Saarland University
Faculty of Mathematics and Computer Science
Department of Mathematics

Random matrices

Lecture notes

Winter 2019/20

(version 2 from May 14, 2020)

Prof. Dr. Roland Speicher

This in an introduction to random matrix theory, giving an impression of some of the most important aspects of this modern subject. In particular, it covers the basic combinatorial and analytic theory around Wigner's semicircle law, featuring also concentration phenomena, and the Tracy–Widom distribution of the largest eigenvalue. The circular law and a discussion of Voiculescu's multivariate extension of the semicircle law, as an appetizer for free probability theory, also make an appearance.

This manuscript here is an updated version of a joint manuscript with Marwa Banna from an earlier version of this course; it relies substantially on the sources given in the literature; in particular, the lecture notes of Todd Kemp were inspiring and very helpful at various places.

The material here was presented in the winter term 2019/20 at Saarland University in 24 lectures of 90 minutes each. The lectures were recorded and can be found online at https://www.math.uni-sb.de/ag/speicher/web_video/index.html.

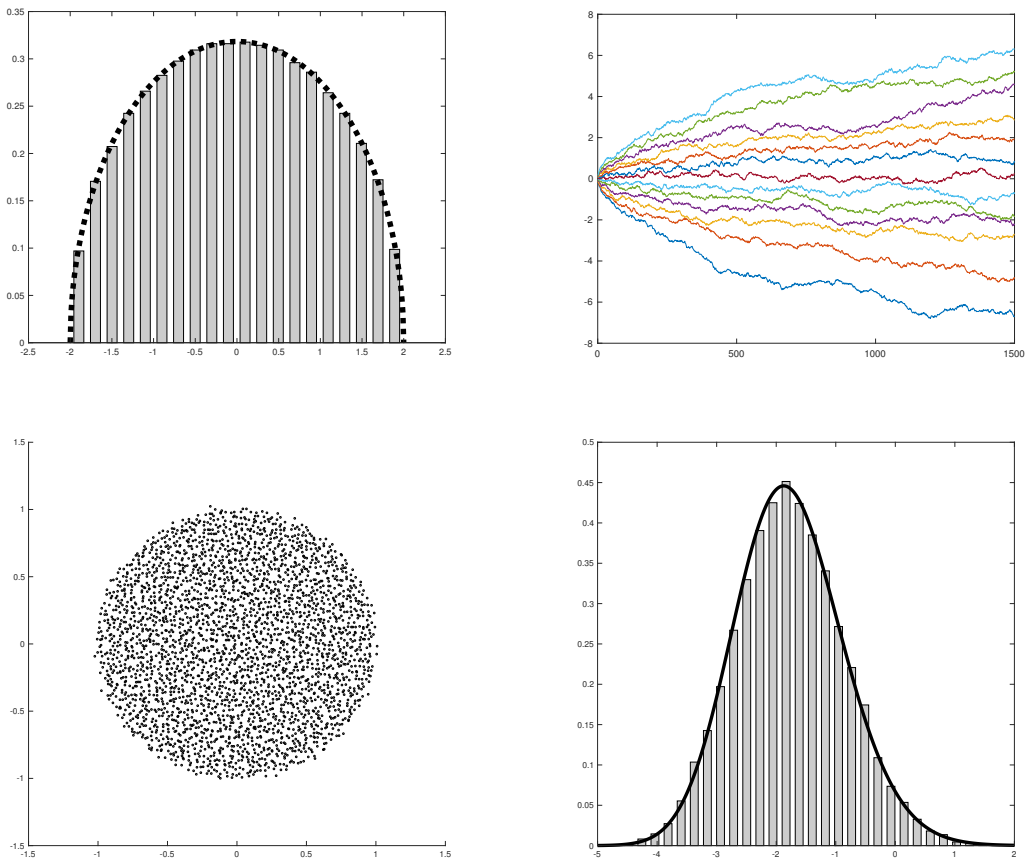


Table of contents

1 Introduction **7**
1.1 Brief history of random matrix theory . . . 7
1.2 What are random matrices and what do we want to know about them? 8
1.3 Wigner's semicircle law . . . 11
1.4 Universality . . . 12
1.5 Concentration phenomena . . . 12
1.6 From histograms to moments . . . 13
1.7 Choice of scaling . . . 14
1.8 The semicircle and its moments . . . 15
1.9 Types of convergence . . . 16

2 Gaussian Random Matrices: Wick Formula and Combinatorial Proof of Wigner's Semicircle **19**
2.1 Gaussian random variables and Wick formula . . . 19
2.2 Gaussian random matrices and genus expansion . . . 23
2.3 Non-crossing pairings . . . 27
2.4 Semicircle law for GUE . . . 29

3 Wigner Matrices: Combinatorial Proof of Wigner's Semicircle Law **33**
3.1 Wigner matrices . . . 33
3.2 Combinatorial description of moments of Wigner matrices . . . 34
3.3 Semicircle law for Wigner matrices . . . 37

4 Analytic Tools: Stieltjes Transform and Convergence of Measures **39**
4.1 Stieltjes transform . . . 40
4.2 Convergence of probability measures . . . 43
4.3 Probability measures determined by moments . . . 46
4.4 Description of weak convergence via the Stieltjes transform . . . 47

5 Analytic Proof of Wigner's Semicircle Law for Gaussian Random Matrices **49**
5.1 GOE random matrices . . . 49

5.2 Stein's identity for independent Gaussian variables 51
5.3 Semicircle law for GOE 52

6 Concentration Phenomena and Stronger Forms of Convergence for the Semicircle Law 55
6.1 Forms of convergence 55
6.2 Markov's and Chebyshev's inequality 56
6.3 Poincaré inequality 58
6.4 Concentration for $\mathrm{tr}[R_A(z)]$ via Poincaré inequality 63
6.5 Logarithmic Sobolev inequalities 65

7 Analytic Description of the Eigenvalue Distribution of Gaussian Random Matrices 67
7.1 Joint eigenvalue distribution for GOE and GUE 67
7.2 Rewriting the Vandermonde 72
7.3 Rewriting the GUE density in terms of Hermite kernels 73

8 Determinantal Processes and Non-Crossing Paths: Karlin–McGregor and Gessel–Viennot 81
8.1 Stochastic version à la Karlin–McGregor 81
8.2 Combinatorial version à la Gessel–Viennot 83
8.3 Dyson Brownian motion and non-intersecting paths 86

9 Statistics of the Largest Eigenvalue and Tracy–Widom Distribution 89
9.1 Some heuristics on single eigenvalues 89
9.2 Tracy–Widom distribution 90
9.3 Convergence of the largest eigenvalue to 2 91
9.4 Estimate for fluctuations 95
9.5 Non-rigorous derivation of Tracy–Widom distribution 96
9.6 Proof of the Harer–Zagier recursion 101

10 Statistics of the Longest Increasing Subsequence 107
10.1 Complete order is impossible 107
10.2 Tracy–Widom for the asymptotic distribution of L_n 109
10.3 Very rough sketch of the proof of the Baik, Deift, Johansson theorem 109
10.3.1 RSK correspondence 110
10.3.2 Relation to non-intersecting paths 112

11 The Circular Law **113**
11.1 Circular law for Ginibre ensemble 113
11.2 General circular law 117

12 Several Independent GUEs and Asymptotic Freeness **119**
12.1 The problem of non-commutativity 119
12.2 Joint moments of independent GUEs 120
12.3 The concept of free independence 122

13 Exercises **127**
13.1 Assignment 1 127
13.2 Assignment 2 128
13.3 Assignment 3 129
13.4 Assignment 4 130
13.5 Assignment 5 131
13.6 Assignment 6 133
13.7 Assignment 7 135
13.8 Assignment 8 135
13.9 Assignment 9 137
13.10 Assignment 10 138
13.11 Assignment 11 139

14 Literature **141**

1 Introduction

We start with giving a brief history of the subject and a feeling for some the basic objects, questions, and methods - this is just motivational and should be seen as an appetizer. Rigorous versions of statements will come in later chapters.

1.1 Brief history of random matrix theory

- 1897: the paper *Über die Erzeugung der Invarianten durch Integration* by Hurwitz is, according to Diaconis and Forrester, the first paper on random matrices in mathematics;
- 1928: usually, the first appearance of random matrices, for fixed size N, is attributed to a paper of Wishart in statistics;
- 1955: Wigner introduced random matrices as statistical models for heavy nuclei and studied in particular the asymptotics for $N \to \infty$ (the so-called "large N limit"); this set off a lot of activity around random matrices in physics;
- since 1960's: random matrices have become important tools in physics; in particular in the context of *quantum chaos* and *universality* questions; important work was done by by Mehta and Dyson;
- 1967: the first and influential book on *Random Matrices* by Mehta appeared;
- 1967: Marchenko and Pastur calculated the asymptotics $N \to \infty$ of Wishart matrices;
- $\sim$ 1972: a relation between the statistics of eigenvalues of random matrices and the zeros of the Riemann ζ-function was conjectured by Montgomery and Dyson; with substantial evidence given by numerical calculations of Odlyzko; this made the subject more and more popular in mathematics;
- since 1990's: random matrices are studied more and more extensively in mathematics, in the context of quite different topics, like
  - Tracy-Widom distribution of largest eigenvalue
  - free probability theory
  - universality of fluctuations
  - "circular law"
  - and many more

1.2 What are random matrices and what do we want to know about them?

A *random matrix* is a matrix $A = (a_{ij})_{i,j=1}^N$ where the entries a_{ij} are chosen randomly and we are mainly interested in the *eigenvalues* of the matrices. Often we require A to be selfadjoint, which guarantees that its eigenvalues are real.

Example 1.1. Choose $a_{ij} \in \{-1, +1\}$ with $a_{ij} = a_{ji}$ for all i, j. We consider all such matrices and ask for typical or generic behaviour of the eigenvalues. In a more probabilistic language we declare all allowed matrices to have the same probability and we ask for probabilities of properties of the eigenvalues. We can do this for different sizes N. To get a feeling, let us look at different N.

- For $N = 1$ we have two matrices.

matrix	eigenvalues	probability of the matrix
(1)	$+1$	$\frac{1}{2}$
(-1)	-1	$\frac{1}{2}$

- For $N = 2$ we have eight matrices.

matrix	eigenvalues	probability of the matrix
$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$0, 2$	$\frac{1}{8}$
$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$\frac{1}{8}$
$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	$0, 2$	$\frac{1}{8}$
$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$\frac{1}{8}$
$\begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$\frac{1}{8}$
$\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$	$-2, 0$	$\frac{1}{8}$

$\begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$\frac{1}{8}$
$\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$	$-2, 0$	$\frac{1}{8}$

- For general N, we have $2^{N(N+1)/2}$ matrices, each counting with probability $2^{-N(N+1)/2}$. Of course, there are always very special matrices such as

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix},$$

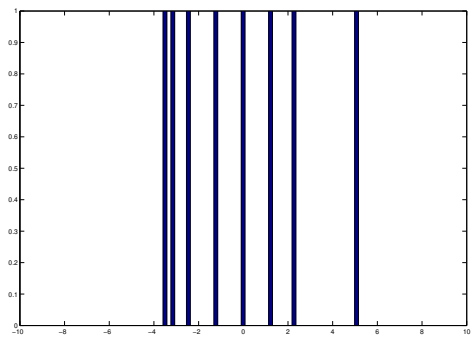
  where all entries are $+1$. Such a matrix has an atypical eigenvalue behaviour (namely, only two eigenvalues, N and 0, the latter with multiplicity $N-1$); however, the probability of such atypical behaviours will become small if we increase N. What we are interested in is the behaviour of most of the matrices for large N.

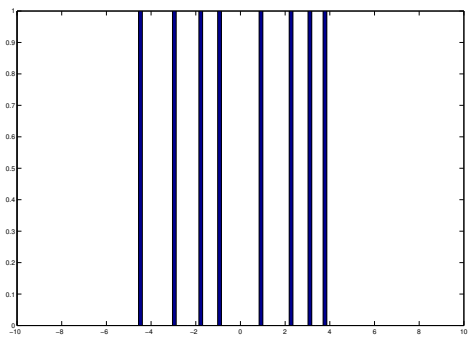
  Question. What is the "typical" behaviour of the eigenvalues?

- Here are two "randomly generated" (by throwing a coin for each of the entries on and above the diagonal) 8×8 symmetric matrices with ± 1 entries ...

$$\begin{pmatrix} -1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & 1 & -1 & -1 \end{pmatrix} \qquad \begin{pmatrix} 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}$$

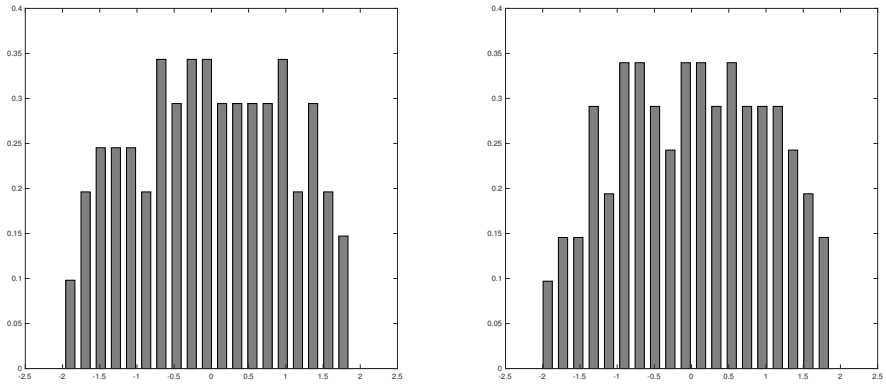
  ... and their corresponding histograms of the 8 eigenvalues



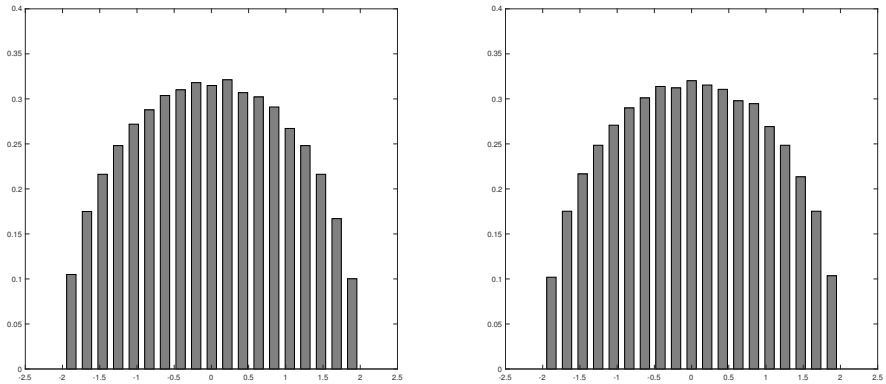


The message of those histograms is not so clear - apart from maybe that degeneration of eigenvalues is atypical. However, if we increase N further then there will appear much more structure.

- Here are the eigenvalue histograms for two "random" 100×100 matrices ...



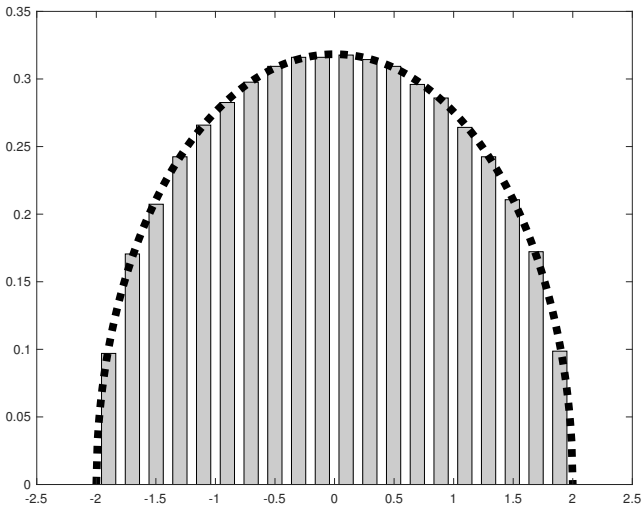
- ... and here for two "random" 3000×3000 matrices ...



To be clear, no coins were thrown for producing those matrices, but we relied on the MATLAB procedure for creating random matrices. Note that we also rescaled our matrices, as we will address in Section 1.7.

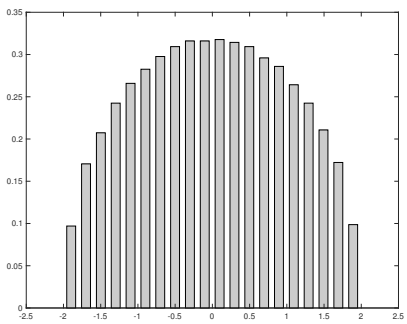
1.3 Wigner's semicircle law

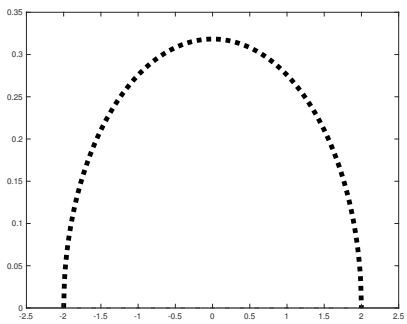
What we see in the above figures is the most basic and important result of random matrix theory, the so-called *Wigner's semicirlce law* ...



... which says that typically the eigenvalue distribution of such a random matrix converges to Wigner's semicircle for $N \to \infty$.

eigenvalue distribution $\quad \to \quad$ semicircle





Note the quite surprising feature that the limit of the random eigenvalue distribution for $N \to \infty$ is a deterministic object - the semicircle distribution. The randomness disappears for large N.

1.4 Universality

This statement is valid much more generally. Choose the a_{ij} not just from $\{-1, +1\}$ but, for example,

- $a_{ij} \in \{1, 2, 3, 4, 5, 6\}$,
- a_{ij} normally (Gauß) distributed,
- a_{ij} distributed according to your favorite distribution,

but still independent (apart from symmetry), then we still have the same result: The eigenvalue distribution typically converges to a semicircle for $N \to \infty$.

1.5 Concentration phenomena

The (quite amazing) fact that the a priori random eigenvalue distribution is, for $N \to \infty$, not random anymore, but concentrated on one deterministic distribution (namely the semicircle) is an example of the general high-dimensional phenomenon of "measure concentration".

Example 1.2. To illustrate this let us give an easy but illustrating example of such a concentration in high dimensions; namely that the volume of a ball is essentially sitting in the surface.

Denote by $B_r(0)$ the ball of radius r about 0 in $\mathbb{R}^n$ and for $0 < \varepsilon < 1$ consider the ε-neighborhood of the surface inside the ball, $B := \{x \in \mathbb{R}^n \mid 1 - \varepsilon \leq \|x\| \leq 1\}$. As we know the volume of balls

$$\mathrm{vol}(B_r(0)) = r^n \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2} - 1\right)!},$$

we can calculate the volume of B as

$$\mathrm{vol}(B) = \mathrm{vol}(B_1(0)) - \mathrm{vol}(B_{1-\varepsilon}(0)) = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2} - 1\right)!} \left(1 - (1 - \varepsilon)^n\right).$$

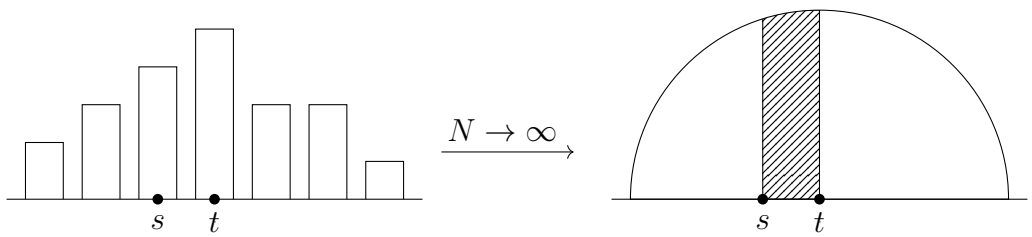
Thus,

$$\frac{\mathrm{vol}(B)}{\mathrm{vol}(B_1(0))} = 1 - (1 - \varepsilon)^n \overset{n \to \infty}{\longrightarrow} 1.$$

This says that in high dimensions the volume of a ball is concentrated in an arbitrarily small neighborhood of the surface. This is, of course, not true in small dimension - hence from our usual 3-dimensional perspective this appears quite counter-intuitive.

1.6 From histograms to moments

Let $A_N = A = (a_{ij})_{i,j=1}^N$ be our selfadjoint matrix with $a_{ij} = \pm 1$ randomly chosen. Then we typically see for the eigenvalues of A:



This convergence means

$$\frac{\#\{\text{eigenvalues in } [s,t]\}}{N} \xrightarrow{N\to\infty} \int_s^t \mathrm{d}\mu_W = \int_s^t p_W(x)\,\mathrm{d}x,$$

where μ_W is the semicircle distribution, with density p_W.

The left-hand side of this is difficult to calculate directly, but we note that the above statement is the same as

$$\frac{1}{N}\sum_{i=1}^N 1_{[s,t]}(\lambda_i) \xrightarrow{N\to\infty} \int_{\mathbb{R}} 1_{[s,t]}(x)\,\mathrm{d}\mu_W(x), \tag{$\star$}$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of A counted with multiplicity and $1_{[s,t]}$ is the characteristic function of $[s,t]$, i.e,

$$1_{[s,t]}(x) = \begin{cases} 1, & x \in [s,t], \\ 0, & x \notin [s,t]. \end{cases}$$

Hence in $(\star)$ we are claiming that

$$\frac{1}{N}\sum_{i=1}^N f(\lambda_i) \xrightarrow{N\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_W(x)$$

for all $f = 1_{[s,t]}$. It is easier to calculate this for other functions f, in particular, for f of the form $f(x) = x^n$, i.e.,

$$\frac{1}{N}\sum_{i=1}^N \lambda_i^n \xrightarrow{N\to\infty} \int_{\mathbb{R}} x^n\,\mathrm{d}\mu_W(x); \tag{$\star\star$}$$

the latter are the *moments* of μ_W. (Note that μ_W must necessarily be a probability measure.)

We will see later that in our case the validity of $(\star)$ for all $s < t$ is equivalent to the validity of $(\star\star)$ for all n. Hence we want to show $(\star\star)$ for all n.

Remark 1.3. The above raises of course the question: What is the advantage of $(\star\star)$ over $(\star)$, or of x^n over $1_{[s,t]}$?

Note that $A = A^*$ is selfadjoint and hence can be diagonalized, i.e., $A = UDU^*$, where U is unitary and D is diagonal with $d_{ii} = \lambda_i$ for all i (where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of A, counted with multiplicity). Moreover, we have

$$A^n = (UDU^*)^n = UD^nU^* \qquad \text{with} \qquad D^n = \begin{pmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_N^n \end{pmatrix},$$

hence

$$\sum_{i=1}^N \lambda_i^n = \mathrm{Tr}(D^n) = \mathrm{Tr}(UD^nU^*) = \mathrm{Tr}(A^n)$$

and thus

$$\frac{1}{N}\sum_{i=1}^N \lambda_i^n = \frac{1}{N}\,\mathrm{Tr}(A^n).$$

Notation 1.4. We denote by $\mathrm{tr} = \frac{1}{N}\,\mathrm{Tr}$ the *normalized trace* of matrices, i.e.,

$$\mathrm{tr}\left((a_{ij})_{i,j=1}^N\right) = \frac{1}{N}\sum_{i=1}^N a_{ii}.$$

So we are claiming that for our matrices we typically have that

$$\mathrm{tr}(A_N^n) \xrightarrow{N\to\infty} \int x^n \,\mathrm{d}\mu_W(x).$$

The advantage in this formulation is that the quantity $\mathrm{tr}(A_N^n)$ can be expressed in terms of the entries of the matrix, without actually having to calculate the eigenvalues.

1.7 Choice of scaling

Note that we need to choose the right scaling in N for the existence of the limit $N \to \infty$. For the case $a_{ij} \in \{\pm 1\}$ with $A_N = A_N^*$ we have

$$\mathrm{tr}(A_N^2) = \frac{1}{N}\sum_{i,j=1}^N a_{ij}a_{ji} = \frac{1}{N}\sum_{i,j=1} a_{ij}^2 = \frac{1}{N}N^2 = N.$$

Since this has to converge for $N \to \infty$ we should rescale our matrices

$$A_N \rightsquigarrow \frac{1}{\sqrt{N}} A_N,$$

i.e., we consider matrices $A_N = (a_{ij})_{i,j=1}^N$, where $a_{ij} = \pm\frac{1}{\sqrt{N}}$. For this scaling we claim that we typically have that

$$\operatorname{tr}(A_N^n) \xrightarrow{N\to\infty} \int x^n \, \mathrm{d}\mu_W(x)$$

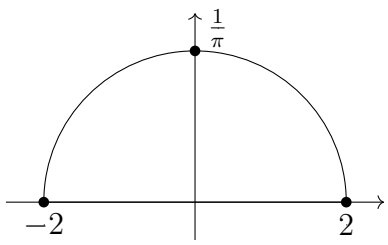
for a deterministic probability measure μ_W.

1.8 The semicircle and its moments

It's now probably time to give the precise definition of the semicircle distribution and, in the light of what we have to prove, also its moments.

Definition 1.5. (1) The (standard) *semicircular distribution* μ_W is the measure on $[-2, 2]$ with density

$$\mathrm{d}\mu_W(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\, \mathrm{d}x.$$



(2) The *Catalan numbers* $(C_k)_{k\geq 0}$ are given by

$$C_k = \frac{1}{k+1}\binom{2k}{k}.$$

They look like this: $1, 1, 2, 5, 14, 42, 132, \ldots$

Theorem 1.6. *(1) The Catalan numbers have the following properties.*

(i) The Catalan numbers satisfy the following recursion:

$$C_k = \sum_{l=0}^{k-1} C_l C_{k-l-1} \qquad (k \geq 1)$$

(ii) The Catalan numbers are uniquely determined by this recursion and by the initial value $C_0 = 1$.

(2) The semicircular distribution μ_W is a probability measure, i.e.,

$$\frac{1}{2\pi}\int_{-2}^{2}\sqrt{4-x^2}\,\mathrm{d}x = 1$$

and its moments are given by

$$\frac{1}{2\pi}\int_{-2}^{2}x^n\sqrt{4-x^2}\,\mathrm{d}x = \begin{cases} 0, & n \text{ odd}, \\ C_k, & n = 2k \text{ even}. \end{cases}$$

Exercises 2 and 3 address the proof of Theorem 1.6.

1.9 Types of convergence

So we are claiming that typically

$$\mathrm{tr}(A_N^2) \to 1, \qquad \mathrm{tr}(A_N^4) \to 2, \qquad \mathrm{tr}(A_N^6) \to 5, \quad \mathrm{tr}(A_N^8) \to 14, \qquad \mathrm{tr}(A_N^{10}) \to 42,$$

and so forth. But what do we mean by "typically"? The mathematical expression for this is "almost surely", but for now let us look on the more intuitive "convergence in probability" for $\mathrm{tr}(A_N^{2k}) \to C_k$.

Denote by Ω_N the set of our considered matrices, that is

$$\Omega_N = \left\{ A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N \,|\, A_N = A_N^* \text{ and } a_{ij} \in \{\pm 1\} \right\}.$$

Then convergence in probability means that for all $\varepsilon > 0$ we have

$$\frac{\#\left\{A_N \in \Omega_N \,\middle|\, \left|\mathrm{tr}\left(A_N^{2k}\right) - C_k\right| > \varepsilon\right\}}{\#\Omega_N} = P\left(A_N \,\middle|\, \left|\mathrm{tr}\left(A_N^{2k}\right) - C_k\right| > \varepsilon\right) \xrightarrow{N\to\infty} 0. \quad (\star)$$

How can we show $(\star)$? Usually, one proceeds as follows.

(1) First we show the weaker form of convergence in average, i.e.,

$$\frac{\sum_{A_N\in\Omega_N}\mathrm{tr}\left(A_N^{2k}\right)}{\#\Omega_N} = \mathbb{E}\left[\mathrm{tr}\left(A_N^{2k}\right)\right] \xrightarrow{N\to\infty} C_k.$$

(2) Then we show that with high probability the derivation from the average will become small as $N \to \infty$.

We will first consider step (1); (2) is a concentration phenomenon and will be treated later.

Note that step (1) is giving us the insight into why the semicircle shows up. Step (2) is more of a theoretical nature, adding nothing to our understanding of the semicircle, but making the (very interesting!) statement that in high dimensions the typical behaviour is close to the average behaviour.

2 Gaussian Random Matrices: Wick Formula and Combinatorial Proof of Wigner's Semicircle

We want to prove convergence of our random matrices to the semicircle by showing

$$\mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \xrightarrow{N\to\infty} C_k,$$

where the C_k are the Catalan numbers.

Up to now our matrices were of the form $A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N$ with $a_{ij} \in \{-1, 1\}$. From an analytic and also combinatorial point of view it is easier to deal with another choice for the a_{ij}, namely we will take them as Gaussian (aka normal) random variables; different a_{ij} will still be, up to symmetry, independent. If we want to calculate the expectations $\operatorname{tr}(A^{2k})$, then we should of course understand how to calculate moments of independent Gaussian random variables.

2.1 Gaussian random variables and Wick formula

Definition 2.1. A *standard Gaussian* (or *normal*) random variable X is a real-valued Gaussian random variable with mean 0 and variance 1, i.e., it has distribution

$$\mathbb{P}\left[t_1 \le X \le t_2\right] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}}\,\mathrm{d}t$$

and hence its moments are given by

$$\mathbb{E}\left[X^n\right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^n e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Proposition 2.2. *The moments of a standard Gaussian random variable are of the form*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^n e^{-\frac{t^2}{2}} \, \mathrm{d}t = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even}, \end{cases}$$

where the "double factorial" is defined, for m odd, as

$$m!! = m(m-2)(m-4)\cdots 5 \cdot 3 \cdot 1.$$

You are asked to prove this in Exercise 5.

Remark 2.3. From an analytic point of view it is surprising that those integrals evaluate to natural numbers. They actually count interesting combinatorial objects,

$$\mathbb{E}\left[X^{2k}\right] = \#\{\text{pairings of } 2k \text{ elements}\}.$$

Definition 2.4. (1) For a natural number $n \in \mathbb{N}$ we put $[n] = \{1, \ldots, n\}$.

(2) A *pairing* π of $[n]$ is a decomposition of $[n]$ into disjoints subsets of size 2, i.e., $\pi = \{V_1, \ldots, V_k\}$ such that for all $i, j = 1, \ldots, k$ with $i \neq j$, we have:
- $V_i \subset [n]$
- $\#V_i = 2$
- $V_i \cap V_j = \emptyset$
- $\bigcup_{i=1}^{k} V_i = [n]$

Note that necessarily $k = \frac{n}{2}$.

(3) The set of all pairings of $[n]$ is denoted by

$$\mathcal{P}_2(n) = \{\pi \mid \pi \text{ is a pairing of } [n]\}.$$

Proposition 2.5. *(1) We have*

$$\#\mathcal{P}_2(n) = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even}. \end{cases}$$

(2) Hence for a standard Gaussian variable X we have

$$\mathbb{E}\left[X^n\right] = \#\mathcal{P}_2(n).$$

Proof. (1) Count elements in $\mathcal{P}_2(n)$ in a recursive way. Choose the pair which contains the element 1, for this we have $n-1$ possibilities. Then we are left with choosing a pairing of the remaining $n-2$ numbers. Hence we have

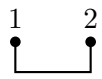
$$\#\mathcal{P}_2(n) = (n-1) \cdot \#\mathcal{P}_2(n-2).$$

Iterating this and noting that $\#\mathcal{P}_2(1) = 0$ and $\#\mathcal{P}_2(2) = 1$ gives the desired result.

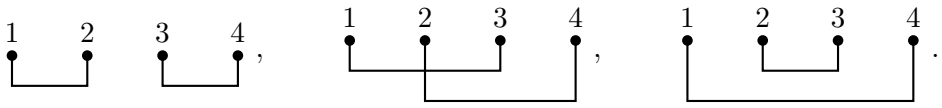
(2) Follows from (i) and Proposition 2.2.

