ij} perform now white noise driven, independent random walk, such that

$$\delta H_{ij} = -aH_{ij}\delta t, \quad \left\langle |\delta H_{ij}|^2 \right\rangle = \frac{1+\delta_{ij}}{N} \delta t \tag{7.8}$$

Introducing the probability $p(x_{ii}, t)$ that the diagonal entry H_{ii} is equal to x_{ii} at time t, and, nalogically, similar object $p_(x_{ij}, t)p(y_{ij}, t)$ for of-diagonal elements H_{ij}, we can write now the evolution equations for these probabilities

$$\begin{aligned}
\partial_t p(x_{ii}, t) &= \left(\frac{1}{2N} \partial^2_{x_{ii}} + a\partial_{x_{ii}} x_{ii} \right) p(x_{ii}, t) \\
\partial_t p(x_{ij}, t) &= \left(\frac{1}{4N} \partial^2_{x_{ij}} + a\partial_{x_{ij}} x_{ij} \right) p(x_{ij}, t) \qquad (i < j) \\
\partial_t p(y_{ij}, t) &= \left(\frac{1}{4N} \partial^2_{y_{ij}} + a\partial_{y_{ij}} y_{ij} \right) p(y_{ij}, t) \qquad (i < j)
\end{aligned} \tag{7.9}$$

These equations are Smoluchowski-Fokker-Planck (hereafter SFP)equations with harmonic potential of strength a confining the diffusion of the matrix elements x_{ij} and y_{ij}, i.e. they represent Ornstein-Uhlenbeck process. Since each probability in (7.9) is independent, probability for the whole matrix H is just the product of individual probabilities, i.e. $P(H, t) \equiv$

$\prod_{k=1}^{N} p(x_{kk},t) \prod_{i<j;i,j=1}^{N} p(x_{ij},t)p(y_{ij},t)$. Probability $P(H,t)$ fulfills therefore the following evolution equation

$$\partial_t P(H,t) = L(H)P(H,t) \tag{7.10}$$

where SFP operator reads explicitly

$$\begin{aligned} L(H) &= \sum_{i=1}^{N} \left(\frac{1}{2N} \partial_{x_{ii}}^2 + a \partial_{x_{ii}} x_{ii} \right) \\ &+ \frac{1}{4N} \sum_{i<j}^{N} \left(\partial_{x_{ij}}^2 + \partial_{y_{ij}}^2 \right) \\ &+ a \sum_{i<j}^{N} \left(\partial_{x_{ij}} x_{ij} + \partial_{y_{ij}} y_{ij} \right) \end{aligned} \tag{7.11}$$

The solution of the SFP equation (7.10) reads therefore

$$P(H,t) = C \exp\left(-\frac{Na}{1-e^{-2at}} \mathrm{tr}(H - H_0 e^{-at})^2 \right) \tag{7.12}$$

where $H_0 = H(t=0)$ is the initial condition of the evolution and C is an irrelevant here normalization constant. Note, that choosing $a = \frac{1}{2}$ and performing the limit $t \to \infty$ we recover the GUE probability (1.15).

7.1.2 Evolution of the characteristic determinant

We define now time-dependent characteristic determinant

$$\pi_N(z,t) = \langle \det(z-H) \rangle_t \tag{7.13}$$

where $\langle ... \rangle_t$ denotes averaging with respect to time-dependent measure (7.12). A simple trick based on Grassmann variables (see Appendix A2) allows to write down the dynamical equation for $\pi_N(z,t)$

$$\partial_t \pi_N(z,t) = -\frac{1}{2N} \partial_{zz} \pi_N(z,t) + az \partial_z \pi_N(z,t) - aN \pi_N(z,t) \tag{7.14}$$

A smart change of variables simplifies this equation considerably. Applying so-called Lamperti transformations

$$\begin{aligned} &\tilde{z} = e^{at} z \\ &\tilde{t} = \frac{1}{2a} \left(e^{2at} - 1 \right) \\ &\tilde{\pi}_N(\tilde{z},\tilde{t})(1+2a\tilde{t})^{-N/2} = \pi_N(z,t) \end{aligned} \tag{7.15}$$

after straightforward, but time-consuming calculations we arrive at the following result in "tilded" variables.

$$\partial_{\tilde{t}}\tilde{\pi}_N = -\frac{1}{2N}\partial_{\tilde{z}\tilde{z}}\tilde{\pi}_N \tag{7.16}$$

This equation looks like a diffusion equation on complex plane $\tilde{z}$, but with unusual sign of the diffusion constant. Moreover, it looks like a diffusion constant vanishes naively in the large N limit! To unravel these puzzles, we apply the complex version of so-called Cole-Hopf transformation

$$\tilde{f}_N = \frac{1}{N}\partial_{\tilde{z}} \ln \tilde{\pi}_N \tag{7.17}$$

This transformation maps our diffusion equation onto *complex* viscid Burgers equation

$$\partial_{\tilde{t}}\tilde{f}_N + \tilde{f}_N\partial\tilde{z}\tilde{f}_N = \nu_s\partial_{\tilde{z}\tilde{z}}\tilde{f}_N \tag{7.18}$$

where the *negative* spectral viscosity $\nu_s = -\frac{1}{2N}$. Since similar equation on the real axis is one of the benchmarks for describing shock waves in toy-models of turbulence, we expect that the flow of the eigenvalues described by (7.16) or (7.18) may be understood using the "hydrodynamic" concepts. To achieve this goal, we first review the basic facts on the Burgers equation.

7.1.3 Burgers equation trivia

7.1.4 Spectral shock waves in random matrix theory

7.1.5 Universal kernels - edge singularity

7.1.6 Universal kernels - bulk

7.1.7 Analogies to diffraction phenomena

7.1.8 Shock waves in other ensembles

7.1.9 Relation to Itzykson-Zuber formula

Chapter 8

Beyond Hermiticity - complex spectra

iopams cite url color
graphicx epstopdf amssymb color
multirow graphicx hyperref graphicx [normalem]ulem color enumitem fancyhdr braket

8.1 Introduction

Contrary to the case of the hermitian ensembles, the spectrum of non-hermitian matrices is genuinely complex. Let us define the simplest example, the so-called complex Ginibre Ensemble where each element of the $N \times N$ matrix X is drawn from a complex Gaussian distribution. That is, each entry $X_{ij} = x_{ij} + iy_{ij}$ consists of x_{ij} and y_{ij} drawn from standard Gaussian distributions. Note that all moments $\langle \mathrm{Tr} X^n \rangle$ (and $\left\langle \mathrm{Tr}(X^\dagger)^n \right\rangle$) vanish because $\langle \mathrm{Tr} X^2 \rangle = 0$. The only non vanishing moments are the mixed ones, i.e. $\left\langle \mathrm{Tr}(XX^\dagger)^n \right\rangle$. As we will show, in the large N limit, the eigenvalues condense uniformly on the centered disc on the complex plane. Therefore, the spectrum exhibits a jump at the rim, contrary to the hermitian cases, when the real spectrum is continuous at the endpoints, and only the derivatives of the spectrum are discontinuous. Moreover, the spectrum is non-analytic inside the disc, which seems to disqualify all the methods based on analyticity of the complex variable z. This is best visible, when we try to repeat the hermitian construction for the Green's function $G(z) = \frac{1}{N}\left\langle \mathrm{Tr}\frac{1}{z-X} \right\rangle$. Since all

moments vanish, such a Green's function is simply equal to $G(z) = \frac{1}{z}$, and does not reflect correctly the spectral properties of the ensemble. Similarly, the characteristic determinant is trivial, $\langle \det(z - X) \rangle = z^N$.

The way out, based on an electrostatic analogy, was suggested a long time ago [**?**]. We define an electrostatic potential

$$\Phi \equiv \Phi(z) = \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{N} \left\langle \mathrm{Tr} \ln[|z - X|^2 + \epsilon^2] \right\rangle \ , \tag{8.1}$$

where we use a short-hand notation: $|z - X|^2 + \epsilon^2 = (z\mathbf{1}_N - X)(\bar{z}\mathbf{1}_N - X^\dagger) + \epsilon^2 \mathbf{1}_N$, where $\mathbf{1_N}$ is the N-dimensional identity matrix. Then, we calculate the "electric field" as a gradient of the electrostatic potential,

$$G(z, \bar{z}) = \partial_z \Phi = \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{N} \left\langle \mathrm{Tr} \frac{\bar{z} - X^\dagger}{|z - X|^2 + \epsilon^2} \right\rangle . \tag{8.2}$$

The electric field plays the role of the correct Green's function. Indeed, applying the Gauss law, in the next step,

$$\rho = \frac{1}{\pi} \partial_{\bar{z}} G = \frac{1}{\pi} \partial_{z\bar{z}} \Phi = \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{1}{\pi N} \left\langle \mathrm{Tr} \frac{\epsilon^2}{[|z - X|^2 + \epsilon^2]^2} \right\rangle , \tag{8.3}$$

we recover the spectral density $\rho(z, \bar{z}) = \frac{1}{N} \langle \sum_{i=1}^N \delta^{(2)}(z - \lambda_i) \rangle$, using the known representation of the two-dimensional delta function $\delta^{(2)}(z) = \lim_{\epsilon \to 0} \frac{1}{\pi} \frac{\epsilon^2}{[|z|^2 + \epsilon^2]^2}$. Note that the Gauss law implies the non-analyticity of $G(z, \bar{z})$. It is crucial that the limit $N \to \infty$ is taken first, before taking the infinitesimal regulator ϵ to zero, since only such order provides the necessary coupling between X and $X^\dagger$, reflected in non-vanishing mixed moments. If one took the limits in an opposite order, X and $X^\dagger$ would decouple, and we would obtain a trivial result $G(z) = 1/z$. The bad news, however, is that the Green's function $G(z, \bar{z})$ (8.2) is given by a very complicated expression, without any similarity to the standard form of the resolvent.

One may bypass the difficulty by relying on the the algebraic construction for the so-called generalized Green's functions proposed some time ago [**?**, **?**]. First, we notice that

$$\mathrm{Tr} \ln[|z - X|^2 + \epsilon^2] = \ln \det[|z - X|^2 + \epsilon^2] = \ln \det \begin{pmatrix} z - X & i\epsilon \\ i\epsilon & \bar{z} - X^\dagger \end{pmatrix} , \tag{8.4}$$

where the argument of the last determinant is a $2N \times 2N$ matrix, built out of four $N \times N$ blocks. Let us now define a new operation called a block-trace,

defined as $\mathrm{bTr} \equiv \mathbf{1}_2 \otimes \mathrm{Tr}_{N\times N}$, which acts in the following way:

$$\mathrm{bTr}\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{2N\times 2N} \equiv \begin{pmatrix} \mathrm{Tr}A & \mathrm{Tr}B \\ \mathrm{Tr}C & \mathrm{Tr}D \end{pmatrix}_{2\times 2}, \tag{8.5}$$

converting a $2N \times 2N$ block matrix into a 2×2 matrix built out of ordinary traces. Additionally, we define another pair of block matrices

$$Q = \begin{pmatrix} z & i\bar{w} \\ iw & \bar{z} \end{pmatrix}, \qquad \mathcal{X} = \begin{pmatrix} X & 0 \\ 0 & X^\dagger \end{pmatrix}. \tag{8.6}$$

We are now ready to propose the construction of the generalized resolvent (2×2 matrix)

$$\mathcal{G}(z,w) \equiv \begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{1\bar{1}} \\ \mathcal{G}_{\bar{1}1} & \mathcal{G}_{\bar{1}\bar{1}} \end{pmatrix} = \frac{1}{N}\left\langle \mathrm{bTr}\frac{1}{Q-\mathcal{X}} \right\rangle, \tag{8.7}$$