□

Example 2.6. Usually we draw our partitions by connecting the elements in each pair. Then $\mathbb{E}[X^2] = 1$ corresponds to the single partition



and $\mathbb{E}[X^4] = 3$ corresponds to the three partitions



Remark 2.7 (Independent Gaussian random variables). We will have several, say two, Gaussian random variables X, Y and have to calculate their joint moments. The random variables are independent; this means that their joint distribution is the product measure of the single distributions,

$$\mathbb{P}[t_1 \leq X \leq t_2, s_1 \leq Y \leq s_2] = \mathbb{P}[t_1 \leq X \leq t_2] \cdot \mathbb{P}[s_1 \leq Y \leq s_2],$$

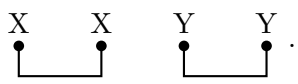
so in particular, for the moments we have

$$\mathbb{E}[X^n Y^m] = \mathbb{E}[X^n] \cdot \mathbb{E}[Y^m].$$

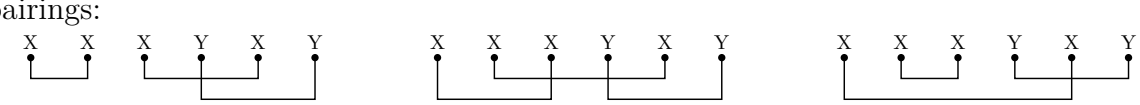
This gives then also a combinatorial description for their mixed moments:

$$\begin{aligned}
\mathbb{E}[X^n Y^m] &= \mathbb{E}[X^n] \cdot \mathbb{E}[Y^m] \\
&= \#\{\text{pairings of } \underbrace{X \cdots X}_{n}\} \cdot \#\{\text{pairings of } \underbrace{Y \cdots Y}_{m}\} \\
&= \#\{\text{pairings of } \underbrace{X \cdots X}_{n}\underbrace{Y \cdots Y}_{m} \text{ which connect } X \text{ with } X \text{ and } Y \text{ with } Y\}.
\end{aligned}$$

Example. We have $\mathbb{E}[XXYY] = 1$ since the only possible pairing is



On the other hand, $\mathbb{E}[XXXYXY] = 3$ since we have the following three possible pairings:



Consider now $x_1, \ldots, x_n \in \{X, Y\}$. Then we still have

$$\mathbb{E}[x_1 \ldots x_n] = \#\{\text{pairings which connect } X \text{ with } X \text{ and } Y \text{ with } Y\}.$$

Can we decide in a more abstract way whether $x_i = x_j$ or $x_i \neq x_j$? Yes, we can read this from the corresponding second moment, since

$$\mathbb{E}[x_i x_j] = \begin{cases} \mathbb{E}[x_i^2] = 1, & x_i = x_j, \\ \mathbb{E}[x_i]\,\mathbb{E}[x_j] = 0, & x_i \neq x_j. \end{cases}$$

Hence we have:

$$\mathbb{E}[x_1 \cdots x_n] = \sum_{\pi \in \mathcal{P}_2(n)} \prod_{(i,j) \in \pi} \mathbb{E}[x_i x_j]$$

Theorem 2.8 (Wick 1950, physics; Isserlis 1918, statistics)**.** *Let $Y_1, \ldots, Y_p$ be independent standard Gaussian random variables and consider $x_1, \ldots, x_n \in \{Y_1, \ldots, Y_p\}$. Then we have the* Wick formula

$$\mathbb{E}[x_1 \cdots x_n] = \sum_{\pi \in \mathcal{P}_2(n)} \mathbb{E}_\pi[x_1, \ldots, x_n],$$

where, for $\pi \in \mathcal{P}_2(n)$, we use the notation

$$\mathbb{E}_\pi[x_1, \ldots, x_n] = \prod_{(i,j) \in \pi} \mathbb{E}[x_i x_j].$$

Note that the Wick formula is linear in the x_i, hence it remains valid if we replace the x_i by linear combinations of the x_j. In particular, we can go over to complex Gaussian variables.

Definition 2.9. A *standard complex Gaussian* random variable Z is of the form

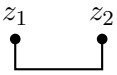
$$Z = \frac{X + iY}{\sqrt{2}},$$

where X and Y are independent standard real Gaussian variables.

Remark 2.10. Let Z be a standard complex Gaussian, i.e., $Z = \frac{X+iY}{\sqrt{2}}$. Then we have $\bar{Z} = \frac{X-iY}{\sqrt{2}}$ and the first and second moments are given by

- $\mathbb{E}[Z] = 0 = \mathbb{E}\left[\bar{Z}\right]$
- $\mathbb{E}[Z^2] = \mathbb{E}[ZZ] = \frac{1}{2}\left[\mathbb{E}[XX] - \mathbb{E}[YY] + i\left(\mathbb{E}[XY] + \mathbb{E}[YX]\right)\right] = 0$

- $\mathbb{E}\left[\bar{Z}^2\right] = 0$
- $\mathbb{E}\left[|Z|^2\right] = \mathbb{E}\left[Z\bar{Z}\right] = \frac{1}{2}\left[\mathbb{E}\left[XX\right] + \mathbb{E}\left[YY\right] + i\left(\mathbb{E}\left[YX\right] - \mathbb{E}\left[YX\right]\right)\right] = 1$

Hence, for $z_1, z_2 \in \{Z, \bar{Z}\}$ and $\pi =$ 

we have

$$\mathbb{E}\left[z_1 z_2\right] = \begin{cases} 1, & \pi \text{ connects } Z \text{ with } \bar{Z}, \\ 0, & \pi \text{ connects } (Z \text{ with } Z) \text{ or } (\bar{Z} \text{ with } \overline{Z}). \end{cases}$$

Theorem 2.11. *Let $Z_1, \ldots, Z_p$ be independent standard complex Gaussian random variables and consider $z_1, \ldots, z_n \in \{Z_1, \bar{Z}_1, \ldots, Z_p, \bar{Z}_p\}$. Then we have the Wick formula*

$$\begin{aligned} \mathbb{E}\left[z_1 \cdots z_n\right] &= \sum_{\pi \in \mathcal{P}_2(n)} \mathbb{E}_\pi\left[z_1, \ldots, z_n\right] \\ &= \#\{\textit{pairings of } [n] \textit{ which connect } Z_i \textit{ with } \bar{Z}_i\}. \end{aligned}$$

2.2 Gaussian random matrices and genus expansion

Now we are ready to consider random matrices with such complex Gaussians as entries.

Definition 2.12. A *Gaussian random matrix* is of the form $A_N = \frac{1}{\sqrt{N}}\left(a_{ij}\right)_{i,j=1}^N$, where

- $A_N = A_N^*$, i.e., $a_{ij} = \bar{a}_{ji}$ for all i, j,
- $\{a_{ij} \mid i \geq j\}$ are independent,
- each a_{ij} is a standard Gaussian random variable, which is complex for $i \neq j$ and real for $i = j$.

Remark 2.13. (1) More precisely, we should address the above as *selfadjoint* Gaussian random matrices.

(2) Another common name for those random matrices is GUE, which stands for *Gaussian unitary ensemble*. "Unitary" corresponds here to the fact that the entries are complex, since such matrices are invariant under unitary transformations. With GUE(N) we denote the GUE of size $N \times N$. There are also real and quaternionic versions, *Gaussian orthogonal ensembles* GOE, and *Gaussian symplectic ensembles* GSE.

(3) Note that we can also express this definition in terms of the Wick formula 2.11 as

$$\mathbb{E}\left[a_{i(1)j(1)}\cdots a_{i(n)j(n)}\right] = \sum_{\pi\in\mathcal{P}_2(n)} \mathbb{E}_\pi\left[a_{i(1)j(1)},\ldots,a_{i(n)j(n)}\right],$$

for all n and $1 \le i(1), j(1), \ldots, i(n), j(n) \le N$, and where the second moments are given by

$$\mathbb{E}\left[a_{ij}a_{kl}\right] = \delta_{il}\delta_{jk}.$$

So we have for example for the fourth moment

$$\begin{aligned}\mathbb{E}\left[a_{i(1)j(1)}a_{i(2)j(2)}a_{i(3)j(3)}a_{i(4)j(4)}\right] =& \delta_{i(1)j(2)}\delta_{j(1)i(2)}\delta_{i(3)j(4)}\delta_{j(3)i(4)}\\ &+\delta_{i(1)j(3)}\delta_{j(1)i(3)}\delta_{i(2)j(4)}\delta_{j(2)i(4)}\\ &+\delta_{i(1)j(4)}\delta_{j(1)i(4)}\delta_{i(2)j(3)}\delta_{j(2)i(3)},\end{aligned}$$

and more concretely, $\mathbb{E}\left[a_{12}a_{21}a_{11}a_{11}\right] = 1$ and $\mathbb{E}\left[a_{12}a_{12}a_{21}a_{21}\right] = 2$.

Remark 2.14 (Calculation of $\mathbb{E}\left[\mathrm{tr}(A_N^m)\right]$). For our Gaussian random matrix we want to calculate their moments

$$\mathbb{E}\left[\mathrm{tr}(A_N^m)\right] = \frac{1}{N}\frac{1}{N^{m/2}}\sum_{i(1),\ldots,i(m)=1}^{N} \mathbb{E}\left[a_{i(1)i(2)}a_{i(2)i(3)}\cdots a_{i(m)i(1)}\right].$$

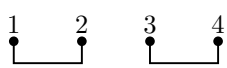
Let us first consider small examples before we treat the general case:

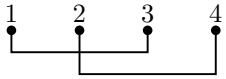
(1)

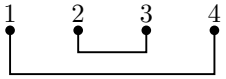
$$\mathbb{E}\left[\mathrm{tr}(A_N^2)\right] = \frac{1}{N^2}\sum_{i,j=1}^{N}\underbrace{\mathbb{E}\left[a_{ij}a_{ji}\right]}_{=1} = \frac{1}{N^2}N^2 = 1,$$

and hence: $\mathbb{E}\left[\mathrm{tr}(A_N^2)\right] = 1 = C_1$ for all N.

(2) We consider the partitions

$\pi_1 =$ 

, $\pi_2 =$ 

, $\pi_3 =$ 

.

With this, we have

$$\mathbb{E}\left[\mathrm{tr}(A_N^4)\right] = \frac{1}{N^3}\sum_{i,j,k,l=1}^{N}\underbrace{\mathbb{E}\left[a_{ij}a_{jk}a_{kl}a_{li}\right]}_{=\mathbb{E}_{\pi_1}[\ldots]+\mathbb{E}_{\pi_2}[\ldots]+\mathbb{E}_{\pi_3}[\ldots]}$$

and calculate

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_1}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1\\ i=k}}^{N} 1 = N^3,$$

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_2}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1\\ j=l}}^{N} 1 = N^3,$$

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_3}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1\\ i=l, j=k, j=i, k=l}}^{N} 1 = \sum_{i=1}^{N} 1 = N,$$

hence

$$\mathbb{E}\left[\operatorname{tr}(A_N^4)\right] = \frac{1}{N^3}\left(N^3 + N^3 + N\right) = 2 + \frac{1}{N^2}.$$

So we have

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}(A_N^4)\right] = 2 = C_2.$$

(3) Let us now do the general case.

$$\begin{aligned}\mathbb{E}\left[a_{i(1)i(2)}a_{i(2)i(3)}\cdots a_{i(m)i(1)}\right] &= \sum_{\pi\in\mathcal{P}_2(m)} \mathbb{E}_\pi\left[a_{i(1)i(2)}, a_{i(2)i(3)}, \ldots, a_{i(m)i(1)}\right]\\ &= \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(k,l)\in\pi} \mathbb{E}\left[a_{i(k)i(k+1)}a_{i(l)i(l+1)}\right].\end{aligned}$$

We use the notation $[i = j] = \delta_{ij}$ and, by identifying a pairing π with a permutation $\pi \in S_m$ via

$$(k,l) \in \pi \leftrightarrow \pi(k) = l, \pi(l) = k,$$

find that

$$\begin{aligned}\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] &= \frac{1}{N^{m/2+1}} \sum_{i(1),\ldots,i(m)=1}^{N} \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(k,l)\in\pi} \mathbb{E}\left[a_{i(k)i(k+1)}a_{i(l)i(l+1)}\right]\\ &= \frac{1}{N^{m/2+1}} \sum_{\pi\in\mathcal{P}_2(m)} \sum_{i(1),\ldots,i(m)=1}^{N} \prod_k \Big[i(k) = i(\underbrace{\pi(k)+1}_{\gamma\pi(k)})\Big],\end{aligned}$$

where $\gamma = (1, 2, \ldots, m) \in S_m$ is the shift by 1 modulo m. The above product is different from 0 if and only if $i\colon [m] \to [N]$ is constant on the cycles of

$\gamma\pi \in S_m$. Thus we get finally

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \frac{1}{N^{m/2+1}} \sum_{\pi \in \mathcal{P}_2(m)} N^{\#(\gamma\pi)},$$

where $\#(\gamma\pi)$ is the number of cycles of the permutation $\gamma\pi$.

Hence we have proved the following.

Theorem 2.15. *Let A_N be a* GUE(N) *random matrix. Then we have for all $m \in \mathbb{N}$,*

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \sum_{\pi \in \mathcal{P}_2(m)} N^{\#(\gamma\pi) - \frac{m}{2} - 1}.$$

Example 2.16. (1) This says in particular that all odd moments are zero, since $\mathcal{P}_2(2k+1) = \emptyset$.

(2) Let $m = 2$, then $\gamma = (1,2)$ and we have only one $\pi = (1,2)$; then $\gamma\pi = \mathrm{id} = (1)(2)$, and thus $\#(\gamma\pi) = 2$ and

$$\#(\gamma\pi) - \frac{m}{2} - 1 = 0.$$

Thus,

$$\mathbb{E}\left[\operatorname{tr}\left(A_N^2\right)\right] = N^0 = 1.$$

(3) Let $m = 4$ and $\gamma = (1,2,3,4)$. Then there are three $\pi \in \mathcal{P}_2(4)$ with the following contributions:

π	$\gamma\pi$	$\#(\gamma\pi) - 3$	contribution
$(1,2)(34)$	$(1,3)(2)(4)$	0	$N^0 = 1$
$(13)(24)$	$(1,4,3,2)$	-2	$N^{-2} = \frac{1}{N^2}$
$(14)(23)$	$(1)(2,4)(3)$	0	$N^0 = 1$

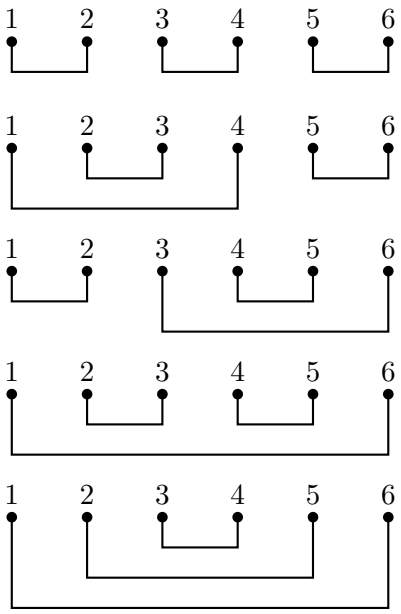
so that

$$\mathbb{E}\left[\operatorname{tr}\left(A_N^4\right)\right] = 2 + \frac{1}{N^2}.$$

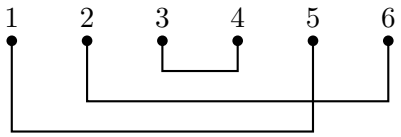
(4) In the same way one can calculate that

$$\mathbb{E}\left[\operatorname{tr}\left(A_N^6\right)\right] = 5 + 10\frac{1}{N^2},$$
$$\mathbb{E}\left[\operatorname{tr}\left(A_N^8\right)\right] = 14 + 70\frac{1}{N^2} + 21\frac{1}{N^4}.$$

(5) For $m = 6$ the following 5 pairings give contribution N^0:

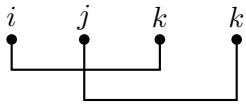


Those are non-crossing pairings, all other pairings $\pi \in \mathcal{P}_2(6)$ have a crossing, e.g.:



2.3 Non-crossing pairings

Definition 2.17. A pairing $\pi \in \mathcal{P}_2(m)$ is *non-crossing* (*NC*) if there are no pairs (i, k) and (j, l) in π with $i < j < k < l$, i.e., we don't have a crossing in π.

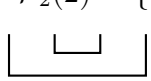
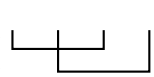


is not allowed!

We put

$$\mathcal{NC}_2(m) = \{\pi \in \mathcal{P}_2(m) \,|\, \pi \text{ is non-crossing}\}\,.$$

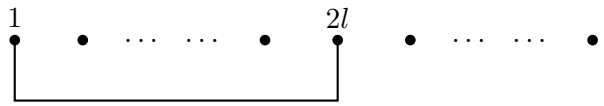
Example 2.18. (1) $\mathcal{NC}_2(2) = \mathcal{P}_2(2) = \{$ $\}$

(2) $\mathcal{NC}_2(4) = \{$,  $\}$ and $\mathcal{P}_2(4) \backslash \mathcal{NC}_2(4) = \{$  $\}$

(3) The 5 elements of $\mathcal{NC}_2(6)$ are given in Example 2.16 (v), $\mathcal{P}_2(6)$ contains 15 elements; thus there are $15 - 5 = 10$ more elements in $\mathcal{P}_2(6)$ with crossings.

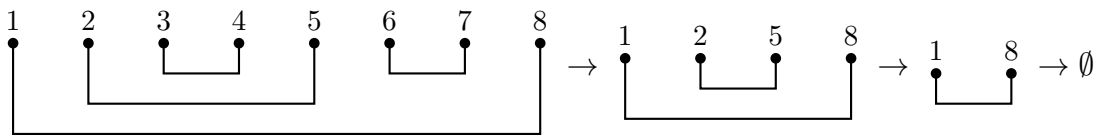
Remark 2.19. Note that NC-pairings have a recursive structure, which usually is crucial for dealing with them.

(1) The first pair of $\pi \in \mathcal{NC}_2(2k)$ must necessarily be of the form $(1, 2l)$ and the remaining pairs can only pair within $\{2, \dots, 2l-1\}$ or within $\{2l+1, \dots, 2l\}$.

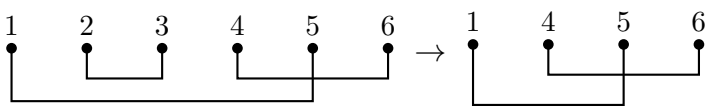


(2) Iterating this shows that we must find in any $\pi \in \mathcal{NC}_2(2k)$ at least one pair of the form $(i, i+1)$ with $1 \leq i \leq 2k-1$. Removing this pair gives a NC-pairing of $2k-2$ points. This characterizes the NC-pairings as those pairings, which can be reduced to the empty set by iterated removal of pairs, which consist of neighbors.

An example for the reduction of a non-crossing pairing is the following.



In the case of a crossing pairing, some reductions might be possible, but eventually one arrives at a point, where no further reduction can be done.



no further reduction possible!

Proposition 2.20. *Consider m even and let $\pi \in \mathcal{P}_2(m)$, which we identify with a permutation $\pi \in S_m$. As before, $\gamma = (1, 2, \dots, m) \in S_m$. Then we have:*

(1) $\#(\gamma\pi) - \frac{m}{2} - 1 \leq 0$ for all $\pi \in \mathcal{P}_2(m)$.

(2) $\#(\gamma\pi) - \frac{m}{2} - 1 = 0$ if and only if $\pi \in \mathcal{NC}_2(m)$.

Proof. First we note that a pair $(i, i+1)$ in π corresponds to a fixed point of $\gamma\pi$. More precisely, in such a situation we have $i+1 \xrightarrow{\pi} i \xrightarrow{\gamma} i+1$ and $i \xrightarrow{\pi} i+1 \xrightarrow{\gamma} i+2$. Hence $\gamma\pi$ contains the cycles $(i+1)$ and $(\dots, i, i+2, \dots)$.

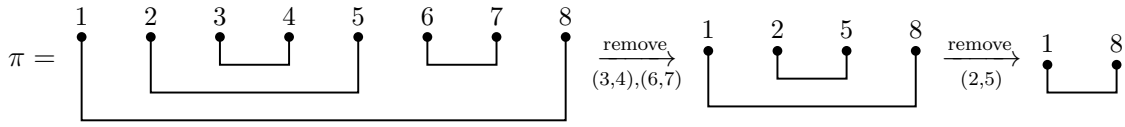
This implication also goes in the other direction: If $\gamma\pi(i+1) = i+1$ then $\pi(i+1) = \gamma^{-1}(i+1) = i$. Since π is a pairing we must then also have $\pi(i) = i+1$ and hence we have the pair $(i, i+1)$ in π.

If we have $(i, i+1)$ in π, we can remove the points i and $i+1$, yielding another pairing $\tilde{\pi}$. By doing so, we remove in $\gamma\pi$ the cycle $(i+1)$ and we remove in the cycle $(\dots, i, i+2, \dots)$ the point i, yielding $\gamma\tilde{\pi}$. We reduce thus m by 2 and $\#(\gamma\pi)$ by 1.

If π is NC we can iterate this until we arrive at $\tilde{\pi}$ with $m = 2$. Then we have $\tilde{\pi} = (1, 2)$ and $\gamma = (1, 2)$ such that $\gamma\tilde{\pi} = (1)(2)$ and $\#(\gamma\tilde{\pi}) = 2$. If $m = 2k$ we did $k - 1$ reductions where we reduced in each step the number of cycles by 1 and at the end we remain with 2 cycles, hence

$$\#(\gamma\pi) = (k-1) \cdot 1 + 2 = k + 1 = \frac{m}{2} + 1.$$

Here is an example for this:



$$\gamma\pi = (1)(268)(35)(4)(7) \quad \longrightarrow \quad (1)(28)(5) \quad \longrightarrow \quad (1)(8)$$

For a general $\pi \in \mathcal{P}_2(m)$ we remove cycles $(i, i+1)$ as long as possible. If π is crossing we arrive at a pairing $\tilde{\pi}$, where this is not possible anymore. It suffices to show that such a $\tilde{\pi} \in \mathcal{P}_2(m)$ satisfies $\#(\gamma\tilde{\pi}) - \frac{m}{2} - 1 < 0$. But since $\tilde{\pi}$ has no cycle $(i, i+1)$, $\gamma\tilde{\pi}$ has no fixed point. Hence each cycle has at least 2 elements, thus

$$\#(\gamma\tilde{\pi}) \leq \frac{m}{2} < \frac{m}{2} + 1.$$

Note that in the above arguments, with $(i, i+1)$ we actually mean $(i, \gamma(i))$; thus also $(1, m)$ counts as a pair of neighbors for a $\pi \in \mathcal{P}_2(m)$, in order to have the characterization of fixed points right. Hence, when reducing a general pairing to one without fixed points we have also to remove such cyclic neighbors as long as possible. $\square$

2.4 Semicircle law for GUE

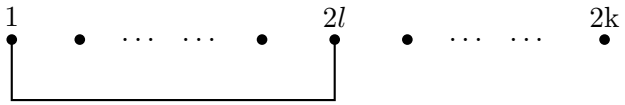
Theorem 2.21 (Wigner's semicircle law for GUE, averaged version)**.** *Let A_N be a Gaussian (GUE) $N \times N$ random matrix. Then we have for all $m \in \mathbb{N}$:*

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \frac{1}{2\pi} \int_{-2}^{2} x^m \sqrt{4 - x^2}\, \mathrm{d}x.$$

Proof. This is true for m odd, since then both sides are equal to zero. Consider $m = 2k$ even. Then Theorem 2.15 and Proposition 2.20 show that

$$\lim_{\mathbb{N}\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \sum_{\pi\in\mathcal{P}_2(m)} \lim_{N\to\infty} N^{\#(\gamma\pi)-\frac{m}{2}-1} = \sum_{\pi\in\mathcal{NC}_2(m)} 1 = \#\mathcal{NC}_2(m).$$

Since the moments of the semicircle are given by the Catalan numbers, it remains to see that $\#\mathcal{NC}_2(2k)$ is equal to the Catalan number C_k. To see this, we now count $d_k := \#\mathcal{NC}_2(2k)$ according to the recursive structure of NC-pairings as in 2.19 (i).



Namely, we can identify $\pi \in \mathcal{NC}_2(2k)$ with $\{(1, 2l)\} \cup \pi_0 \cup \pi_1$, where $l \in \{1, \dots, k\}$, $\pi_0 \in \mathcal{NC}_2(2(l-1))$ and $\pi_1 \in \mathcal{NC}_2(2(k-l))$. Hence we have

$$d_k = \sum_{l=1}^{k} d_{l-1} d_{k-l}, \qquad \text{where } d_0 = 1.$$

This is the recursion for the Catalan numbers, whence $d_k = C_k$ for all $k \in \mathbb{N}$. □

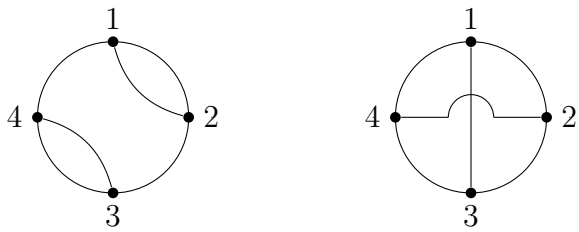
Remark 2.22. (1) One can refine

$$\#(\gamma\pi) - \frac{m}{2} - 1 \le 0 \qquad \text{to} \qquad \#(\gamma\pi) - \frac{m}{2} - 1 = -2g(\pi)$$

for $g(\pi) \in \mathbb{N}_0$. This g has the meaning that it is the minimal genus of a surface on which π can be drawn without crossings. NC pairings are also called *planar*, they correspond to $g = 0$. Theorem 2.15 is usually addressed as *genus expansion*,

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \sum_{\pi\in\mathcal{P}_2(m)} N^{-2g(\pi)}.$$

(2) For example, $(1, 2)(3, 4) \in \mathcal{NC}_2(4)$ has $g = 0$, but the crossing pairing $(1, 3)(2, 4) \in \mathcal{P}_2(4)$ has genus $g = 1$. It has a crossing in the plane but this can be avoided on a torus.



(3) If we denote

$$\varepsilon_g(k) = \#\left\{\pi \in \mathcal{P}_2(2k) \,|\, \pi \text{ has genus } g\right\}$$

then the genus expansion 2.15 can be written as

$$\mathbb{E}\left[\mathrm{tr}(A_N^{2k})\right] = \sum_{g\geq 0} \varepsilon_g(k) N^{-2g}.$$

We know that

$$\varepsilon_g(0) = C_k = \frac{1}{k+1}\binom{2k}{k},$$

but what about the $\varepsilon_g(k)$ for $g > 0$? There does not exist an explicit formula for them, but Harer and Zagier have shown in 1986 that

$$\varepsilon_g(k) = \frac{(2k)!}{(k+1)!(k-2g)!} \cdot \lambda_g(k),$$

where $\lambda_g(k)$ is the coefficient of x^{2g} in

$$\left(\frac{\frac{x}{2}}{\tanh\frac{x}{2}}\right)^{k+1}.$$

We will come back later to this statement of Harer and Zagier; see Theorem 9.2

3 Wigner Matrices: Combinatorial Proof of Wigner's Semicircle Law

Wigner's semicircle law does not only hold for Gaussian random matrices, but more general for so-called Wigner matrices; there we keep the independence and identical distribution of the entries, but allow arbitrary distribution instead of Gaussian. As there is no Wick formula any more, there is no clear advantage of the complex over the real case any more, hence we will consider in the following the real one.

3.1 Wigner matrices

Definition 3.1. Let μ be a probability distribution on $\mathbb{R}$. A corresponding *Wigner random matrix* is of the form $A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N$, where

- $A_N = A_N^*$, i.e., $a_{ij} = a_{ji}$ for all i, j,
- $\{a_{ij} \mid i \geq j\}$ are independent,
- each a_{ij} has distribution μ.

Remark 3.2. (1) In our combinatorial setting we will assume that all moments of μ exist; that the first moment is 0; and the second moment will be normalized to 1. In an analytic setting one can deal with more general situations: usually only the existence of the second moment is needed; and one can also allow non-vanishing mean.

(2) Often one also allows different distributions for the diagonal and the off-diagonal entries.

(3) Even more general, one can give up the assumption of identical distribution of all entries and replace this by uniform bounds on their moments.

(4) We will now try to imitate our combinatorial proof from the Gaussian case also in this more general situation. Without a precise Wick formula for the higher moments of the entries, we will not aim at a precise genus expansion;

it suffices to see that the leading contributions are still given by the Catalan numbers.

3.2 Combinatorial description of moments of Wigner matrices

Consider a Wigner matrix $A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N$, where μ has all moments and

$$\int_{\mathbb{R}} x \, \mathrm{d}\mu(x) = 0, \qquad \int_{\mathbb{R}} x^2 \, \mathrm{d}\mu(x) = 1.$$

Then

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{i_1,\dots,i_m=1}^{N} \mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in \mathcal{P}(m)} \sum_{\substack{i\colon [m] \to [N] \\ \ker i = \sigma}} \mathbb{E}\left[\sigma\right],$$

where we group the appearing indices $(i_1, \dots, i_m)$ according to their "kernel", which is a "partition σ of $\{1, \dots, m\}$.

Definition 3.3. (1) A *partition* σ of $[n]$ is a decomposition of $[n]$ into disjoint, non-empty subsets (of arbitrary size), i.e., $\sigma = \{V_1, \dots, V_k\}$, where

- $V_i \subset [n]$ for all i,
- $V_i \neq \emptyset$ for all i,
- $V_i \cap V_j = \emptyset$ for all $i \neq j$,
- $\bigcup_{i=1}^k V_i = [n]$.

The V_i are called *blocks* of σ. The set of all partitions of $[n]$ is denoted by

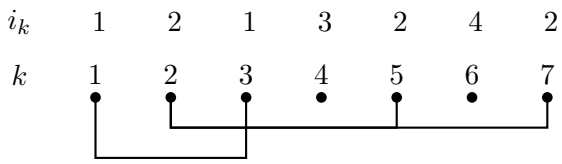
$$\mathcal{P}(n) := \{\sigma \mid \sigma \text{ is a partition of } [n]\}.$$

(2) For a multi-index $i = (i_1, \dots, i_m)$ we denote by $\ker i$ its *kernel*; this is the partition $\sigma \in \mathcal{P}(m)$ such that we have $i_k = i_l$ if and only if k and l are in the same block of σ. If we identify i with a function $i\colon [m] \to [N]$ via $i(k) = i_k$ then we can also write

$$\ker i = \left\{i^{-1}(1), i^{-1}(2), \dots, i^{-1}(N)\right\},$$

where we discard all empty sets.

Example 3.4. For $i = (1, 2, 1, 3, 2, 4, 2)$ we have



such that

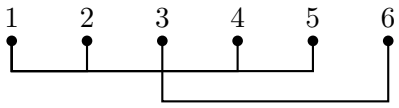
$$\ker i = \{(1, 3), (2, 5, 7), (4), (6)\} \in \mathcal{P}(7).$$

Remark 3.5. The relevance of this kernel in our setting is the following:

For $i = (i_1, \ldots, i_m)$ and $j = (j_1, \ldots, j_m)$ with $\ker i = \ker j$ we have

$$\mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right] = \mathbb{E}\left[a_{j_1 j_2} a_{j_2 j_3} \cdots a_{j_m j_1}\right].$$

For example, for $i = (1, 1, 2, 1, 1, 2)$ and $j = (2, 2, 7, 2, 2, 7)$ we have

$\ker i =$ 

$= \ker j$

and

$$\mathbb{E}\left[a_{11} a_{12} a_{21} a_{11} a_{12} a_{21}\right] = \mathbb{E}\left[a_{11}^2\right] \mathbb{E}\left[a_{12}^4\right] = \mathbb{E}\left[a_{22}^2\right] \mathbb{E}\left[a_{27}^4\right] = \mathbb{E}\left[a_{22} a_{27} a_{72} a_{22} a_{27} a_{72}\right].$$

We denote this common value by

$$\mathbb{E}[\sigma] := \mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right] \qquad \text{if } \ker i = \sigma.$$

Thus we get:

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in \mathcal{P}(m)} \mathbb{E}[\sigma] \cdot \#\left\{i \colon [m] \to [N] \mid \ker i = \sigma\right\}.$$

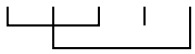
To understand the contribution corresponding to a $\sigma \in \mathcal{P}(m)$ we associate to σ a graph $\mathcal{G}_\sigma$.

Definition 3.6. For $\sigma = \{V_1, \ldots, V_k\} \in \mathcal{P}(m)$ we define a corresponding graph $\mathcal{G}_\sigma$ as follows. The vertices of $\mathcal{G}_\sigma$ are given by the blocks V_p of σ and there is an edge betwees V_p and V_q if there is an $r \in [m]$ such that $r \in V_p$ and $r + 1 (\text{mod } m) \in V_q$.

Another way of saying this is that we start with a graph with vertices $1, 2, \ldots, m$ and edges $(1, 2), (2, 3), (3, 4), \ldots, (m-1, m), (m, 1)$ and then identify vertices according to the blocks of σ. We keep loops, but erase multiple edges.

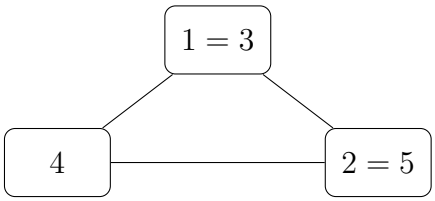
Example 3.7. (1) For

$$\sigma = \{(1,3),(2,5),(4)\} =$$



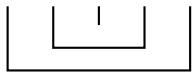
we have

$\mathcal{G}_\sigma =$



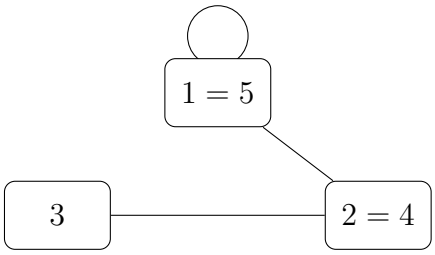
(2) For

$$\sigma = \{(1,5),(2,4),(3)\} =$$



we have

$\mathcal{G}_\sigma =$

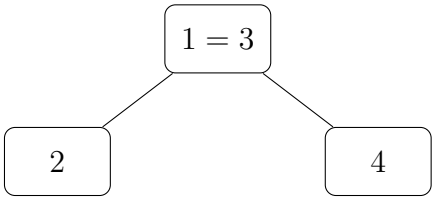


(3) For

$$\sigma = \{(1,3),(2),(4)\} =$$

we have

$\mathcal{G}_\sigma =$



The term $\mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right]$ corresponds now to a walk in $\mathcal{G}_\sigma$, with $\sigma = \ker i$, along the edges with steps

$$i_1 \to i_2 \to i_3 \to \cdots \to i_m \to i_1.$$

Hence we are using each edge in $\mathcal{G}_\sigma$ at least once. Note that different edges in $\mathcal{G}_\sigma$ correspond to independent random variables. Hence, if we use an edge only once in

our walk, then $\mathbb{E}[\sigma] = 0$, because the expectation factorizes into a product with one factor being the first moment of a_{ij}, which is assumed to be zero. Thus, every edge must be used at least twice, but this implies

$$\#\text{ edges in } \mathcal{G}_\sigma \leq \frac{\#\text{ steps in the walk}}{2} = \frac{m}{2}.$$

Since the number of i with the same kernel is

$$\#\{i\colon [m] \to [N] \mid \ker i = \sigma\} = N(N-1)(N-2)\cdots(N-\#\sigma+1),$$

where $\#\sigma$ is the number of blocks in σ, we finally get

$$\mathbb{E}[\operatorname{tr}(A_N^m)] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\substack{\sigma\in\mathcal{P}(m)\\ \#\text{edges}(\mathcal{G}_\sigma)\leq\frac{m}{2}}} \mathbb{E}[\sigma] \underbrace{N(N-1)(N-2)\cdots(N-\#\sigma+1)}_{\sim N^{\#\sigma}\text{ for } N\to\infty}. \qquad (\star)$$

We have to understand what the constraint on the number of edges in $\mathcal{G}_\sigma$ gives us for the number of vertices in $\mathcal{G}_\sigma$ (which is the same as $\#\sigma$). For this, we will now use the following well-known basic result from graph theory.

Proposition 3.8. *Let $\mathcal{G} = (V, E)$ be a connected finite graph with vertices V and edges E. (We allow loops and multi-edges.) Then we have that*

$$\#V \leq \#E + 1$$

and we have equality if and only if $\mathcal{G}$ is a tree*, i.e., a connected graph without cycles.*

3.3 Semicircle law for Wigner matrices

Theorem 3.9 (Wigner's semicircle law for Wigner matrices, averaged version)**.** *Let A_N be a Wigner matrix corresponding to μ, which has all moments, with mean* 0 *and second moment* 1*. Then we have for all $m \in \mathbb{N}$:*

$$\lim_{N\to\infty} \mathbb{E}[\operatorname{tr}(A_N^m)] = \frac{1}{2\pi}\int_{-2}^{2} x^m\sqrt{4-x^2}\,\mathrm{d}x.$$

Proof. From $(\star)$ we get

$$\lim_{N\to\infty} \mathbb{E}[\operatorname{tr}(A_N^m)] = \sum_{\sigma\in\mathcal{P}(m)} \mathbb{E}[\sigma] \lim_{N\to\infty} N^{\#V(\mathcal{G}_\sigma)-\frac{m}{2}-1}.$$

In order to have $\mathbb{E}[\sigma] \neq 0$, we can restrict to σ with $\#E(\mathcal{G}_\sigma) \leq \frac{m}{2}$, which by Proposition 3.8 implies that

$$\#V(\mathcal{G}_\sigma) \leq \#E(\mathcal{G}_\sigma) + 1 \leq \frac{m}{2} + 1.$$

Hence all terms converge and the only contribution in the limit $N \to \infty$ comes from those σ, where we have equality, i.e.,

$$\#V(\mathcal{G}_\sigma) = \#E(\mathcal{G}_\sigma) + 1 = \frac{m}{2} + 1.$$

Thus, $\mathcal{G}_\sigma$ must be a tree and in our walk we use each edge exactly twice (necessarily in opposite directions). For such a σ we have $\mathbb{E}[\sigma] = 1$; thus

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \#\left\{\sigma \in \mathcal{P}(m) \,|\, \mathcal{G}_\sigma \text{ is a tree}\right\}.$$

We will check in Exercise 9 that the latter number is also counted by the Catalan numbers. □

Remark 3.10. Note that our $\mathcal{G}_\sigma$ are not just abstract trees, but they are coming with the walks, which encode

- a starting point, i.e., the $\mathcal{G}_\sigma$ are rooted trees
- a cyclic order of the outgoing edges at a vertex, which gives a planar drawing of our graph

Hence what we have to count are rooted planar trees.

Note also that a rooted planar tree determines uniquely the corresponding walk.

4 Analytic Tools: Stieltjes Transform and Convergence of Measures

Let us recall our setting and goal. We have, for each $N \in \mathbb{N}$, selfadjoint $N \times N$ random matrices, which are given by a probability measure $\mathbb{P}_N$ on the entries of the matrices; the prescription of $\mathbb{P}_N$ should be kind of uniform in N.

For example, for the GUE(N) we have $A = (a_{ij})_{i,j=1}^N$ with the complex variables $a_{ij} = x_{ij} + \sqrt{-1}y_{ij}$ having real part x_{ij} and imaginary part y_{ij}. Since $a_{ij} = \bar{a}_{ji}$, we have $y_{ii} = 0$ for all i and we remain with the N^2 many "free variables" x_{ii} $(i = 1, \ldots, N)$ and x_{ij}, y_{ij} $(1 \leq i < j \leq N)$. All those are independent and Gaussian distributed, which can be written in the compact form

$$d\mathbb{P}(A) = c_N \exp(-N\frac{\mathrm{Tr}(A^2)}{2})dA,$$

where dA is the product of all differentials of the N^2 variables and c_N is a normalization constant, to make $\mathbb{P}_N$ a probability measure.

We want now statements about our matrices with respect to this measure $\mathbb{P}_N$, either in average or in probability. Let us be a bit more specific on this.

Denote by Ω_N the space of our selfadjoint $N \times N$ matrices, i.e.,

$$\Omega_N := \{A = (x_{ij} + \sqrt{-1}y_{ij})_{i,j=1}^N \mid x_{ii} \in \mathbb{R}(i = 1, \ldots, N), x_{ij}, y_{ij} \in \mathbb{R}(i < j)\} \hat{=} \mathbb{R}^{N^2},$$

then, for each $N \in \mathbb{N}$, $\mathbb{P}_N$ is a probability measure on Ω_N.

For $A \in \Omega_N$ we consider its N eigenvalues $\lambda_1, \ldots, \lambda_N$, counted with multiplicity. We encode those eigenvalues in a probability measure μ_A on $\mathbb{R}$:

$$\mu_A := \frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}),$$

which we call the *eigenvalue distribution of* A. Our claim is now that μ_A converges under $\mathbb{P}_N$, for $N \to \infty$ to the semicircle distribution μ_W,

- in average, i.e.,
$$\mu_N := \int_{\Omega_N} \mu_A d\mathbb{P}_N(A) = \mathbb{E}\left[\mu_a\right] \overset{N\to\infty}{\longrightarrow} \mu_W$$
- and, stronger, in probability or almost surely.

So what we have to understand now is:

- What kind of convergence $\mu_N \to \mu$ do we have here, for probability measures on $\mathbb{R}$?
- How can we describe probability measures (on $\mathbb{R}$) and their convergence with analytic tools?

The relevant notions of convergence are the "vague" and "weak" convergence and our analytic tool will be the Stieltjes transform. We start with describing the latter.

4.1 Stieltjes transform

Definition 4.1. Let μ be a Borel measure on $\mathbb{R}$.

(1) μ is *finite* if $\mu(\mathbb{R}) < \infty$.
(2) μ is a *probability measure* if $\mu(\mathbb{R}) = 1$.
(3) For a finite measure μ on $\mathbb{R}$ we define its *Stieltjes transform* S_μ on $\mathbb{C}\backslash\mathbb{R}$ by
$$S_\mu(z) = \int_{\mathbb{R}} \frac{1}{t-z} d\mu(t) \qquad (z \in \mathbb{C}\backslash\mathbb{R}).$$
(4) $-S_\mu = G_\mu$ is also called the *Cauchy transform.*

Theorem 4.2. *The Stieltjes transform has the following properties.*

(1) Let μ be a finite measure on $\mathbb{R}$ and $S = S_\mu$ its Stieltjes transform. Then one has:
  (i) $S : \mathbb{C}^+ \to \mathbb{C}^+$, where $\mathbb{C}^+ := \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$;
  (ii) S is analytic on $\mathbb{C}^+$;
  (iii) $\lim_{y\to\infty} iyS(iy) = -\mu(\mathbb{R})$.

(2) μ can be recovered from S_μ via the Stieltjes inversion formula*: for $a < b$ we have*
$$\lim_{\varepsilon\searrow 0} \frac{1}{\pi} \int_a^b \mathrm{Im}\, S_\mu(x+i\varepsilon)dx = \mu((a,b)) + \frac{1}{2}\mu(\{a,b\}).$$

(3) In particular, we have for two finite measures μ and ν: $S_\mu = S_\nu$ implies that $\mu = \nu$.

Proof. (1) This is Exercise 10.

(2) We have

$$\operatorname{Im} S_\mu(x+i\varepsilon) = \int_{\mathbb{R}} \operatorname{Im}\left(\frac{1}{t-x-i\varepsilon}\right) d\mu(t) = \int_{\mathbb{R}} \frac{\varepsilon}{(t-x)^2+\varepsilon} d\mu(t)$$

and thus

$$\int_a^b \operatorname{Im} S_\mu(x+i\varepsilon)dx = \int_{\mathbb{R}} \int_a^b \frac{\varepsilon}{(t-x)^2+\varepsilon^2} dx d\mu(t).$$

For the inner integral we have

$$\int_a^b \frac{\varepsilon}{(t-x)^2+\varepsilon^2} dx = \int_{(a-b)/\varepsilon}^{(b-t)/\varepsilon} \frac{1}{x^2+1} dx = \tan^{-1}\left(\frac{b-t}{\varepsilon}\right) - \tan^{-1}\left(\frac{a-t}{\varepsilon}\right)$$

$$\xrightarrow{\varepsilon \searrow 0} \begin{cases} 0, & t \notin [a,b] \\ \pi/2, & t \in \{a,b\} \\ \pi, & t \in (a,b) \end{cases}$$

From this the assertion follows.

(3) Now assume that $S_\mu = S_\nu$. By the Stieltjes inversion formula it follows then that $\mu((a,b)) = \nu((a,b))$ for all open intervals such that a and b are atoms neither of μ nor of ν. Since there can only be countably many atoms we can write any interval as

$$(a,b) = \bigcup_{n=1}^{\infty} (a+\varepsilon_n, b-\varepsilon_n),$$

where the sequence $\varepsilon_n \searrow 0$ is chosen such that all $a+\varepsilon_n$ and $b-\varepsilon_n$ are no atoms of μ nor ν. By monotone convergence for measures we get then

$$\mu((a,b) = \lim_{n\to\infty} \mu((a+\varepsilon_n, b-\varepsilon_n)) = \lim_{n\to\infty} \nu((a+\varepsilon_n, b-\varepsilon_n)) = \nu((a,b)).$$

□

Remark 4.3. If we put $\mu_\varepsilon = p_\varepsilon \lambda$ (where λ is Lebesgue measure) with density

$$p_\varepsilon(x) := \frac{1}{\pi} \operatorname{Im} S_\mu(x+i\varepsilon) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\varepsilon}{(t-x)^2+\varepsilon^2} d\mu(t),$$

then $\mu_\varepsilon = \gamma_\varepsilon * \mu$, where γ_ε is the Cauchy distribution, and we have checked explicitly in our proof of the Stieltjes inversion formula the well-known fact that $\gamma_\varepsilon * \mu$ converges weakly to μ for $\varepsilon \searrow 0$. We will talk about weak convergence later, see Definition 4.7.

Proposition 4.4. *Let μ be a compactly supported probability measure, say $\mu([-r,r]) = 1$ for some $r > 0$. Then S_μ has a power series expansion (about ∞) as follows*

$$S_\mu(z) = -\sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}} \qquad \text{for } |z| > r,$$

where $m_n := \int_{\mathbb{R}} t^n d\mu(t)$ are the moments of μ.

Proof. For $|z| > r$ we can expand

$$\frac{1}{t-z} = -\frac{1}{z(1-\frac{t}{z})} = -\frac{1}{z}\sum_{n=0}^{\infty}\left(\frac{t}{z}\right)^n$$

for all $t \in [-r,r]$; the convergence on $[-r,r]$ is uniform, hence

$$S_\mu(z) = \int_{-r}^{r} \frac{1}{t-z} d\mu(t) = -\sum_{n=0}^{\infty}\int_{-r}^{r} \frac{t^n}{z^{n+1}} d\mu(t) = -\sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}}.$$

□

Proposition 4.5. *The Stieltjes transform $S(z)$ of the semicircle distribution, $d\mu_W(t) = \frac{1}{2\pi}\sqrt{t^2-4}dt$, is, for $z \in \mathbb{C}^+$, uniquely determined by*

- *$S(z) \in \mathbb{C}^+$;*
- *$S(z)$ is the solution of the equation $S(z)^2 + zS(z) + 1 = 0$.*

Explicitly, this means

$$S(z) = \frac{-z + \sqrt{z^2-4}}{2} \qquad (z \in \mathbb{C}^+).$$

Proof. By Proposition 4.4, we know that for large $|z|$:

$$S(z) = -\sum_{k=0}^{\infty} \frac{C_k}{z^{2k+1}},$$

where C_k are the Catalan numbers. By using the recursion for the Catalan numbers (see Theorem 1.6), this implies that for large $|z|$ we have $S(z)^2 + zS(z) + 1 = 0$. Since we know that S is analytic on $\mathbb{C}^+$, this equation is, by analytic extension, then valid for all $z \in \mathbb{C}^+$.

This equation has two solutions, $(-z \pm \sqrt{z^2-4})/2$, and only the one with the $+$-sign is in $\mathbb{C}^+$. □

Remark 4.6. Proposition 4.5 gave us the Stieltjes transform of μ_W just from the knowledge of the moments. From $S(z) = (-z - \sqrt{z^2-4})/2$ we can then get the density of μ_W via the Stieltjes inversion formula:

$$\frac{1}{\pi}\operatorname{Im} S(x+i\varepsilon) = \frac{1}{2\pi}\operatorname{Im}\sqrt{(x+i\varepsilon)^2-4} \xrightarrow{\varepsilon\searrow 0} \frac{1}{2\pi}\operatorname{Im}\sqrt{x^4-4} = \begin{cases} 0, & |x|>2 \\ \frac{1}{2\pi}\sqrt{4-x^2}, & |x|\le 2. \end{cases}$$

Thus this analytic machinery gives an effective way to calculate a distribution from its moments (without having to know the density in advance).

4.2 Convergence of probability measures

Now we want to consider the convergence $\mu_N \to \mu$. We can consider (probability) measures from two equivalent perspectives:

(1) measure theoretic perspective: μ gives us the measure (probability) of sets, i.e., $\mu(B)$ for measurable sets B, or just μ(intervals)

(2) functional analytic perspective: μ allows to integrate continuous functions, i.e, it gives us $\int f d\mu$ for continuous f

According to this there are two canonical choices for a notion of convergence:

(1) $\mu_N(B) \to \mu(B)$ for all measurable sets B, or maybe for all intervals B

(2) $\int f d\mu_N \to \int f d\mu$ for all continuous f

The first possibility is problematic in this generality, as it does treat atoms too restrictive.

Example: Take $\mu_N = \delta_{1-1/N}$ and $\mu = \delta_1$. Then we surely want that $\mu_N \to \mu$, but for $B = [1,2]$ we have $\mu_N([1,2]) = 0$ for all N, but $\mu([1,2]) = 1$.

Thus the second possibility above is the better definition. But we have to be careful about which class of continuous functions we allow; we need bounded ones, otherwise $\int f d\mu$ might not exist in general; and, for compactness reasons, it is sometimes better to ignore the behaviour of the measures at infinity.

Definition 4.7. (1) We use the notations

(i) $C_0(\mathbb{R}) := \{f \in C(\mathbb{R}) \mid \lim_{|t|\to\infty} f(t) = 0\}$ are the continuous functions on $\mathbb{R}$ vanishing at infinity

(ii) $C_b(\mathbb{R}) := \{f \in C(\mathbb{R}) \mid \exists M > 0 : |f(t)| \le M \quad \forall t \in \mathbb{R}\}$ are the continuous bounded functions on $\mathbb{R}$

(2) Let μ and μ_N ($N \in N$) be finite measures. Then we say that

(i) μ_N *converges vaguely* to μ, denoted by $\mu_N \xrightarrow{v} \mu$, if

$$\int f(t)d\mu(t) \to \int f(t)d\mu(t) \qquad \text{for all } f \in C_0(\mathbb{R});$$

(ii) μ_N *converges weakly* to μ, denoted by $\mu_N \xrightarrow{w} \mu$, if

$$\int f(t)d\mu_N(t) \to \int f(t)d\mu(t) \qquad \text{for all } f \in C_b(\mathbb{R}).$$

Remark 4.8. (1) Note that weak convergence includes in particular that

$$\mu_N(\mathbb{R}) = \int 1 d\mu_N(t) \to \int 1 d\mu(t) = \mu(\mathbb{R}),$$

and thus the weak limit of probability measures must again be a probability measure. For the vague convergence this is not true; there we can loose mass at infinity.
Example: Consider $\mu_N = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_N$ and $\mu = \frac{1}{2}\delta_1$; then

$$\int f(t)d\mu_N(t) = \frac{1}{2}f(1) + \frac{1}{2}f(N) \to \frac{1}{2}f(1) = \int f(t)d\mu(t)$$

for all $f \in C_0(\mathbb{R})$. Thus the sequence of probability measures $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_N$ converges, for $N \to \infty$, to the finite measure $\frac{1}{2}\delta_1$ with total mass $1/2$.

(2) The relevance of the vague convergence, even if we are only interested in probability measures, is that the probability measures are precompact in the vague topology, but not in the weak topology. E.g., in the above example, $\mu_N = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_N$ has no subsequence which converges weakly (but it has a subsequence, namely itself, which converges vaguely).

Theorem 4.9. *The space of probability measures on $\mathbb{R}$ is precompact in the vague topology: every sequence $(\mu_N)_{N\in\mathbb{N}}$ of probability measures on $\mathbb{R}$ has a subsequence which converges vaguely to a finite measure μ, with $\mu(\mathbb{R}) \leq 1$.*

Proof. (1) From a functional analytic perspective this is a special case of the Banach-Alaoglu theorem; since complex measures on $\mathbb{R}$ are the dual space of the Banach space $C_0(\mathbb{R})$, and its weak* topology is exactly the vague topology.

(2) From a measure theory perspective this is known as Helly's (Selection) Theorem. Here are the main ideas for the proof in this setting.

(i) We describe a finite measure μ by its distribution function F_μ, given by

$$F_\mu : \mathbb{R} \to \mathbb{R}; \quad F_\mu(t) := \mu((-\infty, t]).$$

Such distribution functions can be characterized as functions F with the properties:

- $t \mapsto F(t)$ is non-decreasing

- $F(-\infty) := \lim_{t\to-\infty} F(t) = 0$ and $F(+\infty) := \lim_{t\to\infty} F(t) < \infty$
- F is continuous on the right

(ii) The vague convergence of $\mu_N \xrightarrow{v} \mu$ can also be described in terms of their distribution functions F_N, F; $\mu_N \xrightarrow{v} \mu$ is equivalent to:

$$F_N(t) \to F(t) \qquad \text{for all } t \in \mathbb{R} \text{ at which } F \text{ is continuous.}$$

(iii) Let now a sequence $(\mu_N)_N$ of probability measures be given. We consider the corresponding distribution functions $(F_N)_N$ and want to find a convergent subsequence (in the sense of (ii)) for those.
For this choose a countable dense subset $T = \{t_1, t_2, \dots\}$ of $\mathbb{R}$. Then, by choosing subsequences of subsequences and taking the "diagonal" subsequence, we get convergence for all $t \in T$. More precisely: Choose subsequence $(F_{N_1(m)})_m$ such that

$$F_{N_1(m)}(t_1) \overset{m\to\infty}{\longrightarrow} F_T(t_1),$$

choose then a subsequence $(F_{N_2(m)})_m$ of this such that

$$F_{N_2(m)}(t_1) \overset{m\to\infty}{\longrightarrow} F_T(t_1), \qquad F_{N_2(m)}(t_2) \overset{m\to\infty}{\longrightarrow} F_T(t_2);$$

iterating this gives subsequences $(F_{N_k(m)})_m$ such that

$$F_{N_k(m)}(t_i) \overset{m\to\infty}{\longrightarrow} F_T(t_i) \qquad \text{for all } i = 1, \dots, k.$$

The diagonal subsequence $(F_{N_m(m)})_m$ converges then at all $t \in T$ to $F_T(t)$.
We improve now F_T to the wanted F by

$$F(t) := \inf\{F_T(s) \mid s \in T, s > t\}$$

and show that
- F is a distribution function;
- $F_{N_m(m)}(t) \overset{m\to\infty}{\longrightarrow} F(t)$ at all continuity points of F.

According to (ii) this gives then the convergence $\mu_{N_m(m)} \overset{m\to\infty}{\longrightarrow} \mu$, where μ is the finite measure corresponding to the distribution function F. Note that $F_N(+\infty) = 1$ for all $N \in \mathbb{N}$ gives $F(+\infty) \leq 1$, but we cannot guarantee $F(+\infty) = 1$ in general.

□

Remark 4.10. If we want compactness in the weak topology, then we must control the mass at ∞ in a uniform way. This is given by the notion of tightness. A sequence $(\mu_N)_N$ of probability measures is *tight* if: for all $\varepsilon > 0$ there exists an interval $I = [-R, R]$ such that $\mu_N(I^c) < \varepsilon$ for all N.

Then one has: Any tight sequence of probability measures has a subsequence which converges weakly; the limit is then necessarily a probability measure.

4.3 Probability measures determined by moments

We can now also relate weak convergence to convergence of moments; which shows that our combinatorial approach (using moments) and analytic approach (using Stieltjes transforms) for proving the semicircle law are essentially equivalent. We want to make this more precise in the following.

Definition 4.11. A probability measure μ on $\mathbb{R}$ is *determined by its moments* if
(i) all moments $\int t^k d\mu(t) < \infty$ ($k \in \mathbb{N}$) exist;
(ii) μ is the only probability measure with those moments: if ν is a probability measure and $\int t^k d\nu(t) = \int t^k d\mu(t)$ for all $k \in \mathbb{N}$, then $\nu = \mu$.

Theorem 4.12. *Let μ and μ_N ($N \in \mathbb{N}$) be probability measures for which all moments exist. Assume that μ is determined by its moments. Assume furthermore that we have convergence of moments, i.e.,*

$$\lim_{N\to\infty} \int t^k d\mu_N(t) = \int t^k d\mu(t) \qquad \textit{for all } k \in \mathbb{N}.$$

Then we have weak convergence: $\mu_N \xrightarrow{w} \mu$.

Rough idea of proof. One has to note that convergence of moments implies tightness, which implies the existence of a weakly convergent subsequence, $\mu_{N_m} \to \nu$. Furthermore, the assumption that the moments converge implies that they are uniformly integrable, which implies then that the moments of this subsequence converge to the moments of ν. (These are kind of standard measure theoretic arguments, though a bit involved; for details see the book of Billingsley, in particular, his Theorem 25.12 and its Corollary.) However, the moments of the subsequence converge, as the moments of the whole sequence, by assumption to the moments of μ; this means that μ and ν have the same moments and hence, by our assumption that μ is determined by its moments, we have that $\nu = \mu$.
In the same way all weakly convergent subsequences of $(\nu_N)_N$ must converge to the same μ, and thus the whole sequence must converge weakly to μ.

□

Remark 4.13. (0) Note that in the first version of these notes (and also in the recorded lectures) it was claimed that, under the assumption that the limit is determined by its moments, convergence in moments is equivalent to weak convergence. This is clearly not true as the following simple example shows. Consider

$$\mu_N = (1 - \frac{1}{N})\delta_0 + \frac{1}{N}\delta_N \qquad \text{and} \qquad \mu = \delta_0.$$

Then it is clear that $\mu_N \xrightarrow{w} \mu$, and μ is also determined by its moments. But there is no convergence of moments. For example, the first moment converges, but to the wrong limit

$$\int t d\mu_N(t) = \frac{1}{N} N = 1 \to 1 \neq 0 = \int t d\mu(t),$$

and the other moments explode

$$\int t^k d\mu_N(t) = \frac{1}{N} N^k = N^{k-1} \to \infty \qquad \text{for } k \geq 2.$$

In order to have convergence of moments one needs a uniform integrability assumption; see Billingsley, in particular, his Theorem 25.12 and its Corollary.

(1) Note that there exist measures for which all moments exist but which, however, are not determined by their moments. Weak convergence to them cannot be checked by just looking on convergence of moments.
Example: The log-normal distribution with density

$$d\mu(t) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-(\log x)^2/2} dt \qquad \text{on } [0, \infty)$$

(which is the distribution of e^X for X Gaussian) is not determined by its moments.

(2) Compactly supported measures (like the semicircle) or also the Gaussian distribution are determined by their moments.

4.4 Description of weak convergence via the Stieltjes transform

Theorem 4.14. *Let μ and μ_N ($N \in \mathbb{N}$) be probability measures on $\mathbb{R}$. Then the following are equivalent.*

(i) $\mu_N \xrightarrow{w} \mu$.

(ii) For all $z \in \mathbb{C}^+$ we have: $\lim_{N\to\infty} S_{\mu_N}(z) = S_\mu(z)$.

(iii) There exists a set $D \subset \mathbb{C}^+$, which has an accumulation point in $\mathbb{C}^+$, such that: $\lim_{N\to\infty} S_{\mu_N}(z) = S_\mu(z)$ *for all $z \in D$.*

Proof.

- (i) $\Longrightarrow$ (ii): Assume that $\mu_N \xrightarrow{w} \mu$. For $z \in \mathbb{C}^+$ we consider

$$f_z : \mathbb{R} \to \mathbb{C} \qquad \text{with} \qquad f_z(t) = \frac{1}{t - z}.$$

Since $\lim_{|t|\to\infty} f_z(t) = 0$, we have $f_z \in C_0(\mathbb{R}) \subset C_b(\mathbb{R})$ and thus, by definition of weak convergence:

$$S_{\mu_N}(z) = \int f_z(t) d\mu_N(t) \to \int f_z(t) d\mu(t) = S_\mu(z).$$

- (ii) $\Longrightarrow$ (iii): clear
- (iii) $\Longrightarrow$ (i): By Theorem 4.9, we know that $(\mu_N)_N$ has a subsequence $(\mu_{N(m)})_m$ which converges vaguely to some finite measure ν with $\nu(\mathbb{R}) \leq 1$. Then, as above, we have for all $z \in D$:

$$S_\mu(z) = \lim_{m\to\infty} S_{\mu_{N(m)}}(z) = S_\nu(z).$$

  Thus the analytic functions S_μ and S_ν agree on D and hence, but the identity therem for analytic functions, also on $\mathbb{C}^+$, i.e., $S_\mu = S_\nu$. But this implies, by Theorem 4.2, that $\nu = \mu$.
  Thus the subsequence $(\mu_{N(m)})_m$ converges vaguely to the probability measure μ (and thus also weakly, see Exercise 12). In the same way, any weak cluster point of $(\mu_N)_N$ must be equal to μ, and thus the whole sequence must converge weakly to μ.

□

Remark 4.15. If we only assume that $S_{\mu_N}(z)$ converges to a limit function $S(z)$, then S must be the Stieltjes transform of a measure ν with $\nu(\mathbb{R}) \leq 1$ and we have the vague convergence $\mu_N \xrightarrow{v} \nu$.

5 Analytic Proof of Wigner's Semicircle Law for Gaussian Random Matrices

Now we are ready to give an analytic proof of Wigner's semicircle law, relying on the analytic tools we developed in the last chapter. As for the combinatorial approach, the Gaussian case is easier compared to the general Wigner case and thus we will restrict to this. The difference between real and complex is here not really relevant, instead of GUE we will treat the GOE case.

5.1 GOE random matrices

Definition 5.1. *Real Gaussian random matrices (*GOE*)* are of the form $A_N = (x_{ij})_{i,j=1}^N$, where $x_{ij} = x_{ji}$ for all i, j and $\{x_{ij} \mid i \leq j\}$ are i.i.d. (independent identically distributed) random variables with Gaussian distribution of mean zero and variance $1/N$. More formaly, on the space of symmetric $N \times N$ matrices

$$\Omega_N = \{A_N = (x_{ij})_{i,j=1}^N \mid x_{ij} \in \mathbb{R}, x_{ij} = x_{ji} \, \forall i, j\}$$

we consider the probability measure

$$d\mathbb{P}_N(A_N) = c_N \exp(-\frac{N}{4} \operatorname{Tr}(A_N^2)) \prod_{i \leq j} dx_{ij},$$

with a normalization constant c_N such that $\mathbb{P}_N$ is a probability measure.

Remark 5.2. (1) Note that with this choice of $\mathbb{P}_N$, which is invariant under orthogonal rotations, we have actually different variances on and off the diagonal:

$$\mathbb{E}\left[x_{ij}^2\right] = \frac{1}{N} \quad (i \neq j) \qquad \text{and} \qquad \mathbb{E}\left[x_{ii}^2\right] = \frac{2}{N}.$$

(2) We consider now, for each $N \in \mathbb{N}$, the averaged eigenvalue distribution

$$\mu_N := \mathbb{E}\left[\mu_{A_N}\right] = \int\limits_{\Omega_N} \mu_A dP_N(A).$$

We want to prove that $\mu_N \xrightarrow{w} \mu_W$. According to Theorem 4.14 we can prove this by showing $\lim_{N\to\infty} S_{\mu_N}(z) = S_{\mu_W}(z)$ for all $z \in \mathbb{C}^+$.

(3) Note that

$$S_{\mu_N}(z) = \int\limits_{\mathbb{R}} \frac{1}{t-z} d\mu_N(t) = \mathbb{E}\left[\int\limits_{\mathbb{R}} \frac{1}{t-z} d\mu_{A_N}(t)\right] = \mathbb{E}\left[\mathrm{tr}[(A_N - z1)^{-1}]\right],$$

since, by Assignement ...., $S_{\mu_{A_N}}(t) = \mathrm{tr}[(A_N - z1)^{-1}]$. So what we have to see, is for all $z \in \mathbb{C}^+$:

$$\lim_{N\to\infty} \mathbb{E}\left[\mathrm{tr}[(A_N - z1)^{1-}]\right] = S_{\mu_W}(z).$$

For this, we want to see that $S_{\mu_N}(z)$ satisfies approximately the quadratic equation for $S_{\mu_W}(z)$, from 4.5.

(4) Let us use for the resolvents of our matrices A the notation

$$R_A(z) = \frac{1}{A - z1}, \qquad \text{so that} \qquad S_{\mu_N}(z) = \mathbb{E}\left[\mathrm{tr}(R_{A_N}(z))\right].$$

In the following we will usually suppress the index N at our matrices; thus write just A instead of A_N, as long as the N is fixed and clear.
We have then $(A - z1)R_A(z) = 1$, or $A \cdot R_A(z) - zR_A(z) = 1$, thus

$$R_A(z) = -\frac{1}{z}1 + \frac{1}{z}AR_A(z).$$

Taking the normalized trace and expectation of this yields

$$\mathbb{E}\left[\mathrm{tr}(R_A(z))\right] = -\frac{1}{z} + \frac{1}{z}\mathbb{E}\left[\mathrm{tr}(AR_A(z))\right].$$

The left hand side is our Stieltjes transform, but what about the right hand side; can we relate this also to the Stieltjes transform? Note that the function under the expectation is $\frac{1}{N}\sum_{k,l} x_{kl}[R_A(z)]_{lk}$; thus a sum of terms which are the product of one of our Gaussian variables times a function of all the independent Gaussian variables. There exists actually a very nice and important formula to deal with such expectations of independent Gaussian variables. In a sense, this is the analytic version of the combinatorial Wick formula.

5.2 Stein's identity for independent Gaussian variables

Proposition 5.3 (Stein's identity)**.** *Let $X_1, \dots, X_k$ be independent random variables with Gaussian distribution, with mean zero and variances $\mathbb{E}[X_i] = \sigma_i^2$. Let $h : \mathbb{R}^k \to \mathbb{C}$ be continuously differentiable such that h and all partial derivatives are of polynomial growth. Then we have for $i = 1, \dots, k$:*

$$\mathbb{E}[X_i h(X_1, \dots, X_k)] = \sigma_i^2 \mathbb{E}\left[\frac{\partial h}{\partial x_i}(X_1, \dots, X_k)\right].$$

More explicitly,

$$\int_{\mathbb{R}^k} x_i h(x_1, \dots, x_k) \exp\left(-\frac{x_1^2}{2\sigma_1^2} - \dots - \frac{x_k^2}{2\sigma_k^2}\right) dx_1 \dots dx_k =$$
$$\sigma_i^2 \int_{\mathbb{R}^k} \frac{\partial h}{\partial x_i}(x_1, \dots, x_k) \exp\left(-\frac{x_1^2}{2\sigma_1^2} - \dots - \frac{x_k^2}{2\sigma_k^2}\right) dx_1 \dots dx_k$$

Proof. The main argument happens for $k = 1$. Since $xe^{-x^2/(2\sigma^2)} = [-\sigma^2 e^{-x^2/(2\sigma^2)}]'$ we get by partial integration

$$\int_{\mathbb{R}} xh(x)e^{-x^2/(2\sigma^2)}dx = \int_{\mathbb{R}} h(x)[-\sigma^2 e^{-x^2/(2\sigma^2)}]'dx = \int_{\mathbb{R}} h'(x)\sigma^2 e^{-x^2/(2\sigma^2)}dx;$$

our assumptions on h are just such that the boundary terms vanish.

For general k, we just do partial integration for the i-th coordinate.

□

We want to apply this now to our Gaussian random matrices, with Gaussian random variables x_{ij} $(1 \le i \le j \le N)$ of variance

$$\sigma_{ij}^2 = \begin{cases} \frac{1}{N}, & i \neq j \\ \frac{2}{N}, & i = j \end{cases}$$

and for the function

$$h(x_{ij} \mid i \le j) = h(A) = [R_A(z)]_{lk}, \qquad \text{where} \qquad R_A(z) = \frac{1}{A - z1}.$$

To use Stein's identitiy 5.3 in this case we need the partial derivatives of the resolvents.

Lemma 5.4. *For $A = (x_{ij})_{i,j=1}^N$ with $x_{ij} = x_{ji}$ for all i, j, we have for all i, j, k, l:*

$$\frac{\partial}{\partial x_{ij}}[R_A(z)]_{lk} = \begin{cases} -[R_A(z)]_{li} \cdot [R_A(z)]_{ik}, & i = j, \\ -[R_A(z)]_{li} \cdot [R_A(z)]_{jk} - [R_A(z)]_{lj} \cdot [R_A(z)]_{ik}, & i \neq j. \end{cases}$$

Proof. Note first that

$$\frac{\partial A}{\partial x_{ij}} = \begin{cases} E_{ii}, & i = j \\ E_{ij} + E_{ji}, & i \neq j \end{cases}$$

where E_{ij} is a matrix unit with 1 at position (i, j) and 0 elsewhere.