By construction, $\mathcal{G}_{11}$ is equal to the non-analytic resolvent $G(z,\bar{z})$ (8.2), provided we identify $|w|^2 = \epsilon^2$. Note that the duplication trick allowed us to linearize the problem, since the form of the generalized resolvent (8.7) has formally the form of the standard resolvent for hermitian matrices. One may ask the question, what role is played by the three remaining elements of the matrix $\mathcal{G}$? Let us recall, that the general (non-normal) matrix X is determined in terms of its eigenvalues (Z) and a set of left (L) and right ($|R\rangle$) eigenvectors ($X = \sum_i z_i \, |R_i\rangle \, L_i$), which are bi-orthogonal $L_i|R_j = \delta_{ij}$. By applying a transformation $S = \mathrm{diag}(R, L)$, $S^{-1} = \mathrm{diag}(L^\dagger, R^\dagger)$ (where $L, R, L^\dagger, R^\dagger$ are $N \times N$ matrices built from the corresponding eigenvectors), we notice that

$$\det(Q-\mathcal{X}) = \det[S^{-1}(Q-\mathcal{X})S] = \det\begin{pmatrix} z-Z & i\bar{w}L^\dagger L \\ iwR^\dagger R & \bar{z}-Z^\dagger \end{pmatrix}, \tag{8.8}$$

so the off-diagonal elements of the generalized Green's functions are related to the expectation values of the overlaps of eigenvectors. Indeed, the left-right eigenvector correlator [**?**] reads:

$$O(z,\tau) \equiv \frac{1}{N^2}\left\langle \sum_a O_{aa}\delta^{(2)}(z-z_a) \right\rangle, \tag{8.9}$$

where $O_{ij} = \langle L_i|L_j\rangle \, \langle R_j|R_i\rangle$ is given in the large N limit by the product of off-diagonal elements of $\mathcal{G}$:

$$\lim_{N\to\infty} O(z,\tau) = -\frac{1}{\pi}\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}|_{w=0}. \tag{8.10}$$

as was proven in [**?**]. The appearance of this correlator is a genuine feature of non-hermitian random matrix models, since in the hermitian case left and right eigenvectors coincide and so $O_{ij} = \delta_{ij}$. Finally, for completeness we notice that $\mathcal{G}_{\bar{1}\bar{1}}$ is a complex conjugate of $\mathcal{G}_{11}$ and does not bring any new information.

We would now like to comment on the role of the w variable. In the hermitian case, the method of the resolvent involves the whole complex plane z, despite the fact that the real spectrum comes only as a discontinuity near $z = \lambda \pm i\epsilon$, corresponding to the imaginary part of the resolvent. In the non-hermitian case, the spectrum is complex, but one may be tempted to probe the generalized Green's function with the complex plane w "orthogonal" to the plane z, as schematically depicted on Fig. 8.1. This choice of strategy is reinforced by the above observed coupling of the w plane to eigenvector correlators.

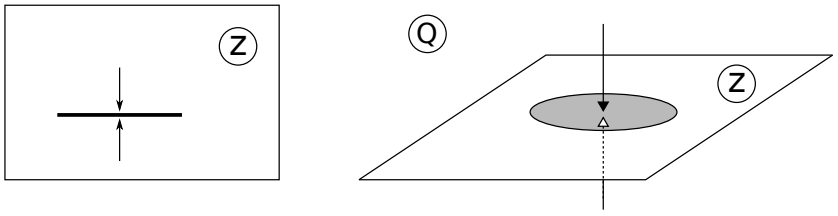


Figure 8.1: Schematic comparison between the domains of hermitian Green's function $G(z)$ and the non-hermitian, generalized Green's function $\mathcal{G}(Q)$. Arrows on the left figure signal the discontinuity of the Green's function when approaching the cut (solid line), arrows on the right figure denote an additional variable w, which in standard approach is treated as only an infinitesimal regulator. Shaded disc represents the non-analytic domain where the eigenvalues condense.

The promotion of the original regulator $i\epsilon$ to a complex variable w has additional advantages. From the algebraic point of view, Q is a quaternion, since $Q = q_0 + i\sigma_j q_j$, where σ_j are Pauli matrices, so $z = q_0 + iq_3$ and $w = q_1 + iq_2$. This fact significantly simplifies the algebraic calculations, since block matrices such as $\mathcal{X}$ and arguments Q naturally appear in non-hermitian random matrix models, e.g. in the generalized Green's function technique [**?**, **?**],in hermitization methods [**?**, **?**, **?**] or in the derivation of the multiplication law for non-hermitian random matrices [**?**]. The above construction was also recently proven rigorously in the mathematical literature [**?**].

8.2 R-transform for nonhermitian ensembles

8.2.1 Large N expansion of the quaternionic Green's function and Feynman diagrams

The procedure for the calculation of the quaternionic Green's function is only slightly modified, compared to the Hermitian case. Again, the unitary invariance of the pdf asserts that the untraced resolvent has a trivial structure $\langle (Q \otimes \mathbf{1} - \mathcal{X})^{-1} \rangle = \mathcal{G} \otimes \mathbf{1}$. We write the geometric series

$$\mathcal{G}(Q) \otimes \mathbf{1} = \mathcal{Q}^{-1} + \left\langle \mathcal{Q}^{-1}\mathcal{X}\mathcal{Q}^{-1} \right\rangle + \left\langle \mathcal{Q}^{-1}\mathcal{X}\mathcal{Q}^{-1}\mathcal{X}\mathcal{Q}^{-1} \right\rangle + \dots, \tag{8.11}$$

where $\mathcal{Q} = Q \otimes \mathbf{1}$. Now, due to the block structure originating from the linearization, all objects in the above expansion, apart from the matrix indices, possess additional indices $(1, \bar{1})$ enumerating the blocks of $\mathcal{X}$ and the elements of the quaternion. The block trace operation taken at the end of the calculations is in fact the partial trace over the matrix space.

To calculate the averages, we decompose the pdf into Gaussian and residual parts. The most general, allowed by the Hermiticity condition, Gaussian part of the potential (**??**) can be written in a convenient form with $\sigma > 0$ and $\tau \in (-1, 1)$

$$P_G(X, X^\dagger) \sim \exp\left[-\frac{N}{\sigma^2} \frac{1}{1-\tau^2} \left(\mathrm{tr} X X^\dagger - \frac{\tau}{2} \mathrm{tr}(X^2 + (X^\dagger)^2) \right) \right]. \tag{8.12}$$

The propagator therefore reads

$$\left\langle \mathcal{X}_{ab}^{\alpha\beta} \mathcal{X}_{cd}^{\mu\nu} \right\rangle_G = \frac{\sigma^2}{N} (1 + (\tau - 1)\delta_{\alpha\mu}) \delta_{\alpha\beta} \delta_{\mu\nu} \delta_{ad} \delta_{bc}, \tag{8.13}$$

where the Greek indices take values from $\{1, \bar{1}\}$ and the Latin ones from $\{1, 2, \dots, N\}$. The residual part is expanded into a power series, bringing additional matrices (vertices in the diagrammatic representation), and all averages are then calculated with respect to the Gaussian measure, which by means of the Wick's theorem reduces to the summation over all possible pairings. The diagrammatic rules are exactly the same as in the Hermitian case, apart from the additional Greek indices, carried by each dot.

8.2.2 Quaternionic R-transform

The structure of the Feynman diagrams is exactly the same as for Hermitian matrices, but now the objects that we calculate are 2×2 matrices. The

Schwinger-Dyson equation relating the quaternionic Green's function with the self-energy composed of 1LI diagrams then reads

$$\mathcal{G}(Q) = [Q - \Sigma(Q)]^{-1}. \tag{8.14}$$

Now, due to the fact that in general X is not related with $X^\dagger$, there are many types of cumulants in the expansion of $\mathcal{G}$, corresponding to the connected averages of different words, separated by a horizontal line (appropriate component of Q^{-1}), e.g. $c_{1\bar{1}1\bar{1}} = \left\langle \frac{1}{N}\mathrm{tr}XX^\dagger XX^\dagger \right\rangle_c \neq \left\langle \frac{1}{N}\mathrm{tr}XXX^\dagger X^\dagger \right\rangle_c = c_{11\bar{1}\bar{1}}$. Remarkably, all possible cumulants are stored in a single object, the quaternionic R-transform, which itself is a 2×2 matrix representation of the quaternion, defined as follows

$$\mathcal{R}(Q)\otimes \mathbf{1} = \langle \mathcal{X}\rangle_c + \langle \mathcal{X}\mathcal{Q}\mathcal{X}\rangle_c + \langle \mathcal{X}\mathcal{Q}\mathcal{X}\mathcal{Q}\mathcal{X}\rangle_c + \dots, \tag{8.15}$$

more explicitly

$$\mathcal{R}(Q)_{\alpha\beta} = c^{(1)}_\alpha \delta_{\alpha\beta} + c^{(2)}_{\alpha\beta} Q_{\alpha\beta} + \sum_{\gamma\in\{1,\bar{1}\}} c^{(3)}_{\alpha\gamma\beta} Q_{\alpha\gamma}Q_{\gamma\beta} + \sum_{\gamma,\epsilon\in\{1,\bar{1}\}} c^{(4)}_{\alpha\gamma\epsilon\beta} Q_{\alpha\gamma}Q_{\gamma\epsilon}Q_{\epsilon\beta} + \dots \ . \tag{8.16}$$

This definition is quite compact and deserves more intuitive explanation. Suppose that we know all cumulants and we want to construct the $\mathcal{R}$-transform. We naturally associate $\bar{1}$ in the index of the cumulant with $\dagger$ in the corresponding expression in X's. The first and the last index of the cumulant give us the appropriate component of $\mathcal{R}$. Starting from the first index, we move towards the rightmost one and each time we make a step between two indices, we pick the component of Q given by the indices we encounter. Q therefore can be considered as a transfer matrix. The cumulant $c^{(4)}_{\alpha\beta\gamma\epsilon} = \left\langle \frac{1}{N}\mathrm{tr}X^\alpha X^\beta X^\gamma X^\epsilon \right\rangle_c$ comes with the expression $Q_{\alpha\beta}Q_{\beta\gamma}Q_{\gamma\epsilon}$ in $\mathcal{R}_{\alpha\beta}$. For example, $c_{111\bar{1}}$ appears with $Q_{11}Q_{11}Q_{1\bar{1}} = z^2(i\bar{w})$ in $\mathcal{R}_{1\bar{1}}$. The ability to store all mixed cumulant in a single object relies on the fact that $\mathcal{Q}$ and $\mathcal{X}$ do not commute. The mapping between $\mathcal{R}$-transform and the cumulants is not bijective, there are different cumulants, which bring the same expression in the components of Q to the quaternionic R-transform[1]. For the one-to-one mapping, one has to consider either different Q's in the expansion (8.15) or a single Q, but with entries from a noncommutative algebra.

The relation between the self-energy and the quaternionic Green's function through the connected diagrams can be expressed via the quaternionic

[1]For example a pair $c^{(6)}_{11\bar{1}1\bar{1}1}$ and $c^{(6)}_{1\bar{1}11\bar{1}1}$

R transform $\Sigma(Q) = \mathcal{R}(\mathcal{G}(Q))$. A direct relation between the generalized Green's function and the quaternionic R-transform can be written in terms of the auxiliary funcion, nicknamed Blue's function, which is the functional inverse of the quaternionic Green's function $\mathcal{B}(\mathcal{G}(Q)) = Q = \mathcal{G}(\mathcal{B}(Q))$. The $\mathcal{R}$-transform is then given by $\mathcal{R}(Q) = \mathcal{B}(Q) - Q^{-1}$. The inverse of Q is understood as the matrix inverse.