We have $R_A(z) \cdot (A - z1) = 1$, which yields by differentiating

$$\frac{\partial R_A(z)}{\partial x_{ij}} \cdot (A - z1) + R_a(z) \cdot \frac{\partial A}{\partial x_{ij}} = 0,$$

and thus

$$\frac{\partial R_A(z)}{\partial x_{ij}} = -R_A(z) \cdot \frac{\partial A}{\partial x_{ij}} \cdot R_A(z).$$

This gives, for $i = j$,

$$\frac{\partial}{\partial x_{ii}}[R_A(z)]_{lk} = -[R_A(z) \cdot E_{ii} \cdot R_A(z)]_{lk} = -[R_A(z)]_{li} \cdot [R_A(z)]_{ik},$$

and, for $i \neq j$,

$$\begin{aligned} \frac{\partial}{\partial x_{ij}}[R_A(z)]_{lk} &= -[R_A(z) \cdot E_{ij} \cdot R_A(z)]_{lk} - [R_A(z) \cdot E_{ji} \cdot R_A(z)]_{lk} \\ &= -[R_A(z)]_{li} \cdot [R_A(z)]_{jk} - [R_A(z)]_{lj} \cdot [R_A(z)]_{ik}. \end{aligned}$$

□

5.3 Semicircle law for GOE

Theorem 5.5. *Let A_N be* GOE *random matrices as in 5.1. Then its averaged eigenvalue distribution $\mu_N := \mathbb{E}[\mu_{A_N}]$ converges weakly to the semicircle distribution: $\mu_N \xrightarrow{w} \mu_W$.*

Proof. By Theorem 4.14, it suffices to show $\lim_{N\to\infty} S_{\mu_N}(z) = S_{\mu_W}(z)$ for all $z \in \mathbb{C}^+$.

In Remark 5.2(4) we have seen that (where we write A instead of A_N)

$$S_{\mu_N}(z) = \mathbb{E}[\mathrm{tr}(R_A(z)]] = -\frac{1}{z} + \frac{1}{z}\mathbb{E}[\mathrm{tr}[AR_A(z)]].$$

Now we calculate, with $A = (x_{ij})_{i,j=1}^N$,

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}[AR_A(z)]\right] &= \frac{1}{N}\sum_{k,l=1}^N \mathbb{E}\left[x_{kl}\cdot[R_A(z)]_{lk}\right] \\
&= \frac{1}{N}\sum_{k,l=1}^N \sigma_{kl}^2 \cdot \mathbb{E}\left[\frac{\partial}{\partial x_{kl}}[R_A(z)]_{lk}\right] \\
&= -\frac{1}{N}\sum_{k,l=1}^N \frac{1}{N}\cdot\Big([R_A(z)]_{lk}\cdot[R_A(z)]_{lk} + [R_A(z)]_{ll}\cdot[R_A(z)]_{kk}\Big).
\end{aligned}$$

(Note that the combination of the different covariances and of the different form of the formula in Lemma 5.4 for on-diagonal and for off-diagonal entries gives in the end the same result for all pairs of (k,l).)

Now note that $(A - z1)$ is symmetric, hence the same is true for its inverse $R_A(z) = (A - z1)^{-1}$ and thus: $[R_A(z)]_{lk} = [R_A(z)]_{kl}$. Thus we get finally

$$\mathbb{E}\left[\operatorname{tr}[AR_A(z)]\right] = -\frac{1}{N}\mathbb{E}\left[\operatorname{tr}[R_A(z)^2]\right] - \mathbb{E}\left[\operatorname{tr}[R_A(z)]\cdot\operatorname{tr}[R_A(z)]\right].$$

To proceed further we need to deal with the two summands on the right hand side; we expect

- the first term, $\frac{1}{N}\mathbb{E}\left[\operatorname{tr}[R_A(z)^2]\right]$, should go to zero, for $N\to\infty$
- the second term, $\mathbb{E}\left[\operatorname{tr}[R_A(z)]\cdot\operatorname{tr}[R_A(z)]\right]$, should be close to its factorized version $\mathbb{E}\left[\operatorname{tr}[R_A(z)]\right]\cdot\mathbb{E}\left[\operatorname{tr}[R_A(z)]\right] = S_{\mu_N}(z)^2$

Both these ideas are correct; let us try to make them rigorous.

- A as a symmetric matrix can be diagonalized by an orthogonal matrix U,

$$A = U\begin{pmatrix}\lambda_1 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \lambda_N\end{pmatrix}U^*, \quad \text{and thus} \quad R_A(z)^2 = \begin{pmatrix}\frac{1}{(\lambda_1-z)^2} & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \frac{1}{(\lambda_N-z)^2}\end{pmatrix}U^*,$$

  which yields

$$|\operatorname{tr}[R_A(z)^2]| \le \frac{1}{N}\sum_{i=1}^N \left|\frac{1}{(\lambda_i - z)^2}\right|.$$

  Note that for all $\lambda\in\mathbb{R}$ and all $z\in\mathbb{C}^+$

$$\left|\frac{1}{\lambda - z}\right| \le \frac{1}{\operatorname{Im} z}$$

  and hence

$$\frac{1}{N}\left|\mathbb{E}\left[\operatorname{tr}(R_A(z)^2)\right]\right| \le \frac{1}{N}\mathbb{E}\left[|\operatorname{tr}[R_A(z)^2]|\right] \le \frac{1}{N}\frac{1}{(\operatorname{Im} z)^2} \to 0 \qquad \text{for } N\to\infty.$$

- By definition of the variance we have

$$\mathbb{V}\mathrm{ar}\left[X\right] := \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2,$$

  and thus

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[X\right]^2 + \mathbb{V}\mathrm{ar}\left[X\right].$$

  Hence we can replace $\mathbb{E}\left[\mathrm{tr}[R_A(z)] \cdot \mathrm{tr}[R_A(z)]\right]$ by

$$\mathbb{E}\left[\mathrm{tr}[R_A(z)]\right]^2 + \mathbb{V}\mathrm{ar}\left[\mathrm{tr}[R_A(z)]\right] = S_{\mu_N}(z)^2 + \mathbb{V}\mathrm{ar}\left[S_{\mu_A}(z)\right].$$

  In the next chapter we will show that we have concentration, i.e., the variance $\mathbb{V}\mathrm{ar}\left[S_{\mu_A}(z)\right]$ goes to zero for $N \to \infty$.

With those two ingredients we have then

$$S_{\mu_N}(z) = -\frac{1}{z} + \frac{1}{z}\mathbb{E}\left[\mathrm{tr}[AR_A(z)]\right] = -\frac{1}{z} - \frac{1}{z}S_{\mu_N}(z)^2 + \varepsilon_N,$$

where $\varepsilon_N \to$ for $N \to \infty$.

Note that, as above, for any Stieltjes transform S_ν we have

$$|S_\nu(z)| = |\int \frac{1}{t-z} d\nu(t)| \leq \int |\frac{1}{t-z}| d\nu(t) \leq \frac{1}{\mathrm{Im}\, z},$$

and thus $(S_{\mu_N}(z))_N$ is a bounded sequence of complex numbers. Hence, by compactness, there exists a convergent subsequence $(S_{\mu_{N(m)}}(z))_m$, which converges to some $S(z)$. This $S(z)$ must then satisfy the limit $N \to \infty$ of the above equation, thus

$$S(z) = -\frac{1}{z} - \frac{1}{z}S(z)^2.$$

Since all $S_{\mu_N}(z)$ are in $\mathbb{C}^+$, the limit $S(z)$ must be in $\overline{\mathbb{C}^+}$, which leaves for $S(z)$ only the possibility that

$$S(z) = \frac{-z + \sqrt{z^2 - 4}}{2} = S_{\mu_W}(z)$$

(as the other solution is in $\mathbb{C}^-$).

In the same way, it follows that any subsequence of $(S_{\mu_N}(z))_N$ has a convergent subsequence which converges to $S_{\mu_W}(z)$; this forces all cluster points of $(S_{\mu_N}(z))_N$ to be $S_{\mu_W}(z)$. Thus the whole sequence converges to $S_{\mu_N}(z)$. This holds for any $z \in \mathbb{C}^+$, and thus implies that $\mu_N \xrightarrow{w} \mu_W$. □

To complete the proof we still have to see the concentration to get the asymptotic vanishing of the variance. We will address such concentration questions in the next chapter.

6 Concentration Phenomena and Stronger Forms of Convergence for the Semicircle Law

6.1 Forms of convergence

Remark 6.1. (1) Recall that our random matrix ensemble is given by probability measures $\mathbb{P}_N$ on sets Ω_N of $N \times N$ matrices and we want to see that μ_{A_N} converges weakly to μ_W, or, equivalently, that, for all $z \in \mathbb{C}^+$, $S_{\mu_{A_N}}(z)$ converges to $S_{\mu_W}(z)$. There are different levels of this convergence with respect to $\mathbb{P}_N$:

(i) *convergence in average*, i.e.,

$$\mathbb{E}\left[S_{\mu_{A_N}}(z)\right] \stackrel{N\to\infty}{\longrightarrow} S_{\mu_W}(z)$$

(ii) *convergence in probability*, i.e.,

$$\mathbb{P}_N\{A_N \mid |S_{\mu_{A_N}}(z) - S_{\mu_W}(z)| \geq \varepsilon\} \stackrel{N\to\infty}{\longrightarrow} 0 \qquad \text{for all } \varepsilon > 0$$

(iii) *almost sure convergence*, i.e.,

$$\mathbb{P}\{(A_N)_N \mid S_{\mu_{A_N}}(z) \text{ does not converge to } S_{\mu_W}(z)\} = 0;$$

instead of making this more precise, let us just point out that this almost sure convergence is guaranteed, by the Borel-Cantelli Lemma, if the convergence in (ii) to zero is sufficiently fast in N, so that for all $\varepsilon > 0$

$$\sum_{N=1}^{\infty} \mathbb{P}_N\{A_N \mid |S_{\mu_{A_N}}(z) - S_{\mu_W}(z)| \geq \varepsilon\} < \infty.$$

(2) Note that we have here convergence of probabilistic quantities to a deterministic limit, thus (ii) and (iii) are saying that for large N the eigenvalue distribution of A_N concentrates in a small neighborhood of μ_W. This is an instance of a quite general "concentration of measure" phenomenon; according

to a dictum of M. Talagrand: "A random variable that depends (in a smooth way) on the influence of many independent variables (but not too much on any of them) is essentially constant."

(3) Note also that many classical results in probability theory (like law of large numbers) can be seen as instances of this, dealing with linear functions. However, this principle also applies to non-linear functions - like in our case, to $\mathrm{tr}[(A - z1)^{-1}]$, considered as function of the entries of A.

(4) Often control of the variance of the considered function is a good way to get concentration estimates. We develop in the following some of the basics for this.

6.2 Markov's and Chebyshev's inequality

Notation 6.2. A *probability space* $(\Omega, \mathbb{P})$ consists of a set Ω, equipped with a σ-algebra of measurable sets, and a probability measure $\mathbb{P}$ on the measurable sets of Ω. A *random variable* X is a measurable function $X : \Omega \to \mathbb{R}$; its *expectation* or *mean* is given by

$$\mathbb{E}[X] := \int_\Omega X(\omega) d\mathbb{P}(\omega),$$

and its *variance* is given by

$$\mathbb{V}\mathrm{ar}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \int_\Omega (X(\omega) - \mathbb{E}[(] X))^2 d\mathbb{P}(\omega).$$

Theorem 6.3 (Markov's Inequality)**.** *Let X be a random variable taking non-negative values. Then, for any $t > 0$,*

$$\mathbb{P}\{\omega \mid X(\omega) \geq t\} \leq \frac{\mathbb{E}[X]}{t}.$$

$\mathbb{E}[X]$ could here also be ∞, but then the statement is not very useful. The Markov inequality only gives useful information if X has finite mean, and then only for $t > \mathbb{E}[X]$.

Proof. Since $X(\omega) \geq 0$ for all $\omega \in \Omega$ we can estimate as follows:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\Omega} X(\omega) d\mathbb{P}(\omega) \\
&= \int_{\{X(\omega)\geq t\}} X(\omega) d\mathbb{P}(\omega) + \int_{\{X(\omega)<t\}} X(\omega) d\mathbb{P}(\omega) \\
&\geq \int_{\{X(\omega)\geq t\}} X(\omega) d\mathbb{P}(\omega) \\
&\geq \int_{\{X(\omega)\geq t\}} t d\mathbb{P}(\omega) \\
&= t \cdot \mathbb{P}\{X(\omega) \geq t\}.
\end{aligned}$$

□

Theorem 6.4 (Chebyshev's Inequality)**.** *Let X be a random variable with finite mean $\mathbb{E}[X]$ and variance $\mathbb{V}ar[X]$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\{\omega \mid |X(\omega) - \mathbb{E}[X]| \geq \varepsilon\} \leq \frac{\mathbb{V}ar[X]}{\varepsilon^2}.$$

Proof. We use Markov's inequality 6.3 for the positive random variable $Y := (X - \mathbb{E}[X])^2$. Note that

$$\mathbb{E}[Y] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{V}\mathrm{ar}[X].$$

Thus we have for $\varepsilon > 0$

$$\begin{aligned}
\mathbb{P}\{\omega \mid |X(\omega - \mathbb{E}[(] X)| \geq \varepsilon\} &= \mathbb{P}\{\omega \mid (X(\omega) - \mathbb{E}[X])^2 \geq \varepsilon^2\} \\
&= \mathbb{P}[\{\omega \mid Y(\omega) \geq \varepsilon^2\} \\
&\leq \frac{\mathbb{E}[Y]}{\varepsilon^2} \\
&= \frac{\mathbb{V}\mathrm{ar}[X]}{\varepsilon^2}.
\end{aligned}$$

□

Remark 6.5. Our goal will thus be to control the variance of $X = f(X_1, \ldots, X_n)$ for $X_1, \ldots, X_n$ independent random variables. (In our case, the X_i will be the entries of the GOE matrix A and f will be the function $f = \mathrm{tr}[(A - z1)^{-1}]$.) A main idea in this context is to have estimates which go over from separate control of each variable to control of all variables together; i.e., which are stable under tensorization. There are two prominent types of such estimates, namely

(i) Poincaré inequality
(ii) LSI=logarithmic Sobolev inequality

We will focus here on (i) and say a few words on (ii) later.

6.3 Poincaré inequality

Definition 6.6. A random variable $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ satisfies a *Poincaré inequality with constant* $c > 0$ if for any differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbb{E}[f(X)^2] < \infty$ we have

$$\mathbb{V}\mathrm{ar}[f(X)] \leq c \cdot \mathbb{E}\left[\|\nabla f(X)\|_2^2\right] \qquad \text{where} \qquad \|\nabla f\|_2^2 = \sum_{i=1}^n (\frac{\partial f}{\partial x_i})^2.$$

Remark 6.7. Let us write this also "explicitly" in terms of the distribution μ of the random variable $X : \Omega \to \mathbb{R}^n$; recall that μ is the push-forward of the probability measure $\mathbb{P}$ under the map X to a probability measure on $\mathbb{R}^n$. In terms of μ we have then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^n} f(x_1, \dots, x_n) d\mu(x_1, \dots, x_n)$$

and the Poincaré inequality asks then for

$$\int_{\mathbb{R}^n} \Big(f(x_1, \dots, x_n) - \mathbb{E}[f(X)]\Big)^2 d\mu(x_1, \dots, x_n)$$
$$\leq c \cdot \sum_{i=1}^n \int_{\mathbb{R}^n} \left(\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)\right)^2 d\mu(x_1, \dots, x_n).$$

Theorem 6.8 (Efron-Stein Inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables and let $f(X_1, \dots, X_n)$ be a square-integrable function of $X = (X_1, \dots, X_n)$. Then we have*

$$\mathbb{V}ar[f(X)] \leq \sum_{i=1}^n \mathbb{E}\left[\mathbb{V}ar^{(i)}[f(X)]\right],$$

where $\mathbb{V}ar^{(i)}$ denotes taking the variance in the i-th variable, keeping all the other variables fixed, and the expectation is then integrating over all the other variables.

Proof. We denote the distribution of X_i by μ_i; this is, for each i, a probability measure on $\mathbb{R}$. Since $X_1, \dots, X_n$ are independent, the distribution of $X = (X_1, \dots, X_n)$ is given by the product measure $\mu_1 \times \cdots \times \mu_n$ on $\mathbb{R}^n$.

Putting $Z = f(X_1, \ldots, X_n)$, we have

$$\mathbb{E}[Z] = \int_{\mathbb{R}^n} f(x_1, \ldots, x_n) d\mu_1 \ldots d\mu_n(x_n)$$

and

$$\mathbb{V}\mathrm{ar}[Z] = \int_{\mathbb{R}^n} \Big(f(x_1, \ldots, x_n) - \mathbb{E}[Z] \Big)^2 d\mu_1 \ldots d\mu_n.$$

We will now do the integration $\mathbb{E}$ by integrating one variable at a time and control each step. For this we write

$$\begin{aligned} Z - \mathbb{E}[Z] = \quad & Z - \mathbb{E}_1[Z] \\ & + \mathbb{E}_1[Z] - \mathbb{E}_{1,2}[Z] \\ & + \mathbb{E}_{1,2}[Z] - \mathbb{E}_{1,2,3}[Z] \\ & \quad \vdots \\ & + \mathbb{E}_{1,2,\ldots,n-1}[Z] - \mathbb{E}[Z], \end{aligned}$$

where $\mathbb{E}_{1,\ldots,k}$ denotes integration over the variables $x_1, \ldots, x_k$, leaving a function of the variables $x_{k+1}, \ldots, x_n$. Thus, with

$$\Delta_i := \mathbb{E}_{1,\ldots,i-1}[Z] - \mathbb{E}_{1,\ldots,i-1,i}[Z]$$

(which is a function of the variables $x_i, x_{i+1}, \ldots, x_n$), we have $Z - \mathbb{E}[Z] = \sum_{i=1}^n \Delta_i$, and thus

$$\begin{aligned} \mathbb{V}\mathrm{ar}[Z] &= \mathbb{V}\mathrm{ar}\left[(Z - \mathbb{E}[Z])^2\right] \\ &= \mathbb{V}\mathrm{ar}\left[\left(\sum_{i=1}^n \Delta_i\right)^2\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[\Delta_i^2\right] + \sum_{i \neq j} \mathbb{E}[\Delta_i \Delta_j]. \end{aligned}$$

Now observe that for all $i \neq j$ we have $\mathbb{E}[\Delta_i \Delta_j] = 0$. Indeed, consider, for example, $n = 2$ and $i = 1$, $j = 2$:

$$\begin{aligned} &\mathbb{E}[\Delta_1 \Delta_2] = \mathbb{E}[(Z - \mathbb{E}_1[Z]) \cdot (\mathbb{E}_1[Z] - \mathbb{E}_{1,2}[Z])] \\ &= \int \left[f(x_1, x_2) - \int f(\tilde{x}_1, x_2) d\mu_1(\tilde{x}_1) \right] \cdot \\ &\qquad \cdot \left[\int f(\tilde{x}_1, x_2) d\mu_1(\tilde{x}_1) - \int f(\tilde{x}_1, \tilde{x}_2) d\mu_1(\tilde{\mu}_1) d\mu_2(\tilde{x}_2) \right] d\mu_1(x_1) d\mu_2(x_2) \end{aligned}$$

Integration with respect to x_1 now affects only the first factor and integrating this gives zero. The general case $i \neq j$ works in the same way. Thus we get

$$\mathbb{V}\mathrm{ar}\,[Z] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2\right].$$

We denote now with $\mathbb{E}^{(i)}$ integration with respect to the variable x_i, leaving a function of the other variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, and

$$\mathbb{V}\mathrm{ar}^{(i)}\,[Z] := \mathbb{E}^{(i)}[(Z - \mathbb{E}^{(i)}[Z])^2].$$

Then we have

$$\Delta_i = \mathbb{E}_{1,\ldots,i-1}[Z] - \mathbb{E}_{1,\ldots,i}[Z] = \mathbb{E}_{1,\ldots,i-1}[Z - \mathbb{E}^{(i)}[Z]],$$

and thus by Jensen's inequality (which is here just the fact that variances are non-negative),

$$\Delta_i^2 \leq \mathbb{E}_{1,\ldots,i-1}[(Z - \mathbb{E}^{(i)}[Z])^2].$$

This gives finally

$$\mathbb{V}\mathrm{ar}\,[Z] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2\right] \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}^{(i)}[(Z - \mathbb{E}^{(i)}[Z])^2]\right],$$

which is the assertion. □

Theorem 6.9. *Let $X_1, \ldots, X_n$ be independent random variables in $\mathbb{R}$, such that each X_i satisfies a Poincaré inequality with constant c_i. Then $X = (X_1, \ldots, X_n)$ satisfies a Poicaré inequality in $\mathbb{R}^n$ iwth constant $c = \max(c_1, \ldots, c_n)$.*

Proof. By the Efron-Stein inequality 6.8, we have

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\,[f(X)] &\leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{V}\mathrm{ar}^{(i)}\,[f(X)]\right] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[c_i \cdot \mathbb{E}^{(i)}\left[\left(\frac{\partial f}{\partial x_i}\right)^2\right]\right] \\
&\leq c \cdot \sum_{i=1}^{n} \mathbb{E}\left[\left(\frac{\partial f}{\partial x_i}\right)^2\right] \\
&= c \cdot \mathbb{E}\left[\|\nabla f(X)\|_2^2\right].
\end{aligned}$$

In the step from the first to the second line we have used, for fixed i, the Poincaré inequality for X_i and the function $x_i \mapsto f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$, for each fixed $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. □

Theorem 6.10 (Gaussian Poincaré Inequality)**.** *Let $X_1, \dots, X_n$ be independent standard Gaussian random variables, $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i^2] = 1$. Then $X = (X_1, \dots, X_n)$ satisfies a Poincaré inequality with constant 1; i.e., for each continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$ we have*

$$\mathbb{V}ar[f(X)] \leq \mathbb{E}\left[\|\nabla f(X)\|^2\right].$$

Remark 6.11. (1) Note the independence of n!

(2) By Theorem 6.9 it suffices to prove the statement for $n = 1$. But even in this one-dimensional case the statement is not obvious. Let us see what we are actually claiming in this case: X is a standard Gaussian random variable and $f : \mathbb{R} \to \mathbb{R}$, and the Poincaré inequality says

$$\mathbb{V}\mathrm{ar}[f(X)] \leq \mathbb{E}\left[f'(X)^2\right].$$

We might also assume that $\mathbb{E}[f(X)] = 0$, then this means explicitly:

$$\int_{\mathbb{R}} f(x)^2 e^{-x^2/2} dx \leq \int_{\mathbb{R}} f'(x)^2 e^{-x^2/2} dx.$$

Proof. As remarked in Remark 6.11, the general case can, by Theorem 6.9, be reduced to the one-dimensional case and, by shifting our function f by a constant, we can also assume that $f(X)$ has mean zero. One possible proof is to approximate X via a central limit theorem by independent Bernoulli variables Y_i.

So let $Y_1, Y_2, \dots$ be independent Bernoulli variables, i.e., $\mathbb{P}[Y_i = 1] = 1/2 = \mathbb{P}[Y_i = -1]$ and put

$$S_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}}.$$

Then, by the central limit theorem, the distribution of S_n converges weakly, for $n \to \infty$, to a standard Gaussian distribution. So we can approximate $f(X)$ by

$$g(Y_1, \dots, Y_n) = f(S_n) = f\left(\frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n)\right).$$

By the Efron-Stein inequality 6.8, we have

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[f(S_n)] &= \mathbb{V}\mathrm{ar}[g(Y_1, \dots, Y_n)] \\
&\leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{V}\mathrm{ar}^{(i)}[g(Y_1, \dots, Y_n)]\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{V}\mathrm{ar}^{(i)}[f(S_n)]\right].
\end{aligned}$$

Put

$$S_n^{[i]} := S_n - \frac{1}{\sqrt{n}} Y_i = \frac{1}{\sqrt{n}}(Y_1 + \cdots Y_{i-1} + Y_{i+1} + \cdots + Y_n).$$

Then

$$\mathbb{E}^{(i)}[f(S_n)] = \frac{1}{2}\left(f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) + f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big)\right)$$

and

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}^{(i)}\left[f(S_n)\right] \\
&= \frac{1}{2}\left\{ \left(f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) - \mathbb{E}^{(i)}[f(S_n)]\right)^2 + \left(f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big) - \mathbb{E}^{(i)}[f(S_n)]\right)^2 \right\} \\
&= \frac{1}{4}\left(f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) - f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big)\right)^2,
\end{aligned}$$

and thus

$$\mathbb{V}\mathrm{ar}\left[f(S_n)\right] \le \frac{1}{4}\sum_{i=1}^{n} \mathbb{E}\left[\left(f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) - f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big)\right)^2 \right].$$

By Taylor's theorem we have now

$$\begin{aligned}
f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) &= f\Big(S_n^{[i]}\Big) + \frac{1}{\sqrt{n}} f'\Big(S_n^{[i]}\Big) + \frac{1}{2n} f''(\xi_+) \\
f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big) &= f\Big(S_n^{[i]}\Big) - \frac{1}{\sqrt{n}} f'\Big(S_n^{[i]}\Big) + \frac{1}{2n} f''(\xi_+)
\end{aligned}$$

We assume that f is twice differentiable and f' and f'' are bounded: $|f'(\xi)| \le K$ and $|f''(\xi)| \le K$ for all $\xi \in \mathbb{R}$. (The general situation can be approximated by this.) Then we have

$$\begin{aligned}
\left(f\Big(S_n^{[i]} + \frac{1}{\sqrt{n}}\Big) - f\Big(S_n^{[i]} - \frac{1}{\sqrt{n}}\Big)\right)^2 &= \left(\frac{2}{\sqrt{n}} f'\Big(S_n^{[i]}\Big) + \frac{1}{2n}\Big(f''(\xi_+) - f''(\xi_-)\Big)\right)^2 \\
&= \frac{4}{n} f'\Big(S_n^{[i]}\Big)^2 + \frac{2}{n^{3/2}} R_1 R_2 + \frac{1}{4n^2} R_2^2,
\end{aligned}$$

where we have put $R_1 := f'\Big(S_n^{[i]}\Big)$ and $R_2 := f''(\xi_+) - f''(\xi_-)$. Note that $|R_1| \le K$ and $|R_2| \le 2K$, and thus

$$\mathbb{V}\mathrm{ar}\left[f(S_n)\right] \le \frac{1}{4} n \left\{ \frac{4}{n}\mathbb{E}\left[f'\Big(S_n^{[i]}\Big)^2 \right] + \frac{2}{n^{3/2}} 2K^2 + \frac{1}{4n^2} 4K^2 \right\}.$$

Note that the first term containing $S_n^{[i]}$ is actually independent of i. Now we take the limit $n \to \infty$ in this inequality; since both S_n and $S_n^{[i]}$ converge to our standard Gaussian variable X we obtain finally the wanted

$$\mathbb{V}\mathrm{ar}\left[f(X)\right] \leq \mathbb{E}\left[f'(X)^2\right].$$

□

6.4 Concentration for $\mathrm{tr}[R_A(z)]$ via Poincaré inequality

We apply this Gaussian Poincaré inequality now to our random matrix setting $A = (x_{ij})_{i,j=1}^N$, where $\{x_{ijj} \mid i \leq j\}$ are independent Gaussian random variables with $\mathbb{E}\left[x_{ij}\right] = 0$ and $\mathbb{E}\left[x_{ii}\right] = 2/N$ on the diagonal and $\mathbb{E}\left[x_{ij}\right] = 1/N$ $(i \neq j)$ off the diagonal. Note that by a change of variable the constant in the Poincare inequality for this variances is given by $\max\{\sigma_{ij}^2 \mid i \leq j\} = 2/N$. Thus we have in our setting for nice real-valued f:

$$\mathbb{V}\mathrm{ar}\left[f(A)\right] \leq \frac{2}{N} \cdot \mathbb{E}\left[\|f(A)\|_2^2\right].$$

We take now

$$g(A) := \mathrm{tr}[(A - z1)^{-1}] = \mathrm{tr}[R_A(z)]$$

and want to control $\mathbb{V}\mathrm{ar}\left[g(A_N\right]$ for $N \to \infty$. Note that g is complex-valued (since $z \in \mathbb{C}^+$), but we can estimate

$$|\mathbb{V}\mathrm{ar}\left[g(A)\right]| = |\mathbb{V}\mathrm{ar}\left[\mathrm{Re}\, g(A) + \sqrt{-1}\,\mathrm{Im}\, g(A)\right]| \leq 2\Big(\mathbb{V}\mathrm{ar}\left[\mathrm{Re}\, g(A)\right] + \mathbb{V}\mathrm{ar}\left[\mathrm{Im}\, g(A)\right]\Big).$$

Thus it suffices to estimate the variance of real and imaginary part of $g(A)$.

We have, for $i < j$,

$$\begin{aligned}
\frac{\partial g(A)}{\partial x_{ij}} &= \frac{\partial}{\partial x_{ij}} \mathrm{tr}[R_A(z)] \\
&= \frac{1}{N} \sum_{k=1}^N \frac{\partial [R_A(z)]_{kk}}{\partial x_{ij}} \\
&= -\frac{1}{N} \sum_{k=1}^N \Big([R_A(z)]_{ki} \cdot [R_A(z)]_{jk} + [R_A(z)]_{kj} \cdot [R_A(z)]_{ik}\Big) \qquad \text{by Lemma 5.4}
\end{aligned}$$

$$= -\frac{2}{N}\sum_{k=1}^{N}[R_A(z)]_{ik}\cdot[R_A(z)]_{kj} \qquad \text{since } R_A(z) \text{ is symmetric, see proof of 5.5}$$
$$= -\frac{2}{N}[R_A(z)^2]_{ij},$$

and the same for $i = j$ with $2/N$ replaced by $1/N$.

Thus we get for $f(A) := \operatorname{Re} g(A) = \operatorname{Re}\operatorname{tr}[R_A(z)]$:

$$|\frac{\partial f(A)}{\partial x_{ij}}| = |\operatorname{Re}\frac{\partial g(A)}{\partial x_{ij}}| \leq \frac{2}{N}|[R_A(z)^2]_{ij}| \leq \frac{2}{N}\|R_A(z)^2\| \leq \frac{2}{N\cdot(\operatorname{Im} z)^2},$$

where in the last step we used the usual estimate for resolvents as in the proof of Theorem 5.5. Hence we have

$$|\frac{\partial f(A)}{\partial x_{ij}}|^2 \leq \frac{4}{N^2\cdot(\operatorname{Im} z)^4},$$

and thus our Gaussian Poincaré inequality 6.10 (with constant $2/N$) yields

$$\mathbb{V}\mathrm{ar}\,[f(A)] \leq \frac{2}{N}\cdot\sum_{i\leq j}|\frac{\partial f(A)}{\partial x_{ij}}|^2 \leq \frac{8}{N\cdot(\operatorname{Im} z)^4}.$$

The same estimate holds for the imaginary part and thus, finally, we have for the variance of the trace of the resolvent:

$$\mathbb{V}\mathrm{ar}\,[\operatorname{tr}[R_A(z)]] \leq \frac{32}{N\cdot(\operatorname{Im} z)^4}.$$

The fact that $\mathbb{V}\mathrm{ar}\,[\operatorname{tr}[R_A(z)]]$ goes to zero for $N \to \infty$ closes the gap in our proof of Theorem 5.5. Furthermore, it also improves the type of convergence in Wigner's semicircle law.

Theorem 6.12. *Let A_N be* GOE *random matrices as in 5.1. Then the eigenvalue distribution μ_{A_N} converges in probability to the semicircle distribution. Namely, for each $z \in \mathbb{C}^+$ and all $\varepsilon > 0$ we have*

$$\lim_{N\to\infty}\mathbb{P}_N\{A_N \mid |S_{\mu_{A_N}}(z) - S_{\mu_W}(z)| \geq \varepsilon\} = 0.$$

Proof. By the Chebyshev inequality 6.4, our above estimate for the variance implies for any $\varepsilon > 0$ that

$$\mathbb{P}_N\{A_N \mid |\operatorname{tr}[R_{A_N}(z)] - \mathbb{E}\left[\operatorname{tr}[R_{A_N}(z)]\right]| \geq \varepsilon\} \leq \frac{\mathbb{V}\mathrm{ar}\left[\operatorname{tr}[R_{A_N}]\right]}{\varepsilon^2}$$
$$\leq \frac{32}{N \cdot (\operatorname{Im} z)^4 \cdot \varepsilon^2} \overset{N\to\infty}{\longrightarrow} 0.$$

Since we already know, by Theorem 5.5, that $\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}[R_{A_N}(z)]\right] = S_{\mu_W}(z)$, this gives the assertion. □

Remark 6.13. Note that our estimate $\mathbb{V}\mathrm{ar}\left[...\right] \sim 1/N$ is not strong enough to get almost sure convergence; one can, however, improve our arguments to get $\mathbb{V}\mathrm{ar}\left[...\right] \sim 1/N^2$, which implies then also almost sure convergence.

6.5 Logarithmic Sobolev inequalities

One actually has typcially even exponential convergence in N. Such stronger concentration estimates rely usually on so called logarithmic Sobolev inequalities

Definition 6.14. A probability measure on $\mathbb{R}^n$ satisfies a *logarithmic Sobolev inequality (LSI)* with constant $c > 0$, if for all nice f:

$$\mathrm{Ent}_\mu(f^2) \leq 2c \int_{\mathbb{R}^n} \|\nabla f\|_2^2 d\mu,$$

where

$$\mathrm{Ent}_\mu(f) := \int_{\mathbb{R}^n} f \log f d\mu - \int_{\mathbb{R}^n} f d\mu \cdot \log \int_{\mathbb{R}^n} f d\mu$$

is an entropy like quantity.