The R-transform for non-Hermitian matrices was discovered in [**?**, **?**] as a function generating all 1LI diagrams. The quaternionic structure was discovered much later [**?**].

We remark here that the mixed moments are encoded in the same way in the quaternionic moment generating function

$$\tilde{\mathcal{M}}(Q) = Q^{-1}\mathcal{G}(Q^{-1})Q^{-1} - Q^{-1}. \tag{8.17}$$

We conclude this section by a comparison of two formalisms (Table 8.1), using a calligraphic notation in the case of non-Hermitian analogues of Hermitian entries.

	Hermitian	Non-Hermitian
Spectrum	real	complex
Green's function	complex-valued $G(z) = \frac{1}{N}\left\langle \mathrm{Tr}(z-H)^{-1}\right\rangle$	quaternion-valued $\mathcal{G}(Q) = \frac{1}{N}\left\langle \mathrm{bTr}(Q-\mathcal{X})^{-1}\right\rangle$
Moments	$\tilde{M}(z) = \frac{1}{z}G\left(\frac{1}{z}\right)\frac{1}{z} - \frac{1}{z}$	$\tilde{\mathcal{M}}(Q) = Q^{-1}\mathcal{G}(Q^{-1})Q^{-1} - Q^{-1}$
Cumulants	$R(z) = \sum_n \kappa_n z^{n-1}$	$[\mathcal{R}(Q)]_{\alpha\beta} = \sum_{k,\{i_1\ldots i_{k-2}\}} c^{(k)}_{\alpha i_1\ldots i_{k-2}\beta} Q_{\alpha i_1}\ldots Q_{i_{k-2}\beta}$
S-D eqs.	$R[G(z)] + \frac{1}{G(z)} = z$	$\mathcal{R}[\mathcal{G}(Q)] + [\mathcal{G}(Q)]^{-1} = Q$

Table 8.1: Comparison between corresponding quantities in Hermitian versus non-Hermitian ensembles.

8.3 Single ring theorem

In this section we discuss another class of random matrices, the spectrum of which, despite being complex, is effectively one dimensional, because the spectral problem has an azimuthal symmetry. We consider matrices generated according to the probability distribution function $P(X, X^\dagger) \sim \exp(-N\mathrm{tr}V(XX^\dagger))$. The symmetry in this case is enhanced from $U(N)$ to

$U(N) \times U(N)$. The spectrum of X is rotationally symmetric on the complex plane $\rho(z, \bar{z}) = \rho_r(|z|)$ and the entire information is encoded in the radial cumulative distribution function $F(s) = \int_{|z| \leq s} \rho(z, \bar{z}) d^2 z = 2\pi \int_0^s s' \rho_r(s') ds'$. Such matrices are the natural extensions of the so called isotropic complex random variables, the distribution of which depends only on their modulus. Moreover, the symmetry transformations can bring the matrix to the diagonal form with the singular values on the diagonal. It is natural, therefore, to expect all spectral properties of X to be determined by its singular values.

In the free probability community such objects are called **R**-diagonal, the meaning of this notion shall become clear next. The **R**-diagonal operators have a polar decomposition $X = PU$, where P is Hermitian positive definite, U is Haar unitary and P and U are mutually free. In the limit $N \to \infty$ bi-unitarily invariant random matrices become the R-diagonal operators, which was shown by Hiai and Petz [**?**, Theorem 4.4.5].

The first relation between the Green's function of $XX^\dagger$, encoding the distribution of the squares of the singular values, and the radial cumulative distribution was found by Feinberg and Zee [**?**]. They also found a very intriguing property that the support of the spectrum of such matrix is either a disc or an annulus, which bears the name of the single ring theorem. Later, Haagerup and Larsen [**?**] within the framework of free probability derived a simple relation between the S-transform of $XX^\dagger$ and the radial cumulative distribution function

$$S_{XX^\dagger}(F(s) - 1) = \frac{1}{s^2} \tag{8.18}$$

The internal and external radii are given by

$$r_{ext}^2 = \int x \rho_{XX^\dagger}(x) dx, \qquad r_{in}^{-2} = \int x^{-1} \rho_{XX^\dagger}(x) dx, \tag{8.19}$$

Recently [**?**], this theorem has been extended to describe also the eigenvector correlation function

$$O(s) = \frac{F(s)(1 - F(s))}{\pi s^2}. \tag{8.20}$$

We remark here that the expression on the right-hand-side of the equality above has already appeared in the pioneering paper by Feinberg and Zee [**?**], but without any connection to eigenvectors.

Let us provide few arguments in the case of bi-unitary ensembles we expect such a tremendous calculational simplification. First, we note that due to the particular form of the potential, namely that X and $X^\dagger$ appear

alternately, and the planarity of diagrams in the large N limit, the structure of diagrams simplifies considerably.

The only non-vanishing in the large N limit connected diagrams correspond to the expressions where X and $X^\dagger$ are alternating, i.e. $\alpha_n := \left\langle \frac{1}{N}\mathrm{tr}(XX^\dagger)^n \right\rangle_c = c^{(2n)}_{1\bar{1}1\bar{1}\ldots 1\bar{1}} = c^{(2n)}_{\bar{1}1\bar{1}1\ldots\bar{1}1}$. Let us consider a generating function for all such cumulants $A(x) := \sum_{k=1}^{\infty} \alpha_k z^{k-1}$. In the free probability context it is also known as the determining sequence [**?**]. Due to the very simple structure of non-vanishing cumulants, only one (commuting) variable is sufficient to encode all cumulants.

According to the prescription (8.16), the quaternionic R-transform for biunitarily invariant random matrices reads

$$\mathcal{R} = A(-|w|^2) \begin{pmatrix} 0 & i\bar{w} \\ iw & 0 \end{pmatrix}. \tag{8.21}$$

The R-transform depends only on the off-diagonal elements of the quaternion, which is a consequence of the fact that $X^\dagger$ has to be sandwiched between X's and vice versa. In the 'hermitization' approach to non-Hermitian matrices, the corresponding R-transform is diagonal, hence the name. Second, there has to exist a relation between R-transform of $XX^\dagger$ and the quaternionic R-transform of X.

In both cases we consider moments like $\left\langle \frac{1}{N}\mathrm{tr}(X^\dagger X)^k \right\rangle$ in the moment expansion of the Green's function of $XX^\dagger$. The average is taken over the probability measure proportional to $\exp(-N\mathrm{tr}V(X^\dagger X))dXdX^\dagger$. There are two ways of calculating such an object.

First, making use of the symmetry of the potential, one changes the integration measure from $dXdX^\dagger$ into $d(XX^\dagger)$ and uses the tools for Hermitian matrices. In this approach, however, the resulting Jacobian modifies the form of the potential, which changes the structure of the vertices in the expansion of the Green's function.

In the diagrammatic approach for non-Hermitian problems, we circumvent the calculation of the Jacobian, calculating the averages with respect to the original measure. Combining both both diagrammatic calculi, we arrive [[**?**]] at the following Schwinger-Dyson equation between the R-transform for $XX^\dagger$ and the quaternionic R transform for X

$$R_{XX^\dagger}(G_{XX^\dagger}(z)) = zG_{X^\dagger X}(z)A\left(zG_{X^\dagger X}(z)G_{XX^\dagger}(z)\right). \tag{8.22}$$

Knowing that for square matrices $G_{XX^\dagger} = G_{X^\dagger X}$, which follows from cyclic property of the trace, we can associate these objects and skip the subscript

for simplicity. Let us also make a substitution $z \to B(z)$ and make use of the fact that B is a functional inverse of G, to finally obtain

$$R(z) = zB(z)A\left(z^2 B(z)\right). \tag{8.23}$$

This relation is convenient if knowing the quaternionic R-transform of X one wants to calculate the complex R-transform of $XX^\dagger$. To invert this relation, let us introduce an auxiliary variable $y := z^2 B(z)$. Note that $zR(z) + 1 = zB(z) = yA(y) + 1$. Introducing variable t given by $z = tS(t)$ and using the generic relation between the R and S transforms $R(z) = 1/S(zR(z))$ we arrive at $t = yA(y)$ and $y = t(t+1)S(t)$, from which we finally read out the S transform of $XX^\dagger$ expressed in terms of the determining sequence of X via

$$S(zA(z)) = \frac{1}{A(z)(1 + zA(z))}. \tag{8.24}$$

This relation is the cornerstone for the derivation of the Haagerup-Larsen theorem, as we demonstrate now.

Making use of the form of the quaternionic R-transform (8.21) and the relation $\mathcal{R}(\mathcal{G}) + \mathcal{G}^{-1} = \mathcal{B}(\mathcal{G}) = Q$, we obtain the matrix equation

$$A(\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}) \begin{pmatrix} 0 & \mathcal{G}_{1\bar{1}} \\ \mathcal{G}_{\bar{1}1} & 0 \end{pmatrix} + \frac{1}{\mathcal{G}_{11}\mathcal{G}_{\bar{1}\bar{1}} - \mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}} \begin{pmatrix} \mathcal{G}_{\bar{1}\bar{1}} & -\mathcal{G}_{1\bar{1}} \\ -\mathcal{G}_{\bar{1}1} & \mathcal{G}_{11} \end{pmatrix} = \begin{pmatrix} z & i\bar{w} \\ iw & \bar{z} \end{pmatrix}. \tag{8.25}$$

We are now interested in calculating the spectral density and the eigenvector correlator, so we set $|w| \to 0$. Let us first consider the upper-left component of matricial eq. (8.25):

$$\frac{\mathcal{G}_{\bar{1}\bar{1}}}{\mathcal{G}_{11}\mathcal{G}_{\bar{1}\bar{1}} - \mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}} = z. \tag{8.26}$$

The $\bar{1}\bar{1}$ component gives the complex conjugate of the above. Combining them together we easily deduce that $z\mathcal{G}_{11}$ =F:=zG_{11} [**?**], we immediately obtain from (8.26) that

$$O(z, \bar{z}) = -\frac{1}{\pi}\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1} = \frac{F - F^2}{\pi |z|^2}. \tag{8.27}$$

Considering now the $1\bar{1}$ component we obtain two possibilities. First $\mathcal{G}_{1\bar{1}} = 0$ and in consequence $\mathcal{G}_{11} = \frac{1}{z}$, which is a trivial solution, valid outside the spectrum. Second, assuming $\mathcal{G}_{1\bar{1}} \neq 0$ we use (8.49) to arrive at

$$A(\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1})\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1} = F - 1. \tag{8.28}$$

Now, evaluating the S transform at both sides of the equation above, exploiting the relation between S and A (8.24), and using (8.49) once again, we finally get

$$S(F-1) = \frac{1}{A(\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1})(1+\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}A(\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}))} = \frac{1}{|z|^2}, \tag{8.29}$$

which is the statement of the original formulation of the Haagerup-Larsen theorem. One deduces that F depends on z and $\bar{z}$ only through their modulus, therefore the spectral density can be calculated by [**?**]

$$\rho(z,\bar{z}) = \frac{1}{\pi}\partial_{\bar{z}}\mathcal{G}_{11} = \frac{1}{2\pi|z|}F'(|z|). \tag{8.30}$$

Moreover, outside the support of the spectral density, the trivial solution gives $F = 1$, therefore F is indeed the radial cumulative distribution function. The inner and outer radii of the spectrum can be calculated by imposing $F = 1$ (outer) or that F is equal to the fraction of the zero modes (inner) and solving the resulting equation for $|z|$. In general, (8.29) can yield several solutions for F. The uniqueness of the radial cumulative distribution function has been shown within the framework of the analytic subordination function theory [**?**].

Multiplying (8.29) by $(F-1)$, evaluating the R transform on both sides of the equation and making use of (**??**), we obtain

$$R_{XX^\dagger}\left(\frac{F-1}{r^2}\right) = r^2. \tag{8.31}$$

Substituting $F = 1$ and taking into account that $R(0) = \kappa_1 = m_1$, we relate the external radius with the first moment of $XX^\dagger$.

Using the relation $S_X(z)S_{X^{-1}}(-1-z) = 1$ [**?**] and performing analogous computations, we arrive at

$$R_{(XX^\dagger)^{-1}}(-Fr^2) = \frac{1}{r^2}. \tag{8.32}$$

Substitution $F = 0$ relates the first inverse moment of $XX^\dagger$ with the internal radius.

8.4 Applications

8.4.1 Quaternion embeddings

The quaternion embedding for arbitrary hermitian random matrix H is given in terms hermitian Green's function according to the prescription

$$\mathcal{G}_H(Q) = \gamma_H(q,\bar{q})\mathbf{1}_2 - \gamma_H^{'}(q,\bar{q})Q^{\dagger} \tag{8.33}$$

where $q,\bar{q}$ is a pair of two (conjugated) eigenvalues of quaternion Q and scalar coefficients read

$$\begin{aligned} \gamma_H(q,\bar{q}) &= \frac{qG_H(q) - \bar{q}G_H(\bar{q})}{q-\bar{q}} \\ \gamma_H^{'}(q,\bar{q}) &= \frac{G_H(q) - G_H(\bar{q})}{q-\bar{q}} \end{aligned} \tag{8.34}$$

The corresponding R transform reads accordingly

$$\mathcal{R}_H(Q) = \beta_H(q,\bar{q})\mathbf{1}_2 - \beta_H^{'}(q,\bar{q})Q^{\dagger} \tag{8.35}$$

where scalar coefficients read

$$\begin{aligned} \beta_H(q,\bar{q}) &= \frac{qR_H(q) - \bar{q}R_H(\bar{q})}{q-\bar{q}} \\ \beta_H^{'}(q,\bar{q}) &= \frac{R_H(q) - R_H(\bar{q})}{q-\bar{q}} \end{aligned} \tag{8.36}$$

One can check by explicit, but lenghty calculation, that $\mathcal{R}[\mathcal{G}(Q)] + 1/\mathcal{G}(Q) = Q$.

Similarly, one can generalize the scaling relations. From the definition of the quaternion Greens function we read, that

$$\mathcal{G}_{gX}(Q) = \mathcal{G}_X[\mathrm{diag}(1/g, 1/\bar{g}) \cdot Q] \cdot \mathrm{diag}(1/g, 1/\bar{g}) \tag{8.37}$$

where g is any, non-zero *complex number*. The corresponding R-transform read

$$\mathcal{R}_{gX}(Q) = \mathrm{diag}(g,\bar{q}) \cdot \mathcal{R}_X[Q \cdot \mathrm{diag}(g,\bar{g})] \tag{8.38}$$

The order of multiplication is important, since 2 by 2 matrices usually do not commute. Above relations are obvious generalizations of hermitian properties under the scaling by the *real* number g for Green's and R-transforms: $G_{gH}(z) = \frac{1}{g}G_H(\frac{1}{g}z)$ and $R_{gH}(z) = gR_H(gz)$.

Note that above relations allow us to calculate spectral properties of any ensemble of the type $X = \alpha_1 H_1 + \alpha_2 H_2$, for any *complex* α_i, provided hermitian H_1, H_2 are mutually free.

As an example we consider the case $X = \frac{1}{\sqrt{2}}(H_1 + iH_2)$, where H_i are free GUE ensembles. This is precisely the case of the Ginibre ensemble, since X is complex, and the potential reads

$$V(XX^\dagger) = \mathrm{tr}XX^\dagger = \frac{1}{2}\mathrm{tr}H_1^2 + \frac{1}{2}\mathrm{tr}H_2^2 \tag{8.39}$$

Let us now use the addition law for $X = \frac{1}{\sqrt{2}}H_1 + \frac{i}{\sqrt{2}}H_2$. Since for GUE $R(z) = z$, the quaternion embedding reads $\mathcal{R}(Q) = Q$ For the first ensemble, $\mathcal{R}_{\frac{1}{\sqrt{2}}H_1}(Q) = \frac{1}{\sqrt{2}}\mathbf{1}_2 Q \frac{1}{\sqrt{2}}\mathbf{1}_2 = \frac{1}{2}Q$. For the second ensemble $\mathcal{R}_{\frac{i}{\sqrt{2}}H_2}(Q) = \frac{1}{\sqrt{2}}i\sigma_3 Q \frac{1}{\sqrt{2}}i\sigma_3 = -\frac{1}{2}\sigma_3 Q \sigma_3$, where σ_3 is the third Pauli matrix. The addition law reads

$$\mathcal{R}_X(Q) = \mathcal{R}_{\frac{1}{\sqrt{2}}H_1}(Q) + \mathcal{R}_{\frac{i}{\sqrt{2}}H_2}(Q) = \frac{1}{2}(Q - \sigma_3 Q \sigma_3) \tag{8.40}$$

Substituting $Q \to \mathcal{G}$ and using the fundamental equation $\mathcal{R}(\mathcal{G}) + 1/\mathcal{G} = Q$, we arrive at the 2 by 2 matricial equation

$$\frac{1}{2}\begin{pmatrix} \mathcal{G}_{11} & \mathcal{G}_{1\bar{1}} \\ \mathcal{G}_{\bar{1}1} & \mathcal{G}_{\bar{1}\bar{1}} \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \mathcal{G}_{11} & -\mathcal{G}_{1\bar{1}} \\ -\mathcal{G}_{\bar{1}1} & \mathcal{G}_{\bar{1}\bar{1}} \end{pmatrix} + \frac{1}{\mathcal{G}_{11}\mathcal{G}_{\bar{1}\bar{1}} - \mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}} \begin{pmatrix} \mathcal{G}_{\bar{1}\bar{1}} & -\mathcal{G}_{1\bar{1}} \\ -\mathcal{G}_{\bar{1}1} & \mathcal{G}_{11} \end{pmatrix} = \begin{pmatrix} z & i\bar{w} \\ iw & \bar{z} \end{pmatrix}. \tag{8.41}$$

Note that on the r.h.s. we can now safely put $w = 0$. The equation in the upper-right corner reads

$$\mathcal{G}_{1\bar{1}}(1 - \frac{1}{\mathrm{det}\mathcal{G}_X}) = 0 \tag{8.42}$$

where $\mathrm{det}\mathcal{G}_X = \mathcal{G}_{11}\mathcal{G}_{\bar{1}\bar{1}} - \mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1}$, which means either

$$\mathcal{G}_{1\bar{1}} = 0, \quad \text{then} \quad \mathcal{G}_{11} = \frac{1}{\mathrm{z}} \tag{8.43}$$

or

$$\mathrm{det}\mathcal{G}_X = 1 \quad \text{then} \quad \mathcal{G}_{11} = \bar{z} \tag{8.44}$$

We explicitly see two kinds of solutions, the holomorphic one and the non-holomorphic one. The solutions match on the border line of the island of eigenvalues

$$\frac{1}{z} = \bar{z} \;\; i.e. \quad x^2 + y^2 = 1 \tag{8.45}$$

i.e. on the circle of radius 1. Outside the circle, there are no eigenvalues, inside the circle, we read the eigenvalues density from the Gauss law

$$\rho_X \lambda = \frac{1}{\pi}\partial_{\bar{z}}\mathcal{G}_{11} = \frac{1}{\pi}\partial_{\bar{z}}\bar{z} = \frac{1}{\pi} \tag{8.46}$$

which completes the solution of the large N spectral behavior of the Ginibre ensemble.

8.4.2 Quaternionic R transform of Haar unitary matrix.

We consider a unitary matrix $UU^\dagger = \mathbf{1}$, the spectral density of which is uniform on the unit circle. Due to unitarity, $R_{UU^\dagger}(z) = 1$ and from (**??**) $S(z) = 1$. Substituting $z \to xA(x)$ and using (8.24), we obtain the quadratic equation

$$xA^2(x) + A(x) - 1 = 0. \tag{8.47}$$

Knowing that $A(0) = c_{1\bar{1}} = \left\langle \frac{1}{N}\mathrm{tr}UU^\dagger \right\rangle_c = 1$, we choose the appropriate branch of the solution. From (8.21) we deduce the quaternionic R transform, which reads

$$\mathcal{R}(Q) = \frac{1-\sqrt{1-4|w|^2}}{2|w|^2}\begin{pmatrix} 0 & i\bar{w} \\ iw & 0 \end{pmatrix}. \tag{8.48}$$

The same result was derived in [**?**] using a different technique. We remark that the free cumulants are $\alpha_n = (-1)^{n-1}C_{n-1}$, where $C_{n-1} = \frac{1}{n+1}\binom{2n}{n}$ are the Catalan numbers, in agreement with [**?**]. Note that now we can "add" any combination of free Wishart, Haar, deterministic or random ensembles, using their quaternionic embeddings.

8.4.3 Spectral properties of Ginibre ensemble from Haagerup-Larsen theorem

. Let us demonstrate the power of Haagerup-Larsen theorem on the example of the Ginibre ensemble. In this case, $XX^\dagger$ is the Wishart ensemble, so $S_{XX^\dagger}(z) = \frac{1}{1+z}$, since the rectangularity $r = 1$. Using the Haagerup Larsen theorem $S_{XX^\dagger}(F(s) - 1) = \frac{1}{s^2}$ we immediately read that $F(s) = s^2$, so the radial spectral density $\rho(s) = \frac{1}{2\pi s}\partial_s F(s) = \frac{1}{\pi}$, where $s = \sqrt{z\bar{z}}$. The outer

radius, corresponding to $F(s) = 1$ reads one, the inner one is zero. As a bonus we obtain the condition for the eigenvector correlator. Since

$$O(z, \bar{z}) = -\frac{1}{\pi}\mathcal{G}_{1\bar{1}}\mathcal{G}_{\bar{1}1} = \frac{F - F^2}{\pi|z|^2}. \tag{8.49}$$

we read $O_X(s) = \frac{1}{\pi}(1 - s^2)$, the result, which was obtained almost 25 years after the Ginibre paper. We were able to make this spectacular shortcut due to the fact, that the Ginibre ensemble is R-diagonal, i.e. can be rewritten as $X = PU$, where U is a Haar measure, and P is the positive part of the Wigner semicircle (Wigner's quatercircle).

8.4.4 Product and quotient of free Ginibre ensembles X_1 and X_2

We can now easily find the free analogs of the product and ratio of two random variables from the Gaussian distribution, which we originally have solved using the Mellin transform. In the case of the product, we use the Haagerup-Larsen theorem for $X = X_1X_2$

$$\frac{1}{s^2} = S_{XX^\dagger}(F(s) - 1) = S_{X_1X_1^\dagger}(F(s) - 1)S_{X_2X_2^\dagger}(F(s) - 1) = \frac{1}{F(s)^2} \tag{8.50}$$

hence $F(s) = s$ and the radial spectral density $\rho_{X_1X_2}(s) = \frac{1}{\pi s}$. In the case of the quotient, we use the fact that $S_{H^{-1}}(z)S_H(-z - 1) = 1$, we read the S-transform for inverse Wishart as $-z$. Then, for $Y = X_1/X_2$

$$\frac{1}{s^2} = S_{YY^\dagger}(F(s) - 1) = S_{X_1X_1^\dagger}(F(s) - 1)S_{[X_2X_2^\dagger]^{-1}}(F(s) - 1) = \frac{1 - F(s)}{F(s)} \tag{8.51}$$

hence $F(s) = \frac{1}{1+s^2}$, so the radial spectral density reads $\rho_{X_1/X_2}(s) = \frac{1}{\pi}\frac{1}{(1+s^2)^2}$. Note the qualitative similarity to the corresponding cases of the classical pdfs for the product and ratio of the Gaussian random variables.