Remark 6.15.
(1) As for Poincaré inequalities, logarithmic Sobolev inqualities are stable under tensorization and Gaussian measures satisfy LSI.
(2) From a logarithmic Sobolev inequality one can then derive a concentration inequality for our random matrices of the form

$$P_N\{A_N \mid |\operatorname{tr}[R_{A_N}(z)] - \mathbb{E}\left[\operatorname{tr}[R_{A_N}(z)]\right]| \geq \varepsilon\} \leq const \cdot \exp\left(-\frac{N^2\varepsilon^2}{2} \cdot (\operatorname{Im} z)^4\right).$$

7 Analytic Description of the Eigenvalue Distribution of Gaussian Random Matrices

In Exercise 7 we showed that the joint distribution of the entries $a_{ij} = x_{ij} + \sqrt{-1}y_{ij}$ of a GUE $A = (a_{ij})_{i,j=1}^N$ has density

$$c \cdot \exp\left(-\frac{N}{2} \operatorname{Tr} A^2\right) \mathrm{d}A.$$

This clearly shows the invariance of the distribution under unitary transformations: Let U be a unitary $N \times N$ matrix and let $B = U^* A U = (b_{ij})_{i,j=1}^N$. Then we have $\operatorname{Tr} B^2 = \operatorname{Tr} A^2$ and the volume element is invariant under unitary transformations, $\mathrm{d}B = \mathrm{d}A$. Therefore, for the joint distributions of entries of A and of B, respectively, we have

$$c \cdot \exp\left(-\frac{N}{2} \operatorname{Tr} B^2\right) \mathrm{d}B = c \cdot \exp\left(-\frac{N}{2} \operatorname{Tr} A^2\right) \mathrm{d}A.$$

Thus the joint distribution of entries of a GUE is invariant under unitary transformations, which explains the name **G**aussian **U**nitary **E**nsemble. What we are interested in, however, are not the entries but the eigenvalues of our matrices. Thus we should transform this density from entries to eigenvalues. Instead of GUE, we will mainly consider the real case, i.e., GOE.

7.1 Joint eigenvalue distribution for GOE and GUE

Let us recall the definition of GOE, see also Definition 5.1.

Definition 7.1. A *Gaussian orthogonal random matrix* (GOE) $A = (x_{ij})_{i,j=1}^N$ is given by real-valued entries x_{ij} with $x_{ij} = x_{ji}$ for all $i, j = 1, \ldots, N$ and joint distribution

$$c_N \exp\left(-\frac{N}{4} \operatorname{Tr} A^2\right) \prod_{i \geq j} \mathrm{d}x_{ij}.$$

With GOE(N) we denote the GOE of size $N \times N$.

Remark 7.2. (1) This is clearly invariant under orthogonal transformation of the entries.

(2) This is equivalent to independent real Gaussian random variables. Note, however, that the variance for the diagonal entries has to be chosen differently from the off-diagonals; see Remark 5.2. Let us check this for $N = 2$ with

$$A = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}. \end{pmatrix}$$

Then

$$\begin{aligned} \exp\left(-\frac{N}{4}\operatorname{Tr}\begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}. \end{pmatrix}^2\right) &= \exp\left(-\frac{N}{4}\left(x_{11}^2 + 2x_{12}^2 + x_{22}^2\right)\right) \\ &= \exp\left(-\frac{N}{4}x_{11}^2\right)\exp\left(-\frac{N}{2}x_{12}^2\right)\exp\left(-\frac{N}{4}x_{22}^2\right); \end{aligned}$$

those give the density of a Gaussian of variance $1/N$ for x_{11} and x_{22} and of variance $1/N$ for x_{12}.

(3) From this one can easily determine the normalization constant c_N (as a function of N).

Since we are usually interested in functions of the eigenvalues, we will now transform this density to eigenvalues.

Example 7.3. As a warmup, let us consider the GOE(2) case,

$$A = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \qquad \text{with density} \qquad p(A) = c_2 \exp\left(-\frac{N}{4}\operatorname{Tr} A^2\right).$$

We parametrize A by its eigenvalues λ_1 and λ_2 and an angle θ by diagonalization $A = O^T D O$, where

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \qquad \text{and} \qquad O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix};$$

explicityly

$$\begin{aligned} x_{11} &= \lambda_1 \cos^2\theta + \lambda_2 \sin^2\theta, \\ x_{12} &= (\lambda_1 - \lambda_2)\cos\theta\sin\theta, \\ x_{22} &= \lambda_1 \sin^2\theta + \lambda_2 \cos^2\theta. \end{aligned}$$

Note that O and D are not uniquely determined by A. In particular, if $\lambda_1 = \lambda_2$ then any orthogonal O works. However, this case has probability zero and thus can be ignored (see Remark 7.4). If $\lambda_1 \neq \lambda_2$, then we can choose $\lambda_1 < \lambda_2$; O contains then the normalized eigenvectors for λ_1 and λ_2. Those are unique up to a sign, which can be fixed by requiring that $\cos\theta \geq 0$. Hence θ is not running from $-\pi$ to π, but instead it can be restricted to $[-\pi/2, \pi/2]$. We will now transform

$$p(x_{11}, x_{22}, x_{12})\,\mathrm{d}x_{11}\,\mathrm{d}x_{22}\,\mathrm{d}x_{12} \to q(\lambda_1, \lambda_2, \theta)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta$$

by the change of variable formula $q = p\,|\det \mathrm{D}F|$, where $\mathrm{D}F$ is the Jacobian of

$$F\colon (x_{11}, x_{22}, x_{12}) \mapsto (\lambda_1, \lambda_2, \theta).$$

We calculate

$$\det \mathrm{D}F = \det \begin{pmatrix} \cos^2\theta & \sin^2\theta & -2(\lambda_1 - \lambda_2)\sin\theta\cos\theta \\ \cos\theta\sin\theta & -\cos\theta\sin\theta & (\lambda_1 - \lambda_2)\left(-\sin^2\theta + \cos^2\theta\right) \\ \sin^2\theta & \cos^2\theta & 2(\lambda_1 - \lambda_2)\sin\theta\cos\theta \end{pmatrix} = -(\lambda_1 - \lambda_2),$$

and hence $|\det \mathrm{D}F| = |\lambda_1 - \lambda_2|$. Thus,

$$q(\lambda_1, \lambda_2, \theta) = c_2 e^{-\frac{N}{4}(\mathrm{Tr}(A^2))}\,|\lambda_1 - \lambda_2| = c_2 e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,.$$

Note that q is independent of θ, i.e., we have a uniform distribution in θ. Consider a function $f = f(\lambda_1, \lambda_2)$ of the eigenvalues. Then

$$\begin{aligned} \mathbb{E}\left[f(\lambda_1, \lambda_2)\right] &= \int\int\int q(\lambda_1, \lambda_2, \theta) f(\lambda_1, \lambda_2)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int\int_{\lambda_1 < \lambda_2} f(\lambda_1, \lambda_2) c_2 e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta \\ &= \pi c_2 \int\int_{\lambda_1 < \lambda_2} f(\lambda_1, \lambda_2) e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2. \end{aligned}$$

Thus, the density for the joint distribution of the eigenvalues on $\{(\lambda_1, \lambda_2);\ \lambda_1 < \lambda_2\}$ is given by

$$\tilde{c}_2 \cdot e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|$$

with $\tilde{c}_2 = \pi c_2$.

Remark 7.4. Let us check that the probability of $\lambda_1 = \lambda_2$ is zero.

λ_1, λ_2 are the solutions of the characteristic equation

$$\begin{aligned} 0 = \det(\lambda I - A) &= (\lambda - x_{11})(\lambda - x_{22}) - x_{12}^2 \\ &= \lambda^2 - (x_{11} + x_{22})\lambda + (x_{11}x_{22} - x_{12}^2) \\ &= \lambda^2 - b\lambda + c. \end{aligned}$$

Then there is only one solution if and only if the discriminant $d = b^2 - 4ac$ is zero. However,

$$\{(x_{11}, x_{22}, x_{12});\ d(x_{11}, x_{22}, x_{12}) = 0\}$$

is a two-dimensional surface in $\mathbb{R}^3$, i.e., its Lebesgue measure is zero.

Now we consider general GOE(N).

Theorem 7.5. *The joint distribution of the eigenvalues of a* GOE(N) *is given by a density*

$$\tilde{c}_N e^{-\frac{N}{4}(\lambda_1^2 + \cdots + \lambda_N^2)} \prod_{k<l} (\lambda_l - \lambda_k)$$

restricted on $\lambda_1 < \cdots < \lambda_N$.

Proof. In terms of the entries of the GOE matrix A we have density

$$p\left(x_{kl} \,|\, k \geq l\right) = c_N e^{-\frac{N}{4} \operatorname{Tr} A^2},$$

where $A = (x_{kl})_{k,l=1}^N$ with x_{kl} real and $x_{kl} = x_{lk}$ for all l, k. Again we diagonalize $A = O^T D O$ with O orthogonal and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ with $\lambda_1 \leq \cdots \leq \lambda_N$. As before, degenerated eigenvalues have probability zero, hence this case can be neglected and we assume $\lambda_1 < \cdots < \lambda_N$. We parametrize O via $O = e^{-H}$ by a *skew-symmetric* matrix H, that is, $H^T = -H$, i.e., $H = (h_{ij})_{i,j=1}^N$ with $h_{ij} \in \mathbb{R}$ and $h_{ij} = -h_{ji}$ for all i, j. In particular, $h_{ii} = 0$ for all i. We have

$$O^T = \left(e^{-H}\right)^T = e^{-H^T} = e^H$$

and thus O is indeed orthogonal:

$$O^T O = e^H e^{-H} = e^{H-H} = e^0 = I = O O^T.$$

$O = e^{-H}$ is actually a parametrization of the Lie group $\mathrm{SO}(N)$ by the Lie algebra $\mathrm{so}(N)$ of skew-symmetric matrices.

Note that our parametrization $A = e^H D e^{-H}$ has the right number of parameters. For A we have the variables $\{x_{ij};\ j \leq i\}$ and for $e^H D e^{-H}$ we have the N eigenvalues

$\{\lambda_1, \dots, \lambda_N\}$ and the $\frac{1}{2}(N^2 - N)$ many parameters $\{h_{ij};\, i > j\}$. In both cases we have $\frac{1}{2}N(N+1)$ many variables. This parametrization is locally bijective; so we need to compute the Jacobian of the map $S\colon A \mapsto e^H D e^{-H}$. We have

$$\begin{aligned} \mathrm{d}A &= (\,\mathrm{d}e^H) D e^{-H} + e^H (\,\mathrm{d}D) e^{-H} + e^H D (\,\mathrm{d}e^{-H}) \\ &= e^H \left[e^{-H} (\,\mathrm{d}e^H) D + \,\mathrm{d}D - D (\,\mathrm{d}e^{-H}) e^H \right] e^{-H}. \end{aligned}$$

This transports the calculation of the derivative at any arbitrary point e^H to the identity element $I = e^0$ in the Lie group. Since the Jacobian is preserved under this transformation, it suffices to calculate the Jacobian at $H = 0$, i.e., for $e^H = I$ and $\mathrm{d}e^H = \mathrm{d}H$. Then

$$\mathrm{d}A = \,\mathrm{d}H \cdot D - D \cdot \,\mathrm{d}H + \,\mathrm{d}D,$$

i.e.,

$$\mathrm{d}x_{ij} = \,\mathrm{d}h_{ij}\lambda_j - \lambda_i \,\mathrm{d}h_{ij} + \,\mathrm{d}\lambda_i \delta_{ij}$$

This means that we have

$$\frac{\partial x_{ij}}{\partial \lambda_k} = \delta_{ij}\delta_{ik} \qquad \text{and} \qquad \frac{\partial x_{ij}}{\partial h_{kl}} = \delta_{ik}\delta_{jl}(\lambda_l - \lambda_k).$$

Hence the Jacobian is given by

$$J = \det DS = \prod_{k<l} (\lambda_l - \lambda_k).$$

Thus,

$$\begin{aligned} q(\lambda_1, \dots, \lambda_N, h_{kl}) = p(x_{ij} \,|\, i \geq j) J &= c_N e^{-\frac{N}{4} \operatorname{Tr} A^2} \prod_{k<l} (\lambda_l - \lambda_k) \\ &= c_N e^{-\frac{N}{4}(\lambda_1^2 + \dots + \lambda_N^2)} \prod_{k<l} (\lambda_l - \lambda_k). \end{aligned}$$

This is independent of the "angles" h_{kl}, so integrating over those variables just changes the constant c_N into another constant $\tilde{c}_N$. □

In a similar way, the complex case can be treated; see Exercise 19. One gets the following.

Theorem 7.6. *The joint distribution of the eigenvalues of a* GUE(N) *is given by a density*

$$\hat{c}_N e^{-\frac{N}{2}(\lambda_1^2 + \dots + \lambda_N^2)} \prod_{k<l} (\lambda_l - \lambda_k)^2$$

restricted on $\lambda_1 < \dots < \lambda_N$.

7.2 Rewriting the Vandermonde

Definition 7.7. The function

$$\Delta(\lambda_1, \dots, \lambda_N) = \prod_{\substack{k,l=1 \\ k<l}}^{N} (\lambda_l - \lambda_k)$$

is called the *Vandermonde determinant.*

Proposition 7.8. *For $\lambda_1, \dots, \lambda_N \in \mathbb{R}$ we have that*

$$\Delta(\lambda_1, \dots, \lambda_N) = \det \left(\lambda_j^{i-1}\right)_{i,j=1}^{N} = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_N \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \dots & \lambda_N^{N-1} \end{pmatrix}.$$

Proof. $\det(\lambda_j^{i-1})$ is a polynomial in $\lambda_1, \dots, \lambda_N$. If $\lambda_l = \lambda_k$ for some $l, k \in \{1, \dots, N\}$ then $\det(\lambda_j^{i-1}) = 0$. Thus $\det(\lambda_j^{i-1})$ contains a factor $\lambda_l - \lambda_k$ for each $k < l$, hence $\Delta(\lambda_1, \dots, \lambda_N)$ divides $\det(\lambda_j^{i-1})$.

Since $\det(\lambda_j^{i-1})$ is a sum of products with one factor from each row, we have that the degree of $\det(\lambda_j^{i-1})$ is equal to

$$0 + 1 + 2 + \cdots + (N-1) = \frac{1}{2}N(N-1),$$

which is the same as the degree of $\Delta(\lambda_1, \dots, \lambda_N)$. This shows that

$$\Delta(\lambda_1, \dots, \lambda_N) = c \cdot \det(\lambda_j^{i-1}) \qquad \text{for some } c \in \mathbb{R}.$$

By comparing the coefficient of $1 \cdot \lambda_2 \cdot \lambda_3^2 \cdots \lambda_N^{N-1}$ on both sides one can check that $c = 1$.

□

The advantage of being able to write our density in terms of a determinant comes from the following observation: In $\det(\lambda_j^{i-1})$ we can add arbitrary linear combinations of smaller rows to the k-th row without changing the value of the determinant, i.e., we can replace λ^k by any arbitrary monic polynomial $p_k(\lambda) = \lambda^k + \alpha_{k-1}\lambda^{k-1} + \cdots + \alpha_1\lambda + \alpha_0$ of degree k. Hence we have the following statement.

Proposition 7.9. *Let $p_0, \dots, p_{N-1}$ be monic polynomials with $\deg p_k = k$. Then we have*

$$\det\left(p_{i-1}(\lambda_j)\right)_{i,j=1}^{N} = \Delta(\lambda_1, \dots, \lambda_N) = \prod_{\substack{k,l=1 \\ k<l}}^{N} (\lambda_l - \lambda_k).$$

7.3 Rewriting the GUE density in terms of Hermite kernels

In the following, we will make a special choice for the p_k. We will choose them as the Hermite polynomials, which are orthogonal with respect to the Gaussian distribution $\frac{1}{c}e^{-\frac{1}{2}\lambda^2}$.

Definition 7.10. The *Hermite polynomials* H_n are defined by the following requirements.

(i) H_n is a monic polynomial of degree n.

(ii) For all $n, m \geq 0$:

$$\int_{\mathbb{R}} H_n(x)\overline{H_m(x)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \delta_{nm}n!$$

Remark 7.11. (1) One can get the $H_n(x)$ from the monomials $1, x, x^2, \ldots$ via Gram-Schmidt orthogonalization as follows.

- We define an inner product on the polynomials by

$$\langle f, g\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(x)\overline{g(x)}e^{-\frac{1}{2}x^2}\,\mathrm{d}x.$$

- We put $H_0(x) = 1$. This is monic of degree 0 with

$$\langle H_0, H_0\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 1 = 0!.$$

- We put $H_1(x) = x$. This is monic of degree 1 with

$$\langle H_1, H_0\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} xe^{-\frac{1}{2}x^2}\,\mathrm{d}x = 0$$

  and

$$\langle H_1, H_1\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^2e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 1 = 1!.$$

- For H_2, note that

$$\langle x^2, H_1\rangle = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^3e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 0$$

and

$$\langle x^2, H_0\rangle = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 1.$$

Hence we set $H_2(x) := x^2 - H_0(x) = x^2 - 1$. Then we have

$$\langle H_2, H_0\rangle = 0 = \langle H_2, H_1\rangle$$

and

$$\begin{aligned}\langle H_2, H_2\rangle &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (x^2-1)^2 e^{-\frac{1}{2}x^2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (x^4 - 2x^2 + 1) e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 3 - 2 + 1 = 2!\end{aligned}$$

- Continue in this way.

Note that the H_n are uniquely determined by the requirements that H_n is monic and that $\langle H_m, H_n\rangle = 0$ for all $m \neq n$. That we have $\langle H_n, H_n\rangle = n!$, is then a statement which has to be proved.

(2) The Hermite polynomials satisfy many explicit relations; important is the three-term recurrence relation

$$xH_n(x) = H_{n+1}(x) + nH_{n-1}(x)$$

for all $n \geq 1$; see Exercise 22.

(3) The first few H_n are

$$\begin{aligned}H_0(x) &= 1,\\ H_1(x) &= x,\\ H_2(x) &= x^2 - 1,\\ H_3(x) &= x^3 - 3x,\\ H_4(x) &= x^4 - 6x^2 + 3.\end{aligned}$$

(4) By Proposition 7.9, we can now use the H_n for writing our Vandermonde determinant as

$$\Delta(\lambda_1, \ldots, \lambda_N) = \det\left(H_{i-1}(\lambda_j)\right)_{i,j=1}^N.$$

We want to use this for our GUE(N) density

$$\begin{aligned} q(\lambda_1, \dots, \lambda_N) &= \hat{c}_N e^{-\frac{N}{2}(\lambda_1^2 + \dots + \lambda_N^2)} \Delta(\lambda_1, \dots, \lambda_N)^2 \\ &= \hat{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \Delta\left(\frac{\mu_1}{\sqrt{N}}, \dots, \frac{\mu_N}{\sqrt{N}}\right)^2 . \\ &= \hat{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \Delta(\mu_1, \dots, \mu_N)^2 \left(\frac{1}{\sqrt{N}}\right)^{N(N-1)} , \end{aligned}$$

where the $\mu_i = \sqrt{N}\lambda_i$ are the eigenvalues of the "unnormalized" GUE matrix $\sqrt{N}A_N$. It will be easier to deal with those. We now will also go over from ordered eigenvalues $\lambda_1 < \lambda_2 < \dots < \lambda_N$ to unordered eigenvalues $(\mu_1, \dots, \mu_N) \in \mathbb{R}^N$. Since in the latter case each ordered tuple shows up $N!$ times, this gives an additional factor $N!$ in our density. We collect all these N-dependent factors in our constant $\tilde{c}_N$. So we now have the density

$$\begin{aligned} p(\mu_1, \dots, \mu_N) &= \tilde{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \Delta(\mu_1, \dots, \mu_N)^2 \\ &= \tilde{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \left[\det\left(H_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2 \\ &= \tilde{c}_N \left[\det\left(e^{-\frac{1}{4}\mu_j^2} H_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2 . \end{aligned}$$

Definition 7.12. The *Hermite functions* Ψ_n are defined by

$$\Psi_n(x) = (2\pi)^{-\frac{1}{4}} (n!)^{-\frac{1}{2}} e^{-\frac{1}{4}x^2} H_n(x).$$

Remark 7.13. (1) We have

$$\int_{\mathbb{R}} \Psi_n(x)\Psi_m(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n!m!}} \int_{\mathbb{R}} e^{-\frac{1}{4}x^2} H_n(x) H_m(x)\,\mathrm{d}x = \delta_{nm},$$

i.e., the Ψ_n are orthonormal with respect to the Lebesgue measure. Actually, they form an orthonormal Hilbert space basis of $L^2(\mathbb{R})$.

(2) Now we can continue the calculation

$$p(\mu_1, \dots, \mu_N) = c_N \left[\det\left(\Psi_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2$$

with a new constant c_N. Denote $V_{ij} = \Psi_{i-1}(\mu_j)$. Then we have

$$(\det V)^2 = \det V^T \det V = \det(V^T V)$$

such that

$$(V^T V)_{ij} = \sum_{k=1}^N V_{ki} V_{kj} = \sum_{k=1}^N \Psi_{k-1}(\mu_i)\Psi_{k-1}(\mu_j).$$

Definition 7.14. The N-th *Hermite kernel* K_N is defined by

$$K_N(x,y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y).$$

Collecting all our notations and calculations we have thus proved the following.

Theorem 7.15. *The unordered joint eigenvalue distribution of an unnormalized* GUE(N) *is given by the density*

$$p(\mu_1,\dots,\mu_N) = c_N \det\left(K_N(\mu_i,\mu_j)\right)_{i,j=1}^N.$$

Proposition 7.16. K_N *is a* reproducing kernel*, i.e.,*

$$\int_{\mathbb{R}} K_N(x,u)K_N(u,y)\,\mathrm{d}u = K_N(x,y).$$

Proof. We calculate

$$\begin{aligned}
\int_{\mathbb{R}} K_N(x,u)K_N(u,y)\,\mathrm{d}u &= \int_{\mathbb{R}} \left(\sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(u)\right)\left(\sum_{l=0}^{N-1} \Psi_l(u)\Psi_l(y)\right)\mathrm{d}u \\
&= \sum_{k,l=0}^{N-1} \Psi_k(x)\Psi_l(y) \int_{\mathbb{R}} \Psi_k(u)\Psi_l(u)\,\mathrm{d}u \\
&= \sum_{k,l=0}^{N-1} \Psi_k(x)\Psi_l(y)\delta_{kl} \\
&= \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y) \\
&= K_N(x,y).
\end{aligned}$$

□

Lemma 7.17. *Let* $K\colon \mathbb{R}^2 \to \mathbb{R}$ *be a reproducing kernel, i.e.,*

$$\int_{\mathbb{R}} K(x,u)K(u,y)\,\mathrm{d}u = K(x,y).$$

Put $d = \int_{\mathbb{R}} K(x,x)\,\mathrm{d}x$*. Then, for all* $n \geq 2$*,*

$$\int_{\mathbb{R}} \det\left(K(\mu_i,\mu_j)\right)_{i,j=1}^n \mathrm{d}\mu_n = (d-n+1)\cdot \det\left(K(\mu_i,\mu_j)\right)_{i,j=1}^{n-1}.$$

We assume that all those integrals make sense, as it is the case for our Hermite kernels.

Proof. Consider the case $n = 2$. Then

$$\begin{aligned}
&\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} \mathrm{d}\mu_2 \\
&\qquad = K(\mu_1,\mu_1) \int_{\mathbb{R}} K(\mu_2,\mu_2)\,\mathrm{d}\mu_2 - \int_{\mathbb{R}} K(\mu_1,\mu_2)K(\mu_2,\mu_1)\,\mathrm{d}\mu_2 \\
&\qquad = (d-1)K(\mu_1,\mu_1) \\
&\qquad = (d-1)K(\mu_1,\mu_1)\det\left(K(\mu_1,\mu_1)\right).
\end{aligned}$$

For $n = 3$,

$$\begin{aligned}
&\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) & K(\mu_1,\mu_3) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) & K(\mu_2,\mu_3) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) & K(\mu_3,\mu_3) \end{pmatrix} \\
&= \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_1,\mu_3) - \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_2,\mu_3) \\
&\qquad + \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} K(\mu_3,\mu_3),
\end{aligned}$$

with

$$\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} K(\mu_3,\mu_3)\,\mathrm{d}\mu_3 = \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} \cdot d,$$

and

$$\begin{aligned}
&-\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_2,\mu_3)\,\mathrm{d}\mu_3 \\
&\qquad = -\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_3)K(\mu_3,\mu_1) & K(\mu_2,\mu_3)K(\mu_3,\mu_2) \end{pmatrix} \mathrm{d}\mu_3 \\
&\qquad = -\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix},
\end{aligned}$$

and

$$\begin{aligned}
\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_1,\mu_3)\,\mathrm{d}\mu_3 \\
&= \int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_1,\mu_3)K(\mu_3,\mu_1) & K(\mu_1,\mu_3)K(\mu_3,\mu_2) \end{pmatrix} \mathrm{d}\mu_3 \\
&= \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \end{pmatrix} \\
&= -\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix}.
\end{aligned}$$

Putting all terms together gives

$$\int_{\mathbb{R}} \det(K(\mu_i,\mu_j))_{i,j=1}^{3}\,\mathrm{d}\mu_3 = (d-2)\det(K(\mu_i,\mu_j))_{i,j=1}^{2}.$$

The general case works in the same way. □

Iteration of Lemma 7.17 gives then the following.

Corollary 7.18. *Under the assumptions of Lemma 7.17 we have*

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K(\mu_i,\mu_j)\right)_{i,j=1}^{n}\,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n = (d-n+1)(d-n+2)\cdots(d-1)d.$$

Remark 7.19. We want to apply this to the Hermite kernel $K = K_N$. In this case we have

$$\begin{aligned}
d &= \int_{\mathbb{R}} K_N(x,x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(x)\,\mathrm{d}x \\
&= \sum_{k=0}^{N-1} \int_{\mathbb{R}} \Psi_k(x)\Psi_k(x)\,\mathrm{d}x \\
&= N,
\end{aligned}$$

and thus, since now $d = N = n$,

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K_N(\mu_i,\mu_j)\right)_{i,j=1}^{N}\,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n = N!.$$

This now allows us to determine the constant c_N in the density $p(\mu_1, \dots, \mu_n)$ in Theorem 7.15. Since p is a probability density on $\mathbb{R}^N$, we have

$$\begin{aligned} 1 &= \int\limits_{\mathbb{R}^N} p(\mu_1, \dots, \mu_n) \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_N \\ &= c_N \int\limits_{\mathbb{R}} \cdots \int\limits_{\mathbb{R}} \det \left(K_N(\mu_i, \mu_j) \right)_{i,j=1}^N \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_N \\ &= c_N N!, \end{aligned}$$

and thus $c_N = \frac{1}{N!}$.

Theorem 7.20. *The unordered joint eigenvalue distribution of an unnormalized* GUE(N) *is given by a density*

$$p(\mu_1, \dots, \mu_N) = \frac{1}{N!} \det \left(K_N(\mu_i, \mu_j) \right)_{i,j=1}^N ,$$

where K_N is the Hermite kernel

$$K_N(x, y) = \sum_{k=0}^{N-1} \Psi_k(x) \Psi_k(x).$$

Theorem 7.21. *The averaged eigenvalue density of an unnormalized* GUE(N) *is given by*

$$p_N(\mu) = \frac{1}{N} K_N(\mu, \mu) = \frac{1}{N} \sum_{k=0}^{N-1} \Psi_k(\mu)^2 = \frac{1}{\sqrt{2\pi}} \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{k!} H_k(\mu)^2 e^{-\frac{\mu^2}{2}}.$$

Proof. Note that $p(\mu_1, \dots, \mu_N)$ is the probability density to have N eigenvalues at the positions $\mu_1, \dots, \mu_N$. If we are integrating out $N-1$ variables we are left with the probability for one eigenvalue (without caring about the others). With the notation $\mu_N = \mu$ we get

$$\begin{aligned} p_N(\mu) &= \int\limits_{\mathbb{R}^{N-1}} p(\mu_1, \dots, \mu_{N-1}, \mu) \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{N-1} \\ &= \frac{1}{N!} \int\limits_{\mathbb{R}^{N-1}} \det \left(K_N(\mu_i, \mu_j) \right)_{i,j=1}^N \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{N-1} \\ &= \frac{1}{N!} (N-1)! \det(K_N(\mu, \mu)) \\ &= \frac{1}{N} K_N(\mu, \mu). \end{aligned}$$

□

8 Determinantal Processes and Non-Crossing Paths: Karlin–McGregor and Gessel–Viennot

Our probability distributions for the eigenvalues of GUE have a determinantal structure, i.e., are of the form

$$p(\mu_1, \dots, \mu_n) = \frac{1}{N!} \det \left(K_N(\mu_i, \mu_j) \right)_{i,j=1}^N .$$

They describe N eigenvalues which repel each other (via the factor $(\mu_i - \mu_j)^2$). If we consider corresponding processes, then the paths of the eigenvalues should not cross; for this see also Section 8.3. There is a quite general relation between determinants as above and non-crossing paths. This appeared in fundamental papers in different contexts:

- in a paper by Karlin and McGregor, 1958, in the context of Markov chains and Brownian motion
- in a paper of Lindström, 1973, in the context of matroids
- in a paper of Gessel and Viennot, 1985, in combinatorics

8.1 Stochastic version à la Karlin–McGregor

Consider a random walk on the integers $\mathbb{Z}$:

- Y_k: position at time k
- $\mathbb{Z}$: possible positions
- Transition probability (to the two neighbors) might depend on position:

$$i - 1 \xleftarrow{q_i} i \xrightarrow{p_i} i + 1, \qquad q_i + p_1 = 1$$

We now consider n copies of such a random walk, which at time $k = 0$ start at different positions x_i. We are interested in the probability that the paths don't

cross. Let x_i be such that all distances are even, i.e., if two paths cross they have to meet.

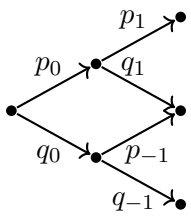
Theorem 8.1 (Karlin–McGregor)**.** *Consider n copies of Y_k, i.e., $(Y_k^{(1)}, \dots, Y_k^{(n)})$ with $Y_0^{(i)} = x_i$, where $x_1 > x_2 > \cdots > x_n$. Consider now $t \in \mathbb{N}$ and $y_1 > y_2 > \cdots > y_n$. Denote by*

$$P_t(x_i, y_j) = \mathbb{P}\left[Y_t = y_j \,|\, Y_0 = x_i\right]$$

the probability of one random walk to get from x_i to y_j in t steps. Then we have

$$\mathbb{P}\left[Y_t^{(i)} = y_i \text{ for all } i, Y_s^{(1)} > Y_s^{(2)} > \cdots > Y_s^{(n)} \text{ for all } 0 \le s \le t\right] = \det\left(P_t(x_i, y_j)\right)_{i,j=1}^n.$$

Example 8.2. For one symmetric random walk Y_t we have the following probabilities to go in two steps from 0 to -2,0,2:



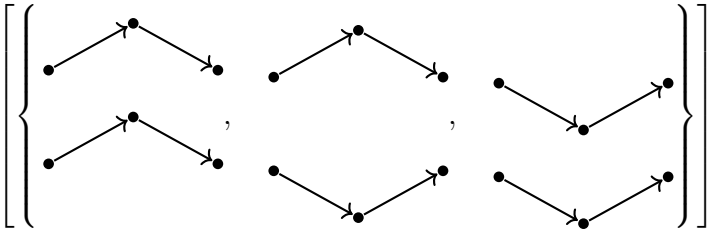
$p_0 p_1 = \frac{1}{4}$

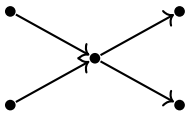
$p_0 q_1 + q_0 p_{-1} = \frac{1}{2}$

$q_0 q_{-1} = \frac{1}{4}$

Now consider two such symmetric random walks and set $x_1 = 2 = y_1$, $x_2 = 0 = y_2$. Then

$$\mathbb{P}\left[Y_2^{(1)} = 2 = Y_0^{(1)}, Y_2^{(2)} = 0 = Y_0^{(2)}, Y_1^{(1)} > Y_1^{(2)}\right]$$

$= \mathbb{P}\Bigg[$  $\Bigg] = \frac{3}{16}.$

Note that  is not allowed.

Theorem 8.1 says that we also obtain this probability from the transition probabilities of one random walk as

$$\det\begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix} = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}.$$

Proof of Theorem 8.1. Let Ω_{ij} be the set of all possible paths in t steps from x_i to y_j. Denote by $\mathbb{P}[\pi]$ the probability for such a path $\pi \in \Omega_{ij}$. Then we have

$$P_t(x_i, y_j) = \sum_{\pi \in \Omega_{ij}} \mathbb{P}[\pi]$$

and we have to consider the determinant

$$\det\left(P_t(x_i, y_j)\right)_{i,j=1}^n = \det\left(\sum_{\pi \in \Omega_{ij}} \mathbb{P}[\pi]\right)_{i,j=1}^n .$$

Let us consider the case $n = 2$:

$$\det\begin{pmatrix} \sum_{\pi \in \Omega_{11}} \mathbb{P}[\pi] & \sum_{\pi \in \Omega_{12}} \mathbb{P}[\pi] \\ \sum_{\pi \in \Omega_{21}} \mathbb{P}[\pi] & \sum_{\pi \in \Omega_{22}} \mathbb{P}[\pi] \end{pmatrix} = \sum_{\pi \in \Omega_{11}} \mathbb{P}[\pi] \cdot \sum_{\sigma \in \Omega_{22}} \mathbb{P}[\sigma] - \sum_{\pi \in \Omega_{12}} \mathbb{P}[\pi] \cdot \sum_{\sigma \in \Omega_{21}} \mathbb{P}[\sigma]$$

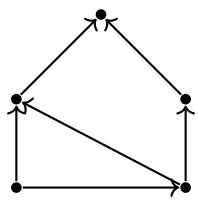
Here, the first term counts all pairs of paths $x_1 \to y_1$ and $x_2 \to y_2$; hence non-crossing ones, but also crossing ones. However, such a crossing pair of paths is, via the "reflection principle" (where we exchange the parts of the two paths after their first crossing), in bijection with a pair of paths from $x_1 \to y_2$ and $x_2 \to y_1$; this bijection also preserves the probabilities.