Generalization for the arbitrary strings of similar matrices is also trivial, due to Haagerup-Larsen theorem. For example, let us consider a product of k independent Ginibre matrices. The multiplication law leads to $S_k = (1+z)^{-k}$. Using the relation (8.24) we obtain the algebraic equation for the determining sequence

$$(zA_k(z) + 1)^{k-1} = A_k(z). \tag{8.52}$$

The solution can be written in a power series [**?**]

$$A_k(z) = \sum_{n=1}^{\infty} A_{n-1}(k-1, k-1) z^{n-1}, \tag{8.53}$$

where

$$A_n(p, r) = \frac{r}{np + r} \binom{np + r}{m} = \frac{r}{n!} \prod_{i=1}^{n-1} (mp + r - i) \tag{8.54}$$

are the two parameter Fuss-Catalan numbers, also known as Raney numbers. The **R**-diagonal cumulants are therefore $\alpha_n^{(k)} = A_{n-1}(k-1, k-1) = A_n(k, 1)$. Such numbers has appeared in the free probability many times [**?**, **?**, **?**] and densities associated with them have been extensively studied [**?**, **?**].

8.4.5 Tetilla law

One of many measures of non-normality of a matrix (see e.g. [**?**]) is defined through the spectral properties of a Hermitian matrix $C := XX^\dagger - X^\dagger X$. Usually it is a square root of its Frobenius norm or the square root of the largest eigenvalue. Substituting the polar decomposition $X = PU$, one obtains $C = P^2 + U(-P^2)U^\dagger$. The unitary matrices assert that in the large N limit the summands are free and their addition reduces to the addition of the corresponding R transforms

$$R_C(z) = R_{P^2}(z) + R_{-P^2}(z) = R_{P^2}(z) - R_P^2(-z), \tag{8.55}$$

since $R_{aX}(z) = aR_X(az)$.

In the simplest instance, the Ginibre matrix, the R-transform of the commutator reads $R_C(z) = \frac{2z}{1-z^2}$. The corresponding Green's function is therefore given by the cubic (Cardano) equation. Spectral distribution obtained from this equation is known as the Tetilla law (due to the similarity of the distribution to the profile of a certain Spanish cheese, see the figure X), derived for the first time in [**?**]. It was also proven to be the limiting law for the anticommutator of Hermitian Wigner matrices [**?**].

8.5 Diffusion of the nonhermitian ensembles

8.6 Unexpected links

Several unexpected links between the static hermitian and non-hermitian random matrix models were noted in the past [**?**]. The spectrum of hermitian matrices is real, but the main tool relies on introducing the complex valued resolvent (Green's function), whose discontinuities allow to infer the spectral function, using the theory of analytic functions. In the large N limit, a particular transform, known as the R-transform, related to the Green's function by the functional inverse as $R[G(z)]+1/G(z) = z$, plays the role of the analog of generating function of classical cumulants in the matrix-valued probability calculus. The R-transform constitutes the cornerstone of the free probability theory [**?**] and generates matrix-valued analogues of classical central limit theorems. In the case of non-hermitian matrices, the spectrum is complex, but the regulator ϵ^2 in the logarithmic potential (8.1) behaves as the tip of an iceberg, pointing at a hidden algebraic structure. Indeed, in order to maintain the analogy to the hermitian case, one has to embed the structure of the generalized Green's functions in the algebra of quaternions. In such a way, a second complex variable w, "perpendicular" to z emerges. In the large N limit, one can adapt the Voiculescu construction for the R-transform by defining the quaternion valued functional inverse $\mathcal{R}[\mathcal{G}(Q)] + 1/\mathcal{G}(Q) = Q$ and thus allowing for non-hermitian and non-commuting convolution of random matrices [**?**, **?**]. Surprisingly, the links between the hermitian and non-hermitian random matrix models stretch out to the area of dynamic processes. In the case of the Gaussian randomness, the exact diffusion equation for the averaged characteristic polynomial finds its exact analogue for the averaged characteristic polynomial valued in the algebra of quaternions. It turns out, that the "hidden" variable w, ignored in standard treatment of non-hermitian random matrix models, plays a crucial role in determining the two-dimensional pattern of the spectral evolution. In the large N limit, hermitian and non-hermitian Smoluchowski-Fokker-Planck equations take the surprisingly similar form of a Burgers-like structure. In general, the Voiculescu equation $\partial_\tau G + R(G)\partial_z G = 0$ is replaced by its quaternionic counterpart [**?**]

	GUE	GE
Spectral density	real	complex
Resolvent	complex-valued $G(z) = \frac{1}{N}\langle \mathrm{Tr}(z-H)^{-1}\rangle$	quaternion-valued $\mathcal{G}(Q) = \frac{1}{N}\langle \mathrm{bTr}(Q-\mathcal{X})^{-1}\rangle$
Determinant	$U(z,\tau) = \langle \det(z-H)\rangle$	$D(Q,\tau) = \langle \det(Q-\mathcal{X})\rangle$
Diffusion eq.	$\partial_\tau U = -\frac{1}{2N}\partial_{zz}U$	$\partial_\tau D = +\frac{1}{N}\partial_{w\bar{w}}D$
Viscosity	negative	positive
Universal behavior	oscillatory (Airy kernel)	smooth (Erfc)
R-transform	$R_{GUE}(G) = G$	$\mathcal{R}_{GG}(\mathcal{G}) = \begin{pmatrix} 0 & \mathcal{G}_{1\bar{1}} \\ \mathcal{G}_{\bar{1}1} & 0 \end{pmatrix}$
Voiculescu equation	$\frac{\partial G}{\partial \tau} + R(G)\frac{\partial G}{\partial z} = 0$	$\frac{\partial \mathcal{G}_{ab}}{\partial \tau} + \sum_{c,d=1}^{2} \mathcal{R}[\mathcal{G}]_{cd}\frac{\partial \mathcal{G}_{ab}}{\partial Q_{cd}} = 0$
Pre-shock waves	Flow of eigenvalues	Flow of eigenvector correlators

Table 8.2: Comparison of links between GUE and GE.

$$\frac{\partial \mathcal{G}_{ab}}{\partial \tau} + \sum_{c,d=1}^{2} \mathcal{R}[\mathcal{G}]_{cd}\frac{\partial \mathcal{G}_{ab}}{\partial Q_{cd}} = 0, \tag{8.56}$$

where latin indices label the two-by two quaternionic structure of Q, $\mathcal{G}$ and $\mathcal{R}$. In both cases, singularities emerge. However, in the hermitian case, singularities appear in the flow of the eigenvalues, whereas in the case of non-hermitian ensembles, singularities appear in the flow of a certain correlator of left and right eigenvectors. In both cases, finite N effects can be taken into account as an appearance of spectral viscosity proportional to $1/N$. There is however a crucial difference in the sign - positive spectral viscosity smoothens the edge of the Ginibre spectrum, yielding universal behavior given by the Erfc function, whereas negative spectral viscosity in the GUE triggers violent oscillations, leading to the formation of the so-called Airy kernel. Resolving the deep reasons for these links still remains one of the challenges of random matrix models. We summarize the unexpected links between the GUE and GE ensembles in Table I.

Chapter 9

Sample applications in finances

9.1 Modeling a financial market

Financial market is a part of the econosystem which is easiest to quantify. We shall use a simplified picture of this market in which the only objects are the prices of assets, asset being the name commonly used to describe a financial instrument, which can be bought or sold, like currencies, bonds, shares *etc.* In the following we shall understand assets solely as shares. Asset (or stock) prices $S_i(t)$ are functions of time. A typical time step ε, when the price is changed is as short as few seconds. It will be the dynamics of price changes, which we shall discuss in this chapter.

We define the instantaneous returns, which we shall alternatively call relative price changes of the asset in the period from τ to $\tau + \varepsilon$

$$x_i(\tau;\varepsilon) = \log S_i(\tau+\varepsilon) - \log S_i(\tau). \tag{9.1}$$

Again the crucial ingredient of this analysis is the assumption about the multiplicative nature of price changes. The definition of return is independent of the unit in which the price is given and seems the best to capture the essential properties of the price system. Return $x_i(\tau;\varepsilon)$ can be any positive real number. Obviously the return over a larger time interval is a sum of all changes over its subintervals

$$x_i(\tau;\varepsilon_1+\varepsilon_2) = x_i(t;\varepsilon_1) + x_i(t+\varepsilon_1;\varepsilon_2)\,. \tag{9.2}$$

Financial databases contain huge number of time series of asset prices, sampled at various frequencies. Phenomenologically one can observe that prices

behave in a random way: relative price changes $x_i(t, \varepsilon)$ fluctuate. The empirically measured time correlations show that these fluctuations have a rather short autocorrelation time, typically of the order of several minutes. Longer autocorrelation times were observed for the absolute values of fluctuations.

If the frequency of sampling ε is chosen larger than the autocorrelation time ε_0, corresponding price changes can be viewed as independent random variables. The simplest assumption one can make is the assumption of stationarity: $x_{it} = x_i(\tau = t * \varepsilon_0; \varepsilon_0)$, where t is an integer, can be interpreted as random numbers generated with the same random number generator, independent of time. One can derive surprisingly strong predictions based on this simple assumption, using very general properties of this random number generator. Let us assume that the generator is characterized by the normalized probability distribution function (pdf) $p(x)$, with a characteristic function $\hat{F}(z)$ defined by the Fourier transform

$$\hat{F}(z) = \int_{-\infty}^{\infty} dx\, p(x) e^{ixz}\,. \tag{9.3}$$

Define a function $\hat{r}(z) = \log \hat{F}(z)$. It is straightforward to see that the sum

$$X_n = \sum_{i=1}^{n} x_i \tag{9.4}$$

of independent random numbers distributed with p is again a random number with a distribution p_n being an n-fold convolution of $p(x)$. In consequence, $\hat{F}_n(z) = \hat{F}^n(z)$ and $\hat{r}_n(z) = n\hat{r}(z)$ where $\hat{r}_n(z) = \log F_n(z)$.

A special role is played by stable distributions, which have the property that the probability distribution of the sum p_n can be mapped into the original distribution by a linear change of the argument

$$dx\, p_n(x) = d(a_n x + b_n)\; p(a_n x + b_n)\,, \tag{9.5}$$

where a_n and b_n are suitable parameters. Saying differently, the stable distributions are self-similar under the convolution which means that the shape of pdf is preserved up to a scale factor and shift. The condition (9.5) can be rewritten as a condition for $\hat{R}(z)$ in the form

$$\hat{R}(z) = n\hat{R}(a_n z) + i b_n z\,. \tag{9.6}$$

As a consequence of stability, if our sampling frequency in the price list is large, say one day, we may expect to a good approximation the relative price changes measured with this frequency to be random numbers obtained from one of the stable distributions.