Those paths, $x_1 \to y_2$ and $x_2 \to y_1$, are counted by the second term in the determinant. Hence the second term cancels out all the crossing terms in the first term, leaving only the non-crossing paths.

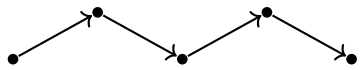
For general n it works in a similar way. □

8.2 Combinatorial version à la Gessel–Viennot

Let G be a weighted directed graph without directed cycles, e.g.



where we have weights $m_{ij} = m_e$ on each edge $i \xrightarrow{e} j$. This gives weights for directed paths

$P =$  via $\quad m(P) = \prod_{e \in P} m_e,$

and then also a weight for connecting two vertices a, b,

$$m(a,b) = \sum_{P\colon a \to b} m(P),$$

where we sum over all directed paths from a to b. Note that this is a finite sum, because we do not have directed cycles in our graph.

Definition 8.3. Consider two n-tuples of vertices $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$. A *path system* $P\colon A \to B$ is given by a permutation $\sigma \in S_n$ and paths $P_i\colon a_i \to b_{\sigma(i)}$ for $i = 1, \dots, n$. We also put $\sigma(P) = \sigma$ and $\operatorname{sgn} P = \operatorname{sgn} \sigma$. A *vertex-disjoint* path system is a path system $(P_1, \dots, P_n)$, where the paths $P_1, \dots, P_n$ do not have a common vertex.

Lemma 8.4 (Gessel–Viennot)**.** *Let G be a finite acyclic weighted directed graph and let $A = (a_1, \dots, a_n)$ and $B = (b_1, \dots, b_n)$ be two n-sets of vertices. Then we have*

$$\det\left(m(a_i, b_j)\right)_{i,j=1}^n = \sum_{\substack{P\colon A \to B \\ \textit{vertex-disjoint}}} \operatorname{sgn} \sigma(P) \prod_{i=1}^n m(P_i).$$

Proof. Similar as the proof of Theorem 8.1; the crossing paths cancel each other out in the determinant. □

This lemma can be useful in two directions. Whereas in the stochastic setting one uses mainly the determinant to count non-crossing paths, one can also count vertex-disjoint path systems to calculate determinants. The following is an example of this.

Example 8.5. Let C_n be the Catalan numbers

$$C_0 = 1, C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14, \dots$$

and consider

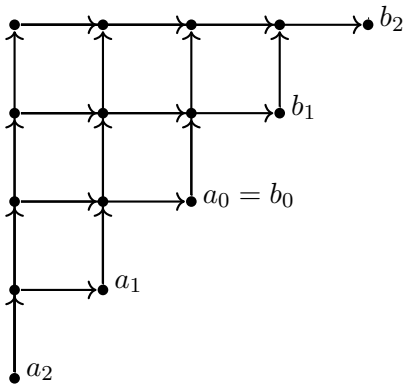
$$M_n = \begin{pmatrix} C_0 & C_1 & \cdots & C_n \\ C_1 & C_2 & \cdots & C_{n+1} \\ \vdots & & \ddots & \vdots \\ C_n & C_{n+1} & \cdots & C_{2n} \end{pmatrix}.$$

Then we have

$$\det M_0 = \det(1) = 1,$$
$$\det M_1 = \det \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = 2 - 1 = 1,$$
$$\det M_2 = \det \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 5 & 14 \end{pmatrix} = 28 + 10 + 10 - 8 - 14 - 25 = 1.$$

This is actually true for all n: $\det M_n = 1$. This is not obvious directly, but follows easily from Gessel–Viennot, if one chooses the right setting.

Let us show it for M_2. For this, consider the graph



The possible directions in the graph are up and right, and all weights are chosen as 1. Paths in this graph correspond to Dyck graphs, and thus the weights for connecting the a's with the b's are counted by Catalan numbers; e.g.,

$$m(a_0, b_0) = C_0,$$
$$m(a_0, b_1) = C_1,$$
$$m(a_0, b_2) = C_2,$$
$$m(a_2, b_2) = C_4.$$

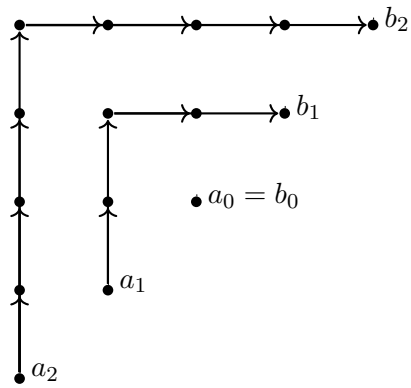
Thus

$$M_2 = \det \begin{pmatrix} m(a_0, b_0) & m(a_0, b_1) & m(a_0, b_2) \\ m(a_1, b_0) & m(a_1, b_1) & m(a_1, b_2) \\ m(a_2, b_0) & m(a_2, b_1) & m(a_2, b_2) \end{pmatrix}$$

and hence, by Gessel-Viennot,

$$\det M_2 = \det \left(m(a_i, b_j)\right)_{i,j=0}^{2} = \sum_{\substack{P\colon (a_0,a_1,a_2) \to (b_0,b_1,b_2) \\ \text{vertex-disjoint}}} 1 = 1,$$

since there is only one such vertex-disjoint system of three paths, corresponding to $\sigma = \mathrm{id}$. This is given as follows; note that the path from a_0 to b_0 is actually a path with 0 steps.

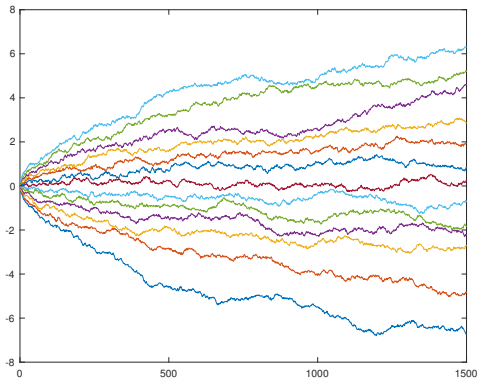


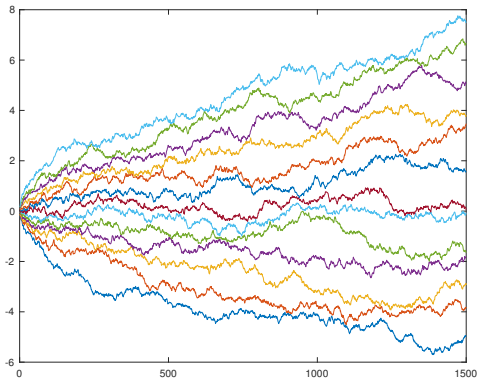
8.3 Dyson Brownian motion and non-intersecting paths

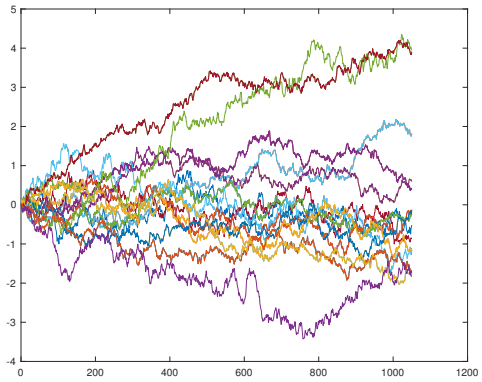
We have seen that the eigenvalues of random matrices repel each other. This becomes even more apparent when we consider process versions of our random matrices, where the eigenvalue processes yield then non-intersecting paths. Those process versions of our Gaussian ensembles are called *Dyson Brownian motions*. They are defined as $A_N(t) := (a_{ij}(t))_{i,j=1}^N$ $(t \geq 0)$, where each $a_{ij}(t)$ is a classical Brownian motion (complex or real) and they are independent, apart from the symmetry condition $a_{ij}(t) = \bar{a}_{ji}(t)$ for all $t \geq 0$ and all $i, j = 1, \ldots, N$. The eigenvalues $\lambda_1(t), \ldots, \lambda_N(t)$ of $A_N(t)$ give then N non-intersecting Brownian motions.

Here are plots for discretized random walk versions of the Dyson Brownian motion, corresponding to GOE(13), GUE(13) and, for comparision, also 13 independent

Brownian motions; see also Exercise 24. Guess which is which!



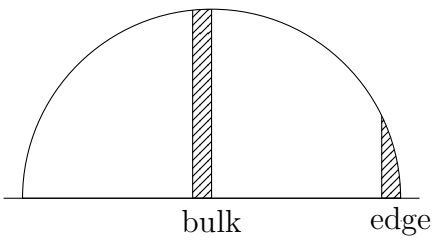




9 Statistics of the Largest Eigenvalue and Tracy–Widom Distribution

Consider GUE(N) or GOE(N). For large N, the eigenvalue distribution is close to a semircircle with density

$$p(x) = \frac{1}{2\pi}\sqrt{4 - x^2}.$$



We will now zoom to a microscopic level and try to understand the behaviour of a single eigenvalue. The behaviour in the *bulk* and at the *edge* is different. We are particularly interested in the largest eigenvalue. Note that at the moment we do not even know whether the largest eigenvalue sticks close to 2 with high probability. Wigner's semicircle law implies that it cannot go far below 2, but it does not prevent it from being very large. We will in particular see that this cannot happen.

9.1 Some heuristics on single eigenvalues

Let us first check heuristically what we expect as typical order of fluctuations of the eigenvalues. For this we assume (without justification) that the semicircle predicts the behaviour of eigenvalues down to the microscopic level.

Behaviour in the bulk: In $[\lambda, \lambda+t]$ there should be $\sim tp(\lambda)N$ eigenvalues. This is of order 1 if we choose $t \sim 1/N$. This means that eigenvalues in the bulk have

for their position an interval of size $\sim 1/N$, so this is a good guess for the order of fluctuations for an eigenvalue in the bulk.

Behaviour at the edge: In $[2-t, 2]$ there should be roughly

$$N \int_{2-t}^{2} p(x)\,\mathrm{d}x = \frac{N}{2\pi} \int_{2-t}^{2} \sqrt{(2-x)(2+x)}\,\mathrm{d}x$$

many eigenvalues. To have this of order 1, we should choose t as follows:

$$1 \approx \frac{N}{2\pi} \int_{2-t}^{2} \sqrt{(2-x)(2+x)}\,\mathrm{d}x \approx \frac{N}{\pi} \int_{2-t}^{2} \sqrt{2-x}\,\mathrm{d}x = \frac{N}{\pi}\frac{2}{3} t^{\frac{3}{2}}$$

Thus $1 \sim t^{3/2} N$, i.e., $t \sim N^{-2/3}$. Hence we expect for the largest eigenvalue an interval or fluctuations of size $N^{-2/3}$. Very optimistically, we might expect

$$\lambda_{\max} \approx 2 + N^{-2/3} X,$$

where X has N-independent distribution.

9.2 Tracy–Widom distribution

This heuristics (at least its implication) is indeed true and one has that the limit

$$F_\beta(x) := \lim_{N\to\infty} \mathbb{P}\left[N^{2/3}(\lambda_{\max} - 2) \le x\right]$$

exists. It is called the *Tracy–Widom distribution.*

Remark 9.1. (1) Note the parameter β! This corresponds to:

ensemble	β	repulsion
GOE	1	$(\lambda_i - \lambda_j)^1$
GUE	2	$(\lambda_i - \lambda_j)^2$
GSE	4	$(\lambda_i - \lambda_j)^4$

It turns out that the statistics of the largest eigenvalue is different for real, complex, quaternionic Gaussian random matrices. The behaviour on the microscopic level is more sensitive to the underlying symmetry than the macroscopic behaviour; note that we get the semicircle as the macroscopic limit for all three ensembles. (In models in physics the choice of β corresponds often to underlying physical symmetries; e.g., GOE is used to describe systems which have a time-reversal symmetry.)

(2) On the other hand, when β is fixed, there is a large universality class for the corresponding Tracy–Widom distribution. F_2 shows up as limiting fluctuations for
   (a) largest eigenvalue of GUE (Tracy, Widom, 1993),
   (b) largest eigenvalue of more general Wigner matrices (Soshnikov, 1999),
   (c) largest eigenvalue of general unitarily invariant matrix ensembles (Deift et al., 1994-2000),
   (d) length of the longest increasing subsequence of random permutations (Baik, Deift, Johansson, 1999; Okounkov, 2000),
   (e) arctic cicle for Aztec diamond (Johansson, 2005),
   (f) various growth processes like ASEP ("asymmetric single exclusion process"), TASEP ("totally asymmetric ...").

(3) There is still no uniform explanation for this universality. The feeling is that the Tracy-Widom distribution is somehow the analogue of the normal distribution for a kind of central limit theorem, where independence is replaced by some kind of dependence. But no one can make this precise at the moment.

(4) Proving Tracy–Widom for GUE is out of reach for us, but we will give some ideas. In particular, we try to derive rigorous estimates which show that our $N^{-2/3}$-heuristic is of the right order and, in particular, we will prove that the largest eigenvalue converges almost surely to 2.

9.3 Convergence of the largest eigenvalue to 2

We want to derive an estimate, in the GUE case, for the probability $\mathbb{P}[\lambda_{\max} \geq 2+\varepsilon]$, which is compatible with our heuristic that ε should be of the order $N^{-2/3}$. We will refine our moment method for this. Let A_N be our normalized GUE(N). We have for all $k \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}[\lambda_{\max} \geq 2+\varepsilon] &= \mathbb{P}\left[\lambda_{\max}^{2k} \geq (2+\varepsilon)^{2k}\right] \\
&\leq \mathbb{P}\left[\sum_{j=1}^{N} \lambda_j^{2k} \geq (2+\varepsilon)^{2k}\right] \\
&= \mathbb{P}\left[\operatorname{tr} A_N^{2k} \geq \frac{(2+\varepsilon)^{2k}}{N}\right] \\
&\leq \frac{N}{(2+\varepsilon)^{2k}} \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right].
\end{aligned}$$

In the last step we used Markov's inequality 6.3; note that we have even powers, and hence the random variable $\operatorname{tr}(A_N^{2k})$ is positive.

In Theorem 2.15 we calculated the expectation in terms of a genus expansion as

$$\mathbb{E}\left[\operatorname{tr}(A_N^{2k})\right] = \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)-k-1} = \sum_{g \geq 0} \varepsilon_g(k) N^{-2g},$$

where

$$\varepsilon_g(k) = \#\left\{\pi \in \mathcal{P}_2(2k) \mid \pi \text{ has genus } g\right\}.$$

The inequality

$$\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] \leq \frac{N}{(2+\varepsilon)^{2k}} \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right]$$

is useless if k is fixed for $N \to \infty$, because then the right hand side goes to ∞. Hence we also have to scale k with N (we will use $k \sim N^{2/3}$), but then the sub-leading terms in the genus expansion become important. Up to now we only know that $\varepsilon_0(k) = C_k$, but now we need some information on the other $\varepsilon_g(k)$. This is provided by a theorem of Harer and Zagier.

Theorem 9.2 (Harer–Zagier, 1986)**.** *Let us define b_k by*

$$\sum_{g \geq 0} \varepsilon_g(k) N^{-2g} = C_k b_k,$$

where C_k are the Catalan numbers. (Note that the b_k depend also on N, but we suppress this dependency in the notation.) Then we have the recursion formula

$$b_{k+1} = b_k + \frac{k(k+1)}{4N^2} b_{k-1}$$

for all $k \geq 2$.

We will prove this later; see Section 9.6. For now, let us just check it for small examples.

Example 9.3. From Remark 2.16 we know

$$\begin{aligned}
C_1 b_1 &= \mathbb{E}\left[\operatorname{tr} A_N^2\right] = 1, \\
C_2 b_2 &= \mathbb{E}\left[\operatorname{tr} A_N^4\right] = 2 + \frac{1}{N^2}, \\
C_3 b_3 &= \mathbb{E}\left[\operatorname{tr} A_N^6\right] = 5 + \frac{10}{N^2}, \\
C_4 b_4 &= \mathbb{E}\left[\operatorname{tr} A_N^8\right] = 14 + \frac{70}{N^2} + \frac{21}{N^4},
\end{aligned}$$

which gives

$$b_1 = 1, \qquad b_2 = 1 + \frac{1}{2N^2}, \qquad b_3 = 1 + \frac{2}{N^2}, \qquad b_4 = 1 + \frac{5}{N^2} + \frac{3}{2N^4}.$$

We now check the recursion from Theorem 9.2 for $k = 3$:

$$\begin{aligned} b_3 + \frac{k(k+1)}{4N^2} b_2 &= 1 + \frac{2}{N^2} + \frac{12}{4N^2}\left(1 + \frac{1}{2N^2}\right) \\ &= 1 + \frac{5}{N^2} + \frac{3}{2N^4} \\ &= b_4 \end{aligned}$$

Corollary 9.4. *For all $N, k \in \mathbb{N}$ we have for a* GUE(N) *matrix A_N that*

$$\mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \leq C_k \exp\left(\frac{k^3}{2N^2}\right).$$

Proof. Note that, by definition, $b_k > 0$ for all $k \in \mathbb{N}$ and hence, by Theorem 9.2, $b_{k+1} > b_k$. Thus,

$$b_{k+1} = b_k + \frac{k(k+1)}{4N^2} b_{k-1} \leq b_k \left(1 + \frac{k(k+1)}{4N^2}\right) \leq b_k \left(1 + \frac{k^2}{2N^2}\right);$$

iteration of this yields

$$\begin{aligned} b_k &\leq \left(1 + \frac{(k-1)^2}{2N^2}\right)\left(1 + \frac{(k-2)^2}{2N^2}\right) \cdots \left(1 + \frac{1^2}{2N^2}\right) \\ &\leq \left(1 + \frac{k^2}{2N^2}\right)^k \\ &\leq \exp\left(\frac{k^2}{2N^2}\right)^k && \text{since } 1 + x \leq e^x \\ &= \exp\left(\frac{k^3}{2N^2}\right) \end{aligned}$$

□

We can now continue our estimate for the largest eigenvalue.

$$\begin{aligned}
\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] &\leq \frac{N}{(2+\varepsilon)^{2k}} \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \\
&= \frac{N}{(2+\varepsilon)^{2k}} C_k b_k \\
&\leq \frac{N}{(2+\varepsilon)^{2k}} C_k \exp\left(\frac{k^3}{2N^2}\right) \\
&\leq \frac{N}{(2+\varepsilon)^{2k}} \frac{4^k}{k^{3/2}} \exp\left(\frac{k^3}{2N^2}\right).
\end{aligned}$$

For the last estimate, we used (see Exercise 26)

$$C_k \leq \frac{4^k}{\sqrt{\pi} k^{3/2}} \leq \frac{4^k}{k^{3/2}}.$$

Let us record our main estimate in the following proposition.

Proposition 9.5. *For a normalized* GUE(N) *matrix A_N we have for all $N, k \in \mathbb{N}$ and all $\varepsilon > 0$*

$$\mathbb{P}\left[\lambda_{\max}(A_N) \geq 2+\varepsilon\right] \leq \frac{N}{(2+\varepsilon)^{2k}} \frac{4^k}{k^{3/2}} \exp\left(\frac{k^3}{2N^2}\right).$$

This estimate is now strong enough to see that the largest eigenvalue has actually to converge to 2. For this, let us fix $\varepsilon > 0$ and choose k depending on N as $k_N := \lfloor N^{2/3} \rfloor$, where $\lfloor x \rfloor$ denotes the smallest integer $\geq x$. Then

$$\frac{N}{k_N^{3/2}} \xrightarrow{N\to\infty} 1 \qquad \text{and} \qquad \frac{k_N^3}{2N^2} \xrightarrow{N\to\infty} \frac{1}{2}.$$

Hence

$$\limsup_{N\to\infty} \mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] \leq \lim_{N\to\infty} \left(\frac{2}{2+\varepsilon}\right)^{2k_N} \cdot 1 \cdot e^{1/2} = 0,$$

and thus for all $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] = 0.$$

This says that the largest eigenvalue $\lambda_{\max}$ of a GUE converges in probability to 2.

Corollary 9.6. *For a normalized* GUE(N) *matrix A_N we have that its largest eigenvalue converges in probability, and also almost surely, to* 2*, i.e.,*

$$\lambda_{\max}(A_N) \xrightarrow{N\to\infty} 2 \qquad \textit{almost surely.}$$

Proof. The convergence in probability was shown above. For the strenghtening to almost sure convergence one has to use Borel–Cantelli and the fact that

$$\sum_N \left(\frac{2}{2+\varepsilon}\right)^{2k_N} < \infty.$$

See Exercise 28. □

9.4 Estimate for fluctuations

Our estimate from Proposition 9.5 gives also some information about the fluctuations of $\lambda_{\max}$ about 2, if we choose ε also depending on N. Let us use there now

$$k_N = \lfloor N^{2/3} r \rfloor \qquad \text{and} \qquad \varepsilon_N = N^{-2/3} t.$$

Then

$$\frac{N}{k_N^{3/2}} \xrightarrow{N\to\infty} \frac{1}{r^{3/2}} \qquad \text{and} \qquad \frac{k_N^3}{2N^2} \xrightarrow{N\to\infty} \frac{r^3}{2},$$

and

$$\frac{4^{k_N}}{(2+\varepsilon_N)^{2k_N}} = \left(\frac{1}{1+\frac{1}{2N^{2/3}}t}\right)^{2\lfloor N^{2/3} r\rfloor} \xrightarrow{N\to\infty} e^{-rt},$$

and thus

$$\limsup_{N\to\infty} \mathbb{P}\left[\lambda_{\max} \geq 2 + tN^{-2/3}\right] \leq \frac{1}{r^{3/2}} e^{r^3/2} e^{-rt}$$

for arbitrary $r > 0$. We optimize this now by choosing $r = \sqrt{t}$ for $t > 0$ and get

Proposition 9.7. *For a normalized* GUE(N) *matrix A_N we have for all $t > 0$*

$$\limsup_{N\to\infty} \mathbb{P}\left[\lambda_{\max}(A_N) \geq 2 + tN^{-\frac{2}{3}}\right] \leq t^{-3/4} \exp\left(-\frac{1}{2} t^{3/2}\right).$$

Although this estimate does not prove the existence of the limit on the left hand side, it turns out that the right hand side is quite sharp and captures the tail behaviour of the Tracy–Widom distribution quite well.

9.5 Non-rigorous derivation of Tracy–Widom distribution

For determining the Tracy–Widom fluctuations in the limit $N \to \infty$ one has to use the analytic description of the GUE joint density. Recall from Theorem 7.20 that the joint density of the unordered eigenvalues of an unnormalized GUE(N) is given by

$$p(\mu_1, \dots, \mu_N) = \frac{1}{N!} \det \left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^N,$$

where K_N is the Hermite kernel

$$K_N(x, y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y)$$

with the Hermite functions Ψ_k from Definition 7.12. Because K_N is a reproducing kernel, we can integrate out some of the eigenvalues and get a density of the same form. If we are integrate out all but r eigenvalues we get, by Lemma 7.17,

$$\begin{aligned}
\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p(\mu_1, \dots, \mu_N) \, \mathrm{d}\mu_{r+1} \cdots \mathrm{d}\mu_N &= \frac{1}{N!} \cdot 1 \cdot 2 \cdots (N-r) \cdot \det \left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^r \\
&= \frac{(N-r)!}{N!} \det \left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^r \\
&=: p_N(\mu_1, \dots, \mu_r).
\end{aligned}$$

Now consider

$$\begin{aligned}
&\mathbb{P}\left[\mu_{\max}^{(N)} \le t\right] = \mathbb{P}\left[\text{there is no eigenvalue in } (t, \infty)\right] \\
&= 1 - \mathbb{P}\left[\text{there is an eigenvalue in } (t, \infty)\right] \\
&= 1 - \left[N\mathbb{P}\left[\mu_1 \in (t, \infty)\right] - \binom{N}{2}\mathbb{P}\left[\mu_1, \mu_2 \in (t, \infty)\right] + \binom{N}{3}\mathbb{P}\left[\mu_1, \mu_2, \mu_3 \in (t, \infty)\right] - \cdots\right] \\
&= 1 + \sum_{r=1}^N (-1)^r \binom{N}{r} \int_t^\infty \cdots \int_t^\infty p_N(\mu_1, \dots, \mu_r) \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_r \\
&= 1 + \sum_{r=1}^N (-1)^r \frac{1}{r!} \int_t^\infty \cdots \int_t^\infty \det \left(K(\mu_i, \mu_j)\right)_{i,j=1}^r \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_r.
\end{aligned}$$

Does this have a limit for $N \to \infty$?

Note that p is the distribution for a GUE(N) without normalization, i.e., $\mu_{\max}^{(N)} \approx 2\sqrt{N}$. More precisely, we expect fluctutations

$$\mu_{\max}^{(N)} \approx \sqrt{N}\left(2 + tN^{-2/3}\right) = 2\sqrt{N} + tN^{-1/6}.$$

We put

$$\tilde{K}_N(x,y) = N^{-1/6} \cdot K_N\left(2\sqrt{N} + xN^{-1/6}, 2\sqrt{N} + yN^{-1/6}\right)$$

so that we have

$$\mathbb{P}\left[N^{2/3}\left(\frac{\mu_{\max}^{(N)} - 2}{\sqrt{N}} - 2\right) \leq t\right] = \sum_{r=0}^{N} \frac{(-1)^r}{r!} \int_t^\infty \cdots \int_t^\infty \det\left(\tilde{K}(x_i, x_j)\right)_{i,j=1}^r \mathrm{d}x_1 \cdots \mathrm{d}x_r.$$

We expect that the limit

$$F_2(t) := \lim_{N\to\infty} \mathbb{P}\left[N^{2/3}\left(\frac{\mu_{\max}^{(N)} - 2}{\sqrt{N}} - 2\right) \leq t\right]$$

exists. For this, we need the limit $\lim\limits_{N\to\infty} \tilde{K}_N(x,y)$. Recall that

$$K_N(x,y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y).$$

As this involves Ψ_k for all $k = 0, 1, \ldots, N-1$ this is not amenable to taking the limit $N \to \infty$. However, by the Christoffel–Darboux identity for the Hermite functions (see Exercise 27))

$$\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(x-y)\ (n-1)!}$$

and with

$$\Psi_k(x) = (2\pi)^{-1/4}(k!)^{-1/2}e^{-\frac{1}{4}x^2}H_k(x),$$

as defined in Definition 7.12, we can rewrite K_N in the form

$$\begin{aligned} K_N(x,y) &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{N-1} \frac{1}{k!} e^{-\frac{1}{4}\left(x^2+y^2\right)} H_k(x)H_k(y) \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}\left(x^2+y^2\right)} \frac{H_N(x)H_{N-1}(y) - H_{N-1}(x)H_N(y)}{(x-y)\ (N-1)!} \\ &= \sqrt{N} \cdot \frac{\Psi_N(x)\Psi_{N-1}(y) - \Psi_{N-1}(x)\Psi_N(y)}{x-y}. \end{aligned}$$

Note that the Ψ_N satisfy the differential equation (see Exercise 30)

$$\Psi_N'(x) = -\frac{x}{2}\Psi_N(x) + \sqrt{N}\Psi_{N-1}(x),$$

and thus

$$\begin{aligned} K_N(x,y) &= \frac{\Psi_N(x)\left[\Psi_N'(y) + \frac{y}{2}\Psi_N(y)\right] - \left[\Psi_N'(x) + \frac{x}{2}\Psi_N(x)\right]\Psi_N(y)}{x-y} \\ &= \frac{\Psi_N(x)\Psi_N'(y) - \Psi_N'(x)\Psi_N(y)}{x-y} - \frac{1}{2}\Psi_N(x)\Psi_N(y). \end{aligned}$$

Now put

$$\widetilde{\Psi}_N(x) := N^{1/12} \cdot \Psi_N\left(2\sqrt{N} + xN^{-1/6}\right),$$

thus

$$\widetilde{\Psi}_N'(x) = N^{1/12} \cdot \Psi_N'\left(2\sqrt{N} + xN^{-1/6}\right) \cdot N^{-1/6} = N^{-1/12} \cdot \Psi_N'\left(2\sqrt{N} + xN^{-1/6}\right).$$

Then

$$\tilde{K}(x,y) = \frac{\widetilde{\Psi}_N(x)\widetilde{\Psi}_N'(y) - \widetilde{\Psi}_N'(x)\widetilde{\Psi}_N(y)}{x-y} - \frac{1}{2N^{1/3}}\widetilde{\Psi}_N'(x)\widetilde{\Psi}_N'(y).$$

One can show, by a quite non-trivial steepest descent method, that $\widetilde{\Psi}_N(x)$ converges to a limit. Let us call this limit the *Airy function*

$$\mathrm{Ai}(x) = \lim_{N\to\infty} \widetilde{\Psi}_N(x).$$

The convergence is actually so strong that also

$$\mathrm{Ai}'(x) = \lim_{N\to\infty} \widetilde{\Psi}_N'(x),$$

and hence

$$\lim_{N\to\infty} \tilde{K}(x,y) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y} =: \mathrm{A}(x,y).$$

A is called the *Airy kernel*.

Let us try, again non-rigorously, to characterize this limit function Ai. For the Hermite functions we have (see Exercise 30)

$$\Psi_N''(x) + \left(N + \frac{1}{2} - \frac{x^2}{4}\right)\Psi_N(x) = 0.$$

For the $\widetilde{\Psi}_N$ we have

$$\widetilde{\Psi}'_N(x) = N^{-1/12} \cdot \Psi'_N\left(2\sqrt{N} + xN^{-1/6}\right)$$

and

$$\widetilde{\Psi}''_N(x) = N^{-1/4} \cdot \Psi''_N\left(2\sqrt{N} + xN^{-1/6}\right).$$

Thus,

$$\begin{aligned}
\widetilde{\Psi}''_N(x) &= -N^{-1/4}\left[N + \frac{1}{2} - \frac{\left(2\sqrt{N} + xN^{-1/6}\right)^2}{4}\right]\Psi_N\left(2\sqrt{N} + xN^{-1/6}\right) \\
&= -N^{-1/3}\left[N + \frac{1}{2} - \frac{4N + 4\sqrt{N}xN^{-1/6} + x^2N^{-1/3}}{4}\right]\widetilde{\Psi}_N(x) \\
&= -N^{-1/3}\left[\frac{1}{2} - \frac{4xN^{1/3} + x^2N^{-1/3}}{4}\right]\widetilde{\Psi}_N(x) \\
&\approx -x\widetilde{\Psi}_N(x) \qquad\qquad \text{for large } N.
\end{aligned}$$

Hence we expect that Ai should satisfy the differential equation

$$\mathrm{Ai}''(x) - x\,\mathrm{Ai}(x) = 0.$$

This is indeed the case, but the proof is again beyond our tools. Let us just give the formal definition of the Airy function and formulate the final result.

Definition 9.8. The *Airy function* $\mathrm{Ai}\colon \mathbb{R} \to \mathbb{R}$ is a solution of the *Airy ODE*

$$u''(x) = xu(x)$$

determined by the following asymptotics as $x \to \infty$:

$$\mathrm{Ai}(x) \sim \frac{1}{2\sqrt{\pi}}x^{-1/4}\exp\left(-\frac{2}{3}x^{3/2}\right).$$

The *Airy kernel* is defined by

$$\mathrm{A}(x,y) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}.$$

Theorem 9.9. *The random variable $N^{2/3}(\lambda_{\max}(A_N) - 2)$ of a normalized* GUE(N) *has a limiting distribution as $N \to \infty$. Its limiting distribution function is*

$$\begin{aligned} F_2(t) :&= \lim_{N\to\infty} \mathbb{P}\left[N^{\frac{2}{3}}\left(\lambda_{\max} - 2\right) \le t\right] \\ &= \sum_{r=0}^{N} \frac{(-1)^r}{r!} \int_t^\infty \cdots \int_t^\infty \det\left(A(x_i, x_j)\right)_{i,j=1}^r \, \mathrm{d}x_1 \cdots \mathrm{d}x_r. \end{aligned}$$

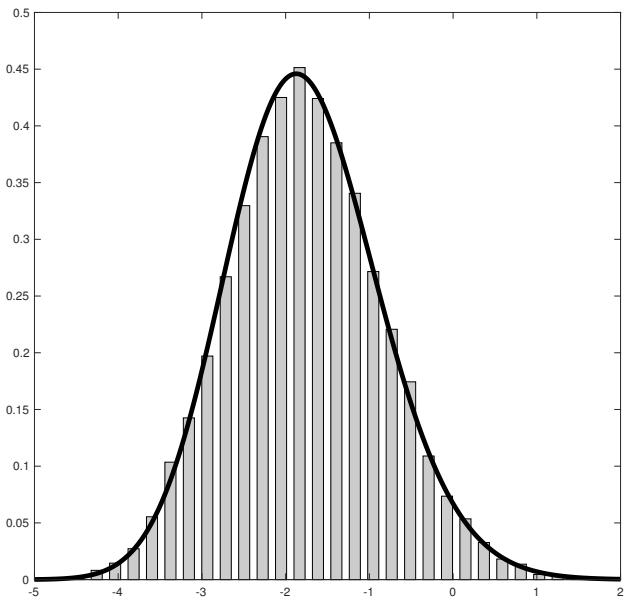
The form of F_2 from Theorem 9.9 is more of a theoretical nature and not very convenient for calculations. A main contribution of Tracy–Widom in this context was that they were able to derive another, quite astonishing, representation of F_2.

Theorem 9.10 (Tracy–Widom, 1994)**.** *The distribution function F_2 satisfies*

$$F_2(t) = \exp\left(-\int_t^\infty (x-t) q(x)^2 \, \mathrm{d}x\right),$$

where q is a solution of the Painlevé II equation $q''(x) - xq(x) + 2q(x)^3 = 0$ with $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$.

Here is a plot of the Tracy–Widom distribution F_2, via solving the Painlevé II equation from above, and a comparision with the histogram for the rescaled largest eigenvalue of 5000 GUE(200); see also Exercise 31.



9.6 Proof of the Harer–Zagier recursion

We still have to prove the recursion of Harer–Zagier, Theorem 9.2.

Let us denote

$$T(k,N) := \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] = \sum_{g\geq 0} \varepsilon_g(k) N^{-2g}.$$

The genus expansion shows that $T(k,N)$ is, for fixed k, a polynomial in N^{-1}. Expressing it in terms of integrating over eigenvalues reveals the surprising fact that, up to a Gaussian factor, it is also a polynomial in k for fixed N. We show this in the next lemma. This is actually the only place where we need the random matrix interpretation of this quantity.

Lemma 9.11. *The expression*

$$N^k \frac{1}{(2k-1)!!} T(k,N)$$

is a polynomial of degree $N-1$ in k.

Proof. First check the easy case $N=1$: $T(k,1) = (2k-1)!!$ is the $2k$-th moment of a normal variable and

$$\frac{T(k,1)}{(2k-1)!!} = 1$$

is a polynomial of degree 0 in k.

For general N we have

$$\begin{aligned}
T(k,N) &= \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \\
&= c_N \int_{\mathbb{R}^N} \left(\lambda_1^{2k} + \cdots + \lambda_N^{2k}\right) e^{-\frac{N}{2}\left(\lambda_1^2+\cdots+\lambda_N^2\right)} \prod_{i<j} (\lambda_i - \lambda_j)^2 \,\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_N \\
&= N c_N \int_{\mathbb{R}^N} \lambda_1^{2k} e^{-\frac{N}{2}\left(\lambda_1^2+\cdots+\lambda_N^2\right)} \prod_{i\neq j} |\lambda_i - \lambda_j| \,\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_N \\
&= N c_N \int_{\mathbb{R}} \lambda_1^{2k} e^{-\frac{N}{2}\lambda_1^2} p_N(\lambda_1) \,\mathrm{d}\lambda_1,
\end{aligned}$$

where p_N is the result of integrating the Vandermonde over $\lambda_2, \ldots, \lambda_N$. It is an even polynomial in λ_1 of degree $2(N-1)$, whose coefficients depend only on N and not on k. Hence

$$p_N(\lambda_1) = \sum_{l=0}^{N-1} \alpha_l \lambda_1^{2l}$$

with α_l possibly depending on N. Thus,

$$\begin{aligned} T(k,N) &= Nc_N \sum_{l=0}^{N-1} \alpha_l \int_{\mathbb{R}} \lambda_1^{2k+2l} e^{-\frac{N}{2}\lambda_1^2} \, \mathrm{d}\lambda_1 \\ &= Nc_N \sum_{l=0}^{N-1} \alpha_l \cdot k_N \cdot (2k+2l-1)!! \cdot N^{-(k+l)}, \end{aligned}$$

since the integral over λ_1 gives the $(2k+2l)$-th moment of a Gauss variable of variance N^{-1}, where k_N contains the N-dependent normalization constants of the Gaussian measure; hence

$$\frac{N^k T(k,N)}{(2k-1)!!}$$

is a linear combination (with N-dependent coefficients) of terms of the form

$$\frac{(2k+2l-1)!!}{(2k-1)!!}.$$

These terms are polynomials in k of degree l. □

We now have that

$$\begin{aligned} N^k \frac{1}{(2k-1)!!} T(k,N) &= N^k \frac{1}{(2k-1)!!} \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)-k-1} \\ &= \frac{1}{N} \frac{1}{(2k-1)!!} \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)} \\ &= \frac{1}{N} \frac{1}{(2k-1)!!} t(k,N), \end{aligned}$$

where the last equality defines $t(k,N)$.

By Lemma 9.11, $t(k,N)/(2k-1)!!$ is a polynomial of degree $N-1$ in k. We interpret it as follows:

$$t(k,N) = \sum_{\pi \in \mathcal{P}_2(2k)} \#\left\{\text{coloring cycles of } \gamma\pi \text{ with at most } N \text{ different colors}\right\}.$$

Let us introduce

$$\tilde{t}(k,L) = \sum_{\pi \in \mathcal{P}_2(2k)} \#\left\{\text{coloring cycles of } \gamma\pi \text{ with exactly } L \text{ different colors}\right\},$$

then we have

$$t(k,N) = \sum_{L=1}^{N} \tilde{t}(k,L)\binom{N}{L},$$

because if we want to use at most N different colors, then we can do this by using exactly L different colors (for any L between 1 and N), and after fixing L we have $\binom{N}{L}$ many possibilities to choose the L colors among the N colors.

This relation can be inverted by

$$\tilde{t}(k,N) = \sum_{L=1}^{N} (-1)^{N-L}\binom{N}{L} t(k,L)$$

and hence $\tilde{t}(k,N)/(2k-1)!!$ is also a polynomial in k of degree $N-1$. But now we have

$$0 = \tilde{t}(0,N) = \tilde{t}(1,N) = \cdots = \tilde{t}(N-2,N),$$

since $\gamma\pi$ has, by Proposition 2.20, at most $k+1$ cycles for $\pi \in \mathcal{P}_2(2k)$; and thus $\tilde{t}(k+1,N) = 0$ if $k+1 < N$, as we need at least N cycles if we want to use N different colors.

So, $\tilde{t}(k,N)/(2k-1)!!$ is a polynomial in k of degree $N-1$ and we know $N-1$ zeros; hence it must be of the form

$$\frac{\tilde{t}(k,N)}{(2k-1)!!} = \alpha_N k(k-1)\cdots(k-N+2) = \alpha_N \binom{k}{N-1}(N-1)!.$$

Hence,

$$t(k,N) = \sum_{L=1}^{N} \binom{N}{L}\binom{k}{L-1}(L-1)!\alpha_L(2k-1)!!.$$

To identify α_N we look at

$$\alpha_{N+1}\binom{N}{N} N!(2N-1)!! = \tilde{t}(N,N+1) = C_N(N+1)!.$$

Note that only the NC pairings can be colored with exactly $N+1$ colors, and for

each such π there are $(N+1)!$ ways of doing so. We conclude

$$\begin{aligned}
\alpha_{N+1} &= \frac{C_N(N+1)!}{N!(2N-1)!!} \\
&= \frac{C_N(N+1)}{(2N-1)!!} \\
&= \frac{1}{N+1}\binom{2N}{N}\frac{N+1}{(2N-1)!!} \\
&= \frac{(2N)!}{N!N!(2N-1)!!} \\
&= \frac{2^N}{N!}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
T(k,N) &= \frac{1}{N^{k+1}}t(k,N) \\
&= \frac{1}{N^{k+1}}\sum_{L=1}^{N}\binom{N}{L}\binom{k}{L-1}(L-1)!\frac{2^{L-1}}{(L-1)!}(2k-1)!! \\
&= (2k-1)!!\frac{1}{N^{k+1}}\sum_{L=1}^{N}\binom{N}{L}\binom{k}{L-1}2^{L-1}.
\end{aligned}$$

To get information from this on how this changes in k we consider a generating function in k,

$$\begin{aligned}
\mathcal{T}(s,N) &= 1 + 2\sum_{k=0}^{\infty}\frac{T(k,N)}{(2k-1)!!}(Ns)^{k+1} \\
&= 1 + 2\sum_{k=0}^{\infty}\sum_{L=1}^{N}\binom{N}{L}\binom{k}{L-1}2^{L-1}s^{k+1} \\
&= \sum_{L=0}^{N}\binom{N}{L}2^L\sum_{k=L-1}^{\infty}\binom{k}{L-1}s^{k+1} \\
&= \sum_{L=0}^{N}\binom{N}{L}2^L\left(\frac{s}{1-s}\right)^L \\
&= \sum_{L=0}^{N}\binom{N}{L}\left(\frac{2s}{1-s}\right)^L
\end{aligned}$$

$$
\begin{aligned}
&= \left(1 + \frac{2s}{1-s}\right)^N \\
&= \left(\frac{1+s}{1-s}\right)^N .
\end{aligned}
$$

Note that (as in our calculation for the α_N)

$$
\frac{1}{(2k-1)!!} = \frac{2^k}{k! C_k (k+1)}
$$

and hence $T(s,N)$ can also be rewritten as a generating function in our main quantity of interest,

$$
b_k^{(N)} = \frac{T(k,N)}{C_k} \qquad \text{(we make now the dependence of } b_k \text{ on } N \text{ explicit)}
$$

as

$$
\begin{aligned}
\mathcal{T}(s,N) &= 1 + 2\sum_{k=0}^{\infty} \frac{T(k,N)}{(k+1)! C_k} 2^k (Ns)^{k+1} \\
&= 1 + \sum_{k=0}^{\infty} \frac{b_k^{(N)}}{(k+1)!} (2Ns)^{k+1}.
\end{aligned}
$$

In order to get a recursion for the $b_k^{(N)}$, we need some functional relation for $\mathcal{T}(s,N)$. Note that the recursion in Harer–Zagier involves b_k, b_{k+1}, b_{k-1} for the same N, thus we need a relation which does not change the N. For this we look on the derivative with respect to s. From

$$
\mathcal{T}(s,N) = \left(\frac{1+s}{1-s}\right)^N
$$

we get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) &= N\left(\frac{1+s}{1-s}\right)^{N-1} \frac{(1-s)+(1+s)}{(1-s)^2} \\
&= 2N\left(\frac{1+s}{1-s}\right)^{N} \frac{1}{(1-s)(1+s)} \\
&= 2N \cdot \mathcal{T}(s,N) \frac{1}{1-s^2},
\end{aligned}
$$

and thus

$$
(1-s^2)\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) = 2N \cdot \mathcal{T}(s,N).
$$

Note that we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) = \sum_{k=0}^{\infty} \frac{b_k^{(N)}}{k!}(2Ns)^k 2N.$$

Thus, by comparing coefficients of s^{k+1} in our differential equation from above, we conclude

$$\frac{b_{k+1}^{(N)}}{(k+1)!}(2N)^{k+2} - \frac{b_{k-1}^{(N)}}{(k-1)!}(2N)^k = 2N\frac{b_{k+1}^{(N)}}{(k+1)!}(2N)^{k+1},$$

and thus, finally,

$$b_{k+1}^{(N)} = b_k^{(N)} + b_{k-1}^{(N)}\frac{(k+1)k}{(2N)^2}.$$

10 Statistics of the Longest Increasing Subsequence

10.1 Complete order is impossible

Definition 10.1. A permutation $\sigma \in S_n$ is said to have an *increasing subsequence of length* k if there exist indices $1 \leq i_1 < \cdots < i_k \leq n$ such that $\sigma(i_1) < \cdots < \sigma(i_k)$. For a *decreasing subsequence of length* k the above holds with the second set of inequalities reversed. For a given $\sigma \in S_n$ we denote the length of an increasing subsequence of maximal length by $L_n(\sigma)$.

Example 10.2. (1) Maximal length is achieved for the identity permutation

$$\sigma = \mathrm{id} = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 1 & 2 & \cdots & n-1 & n \end{pmatrix};$$

this has an increasing subsequence of length n, hence $L_n(\mathrm{id}) = n$. In this case, all decreasing subsequences have length 1.

(2) Minimal length is achieved for the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix};$$

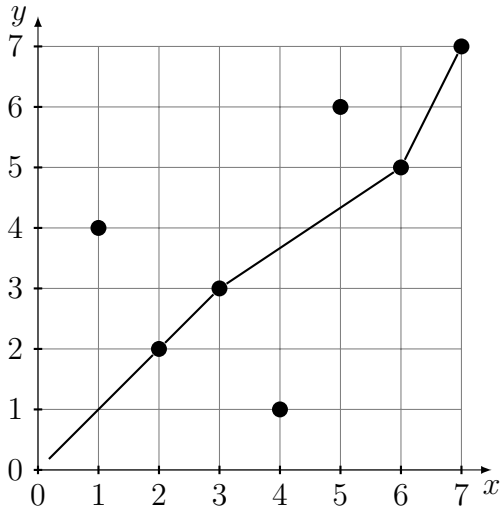
in this case all increasing subsequences have length 1, hence $L_n(\sigma) = 1$; but there is a decreasing subsequence of length n.

(3) Consider a more "typical" permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 1 & 6 & 5 & 7 \end{pmatrix};$$

this has $(2, 3, 5, 7)$ and $(2, 3, 6, 7)$ as longest increasing subsequences, thus $L_7(\sigma) = 4$. Its longest decreasing subsequences are $(4, 2, 1)$ and $(4, 3, 1)$ with

length 3. In the graphical representation



an increasing subsequence corresponds to a path that always goes up.

Remark 10.3. (1) Longest increasing subsequences are relevant for sorting algorithms. Consider a library of n books, labeled bijectively with numbers $1, \ldots, n$, arranged somehow on a single long bookshelf. The configuration of the books corresponds to a permutation $\sigma \in S_n$. How many operations does one need to sort the books in a canonical ascending order $1, 2, \ldots, n$? It turns out that the minimum number is $n - L_n(\sigma)$. One can sort around an increasing subsequence.

Example. Around the longest increasing subsequence $(1, 2, 6, 8)$ we sort

$$\begin{array}{cccccccccc}
 & 4 & 1 & 9 & 3 & 2 & 7 & 6 & 8 & 5 \\
\to & 4 & 1 & 9 & 2 & 3 & 7 & 6 & 8 & 5 \\
\to & 1 & 9 & 2 & 3 & 4 & 7 & 6 & 8 & 5 \\
\to & 1 & 9 & 2 & 3 & 4 & 5 & 7 & 6 & 8 \\
\to & 1 & 9 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\to & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9
\end{array}$$

in $9 - 4 = 5$ operations.

(2) One has situations with only small increasing subsequences, but then one has long decreasing subsequences. This is true in general; one cannot avoid both long decreasing and long increasing subsequences at the same time. According to the slogan

"Complete order is impossible." (Motzkin)

Theorem 10.4 (Erdős, Szekeres, 1935)**.** *Every permutation $\sigma \in S_{n^2+1}$ has a monotone subsequence of length more than n.*

Proof. Write $\sigma = a_1 a_2 \cdots a_{n^2+1}$. Assign labels (x_k, y_k), where x_k is the length of a longest increasing subsequence ending at a_k; and y_k is the length of a longest decreasing subsequence ending at a_k. Assume now that there is no monotone subsequence of length $n+1$. Hence we have for all k: $1 \leq x_k, y_k \leq n$; i.e., there are only n^2 possible labels. By the pigeonhole principle there are $i < j$ with $(x_i, y_i) = (x_j, y_j)$. If $a_i < a_j$ we can append a_j to a longest increasing subsequence ending at a_i, but then $x_j > x_i$. If $a_i > a_j$ we can append a_j to a longest decreasing subsequence ending at a_i, but then $y_j > y_i$. In both cases we have a contradiction. □

10.2 Tracy–Widom for the asymptotic distribution of L_n

We are now interested in the distribution of $L_n(\sigma)$ for $n \to \infty$. This means, we put the uniform distribution on permutations, i.e., $\mathbb{P}[\sigma] = 1/n!$ for all $\sigma \in S_n$, and consider $L_n \colon S_n \to \mathbb{R}$ as a random variable. What is the asymptotic distribution of L_n? This question is called Ulan's problem and was raised in the 1960's. In 1972, Hammersley showed that the limit

$$\Lambda = \lim_{n\to\infty} \frac{\mathbb{E}[L_n]}{\sqrt{n}}$$

exists and that $L_n/\sqrt{n}$ converges to Λ in probability. In 1977, Vershik–Kerov and Logan–Shepp showed independently that $\Lambda = 2$. Then in 1998, Baik, Deift and Johansson proved the asymptotic behaviour of the fluctuations of L_n; quite surprisingly, this is also captured by the Tracy–Widom distribution:

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{L_n - 2\sqrt{n}}{n^{1/6}} \leq t\right] = F_2(t).$$

10.3 Very rough sketch of the proof of the Baik, Deift, Johansson theorem

Again, we have no chance of giving a rigorous proof of the BDJ theorem. Let us give at least a possible route for a proof, which gives also an idea why the statistics of

the length of the longest subsequence could be related to the statistics of the largest eigenvalue.

(1) The RSK correspondence relates permutations to Young diagrams. L_n goes under this mapping to the length of the first row of the diagram.
(2) These Young diagrams correspond to non-intersecting paths.
(3) Via Gessel–Viennot the relevant quantities in terms of NC paths have a determinantal form.
(4) Then one has to show that the involved kernel, suitably rescaled, converges to the Airy kernel.

In the following we want to give some idea of the first two items in the above list; the main (and very hard part of the proof) is to show the convergence to the Airy kernel.

10.3.1 RSK correspondence

RSK stands for Robinson–Schensted–Knuth after papers from 1938, 1961 and 1973. It gives a bijection

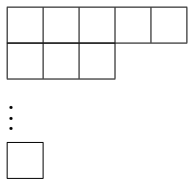
$$S_n \longleftrightarrow \bigcup_{\substack{\lambda \\ \text{Young diagram} \\ \text{of size } n}} (\operatorname{Tab}\lambda \times \operatorname{Tab}\lambda),$$

where $\operatorname{Tab}\lambda$ is the set of Young tableaux of shape λ.

Definition 10.5. (1) Let $n \geq 1$. A *partition* of n is a sequence of natural numbers $\lambda = (\lambda_1, \ldots, \lambda_r)$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \qquad \text{and} \qquad \sum_{i=1}^{r} \lambda_i = n.$$

We denote this by $\lambda \vdash n$. Graphically, a partition $\lambda \vdash n$ is represented by a *Young diagram* with n boxes.



(2) A *Young tableau* of shape λ is the Young diagram λ filled with numbers $1, \ldots, n$ such that in any row the numbers are increasing from left to right and in any column the numbers are increasing from top to bottom. We denote the set of all Young tableaux of shape λ by $\operatorname{Tab}\lambda$.

Example 10.6. (1) For $n = 1$ there is only one Young diagram, $\ydiagram{1}$, and one corresponding Young tableau: $\ytableau{1}$.

For $n = 2$, there are two Young diagrams,

$$\ydiagram{2} \quad \text{and} \quad \ydiagram{1,1},$$

each of them having one corresponding Young tableau

$$\ytableau{1 & 2} \quad \text{and} \quad \ytableau{1 \\ 2}.$$

For $n = 3$, there are three Young diagrams

$$\ydiagram{3} \qquad \ydiagram{2,1} \qquad \ydiagram{1,1,1};$$

the first and the third have only one tableau, but the middle one has two:

$$\ytableau{1 & 2 \\ 3} \qquad \ytableau{1 & 3 \\ 2}$$

(2) Note that a tableau of shape λ corresponds to a walk from $\emptyset$ to λ by adding one box in each step and only visiting Young diagrams. For example, the Young tableau

$$\ytableau{1 & 2 & 4 & 8 \\ 3 & 7 \\ 5 \\ 6}$$

corresponds to the walk

$$\ytableau{1} \to \ytableau{1 & 2} \to \ytableau{1 & 2 \\ 3} \to \ytableau{1 & 2 & 4 \\ 3} \to \ytableau{1 & 2 & 4 \\ 3 \\ 5} \to \ytableau{1 & 2 & 4 \\ 3 \\ 5 \\ 6} \to \ytableau{1 & 2 & 4 \\ 3 & 7 \\ 5 \\ 6} \to \ytableau{1 & 2 & 4 & 8 \\ 3 & 7 \\ 5 \\ 6}$$

Remark 10.7. Those objects are extremely important since they parametrize the irreducible representations of S_n:

$$\lambda \vdash n \longleftrightarrow \text{irreducible representation } \pi_\lambda \text{ of } S_n.$$

Furthermore, the dimension of such a representation π_λ is given by the number of tableaux of shape λ. If one recalls that for any finite group one has the general statement that the sum of the squares of the dimensions over all irreducible representations of the group gives the number of elements in the group, then one has for the symmetric group the statement that

$$\sum_{\lambda \vdash n} (\#Tab\lambda)^2 = \#S_n = n!.$$

This shows that there is a bijection between elements in S_n and pairs of tableaux of the same shape $\lambda \vdash n$. The RSK correspondence is such a concrete bijection, given by an explicit algorithm. It has the property, that L_n goes under this bijection over to the length of the first row of the corresponding Young diagram λ.

Example 10.8. For example, under the RSK correspondence, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 6 & 5 & 1 & 7 \end{pmatrix}$$

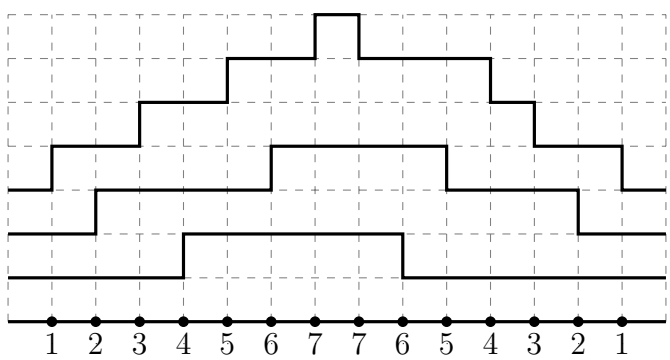
corresponds to the pair of Young tableaux

$$\begin{array}{|c|c|c|c|}\hline 1 & 3 & 5 & 7 \\ \hline 2 & 6 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array}\ , \quad \begin{array}{|c|c|c|c|}\hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 \\ \cline{1-2} 6 \\ \cline{1-1} \end{array}\ .$$

Note that $L_7(\sigma) = 4$ is the length of the first row.

10.3.2 Relation to non-intersecting paths

Pairs $(Q, P) \in \operatorname{Tab}\lambda \times \operatorname{Tab}\lambda$ can be identified with $r = \#\text{rows}(\lambda)$ paths. Q gives the positions of where to go up and P of where to go down; the conditions on the Young tableau guarantee that the paths will be non-intersecting. For example, the pair corresponding to the σ from Example 10.8 above gives the following non-intersecting paths:



11 The Circular Law

11.1 Circular law for Ginibre ensemble

The non-selfadjoint analogue of GUE is given by the Ginibre ensemble, where all entries are independent and complex Gaussians. A standard complex Gaussian is of the form

$$z = \frac{x+iy}{\sqrt{2}},$$

where x and y are independent standard real Gaussians, i.e., with joint distribution

$$p(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}\,\mathrm{d}x\,\mathrm{d}y.$$

If we rewrite this in terms of a density with respect to the Lebesgue measure for real and imaginary part

$$z = \frac{x+iy}{\sqrt{2}} = t_1 + it_2, \qquad \overline{z} = \frac{x-iy}{\sqrt{2}} = t_1 - it_2,$$

we get

$$p(t_1,t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 = \frac{1}{\pi} e^{-(t_1^2+t_2^2)}\,\mathrm{d}t_1\,\mathrm{d}t_2 = \frac{1}{\pi} e^{-|z|^2}\,\mathrm{d}^2 z,$$

where $\mathrm{d}^2 z = \mathrm{d}t_1\,\mathrm{d}t_2$.

Definition 11.1. A *(complex) unnormalized Ginibre ensemble* $A_N = (a_{ij})_{i,j=1}^N$ is given by complex-valued entries with joint distribution

$$\frac{1}{\pi^{N^2}} \exp\left(-\sum_{i,j=1}^N |a_{ij}|^2\right)\,\mathrm{d}A = \frac{1}{\pi^{N^2}} \exp\left(-\operatorname{Tr}(AA^*)\right)\,\mathrm{d}A, \qquad \text{where } \mathrm{d}A = \prod_{i,j=1}^N \mathrm{d}^2 a_{ij}.$$

As for the GUE case, Theorem 7.6, we can rewrite the density in terms of eigenvalues. Note that the eigenvalues are now complex.

Theorem 11.2. *The joint distribution of the complex eigenvalues of an $N \times N$ Ginibre ensemble is given by a density*

$$p(z_1, \dots, z_N) = c_N \exp\left(-\sum_{k=1}^{N} |z_k|^2\right) \prod_{1 \le i < j \le N} |z_i - z_j|^2 .$$

Remark 11.3. (1) Note that typically Ginibre matrices are not normal, i.e., $AA^* \neq A^*A$. This means that one loses the relation between functions in eigenvalues and traces of functions in the matrix. The latter is what we can control, the former is what we want to understand.

(2) As in the selfadjoint case the eigenvalues repel, hence there will almost surely be no multiple eigenvalues. Thus we can also in the Ginibre case diagonalize our matrix, i.e., $A = VDV^{-1}$, where $D = \operatorname{diag}(z_1, \dots, z_N)$ contains the eigenvalues. However, V is now not unitary anymore, i.e., eigenvectors for different eigenvalues are in general not orthogonal. We can also diagonalize A^* via $A^* = (V^{-1})^* D^* V^*$, but since $V^{-1} \neq V^*$ (if A is not normal) we cannot diagonalize A and A^* simultaneously. This means that in general, for example $\operatorname{Tr}(AA^*A^*A)$ has no clear relation to $\sum_{i=1}^N z_i \bar{z}_i \bar{z}_i z_i$. (Note that $\operatorname{Tr}(AA^*A^*A) \neq \operatorname{Tr}(AA^*AA^*)$ if $AA^* \neq A^*A$, but of course $\sum_{i=1}^N z_i \bar{z}_i \bar{z}_i z_i = \sum_{i=1}^N z_i \bar{z}_i z_1 \bar{z}_i$.)

(3) In Theorem 11.2 it seems that we have rewritten the density $\exp(-\operatorname{Tr}(AA^*))$ as $\exp(-\sum_{k=1}^N |z_k|^2)$. However, this is more subtle. On can bring any matrix via a unitary conjugation in a triangular form: $A = UTU^*$, where U is unitary and

$$T = \begin{pmatrix} z_1 & \star & \cdots & \star \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \star \\ 0 & \cdots & 0 & z_n \end{pmatrix}$$

contains on the diagonal the eigenvalues $z_1, \dots, z_n$ of A (this is usually called Schur decomposition). Then $A^* = UT^*U^*$ with

$$T^* = \begin{pmatrix} \bar{z}_1 & 0 & \cdots & 0 \\ \star & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \star & \cdots & \star & \bar{z}_n \end{pmatrix}$$

and

$$\operatorname{Tr}(AA^*) = \operatorname{Tr}(TT^*) = \sum_{k=1}^{N} |z_k|^2 + \sum_{j>i} t_{ij} \bar{t}_{ij}.$$

Integrating out the t_{ij} $(j > i)$ gives then the density for the z_i.

(4) As for the GUE case (Theorem 7.15) we can write the Vandermonde density in a determinantal form. The only difference is that we have to replace the Hermite polynomials $H_k(x)$, which orthogonalize the real Gaussian distribution, by monomials z^k, which orthogonalize the complex Gaussian distribution.

Theorem 11.4. *The joint eigenvalue distribution of the Ginibre ensemble is of the determinantal form* $p(z_1, \dots, z_n) = \frac{1}{N!} \det(K_N(z_i, z_j))_{i,j=1}^N$ *with the kernel*

$$K_N(z, w) = \sum_{k=0}^{N-1} \varphi_k(z)\bar{\varphi}_k(w), \qquad \textit{where} \qquad \varphi_k(z) = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}|z|^2} \frac{1}{\sqrt{k!}} z^k.$$

In particular, for the averaged eigenvalue density of an unnormalized Ginibre eigenvalue matrix we have the density

$$p_N(z) = \frac{1}{N} K_N(z, z) = \frac{1}{N\pi} e^{-|z|^2} \sum_{k=0}^{N-1} \frac{|z|^{2k}}{k!}.$$

Theorem 11.5 (Circular law for the Ginibre ensemble)**.** *The averaged eigenvalue distribution for a normalized Ginibre random matrix* $\frac{1}{\sqrt{N}} A_N$ *converges for* $N \to \infty$ *weakly to the uniform distribution on the unit disc of* $\mathbb{C}$ *with density* $\frac{1}{\pi} 1_{\{z \in \mathbb{C} | |z| \leq 1\}}$.

Proof. The density q_N of the normalized Ginibre is given by

$$q_N(z) = N \cdot p_N(\sqrt{N} z) = \frac{1}{\pi} e^{-N|z|^2} \sum_{k=0}^{N-1} \frac{(N|z|^2)^k}{k!}.$$

We have to show that this converges to the circular density. For $|z| < 1$ we have

$$\begin{aligned} e^{N|z|^2} - \sum_{k=0}^{N-1} \frac{(N|z|^2)^k}{k!} &= \sum_{k=N}^{\infty} \frac{(N|z|^2)^k}{k!} \\ &\leq \frac{(N|z|^2)^N}{N!} \sum_{l=0}^{\infty} \frac{(N|z|^2)^l}{(N+1)^l} \\ &\leq \frac{(N|z|^2)^N}{N!} \frac{1}{1 - \frac{N|z|^2}{N+1}}, \end{aligned}$$

Furthermore, using the lower bound $N! \geq \sqrt{2\pi} N^{N+\frac{1}{2}} e^{-N}$ for $N!$, we calculate

$$\begin{aligned} e^{-N|z|^2} \frac{(N\,|z|^2)^N}{N!} &\leq e^{-N|z|^2} N^N\,|z|^{2N} \frac{1}{\sqrt{2\pi}} \frac{1}{N^{N+\frac{1}{2}}} e^N \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N}} e^{-N|z|^2} e^{N \ln|z|^2} e^N \\ &= \frac{1}{\sqrt{2\pi}} \frac{\exp[N(-\,|z|^2 + \ln|z|^2 + 1)]}{\sqrt{N}} \overset{N\to\infty}{\longrightarrow} 0. \end{aligned}$$

Here, we used that $-\,|z|^2 + \ln|z|^2 + 1 < 0$ for $|z| < 1$. Hence we conclude

$$1 - e^{-N|z|^2} \sum_{k=0}^{N-1} \frac{(N\,|z|^2)^k}{k!} \leq e^{-N|z|^2} \frac{(N\,|z|^2)^N}{N!} \frac{1}{1 - \frac{N|z|^2}{N+1}} \overset{N\to\infty}{\longrightarrow} 0.$$

Similarly, for $|z| > 1$,

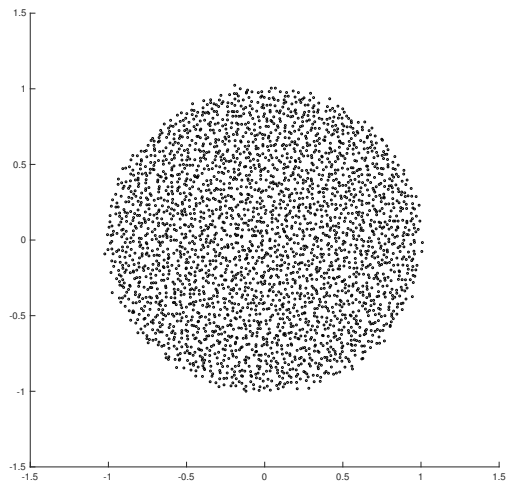
$$\sum_{k=0}^{N-1} \frac{(N\,|z|^2)^k}{k!} \leq \frac{(N\,|z|^2)^{N-1}}{(N-1)!} \sum_{l=0}^{N-1} \frac{(N-1)^l}{(N\,|z|^2)^l} \leq \frac{(N\,|z|)^{N-1}}{(N-1)!} \frac{1}{1 - \frac{N-1}{N|z|^2}},$$

which shows that

$$E^{-N|z|^2} \sum_{k=0}^{N-1} \frac{(N\,|z|^2)^k}{k!} \overset{N\to\infty}{\longrightarrow} 0.$$

□

Remark 11.6. (1) The convergence also holds almost surely. Here is a plot of the 3000 eigenvalues of one realization of a 3000×3000 Ginibre matrix.



(2) The circular law also holds for non-Gaussian entries, but proving this is much harder than the extension for the semicircle law from the Gaussian case to Wigner matrices.

11.2 General circular law

Theorem 11.7 (General circular law)**.** *Consider a complex random matrix* $A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N$*, where the* a_{ij} *are independent and identically distributed complex random variables with variance* 1*, i.e.,* $\mathbb{E}[|a_{ij}|^2] - \mathbb{E}[a_{ij}]^2 = 1$*. Then the eigenvalue distribution of* A_N *converges weakly almost surely for* $N \to \infty$ *to the uniform distribution on the unit disc.*

Note that only the existence of the second moment is required, higher moments don't need to be finite.

Remark 11.8. (1) It took quite a while to prove this in full generality. Here is a bit about the history of the proof.

- 60's, Mehta proved it (see his book) in expectation for Ginibre ensemble;
- 80's, Silverstein proved almost sure convergence for Ginibre;
- 80's, 90's, Girko outlined the main ideas for a proof in the general case;
- 1997, Bai gave the first rigorous proof, under additional assumptions on the distribution
- papers by Tao–Vu, Götze–Tikhomirov, Pan–Zhou and others, weakening more and more the assumptions;
- 2010, Tao–Vu gave final version under the assumption of the existence of the second moment.

(2) For measures on $\mathbb{C}$ one can use $*$-moments or the Stieltjes transform to describe them, but controlling the convergence properties is the main problem.

(3) For a matrix A its $*$-moments are all expressions of the form $\operatorname{tr}(A^{\varepsilon(1)} \cdots A^{\varepsilon(m)})$, where $m \in \mathbb{N}$ and $\varepsilon(1), \ldots, \varepsilon(m) \in \{1, *\}$. The eigenvalue distribution

$$\mu_A = \frac{1}{N}(\delta_{z_1} + \cdots + \delta_{z_N}) \qquad (z_1, \ldots, z_n \text{ are complex eigenvalues of } A)$$

of A is uniquely determined by the knowledge of all $*$-moments of A, but convergence of $*$-moments does not necessarily imply convergence of the eigenvalue distribution.