If the idealized assumption of stationarity holds, we can represent the history of the financial market as a matrix x_{it}, with the times t measured in intervals of the sampling unit ε, corresponding to one day. In this way we lose information about the short time scale fluctuations, but we may expect that for each i the entries x_{it} will represent a sequence of random numbers drawn from the same stable distribution. It is, of course, a crucial question, which stable distribution is realized in practice. We may deduce the properties of this distribution studying a finite sample of x_{it} on a time window T, consisting of many days (say one month).

We start from the simplest and most known stable distribution, which is Gaussian.

9.2 Gaussian world

9.2.1 Pearson estimators

Simplest models assume the distribution to be Gaussian. If this is the case, it can be characterized by two parameters: the shift $\delta_i = x_i \longrightarrow$ and the variance $\sigma_i^2 = 2\gamma_i^2 = (x_i - \delta_i)^2 \longrightarrow$. Both parameters can be easily determined empirically from the data on a time window T by the following estimators

$$\begin{aligned} \tilde{\delta}_i &= \frac{1}{T}\sum_t^T x_{it}\,, \\ \tilde{\sigma}_i^2 &= \frac{1}{T}\sum_t^T \left(x_{it} - \tilde{\delta}_i\right)^2\,. \end{aligned} \tag{9.7}$$

Obviously these numbers would be subject to a statistical error due to the finiteness of the time window. The values of the estimators converge to the exact values $\tilde{\delta}_i \to \delta_i$, $\tilde{\sigma}^2 \to \sigma^2$ only in the limit $T \to \infty$. In the Gaussian world the evolution of the price (or in our case the logarithm of the price) is just a diffusion process with a drift. Knowledge of the parameters of the Gaussian distribution describing price changes in one day can be used to predict the distribution of the relative price changes on a longer time scales.

These will again be given by the Gaussian distribution (due to its stability), but with rescaled variance and shift.

The market consists of many assets (say $i = 1, \ldots, N$). The number of assets in the market is typically a large number (the well-known Standard and Poor index SP500 quotes prices of 500 companies). The market reality is more complex than suggested by the model of independent stationary Gaussian returns discussed above.

The first problem is that the market reality is *not stationary*. One cannot expect that the prices will fluctuate according to the same law over twenty years. In this period many things may happen which may affect performances of individual companies. One has to weaken the stationarity assumption and to substitute it by a sort of quasi-stationarity. In practice this means that the time window T used in the estimators (9.7) should be limited and so should be the future time in which one uses the value of the estimators. Practitioners [**?**] introduce further improvements to the estimators by weighting past events with weight, which gradually decreases with time. Here we shall not discuss this issue further, assuming in what follows a quasi-stationarity.

The second correction which one has to introduce to the model discussed above is that in reality the prices of individual stocks are *mutually correlated* as a result of the existence of the network of inter-company dependencies. Indeed even by a purely statistical analysis of the correlation matrix [**?**] one can observe and determine the statistical correlations of price fluctuations of stock prices of companies from the same industrial sectors. Of course, inter-sector correlations also exist. Further, the stock market is not a closed system. The total capital invested in the market may shift between the stock market and other investments like for instance the real estate. This leads to the observed periods of flows of the capital into the stock market or out of the stock market. As a result the prices may go up or down, depending on whether the market attracts are repulses the capital. This is closely related to the effect known in sociology as herding. The effect of herding is also clearly seen in the statistical analysis of the matrix which shows the occurrence of an eigenvalue in the spectrum of the correlation matrix which is significantly larger than all other. The corresponding eigenvector is interpreted as a vector of correlations of changes of individual prices to the main market tendencies which are often referred to as β-parameters after the Capital Asset Pricing Model [**?**]. We shall come back to this issue later. This discussion shows that a realistic approach should allow to model the inter-company correlations.

A logical generalization of the Gaussian model described above is the

model of correlated asset fluctuations generated from some multidimensional Gaussian distribution. The probability of generating a vector of returns x_{it}, $i = 1, \ldots, N$ at some time t is

$$\prod_i dx_i \; P(x_1, x_2, \ldots, x_N) \sim \prod_i dx_i \; \exp -\frac{1}{2} \sum_{ij} (x_i - \delta_i) C_{ij}^{-1} (x_j - \delta_j) \, . \quad (9.8)$$

The properties of this generator can be assumed, as discussed before, to be constant in the period of time for which the shifts δ_i and the correlation matrix C_{ij} are estimated (quasi-stationarity)

$$\tilde{C}_{ij} = \frac{1}{T} \sum_t^T \left(x_{it} - \tilde{\delta}_i \right) \left(x_{jt} - \tilde{\delta}_j \right) \, . \quad (9.9)$$

The correlations may be both positive or negative. Knowledge of the correlation matrix C_{ij} is crucial in financial engineering, and in the construction of "optimal portfolios" following the Markowitz recipe [**?**].

9.2.2 Portfolio optimization

The main idea in the construction of "optimal portfolios" is to reduce the risk by diversification. The most simplified portfolio is constructed by dividing the total invested capital into fractions p_i which are held in different assets: $\sum_i^N p_i = 1$. The evolution of the return of the portfolio is now given by the stochastic linearized variable $X(\vec{p}) = \sum_i^N p_i x_i$, which produces an instantaneous return $X(\vec{p})_t = \sum_i^N p_i x_{it}$ at time t. The quintessence of the Markowitz idea is to minimize the fluctuations of the random variable $X(\vec{p})$ at a given expected return by optimally choosing the p_i's. The risk is measured by the variance of the stochastic variable $X(\vec{p})$

$$\Sigma^2 = \sum_{ij} p_i C_{ij} p_j \, . \quad (9.10)$$

Clearly, the information encoded in C_{ij} is crucial for the appropriate choice of p_i's. Intuitively, a diversification makes only sense when one diversifies between independent components and one does not gain too much if one redistributes capital between strongly correlated assets which make collective moves on the market.

The covariance matrix contains this precious information about the independent components. The spectrum of eigenvalues tells us about the strength

of fluctuations of individual components, and the corresponding eigenvectors about the participation of different assets in this independent components.

The fundamental question which arises is how good is the estimate $\tilde{C}_{ij}$ given by the equation (9.8) of the underlying covariance matrix (9.9), in particular how good is the risk estimate

$$\tilde{\Sigma}^2 = \sum_{ij} \tilde{p}_i \tilde{C}_{ij} \tilde{p}_j \tag{9.11}$$

of risk (9.10). Although the question looks simple, the answer is not immediate. One can quantify the answer with the help of the random matrix theory.

To start with, consider the simplest case of completely uncorrelated assets which are equally risky. Further, we assume that they all fluctuate symmetrically around zero $\delta_i = 0$ with the same variance $\sigma_i = 1$. The correlation matrix reads in this case $C_{ij} = \delta_{ij}$. The spectrum of eigenvalues of this matrix is $\rho(\lambda) = \delta(\lambda - 1)$ which means that it is entirely localized at unity. For the ideal diversification $p_i = 1/N$ the risk measured by Σ (9.10) is $\Sigma = 1/\sqrt{N}$. What shall we obtain if we use in this case the estimate $\tilde{C}_{ij}$ instead?

The random matrix theory gives a definite answer. The first observation is that the quality of the estimator (9.9) depends on the time T for which we could measure the correlation matrix. The longer time T, the better quality of the information which can be read of from $\tilde{C}_{ij}$: all diagonal elements should approach unity, and off-diagonal ones zero. In reality, as we mentioned, one never has an infinite time T at ones disposal. Geometry of the data matrix $x_{it}, i = 1, \ldots, N, t = 1, \ldots, T$ is finite. It is just a rectangular matrix with the asymmetry parameter $c = N/T < 1$. Such matrices form an ensemble called the Wishart ensemble [**?**]. The case $c > 1$ requires a special treatment and is not relevant in this case. For c larger than zero we expect that the spectrum of the matrix $\tilde{C}$ will be smeared in comparison with the delta spectrum of C. Since matrix $\tilde{C}$ is a Wishart matrix, its spectrum is given by the Marchenko-Pastur formula

$$\tilde{\rho}(\lambda) = \frac{1}{2\pi c} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \tag{9.12}$$

with $\lambda_\pm = (1 \pm \sqrt{c})^2$. Only in the limit $c \to 0$ we get the spectrum peaked at unity.

Although the empirical matrix x_{it} is obtained from a single realization of a random matrix from the Wishart ensemble, its spectral properties are

in general very similar to those described above. This is due to the self-averaging property of large matrices.

We can also explicitly find the estimate of risk (9.11). With the help of Lagrange multiplier μ we define $f(\mu) = \tilde{\Sigma} - \mu(\sum \tilde{p}_i - 1)$. Condition $\frac{\partial f}{\partial \tilde{p}_i} = 0$, together with $\sum \tilde{p}_i = 1$, gives the optimal choice of probabilities $\tilde{p}_i$ which minimizes the risk $\tilde{\Sigma}$ depends on $\tilde{C}_{ij}$

$$\tilde{p}_i = \frac{\sum_j^N \tilde{C}_{ij}^{-1}}{\sum_{jk}^N \tilde{C}_{jk}^{-1}} \,. \tag{9.13}$$

Inserting this solution into the $\Sigma^2 = \sum p_i C_{ij} p_j$ we can calculate the minimal value of the estimated risk

$$\Sigma^2 = \frac{1}{N} \frac{\int d\lambda \; \rho(\lambda)\lambda^{-2}}{\left(\int d\lambda \; \rho(\lambda)\lambda^{-1}\right)^2} \tag{9.14}$$

which eventually gives

$$\Sigma = \frac{1}{\sqrt{N}} \frac{1}{\sqrt{1-c}} \,. \tag{9.15}$$

Note that $\frac{1}{\sqrt{N}}$ is the "true risk". The ratio of estimated risk to true risk explodes when $c \to 1$!

9.2.3 Lesson from data

An example of the eigenvalue spectrum of the empirical covariance matrix $\tilde{C}$ (9.9), is shown in figure 9.1. It is calculated for the SP500 for the period. The data matrix x_{it} has the size $N = 406$ and $T = 1308$ which corresponds to the asymmetry parameter $a = 0.31$. In the spectral analysis of the empirical matrix one usually unifies the scale of return fluctuations of different assets by normalizing them by individual variances σ_i (9.7): $x_{it} \to x_{it}/\sigma_i$ which for each asset produces fluctuations of unit width. For such normalized fluctuations the formula (9.12) tells us that that the random part of the spectrum of the covariance matrix should be concentrated between 0.20 and 2.43. We clearly see the presence of larger eigenvalues in the spectrum presented in the left plot in figure 9.3, which as mentioned, can be attributed to the inter-asset correlations. However, the large eigenvalues disappear when one removes the inter-asset correlation. One can do this by random reshuffling of the time ordering of returns for each individual asset. A random reshuffling does

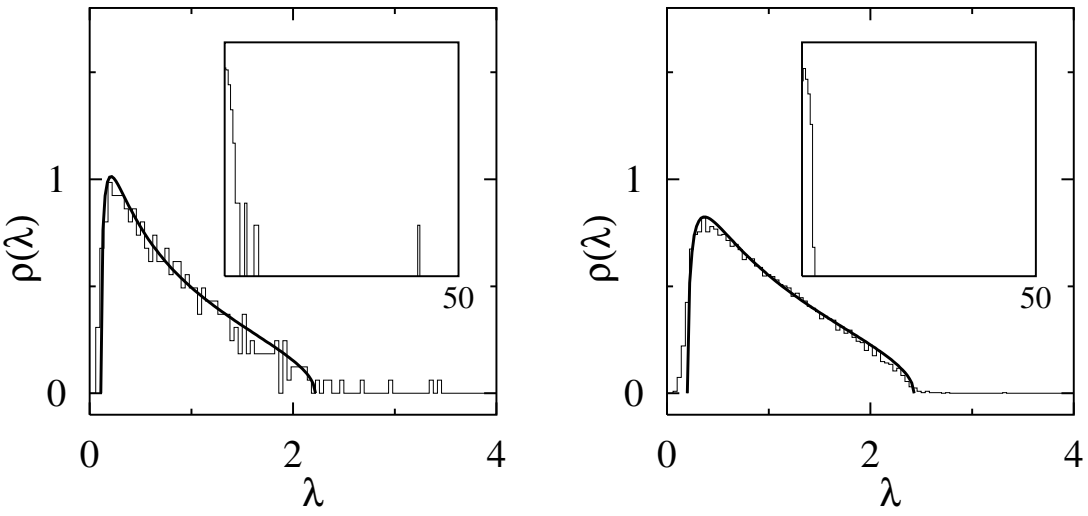


Figure 9.1: The spectrum of the financial covariance matrix for the daily SP500 for $N = 406$ stocks and for $T = 1309$ days from 01.01.1991 to 06.03.1996. The left plot represents the spectrum of the covariance matrix for the normalized returns in the natural time ordering; the right one for the normalized return in the reshuffled ordering. The reshuffling destroys correlations between entries of the matrix $\tilde{C}_{ij}$. The random matrix prediction is plotted in solid line. The large eigenvalues lying outside the random matrix spectrum in the left figure disappear from the spectrum for reshuffled data shown in the right.

not change the content of information stored in each separate row of data but it destroys the statistical information about the correlations between different rows. Indeed as is shown on the right plot in the figure 9.1, the larger eigenvalues disappear from the spectrum. The resulting spectrum of the covariance matrix of such reshuffled data is perfectly described by the random matrix formula (9.7). From this point of view, it is also puzzling how late the random matrices (in our language matricial probabilities) were used for the analysis of financial data. The breakthrough came in 1999, when two groups [**?**, **?**] have analyzed the spectral characteristics of empirical covariances, calculated for all companies belonging to Standard and Poor 500 index, which remained listed from 1991 till 1996. The spectrum of the empirical covariance matrix constructed from this matrix was then confronted

with the analytically known spectrum of a covariance matrix constructed solely from the maximal-entropy (Gaussian) ensemble with the same number of rows and columns.

The unexpected (for many) results showed, that the majority of the spectrum of empirical covariance matrices is populated by noise!

In the case of a Gaussian disorder, 94% of empirical eigenvalues were consistent with random matrix spectra [**?**]. Only few largest eigenvalues did not match the pattern, reflecting the appearance of large clusters of companies, generally corresponding to the sectorization of the market and market itself [**?**].

At the end of this section let us come to the problem of the large eigenvalues observed in the spectra of eigenvalues of the financial covariance matrices $\tilde{C}_{ij}$. The spectra consist typically of the random part (9.12) which is universal as discussed above and few large eigenvalues. Among them one is particularly large. Its value is roughly speaking proportional to the number N of the assets in the market. The corresponding eigenvector contains the contribution from almost all N companies on the market. This eigenvector is called the "market". One can relatively easily understand the source of the appearance of the market in the spectrum in terms of the herding phenomena which we shortly signaled before. Imagine that there is a collective behavior of investors on the market which can be driven by some sociological factors. Mathematically such a collective movement may be in the simplest version modeled by the coupling of the individual prices to some common background, for example by substituting the generator of the vector of prices (9.8) by a new generator of the form

$$\prod_i dx_i \; P(\vec{x}) \sim \prod_i dx_i \; \exp -\frac{1}{2} \sum_{ij} (x_i - \beta_i m_t) C_{ij}^{-1} (x_j - \beta_j m_t) \,, \qquad (9.16)$$

where β_i's are some constants, and m_t is a common random variable describing the market movements. This is the basic idea underlying the CAPM model [**?**] mentioned above. One can check that the largest eigenvalue disappears from the spectrum leaving the remaining part intact if at each t one subtracts from each return the market background represented as the instantaneous average over all companies.

The other large eigenvalues can be attributed to the real strong correlations between companies. The analysis of the eigenvectors allows to divide the market into highly correlated clusters, usually corresponding to compa-

nies from the same industrial sector. For example, one can see that the gold companies form a cluster which is anti-correlated to the market.

The random matrix analysis posed therefore a fundamental question for quantitative finances. If empirical covariance matrices are so "noisy", why there are so valuable for practitioners? Every industrial application of risk measurement depends heavily on covariance matrix formulation. The Markowitz's theory of diversification of investment portfolios depends crucially on the information included in the covariance matrix [**?**]. If indeed the lower part of the covariance matrix spectrum has practically no information, the effects of noise would strongly contaminate the optimal choice of the diversification, resulting in the dangerous underestimation of the risk of the portfolio.

Bouchaud and others [**?**] suggested a way out, simply filtering out the noisy part of the correlation matrix and repeating the Markowitz analysis with refined matrix. This resulted in a better approximation of the risk.

Their analysis did not answer however the fundamental question. If the original matrix is noisy, *ie* has almost no information, how come the covariance matrices form the pillars of quantitative finance?

9.2.4 Denoising by FRV

Let us try to unravel a rather nontrivial relation between the true covariance matrix C and its estimator $\tilde{C}$, using the power of Free Random Variables. Les us consider a Gaussian world with non-trivial cross-covariances and auto-covariances $C_{ia,jb} = C_{ij}A_{ab}$, i.e. the multivariate Gaussian distribution

$$P_{cG}(X)dX = \frac{1}{N_{cG}} \exp\left(-\frac{1}{2}\mathrm{tr}X^{\dagger}C^{-1}XA^{-1}\right) dX \tag{9.17}$$

where normalization constant reads $N_{cG} = (2\pi)^{NT/2}(\det C)^{T/2}(\det A)^{N/2}$. We denote transposition by dagger, to avoid confusion with the notation for the length of the time T. Cross-covariance matrix A is an N by N matrix, and autocorrelation matrix C is a T by T matrix. The Pearson estimator for cross-covariance reads $\tilde{C} = \frac{1}{T}XX^{\dagger}$. We want to calculate the Green's function $G_{\tilde{C}}(z) = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z-\tilde{C}}\right\rangle_{cG}$. The trick is to change the variables $X = \sqrt{C}Y\sqrt{A}$, thanks to the positivity of cross- and auto-correlation matrices. Note that the measure becomes uncorrelated (Gaussian) i.e

$$P_G(Y)dY = \frac{1}{N_G} \exp\left(-\frac{1}{2}\mathrm{tr}Y^{\dagger}Y\right) dY \tag{9.18}$$

where $N_G = (2\pi)^{NT/2}$, but the estimator $\tilde{C}$ is now dressed by the a priori unknown correlation matrices A and C, i.e.

$$\tilde{C} = \frac{1}{T}\sqrt{C}YAY^\dagger\sqrt{C} \tag{9.19}$$

We have therefore rephrased the problem of the resolvent for $\tilde{C}$ as

$$G_{\tilde{C}}(z) = \frac{1}{N}\left\langle \mathrm{tr}\frac{1}{z - \frac{1}{T}\sqrt{C}YAY^\dagger\sqrt{C}}\right\rangle_G \tag{9.20}$$

i.e. the Green's function for complicated matrix, but averaged with respect to trivial (uncorrelated) Gaussian measure. The combination of the cyclic property of trace operation and free random variable calculus will allow us to find an exact relation between the moments of the true covariance matrix and its estimator.

Before we do this, let us recall basic facts of multiplicative properties of free random variables. First, the multiplicative transform S for an ensemble E is defines as $S_E(z) = \frac{1+z}{z}\chi_E(z)$, where $\frac{1}{\chi_E(z)}G_E(\frac{1}{\chi_E(z)}) - 1 = z$. In this book we prefer the notation $M_E(z) = zG_E(z) - 1$ and $N_E(z) = \frac{1}{\chi_E(z)}$. Then $M_E[N_E(z)] = z$, i.e. N_E is the functional inverse of the moment generating function $M_E(z) = \sum_{n=1}^{\infty}\frac{m_n^{(E)}}{z^n}$. The multiplication law for N stems from multiplication law of S-transform and reads

$$\frac{z}{1+z}N_{E_1}(z)\cdot N_{E_2}(z) = N_{E_1\cdot E_2}(z) \tag{9.21}$$

for any free ensembles E_1 and E_2.

Second, we recall that since the S-transform for Wishart ensemble $W = \frac{1}{T}YY^\dagger$ reads $S_W(z) = \frac{1}{1+cz}$, where $1 > c = N/T$, the N-transform reads

$$N_W(z) = \frac{(1+z)(1+cz)}{z} \tag{9.22}$$

Finally, we notice the particular duality property between the N by N Wishart ensemble $W = \frac{1}{T}YY^\dagger$ and the "anti-Wishart" T by T ensemble $\bar{W} = \frac{1}{N}Y^\dagger Y$. Note that matrices $YY^\dagger$ and $Y^\dagger Y$ have identical *non-zero* eigenvalues and the spectrum between the larger one and the smaller one differs only by $T - N$ zero modes. The moments are therefore identical, modulo the rescaling coming from different normalization, i.e. $M_W(z) = cM_{\bar{W}}(cz)$.

We therefore read $N_{\bar{W}}(z) = \frac{1}{c} N_W(\frac{z}{c}) = \frac{(1+z)(r+z)}{rz}$. Finally, we note that for any real number g, $N_{gE} = gN_E(z)$. The above observations allow us to find strikingly simple relation between moments of the true covariance matrix and its estimator. It reads

$$z = cM_{\tilde{C}}(z) N_A[rM_{\tilde{C}}(z)] N_C[M_{\tilde{C}}(z)] \tag{9.23}$$

The proof goes as follows:

- We notice that cyclic property allows us to write $N_{\tilde{C}(z)}$ as a product of two free random matrices $\frac{1}{T} Y A Y^\dagger$ and C, hence

$$N_{\tilde{C}}(z) = N_{\frac{1}{T} Y A Y^\dagger C}(z) = \frac{z}{1+z} N_{\frac{1}{T} Y A Y^\dagger}(z) N_C(z) = ... \tag{9.24}$$

- Next, the cyclic property applied to the first of these matrices implies

$$... = \frac{z}{1+z} N_{\frac{1}{T} Y^\dagger A Y}(cz) N_C(z) = ... \tag{9.25}$$

  where the argument cz appears because the cyclic shift converts an N by N matrix into T by T matrix, which results in rescaling the moments accordingly.

- In the next step, we realize that the first N transform is again the product of two free random matrices, i.e the c-rescaled anti-Wishart $c\frac{1}{N} Y^\dagger Y$ and A, hence

$$... = \frac{z}{1+z} \frac{cz}{1+cz} N_{\frac{1}{T} T^\dagger Y} N_A(cz) N_C(z) = ... \tag{9.26}$$

- Finally, we exploit the explicit form of c-rescaled, non-correlated anti-Wishart, getting the result

$$... = czN_A(cz) N_C(z) \tag{9.27}$$

- The substitution $z \to M_{\tilde{C}}(z)$ yields the announced main formula (9.23).

9.2.5 Example

Let us see how the general formula (9.23) works in the simplified case, when $A = \mathbf{1}_T$, but C is arbitrary. Since $N_A(z) = 1 + 1/z$, the main formula reads

$$M_{\tilde{C}}(z) = M_C(w) \tag{9.28}$$

with $w = \frac{z}{1+rM_{\tilde{C}}(z)}$, i.e. relates the true and estimated moments by involved *conformal mapping* $w = f(z)$. The operational procedure is as follows.

- We exploit the conformal mapping, expanding $M_{\tilde{C}}(x = 1/z) = \sum \tilde{m}_i x^i$ and $M_C(z = 1/x) = \sum m_i x^i$, i.e.

$$\sum_{k=1}^{\infty} \tilde{m}_i x^i = \sum_{k=1}^{\infty} m_i x^i (1 + c \sum_{l=1}^{\infty} \tilde{m}_i x^i)^k \tag{9.29}$$

- The comparison between the lhs and rhs gives us an infinite tower of relations between the true moments and the moments of the estimator, i.e.

$$\begin{aligned} \tilde{m}_1 &= m_1 \,, \\ \tilde{m}_2 &= m_2 + c m_1^2 \,, \\ \tilde{m}_3 &= m_3 + 3 c m_1 m_2 + c^2 m_1^3 \\ &\cdots \end{aligned} \tag{9.30}$$

- We can solve these relations to find the reciprocal relations, i.e. to express the true moments m_i in terms of the moments of an estimator $\tilde{m}_i$.

- Due to the errors, we have always to truncate this infinite hierarchy of equations at some m_{max}, so the spectral parameters we seek can be written down as a vector $\Theta = (\lambda_1, ..., \lambda_{m_{max}}, p_1, ..., p_{m_{max}-1})$, where p_i are the weights $p_i = n_i/N$.

- Then, by definition, the Green's function for the approximant of the true covariance matrix reads

$$G_C(z) = \sum_{i=1}^{m_{max}} \frac{p_i}{z - \lambda_i} = \frac{1}{z}(1 + \sum_{j=1}^{2m_{max}-1} \frac{m_j}{z^j}) = \frac{R(z)}{P(z)} \tag{9.31}$$

  where R and P are polynomials whose coefficients are functions of eigenvalues and weights.

- Writing the polynomials explicitly (with $z = 1/x$) we get

$$1 + R_1 x + R_2 x^2 + ... + R_{m_{max}-1} x^{m_{max}-1} = \\ (1 + P_1 x + P_2 x^2 + ... + P_{m_{max}} x^{m_{max}})(1 + \sum_{k=1}^{2m_{max}-1} m_k x^k) \quad (9.32)$$

- Comparing coefficients at each power of x gives a set of $2m_{max} - 1$ linear equations for P_i and R_i. Then we rephrase the m_i in terms of the measured moments $\tilde{m}_i$, on the basis of conformal transformation relations. Eigenvalues of the covariance matrix correspond to the zeroes of $P(z)$, and their weights (degeneracies) are $p_i = \frac{R(\lambda_i)}{P'(\lambda_i)}$. In practice, it is convenient the use the Padé approximants for the optimal fit of the Θ vector. In figure XX, we show the result of such analysis for the aforementioned SP500 data. One can identify several correlated clusters in the bulk, which, due to the noise, mimic naively the bulk of the spectrum as an uncorrelated Wishart.

9.3 Beyond Gaussian world

We have mentioned, that in the above spectral analysis of the empirical matrix one usually unifies the scale of return fluctuations of different assets by normalizing them by individual variances σ_i (9.7): $x_{it} \to x_{it}/\sigma_i$ which for each asset produces fluctuations of unit width.

The above mentioned normalization of return fluctuation $x_{it} \to x_{it}/\sigma_i$ is natural *if* fluctuations belong to the Gaussian universality class. However, if the underlying distributions governing the return fluctuations have fat tails, this normalization is not appropriate since the variance of the distribution does not exist. In this case the use of the normalization $x_{it} \to x_{it}/\sigma_i$ artificially forces the resulting rescaled quantities to behave as if they belonged to the Gaussian universality class of distributions with the unit variance. This introduces a bias to the analysis in case of non-Gaussian statistics. Indeed, if one skips this normalization one observes that covariance matrices for the original SP500 data as well as for the reshuffled SP500 data both possess large eigenvalues in the spectra (see Fig. 9.3). What is the reason that the reshuffling does not remove them? Is the random matrix prediction (9.12) wrong? The random matrix prediction is not wrong of course but is valid

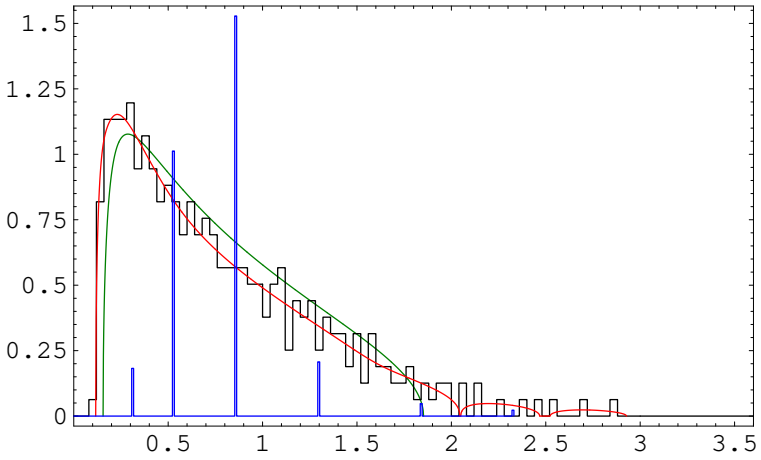


Figure 9.2: Analysis of the eigenvalue distribution presented above. Six eigenvalues were assumed, which resulted in the optimal fit of the experimental distribution. The determined eigenvalues are represented by blue lines and their relative height corresponds to the probability p_j. The red line is the fit of the distribution with six eigenvalues, to be compared with the green line, corresponding to the fit presented before. From Burda Jurkiewicz, 2003

only for matrices from the Gaussian ensemble. The removal of the normalization condition revealed the nature of the randomness of return fluctuations which contain fat tails. For example, the spectra of Lévy random matrices contain fat tails which means that even a completely random matrix may contain large eigenvalues. The main conclusion of this discussion is that the large eigenvalues in the spectrum of financial covariances stem both from inter-asset correlations and from the Lévy statistics of return fluctuations and therefore a proper statistical analysis of financial data, in principle of the eigenvalue content, would require the new methodology based on random matrices with fat tailed spectrum. Such constructions are possible [**?**] , however, they are much more involved at the technical level and go beyond the introductory and exemplary character of this chapter.

9.4 Lagged correlations from FRV

9.4.1 Symmetrization - quartic equation for the resolvent

9.4.2 Singular value analysis - cubic equation for the resolvent

9.4.3 R-diagonal free Poisson - another cubic equation for the resolvent

9.4.4 Whitening - free Jacobi ensemble - quadratic equation for the resolvent

Chapter 10

Sample applications in information theory and machine learning

10.1 Information theory redux

Let x be the value of random variable, which comes from the set of *discrete* values $A \equiv \{a_1, a_2, ..., a_I\}$ ("alphabet"), with corresponding probabilities $P \equiv \{p_1, p_2, ..., p_i\}$. In brief, $P(x = a_i) = p_i$. Obviously, $p_i \geq 0$ and $\sum_A P(x = a_i) = 1$.

We consider the problem of measuring the information content of the outcome variable $x = a_i$ coming from some source. Shannon defined it as

$$h(x) = \log \frac{1}{P(x)} \tag{10.1}$$

Indeed, assuming that:
1. We never loose the information (positivity of h)
2. We gain more information from less probable events ($h(a_k) > h(a_i)$ if $p_k < p_i$
3. For statistically independent "letters" a_i, a_k, information content is additive $h(a_i \cdot a_j) = h(a_i) + h(a_j)$
Shannon has proven that information content of the outcome x is given uniquely by the minus logarithm of probability. In information theory, one uses the binary basis, since "alphabet" is composed of 0 and 1 (bits), but

since the change of the basis of a logarithm is trivial, we will usually omit the subscript of the logarithm.

Let us name the triple *random variable, alphabet, probabilities* as an ensemble X. The entropy of the ensemble X is defined as the average Shannon information content generated by the source and reads

$$H(X) = \sum_x P(x) \log \frac{1}{P(x)} \tag{10.2}$$

For independent random variables x, y probability factorizes, hence addition of information content. This implies that Shannon entropies are also additive for independent ensembles,

$$H(X,Y) = \sum_{x,y} P(x)P(y) \log \frac{1}{P(x)P(y)} = H(X) + H(Y) \tag{10.3}$$

If variables are dependent, we introduce conditional entropy of ensemble X provided $y = b_k$

$$H(X|y = b_k) = \sum_x P(x|y = b_k) \log \frac{1}{P(x|y = b_k)} \tag{10.4}$$

We may generalize now the concept of the conditional entropy of ensemble X given Y, by performing average over y

$$H(X|Y) = \sum_y P(y) \left[\sum_x P(x) \log \frac{1}{P(x|y)} \right] = \sum_{x,y} P(x,y) log \frac{1}{P(x|y)} \tag{10.5}$$

where we used the definition of conditional probability $P(x,y) = P(y)P(x|y) = P(x)P(y|x)$. If we introduce joint entropy $H(X,Y) \equiv \sum_{x,y} P(x,y) \log \frac{1}{P(x,y)}$, definition of conditional probability implies the chain rule

$$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y) \tag{10.6}$$

Finally, we introduce mutual information between ensembles X and Y as

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X) \tag{10.7}$$

which measures averaged reduction of uncertainty about x that results from learning the value of y (first equation) or average amount of information that x brings about y (second equation). Note that using the definition

Figure 10.1: Schematic relation between various entropies

.

of marginal entropy $H(X)$ and conditional entropy $H(X|Y)$ we can write explicitly the mutual entropy as

$$I(X;Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{10.8}$$

Relations between all above entropies can be summarized in the form of cartoon.

In the information theory, switching from discrete random variables to the continuous ones is subtle. Entropy for continuous variables is infinitely large, since the number of possible values in uncountable. However, the mutual information (10.8) can be generalized for the continuous variables

$$I(X;Y) = \int_y dy \int_x dx P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{10.9}$$

10.2 Single input single output Gaussian channel

Let us define now the channel capacity C as

$$C = \max \; I(X;Y)|_{P_X} \tag{10.10}$$

The distribution P_X that maximises the mutual information is called the optimal input distribution.

We consider a Gaussian channel, i.e. the one when the noise is given by the Gaussian. Then the output y is the sum of input signal x and the noise b.

$$y = x + b \tag{10.11}$$

The entropy of the noise is fixed. Since $I(X;Y) = H(Y) - H(Y|X)$, where the last term is just the entropy of the noise, maximization of the mutual information requires the maximization of $H(Y)$, which corresponds also to the Gaussian distribution. We also know that the sum of two variables from

Gaussian distribution has a Gaussian distribution, hence $P(X)$ has to be also Gaussian. This is Shannon's continuous noisy channel coding theorem for Gaussian channels. Since the source has always finite power (here the variance σ_x^2), calculation of $H(X)$ for Gaussian pdf yields

$$H(X) = \frac{1}{2}\log 2\pi e\sigma_x^2 \tag{10.12}$$

Similarly

$$H(b) = \frac{1}{2}\log 2\pi e\sigma_b^2 \tag{10.13}$$

hence entropy of Y is

$$H(Y) = \frac{1}{2}\log 2\pi e\sigma_y^2 = \frac{1}{2}\log 2\pi e(\sigma_x^2 + \sigma_b^2) \tag{10.14}$$

Then maximum of $I(X|Y) = H(Y) - H(b)$ reads

$$C = \frac{1}{2}\log(1 + \frac{P}{N}) \tag{10.15}$$

where P is the input signal *power* σ_x^2 and N is the *noise* power σ_b^2. The ratio P/N is therefore called the signal to noise ratio (SNR or SINR).

10.3 Multiple input multiple output channel

10.4 Deep networks