Example. Consider

$$A_N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \qquad \text{and} \qquad B_N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Then $\mu_{A_N} = \delta_0$, but μ_{B_N} is the uniform distribution on the N-th roots of unity. Hence $\mu_{A_N} \to \delta_0$, whereas μ_{B_N} converges to the uniform distribution on the unit circle. However, the limits of the $*$-moments are the same for A_N and B_N.

(4) For each measure μ on $\mathbb{C}$ one has the Stieltjes transform

$$S_\mu(w) = \int_{\mathbb{C}} \frac{1}{z-w} \, \mathrm{d}\mu(z).$$

This is almost surely defined. However, it is analytic in w only outside the support of μ. In order to recover μ from S_μ one also needs the information about S_μ inside the support. In order to determine and deal with μ_A one reduces it via Girko's "hermitization method"

$$\int_{\mathbb{C}} \log|\lambda - z| \, \mathrm{d}\mu_A(z) = \int_0^t \log t \, \mathrm{d}\mu_{|A-\lambda 1|}(t)$$

to selfadjoint matrices. The left hand side for all λ determines μ_A and the right hand side is about selfadjoint matrices

$$|A - \lambda 1| = \sqrt{(A-\lambda 1)(A-\lambda 1)^*}.$$

Note that the eigenvalues of $|B|$ are related to those of

$$\begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}.$$

In this analytic approach one still needs to control convergence properties. For this, estimates of probabilities of small singular values are crucial.
For more details on this one should have a look at the survey of Bordenave–Chafai, *Around the circular law.*

12 Several Independent GUEs and Asymptotic Freeness

Up to now, we have only considered limits $N \to \infty$ of one random matrix A_N. But often one has several matrix ensembles and would like to understand the "joint" distribution; e.g., in order to use them as building blocks for more complicated random matrix models.

12.1 The problem of non-commutativity

Remark 12.1. (1) Consider two random matrices $A_1^{(N)}$ and $A_2^{(N)}$ such that their entries are defined on the same probability space. What is now the "joint" information about the two matrices which survies in the limit $N \to \infty$? Note that in general our analytical approach breaks down if A_1 and A_2 do not commute, since then we cannot diagonalize them simultaneously. Hence it makes no sense to talk about a joint eigenvalue distribution of A_1 and A_2. The notion μ_{A_1,A_2} has no clear analytic meaning.
What still makes sense in the multivariate case is the combinatorial approach via "mixed" moments with respect to the normalized trace tr. Hence we consider the collection of all *mixed moments* $\operatorname{tr}(A_{i_1}^{(N)} \cdots A_{i_m}^{(N)})$ in A_1 and A_2, with $m \in \mathbb{N}$, $i_1, \ldots, i_m \in \{1, 2\}$, as the joint distribution of A_1 and A_2 and denote this by μ_{A_1,A_2}. We want to understand, in interesting cases, the behavior of μ_{A_1,A_2} as $N \to \infty$.

(2) In the case of one selfadjoint matrix A, the notion μ_A has two meanings:
analytic as $\mu_A = \frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N})$, which is a probability measure on $\mathbb{R}$;
combinatorial, where μ_A is given by all moments $\operatorname{tr}[A^k]$ for all $k \geq 1$.
These two points of view are the same (at least when we restrict to cases where the proabability measure μ is determined by its moments) via

$$\operatorname{tr}(A^k) = \int t^k \, \mathrm{d}\mu_A(t).$$

In the case of two matrices A_1, A_2 the notion μ_{A_1,A_2} has only one meaning,

namely the collection of all mixed moments $\mathrm{tr}[A_{i_1} \cdots A_{i_m}]$ with $m \in \mathbb{N}$ and $i_1, \ldots, i_m \in \{1, 2\}$. If A_1 and A_2 do not commute then there exists no probability measure μ on $\mathbb{R}^2$ such that

$$\mathrm{tr}[A_{i_1} \cdots A_{i_m}] = \int t_{i_1} \cdots t_{i_m} \, \mathrm{d}\mu(t_1, t_2)$$

for all $m \in \mathbb{N}$ and $i_1, \ldots, i_m \in \{1, 2\}$.

12.2 Joint moments of independent GUEs

We will now consider the simplest case of several random matrices, namely r GUEs $A_1^{(N)}, \ldots, A_r^{(N)}$, which we assume to be independent of each other, i.e., we have $A_i^{(N)} = \frac{1}{\sqrt{N}}(a_{kl}^{(i)})_{k,l=1}^N$, where $i = 1, \ldots, r$, each $A_i^{(N)}$ is a GUE(N) and

$$\left\{a_{kl}^{(1)};\, k, l = 1, \ldots, N\right\}, \ldots, \left\{a_{kl}^{(r)};\, k, l = 1, \ldots, N\right\}$$

are independent sets of Gaussian random variables. Equivalently, this can be characterized by the requirement that all entries of all matrices together form a collection of independent standard Gaussian variables (real on the diagonal, complex otherwise). Hence we can express this again in terms of the Wick formula 2.8 as

$$\mathbb{E}\left[a_{k_1 l_1}^{(i_1)} \cdots a_{k_m l_m}^{(i_m)}\right] = \sum_{\pi \in \mathcal{P}_2(m)} \mathbb{E}_\pi\left[a_{k_1 l_1}^{(i_1)}, \ldots, a_{k_m l_m}^{(i_m)}\right]$$

for all $m \in \mathbb{N}$, $1 \leq k_1, l_1, \ldots, k_m, l_m \leq N$ and $1 \leq i_1, \ldots, i_m \leq r$ and where the second moments are given by

$$\mathbb{E}\left[a_{pq}^{(i)} a_{kl}^{(j)}\right] = \delta_{pl}\delta_{qk}\delta_{ij}.$$

Now we can essentially repeat the calculations from Remark 2.14 for our mixed moments:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{tr}(A_{i_1} \cdots A_{i_m})\right] &= \frac{1}{N^{1+\frac{m}{2}}} \sum_{k_1, \ldots, k_m = 1}^{N} \mathbb{E}\left[a_{k_1 k_2}^{(i_1)} a_{k_2 k_3}^{(i_2)} \cdots a_{k_m k_1}^{(i_m)}\right] \\
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{k_1, \ldots, k_m = 1}^{N} \sum_{\pi \in \mathcal{P}_2(m)} \mathbb{E}_\pi\left[a_{k_1 k_2}^{(i_1)}, a_{k_2 k_3}^{(i_2)}, \ldots, a_{k_m k_1}^{(i_m)}\right] \\
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{k_1, \ldots, k_m = 1}^{N} \sum_{\pi \in \mathcal{P}_2(m)} \prod_{(p,q) \in \pi} \mathbb{E}\left[a_{k_p k_{p+1}}^{(i_p)} a_{k_q k_{q+1}}^{(i_q)}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{k_1,\dots,k_m=1}^{N} \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(p,q)\in\pi} [k_p = k_{q+1}]\,[k_q = k_{p+1}]\,[i_p = i_q] \\
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{\substack{\pi\in\mathcal{P}_2(m)\\ (p,q)\in\pi\\ i_p=i_q}} \sum_{k_1,\dots,k_m=1}^{N} \prod_{p} \left[k_p = k_{\gamma\pi(p)}\right] \\
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{\substack{\pi\in\mathcal{P}_2(m)\\ (p,q)\in\pi\\ i_p=i_q}} N^{\#(\gamma\pi)},
\end{aligned}
$$

where $\gamma = (1\ 2 \dots\ m) \in S_m$ is the shift by 1 modulo m. Hence we get the same kind of genus expansion for several GUEs as for one GUE. The only difference is, that in our pairings we only allow to connect the same matrices.

Notation 12.2. For a given $i = (i_1, \dots, i_m)$ with $1 \le i_1, \dots, i_m \le r$ we say that $\pi \in \mathcal{P}_2(m)$ *respects* i if we have $i_p = i_q$ for all $(p, q) \in \pi$. We put

$$\mathcal{P}_2^{[i]}(m) := \{\pi \in \mathcal{P}_2(m) \mid \pi \text{ respects } i\}$$

and also

$$\mathcal{NC}_2^{[i]}(m) := \{\pi \in \mathcal{NC}_2(m) \mid \pi \text{ respects } i\}\,.$$

Theorem 12.3 (Genus expansion of independent GUEs)**.** *Let $A_1, \dots, A_r$ be r independent* GUE(N)*. Then we have for all $m \in \mathbb{N}$ and all $i_1, \dots, i_m \in [r]$ that*

$$\mathbb{E}\left[\operatorname{tr}(A_{i_1} \cdots A_{i_m})\right] = \sum_{\pi\in\mathcal{P}_2^{[i]}(m)} N^{\#(\gamma\pi)-\frac{m}{2}-1}$$

and thus

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}(A_{i_1} \cdots A_{i_m})\right] = \#\mathcal{NC}_2^{[i]}(m).$$

Proof. The genus expansion follows from our computation above. The limit for $N \to \infty$ follows as for Wigner's semicircle law 2.21 from the fact that

$$\lim_{N\to\infty} N^{\#(\gamma\pi)-\frac{m}{2}-1} = \begin{cases} 1, & \pi \in \mathcal{NC}_2(m), \\ 0, & \pi \notin \mathcal{NC}_2(m). \end{cases}$$

The index tuple $(i_1, \dots, i_m)$ has no say in this limit. □

12.3 The concept of free independence

Remark 12.4. We would like to find some structure in those limiting moments. We prefer to talk directly about the limit instead of making asymptotic statements. In the case of one GUE, we had the semicircle μ_W as a limiting analytic object. Now we do not have an analytic object in the limit, but we can organize our distribution as the limit of moments in a more algebraic way.

Definition 12.5. (1) Let $\mathcal{A} = \mathbb{C}\langle s_1, \ldots, s_r\rangle$ be the algebra of polynomials in non-commuting variables $s_1, \ldots, s_r$; this means $\mathcal{A}$ is the free unital algebra generated by $s_1, \ldots, s_r$ (i.e., there are no non-trivial relations between $s_1, \ldots, s_r$ and $\mathcal{A}$ is the linear span of the monomials $s_{i_1} \cdots s_{i_m}$ for $m \geq 0$ and $i_1, \ldots, i_m \in [r]$; multiplication for monomials is given by concatenation).

(2) On this algebra $\mathcal{A}$ we define a unital linear functional $\varphi\colon \mathcal{A} \to \mathbb{C}$ by $\varphi(1) = 1$ and

$$\varphi(s_{i_1} \cdots s_{i_m}) = \lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}(A_{i_1} \cdots A_{i_m})\right] = \#\mathcal{NC}_2^{[i]}(m).$$

(3) We also address $(\mathcal{A}, \varphi)$ as a *non-commutative probability space* and $s_1, \ldots, s_r \in \mathcal{A}$ as *(non-commutative) random variables*. The *moments* of $s_1, \ldots, s_r$ are the $\varphi(s_{i_1} \cdots s_{i_m})$ and the collection of those moments is the *(joint) distribution* of $s_1, \ldots, s_r$.

Remark 12.6. (1) Note that if we consider only one of the s_i, then its distribution is just the collection of Catalan numbers, hence correspond to the semicircle, which we understand quite well.

(2) If we consider all $s_1, \ldots, s_r$, then their joint distribution is a large collection of numbers. We claim that the following theorem discovers some important structure in those.

Theorem 12.7. *Let $\mathcal{A} = \mathbb{C}\langle s_1, \ldots, s_r\rangle$ and let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be defined by $\varphi(s_{i_1} \cdots s_{i_m}) = \#\mathcal{NC}_2^{[i]}(m)$ as before. Then for all $m \geq 1$, $i_1, \ldots, i_m \in [r]$ with $i_1 \neq i_2$, $i_2 \neq i_3$, $\ldots$, $i_{m-1} \neq i_m$ and all polynomials $p_1, \ldots, p_m$ in one variable such that $\varphi\left(p_k(s_{i_k})\right) = 0$ we have:*

$$\varphi\left(p_1(s_{i_1})p_2(s_{i_2}) \cdots p_m(s_{i_m})\right) = 0.$$

In words: the alternating product of centered variables is centered.

We say that $s_1, \ldots, s_r$ are *free (or freely independent)*; in terms of the independent GUE random matrices, we say that $A_1^{(N)}, \ldots, A_r^{(N)}$ are *asymtotially free*. Those notions and the results above are all due to Dan Voiculescu.

Proof. It suffices to prove the statement for polynomials of the form

$$p_k(s_{i_k}) = s_{i_k}^{p_k} - \varphi\left(s_{i_k}^{p_k}\right)$$

for any power p_k, since general polynomials can be written as linear combinations of those. The general statement then follows by linearity. So we have to prove that

$$\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right)\cdots\left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] = 0.$$

We have

$$\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right)\cdots\left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] = \sum_{M\subset[m]}(-1)^{|M|}\prod_{j\in M}\varphi\left(s_{i_j}^{p_j}\right)\varphi\left(\prod_{j\notin M}s_{i_j}^{p_j}\right)$$

with

$$\varphi\left(s_{i_j}^{p_j}\right) = \varphi\left(s_{i_j}\cdots s_{i_j}\right) = \#\mathcal{NC}_2(p_j)$$

and

$$\varphi\left(\prod_{j\notin M}s_{i_j}^{p_j}\right) = \#\mathcal{NC}_2^{[\text{respects indices}]}\left(\sum_{j\notin M}p_j\right).$$

Let us put

$$\begin{aligned} I_1 &= \{1,\dots,p_1\} \\ I_2 &= \{p_1+1,\dots,p_1+p_2\} \\ &\vdots \\ I_m &= \{p_1+p_2+\cdots+p_{m-1}+1,\dots,p_1+p_2+\cdots+p_m\} \end{aligned}$$

and $I = I_1\cup I_2\cup\cdots\cup I_m$. Denote

$$[\dots] = [i_1,\dots,i_1,i_2,\dots,i_2,\dots,i_m,\dots,i_m].$$

Then

$$\prod_{j\in M}\varphi\left(s_{i_j}^{p_j}\right)\varphi\left(\prod_{j\notin M}s_{i_j}^{p_j}\right) = \#\{\pi\in\mathcal{NC}_2^{[\dots]}(I) \mid \text{for all } j\in M \text{ all elements in } I_j \text{ are only paired amongst each other}\}$$

Let us denote

$$\mathcal{NC}_2^{[\dots]}(I:j) := \{\pi\in\mathcal{NC}_2^{[\dots]}(I) \mid \text{elements in } I_j \text{ are only paired amongst each other}\}.$$

Then, by the inclusion-exclusion formula,

$$\varphi\left[\left(s_{i_1}^{p_1}-\varphi\left(s_{i_1}^{p_1}\right)\right)\cdots\left(s_{i_m}^{p_m}-\varphi\left(s_{i_m}^{p_m}\right)\right)\right]=\sum_{M\subset[m]}(-1)^{|M|}\cdot\#\left(\bigcap_{j\in M}\mathcal{NC}_2^{[\cdots]}(I:j)\right)$$
$$=\#\left(\mathcal{NC}_2^{[\cdots]}(I)\backslash\bigcup_j\mathcal{NC}_2^{[\cdots]}(I:j)\right).$$

These are $\pi\in\mathcal{NC}_2^{[\cdots]}(I)$ such that at least one element of each interval I_j is paired with an element from another interval I_k. Since $i_1\neq i_2,\ i_2\neq i_3,\ \ldots,\ i_{m-1}\neq i_m$ we cannot connect neighboring intervals and each interval must be connected to another interval in a non-crossing way. But there is no such π, hence

$$\varphi\left[\left(s_{i_1}^{p_1}-\varphi\left(s_{i_1}^{p_1}\right)\right)\cdots\left(s_{i_m}^{p_m}-\varphi\left(s_{i_m}^{p_m}\right)\right)\right]=\#\left(\mathcal{NC}_2^{[\cdots]}(I)\backslash\bigcup_j\mathcal{NC}_2^{[\cdots]}(I:j)\right)=0,$$

as claimed. □

Remark 12.8. (1) Note that in Theorem 12.7 we have traded the explicit description of our moments for implicit relations between the moments.

(2) For example, the simplest relations from Theorem 12.7 are

$$\varphi\left([s_i^p-\varphi(s_i^p)1][s_j^q-\varphi(s_j^q)1]\right)=0,$$

for $i\neq j$, which can be reformulated to

$$\varphi(s_i^ps_j^q)-\varphi(s_i^p1)\varphi(s_j^q)-\varphi(s_i^p)\varphi(s_j^q1)+\varphi(s_i^p)\varphi(s_j^q)\varphi(1)=0,$$

i.e.,

$$\varphi(s_i^ps_j^q)=\varphi(s_i^p)\varphi(s_j^q).$$

Those relations are quickly getting more complicated. For example,

$$\varphi\left[(s_1^{p_1}-\varphi(s_1^{p_1})1)(s_2^{q_1}-\varphi(s_2^{q_1})1)(s_1^{p_2}-\varphi(s_1^{p_2})1)(s_2^{q_2}-\varphi(s_2^{q_2})1)\right]=0$$

leads to

$$\begin{aligned}\varphi\left(s_1^{p_1}s_2^{q_1}s_1^{p_2}s_2^{q_2}\right)&=\varphi\left(s_1^{p_1+p_2}\right)\varphi\left(s_2^{q_1}\right)\varphi\left(s_2^{q_2}\right)\\&\quad+\varphi\left(s_1^{p_1}\right)\varphi\left(s_1^{p_2}\right)\varphi\left(s_2^{q_1+q_2}\right)\\&\quad-\varphi\left(s_1^{p_1}\right)\varphi\left(s_2^{q_1}\right)\varphi\left(s_1^{p_2}\right)\varphi\left(s_2^{q_2}\right).\end{aligned}$$

These relations are to be considered as non-commutative versions for the factoriziation rules of expectations of independent random variables.

(3) One might ask: What is it good for to find those relations between the moments, if we know the moments in a more explicit form anyhow?
Answer: Those relations occur in many more situations. For example, independent Wishart matrices satisfy the same relations, even though the explicit form of their mixed moments is quite different from the GUE case.
Furthermore, we can control what happens with these relations much better than with the explicit moments if we deform our setting or construct new random matrices out of other ones.
Not to mention that those relations also show up in very different corners of mathematics (like operator algebras).
To make a long story short: Those relations from Theorem 12.7 are really worth being investigated further, not just in a random matrix context, but also for its own sake. This is the topic of a course on *Free Probability Theory*, which can, for example, be found here:
`rolandspeicher.files.wordpress.com/2019/08/free-probability.pdf`

13 Exercises

13.1 Assignment 1

Exercise 1. Make yourself familiar with MATLAB (or any other programming language which allows you to generate random matrices and calculate eigenvalues). In particular, you should try to generate random matrices and calculate and plot their eigenvalues.

Exercise 2. In this exercise we want to derive the explicit formula for the Catalan numbers. We define numbers c_k by the recursion

$$c_k = \sum_{l=0}^{k-1} c_l c_{k-l-1} \tag{13.1}$$

for $k > 0$, with the initial data $c_0 = 1$.

(1) Show that the numbers c_k are uniquely defined by the recursion (13.1) and its initial data.

(2) Consider the (generating) function

$$f(z) = \sum_{k=0}^{\infty} c_k z^k$$

and show that the recursion (13.1) implies the relation

$$f(z) = 1 + z f(z)^2.$$

(3) Show hat f is a power series representation for

$$z \mapsto \frac{1 - \sqrt{1 - 4z}}{2z}.$$

Note: You may use the fact that the formal power series f, defined in (2), has a positive radius of convergence.

(4) Conclude that

$$c_k = C_k = \frac{1}{k+1}\binom{2k}{k}.$$

Exercise 3. Consider the semicircular distribution, given by the density function

$$\frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}, \tag{13.2}$$

where $\mathbb{1}_{[-2,2]}$ denotes the indicator function of the interval $[-2,2]$. Show that (13.2) indeed defines a probability measure, i.e.

$$\frac{1}{2\pi}\int_{-2}^{2}\sqrt{4-x^2}\,\mathrm{d}x = 1.$$

Moreover show that the even moments of the measure are given by the Catalan numbers and the odd ones vanish, i.e.

$$\frac{1}{2\pi}\int_{-2}^{2}x^n\sqrt{4-x^2}\,\mathrm{d}x = \begin{cases} 0 & n \text{ is odd} \\ C_k & n = 2k \end{cases}.$$

13.2 Assignment 2

Exercise 4. Using your favorite programing language or computer algebra system, generate $N \times N$ random matrices for $N = 3, 9, 100$. Produce a plot of the eigenvalue distribution for a single random matrix and as well as a plot for the average over a reasonable number of matrices of given size. The entries should be independent and identically distributed (i.i.d.) according to

(1) the Bernoulli distribution $\frac{1}{2}(\delta_{-1} + \delta_1)$, where δ_x denotes the Dirac measure with atom x.
(2) the normal distribution.

Exercise 5. Prove Proposition 2.2, i.e. compute the moments of a standart Gaussian random variable:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^n e^{-\frac{t^2}{2}}\,\mathrm{d}t = \begin{cases} 0 & n \text{ odd}, \\ (n-1)!! & n \text{ even}. \end{cases}$$

Exercise 6. Let $Z, Z_1, Z_2, \ldots, Z_n$ be independent standard complex Gaussian random variables with mean 0 and $\mathbb{E}[|Z_i|] = 1$ for $i = 1, \ldots, n$.

(1) Show that

$$\mathbb{E}[Z_{i_1}, \dots, Z_{i_r} \bar{Z}_{j_1} \dots \bar{Z}_{j_r}] = \#\{\sigma \in S_r \colon i_k = j_{\sigma(k)} \text{ for } k = 1, \dots, r\}.$$

(2) Show that

$$\mathbb{E}[Z^n \bar{Z}^m] = \begin{cases} 0 & m \neq n, \\ n! & m = n. \end{cases}$$

Exercise 7. Let $A = (a_{ij})_{i,j=1}^N$ be a Gaussian (GUE(N)) random matrix with entries $a_{ii} = x_{ii}$ and $a_{ij} = x_{ij} + \sqrt{-1} y_{ij}$, i.e. the x_{ij}, y_{ij} are real i.i.d. Gaussian random variables, normalized such that $\mathbb{E}[|a_{ij}^2|] = 1/N$. Consider the N^2 random vector

$$(x_{11}, \dots, x_{NN}, x_{12}, \dots, x_{1N}, \dots, x_{N-1N}, y_{12}, \dots, y_{1N}, \dots, y_{N-1N})$$

and show that it has the density

$$C \exp(-N \frac{\mathrm{Tr}(A^2)}{2}) \mathrm{d}A,$$

where C is a constant and

$$\mathrm{d}A = \prod_{i=1}^N \mathrm{d}x_{ii} \prod_{i<y} \mathrm{d}x_{ij} \mathrm{d}y_{ij}.$$

13.3 Assignment 3

Exercise 8. Produce histograms for various random matrix ensembles.

(1) Produce histograms for the averaged situation: average over 1000 realizations for the eigenvalue distribution of a an $N \times N$ Gaussian random matrix (or alternatively ± 1 entries) and compare this with one random realization for $N = 5, 50, 500, 1000$.

(2) Check via histograms that Wigner's semicircle law is insensitive to the common distribution of the entries as long as those are independent; compare typical realisations for $N = 100$ and $N = 3000$ for different distributions of the entries: ± 1, Gaussian, uniform distribution on the interval $[-1, +1]$.

(3) Check what happens when we give up the constraint that the the entries are centered; take for example the uniform distribution on $[0, 2]$.

(4) Check whether the semicircle law is sensitive to what happens on the diagonal of the matrix. Choose one distribution (e.g. Gaussian) for the off-diagonal elements and another distribution for the elements on the diagonal (extreme case: put the diagonal equal to zero).
(5) Try to see what happens when we take a distribution for the entries which does not have finite second moment; for example, the Cauchy distribution.

Exercise 9. In the proof of Theorem 3.9 we have seen that the m-th moment of a Wigner matrix is asymptotically counted by the number of partitions $\sigma \in \mathcal{P}(m)$, for which the corresponding graph $\mathcal{G}_\sigma$ is a tree; then the corresponding walk $i_1 \to i_2 \to \cdots \to i_m \to i_1$ (where $\ker i = \sigma$) uses each edge exactly twice, in opposite directions. Assign to such a σ a pairing by opening/closing a pair when an edge is used for the first/second time in the corresponding walk.
(1) Show that this map gives a bijection between the $\sigma \in \mathcal{P}(m)$ for which $\mathcal{G}_\sigma$ is a tree and non-crossing pairings $\pi \in NC_2(m)$.
(2) Is there a relation between σ and $\gamma\pi$, under this bijection?

Exercise 10. For a probability measure μ on $\mathbb{R}$ we define its Stieltjes transform S_μ by

$$S_\mu(z) := \int_{\mathbb{R}} \frac{1}{t-z} d\mu(t)$$

for all $z \in \mathbb{C}^+ := \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$. Show the following for a Stieltjes transform $S = S_\mu$.
(1) $S : \mathbb{C}^+ \to C^+$.
(2) S is analytic on $\mathbb{C}^+$.
(3) We have

$$\lim_{y\to\infty} iyS(iy) = -1 \qquad \text{and} \qquad \sup_{y>0, x\in\mathbb{R}} y|S(x+iy)| = 1.$$

13.4 Assignment 4

Exercise 11. (1) Let ν be the Cauchy distribution, i.e.,

$$d\nu(t) = \frac{1}{\pi}\frac{1}{1+t^2}dt.$$

Show that the Stieltjes transform of ν is given by

$$S(z) = \frac{1}{-i-z} \qquad \text{for } z \in \mathbb{C}^+.$$

(Note that this formula is not valid in $\mathbb{C}^-$.)
Recover from this the Cauchy distribution via the Stieltjes inversion formula.

(2) Let A be a selfadjoint matrix in $M_N(\mathbb{C})$ and consider its spectral distribution $\mu_A = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues (counted with multiplicity) of A. Prove that for any $z \in \mathbb{C}^+$ the Stieltjes transform S_{μ_A} of μ_A is given by
$$S_{\mu_A}(z) = \mathrm{tr}[(A - zI)^{-1}].$$

Exercise 12. Let $(\mu_N)_{N\in\mathbb{N}}$ be a sequence of probability measures on $\mathbb{R}$ which converges vaguely to μ. Assume that μ is also a probablity measure. Show the following.

(1) The sequence $(\mu_N)_{N\in\mathbb{N}}$ is tight, i.e., for each $\varepsilon > 0$ there is a compact interval $I = [-R, R]$ such that $\mu_N(\mathbb{R}\backslash I) \leq \varepsilon$ for all $N \in \mathbb{N}$.

(2) μ_N converges to μ also weakly.

Exercise 13. The problems with being determined by moments and whether convergence in moments implies weak convergence are mainly coming from the behaviour of our probability measures around infinity. If we restrict everything to a compact interval, then the main statements follow quite easily by relying on the Weierstrass theorem for approximating continuous functions by polynomials. In the following you should not use Theorem 4.12.

In the following let $I = [-R, R]$ be a fixed compact interval in $\mathbb{R}$.

(1) Assume that μ is a probability measure on $\mathbb{R}$ which has its support in I (i.e., $\mu(I) = 1$). Show that all moments of μ are finite and that μ is determined by its moments (among all probability measures on $\mathbb{R}$).

(2) Consider in addition a sequence of probability measures μ_N, such that $\mu_N(I) = 1$ for all N. Show that the following are equivalent:
   - μ_N converges weakly to μ;
   - the moments of μ_N converge to the corresponding moments of μ.

13.5 Assignment 5

In this assignment we want to investigate the behaviour of the limiting eigenvalue distribution of matrices under certain perturbations. In order to do so, it is crucial to deal with different kinds of matrix norms. We recall the most important ones for the following exercises. Let $A \in M_N(\mathbb{C})$, then we define the following norms.

- The spectral norm (or operator norm):
$$\|A\| = \max\{\sqrt{\lambda} \colon \lambda \text{ is an eigenvalue of } AA^*\}.$$
Some of its important properties are:

(i) It is submultiplicative, i.e. for $A, B \in M_N(\mathbb{C})$ one has

$$\|AB\| \leq \|A\| \cdot \|B\|.$$

(ii) It is also given as the operator norm

$$\|A\| = \sup_{\substack{x \in \mathbb{C}^N \\ x \neq 0}} \frac{\|Ax\|_2}{\|x\|_2},$$

where $\|x\|_2$ is here the Euclidean 2-norm of the vector $x \in \mathbb{C}^N$.

- The Frobenius (or Hilbert-Schmidt or L^2) norm:

$$\|A\|_2 = \Big(\mathrm{Tr}(A^*A)\Big)^{1/2} = \sqrt{\sum_{1 \leq i,j \leq N} |a_{ij}|^2}$$

Exercise 14. In this exercise we will prove some useful facts about these norms, which you will have to use in the next exercise when adressing the problem of perturbed random matrices.

Prove the following properties of the matrix norms.

(1) For $A, B \in M_N(\mathbb{C})$ we have $|\,\mathrm{Tr}(AB)| \leq \|A\|_2 \cdot \|B\|_2$.

(2) Let $A \in M_N(\mathbb{C})$ be positive and $B \in M_N(\mathbb{C})$ arbitrary. Prove that

$$|\,\mathrm{Tr}(AB)| \leq \|B\|\,\mathrm{Tr}(A).$$

($A \in M_N(\mathbb{C})$ is positive if there is a matrix $C \in M_N(\mathbb{C})$ such that $A = C^*C$; this is equivalent to the fact that A is selfadjoint and all the eigenvalues of A are positive.)

(3) Let $A \in M_N(\mathbb{C})$ be normal, i.e. $AA^* = A^*A$, and $B \in M_N(\mathbb{C})$ arbitrary. Prove that

$$\max\{\|AB\|_2, \|BA\|_2\} \leq \|B\|_2 \cdot \|A\|$$

Hint: normal matrices are unitarily diagonalizable.

Exercise 15. In this main exercise we want to investigate the behaviour of the eigenvalue distribution of selfadjoint matrices with respect to certain types of perturbations.

(1) Let $A \in M_N(\mathbb{C})$ be selfadjoint, $z \in \mathbb{C}^+$ and $R_A(z) = (A - zI)^{-1}$. Prove that

$$\|R_A(z)\| \leq \frac{1}{\mathrm{Im}(z)}$$

and that $R_A(z)$ is normal.

(2) First we study a general perturbation by a selfadjoint matrix.
Let, for any $N \in \mathbb{N}$, $X_N = (X_{ij})_{i,j=1}^N$ and $Y_N = (Y_{ij})_{i,j=1}^N$ be selfadjoint matrices in $M_N(\mathbb{C})$ and define $\tilde{X}_N = X_N + Y_N$. Show that

$$\left| \operatorname{tr}(R_{\frac{1}{\sqrt{N}} X_N}(z)) - \operatorname{tr}(R_{\frac{1}{\sqrt{N}} \tilde{X}_N})(z) \right| \leq \frac{1}{(\operatorname{Im}(z))^2} \sqrt{\frac{\operatorname{tr}(Y_N^2)}{N}}$$

(3) In this part we want to show that the diagonal of a matrix does not contribute to the eigenvalue distribution in the large N limit, if it is not too ill-behaved. As before, consider a selfadjoint matrix $X_N = (X_{ij})_{i,j=1}^N \in M_N(\mathbb{C})$; let $X_N^D = \operatorname{diag}(X_{11}, \ldots, X_{NN})$ be the diagonal part of X_N and $X_N^{(0)} = X_N - X_N^D$ the part of X_N with zero diagonal. Assume that $\|X_N^D\|_2 \leq N$ for all $N \in \mathbb{N}$. Show that

$$\left| \operatorname{tr}(R_{\frac{1}{\sqrt{N}} X_N}(z)) - \operatorname{tr}(R_{\frac{1}{\sqrt{N}} X_N^{(0)}})(z) \right| \to 0, \quad \text{as } N \to \infty.$$

13.6 Assignment 6

Exercise 16. We will address here concentration estimates for the law of large numbers, and see that control of higher moments allows stronger estimates. Let X_i be a sequence of independent and identically distributed random variables with common mean $\mu = E[X_i]$. We put

$$S_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

(1) Assume that the variance $\mathbb{V}\mathrm{ar}\,[X_i]$ is finite. Prove that we have then the weak law of large numbers, i.e., convergence in probability of S_n to the mean: for any $\epsilon > 0$

$$\mathbb{P}(\omega \mid |S_n(\omega) - \mu| \geq \epsilon) \to 0, \qquad \text{for } n \to \infty.$$

(2) Assume that the fourth moment of the X_i is finite, $\mathbb{E}\,[X_i^4] < \infty$. Show that we have then the strong law of large numbers, i.e.,

$$S_n \to \mu, \qquad \text{almost surely.}$$

(Recall that by Borel–Cantelli it suffices for almost sure convergence to show that

$$\sum_{n=1}^\infty \mathbb{P}(\omega \mid |S_n(\omega) - \mu| \geq \epsilon) < \infty.)$$

One should also note that our assumptions for the weak and strong law of large numbers are far from optimal. Even the existence of the variance is not needed for them, but for proofs of such general versions one needs other tools then our simple consequences of Cheyshev's inequality.

Exercise 17. Let $X_N = \frac{1}{\sqrt{N}}(x_{ij})_{i,j=1}^N$, where the x_{ij} are all (without symmetry condition) independent and identically distributed with standard complex Gaussian distribution. We denote the adjoint (i.e., congugate transpose) of X_N by X_N^*.

(1) By following the ideas from our proof of Wigner's semicircle law for the GUE in Chapter 3 show the following: the averaged trace of any $*$-moment in X_N and X_N^*, i.e.,

$$E[\mathrm{tr}(X_N^{p(1)} \cdots X_N^{p(m)})] \qquad \text{where } p(1), \ldots, p(m) \in \{1, *\}$$

is for $N \to \infty$ given by the number of non-crossing pairings π in $NC_2(m)$ which satisfy the additional requirement that each block of π connects an X with an X^*.

(2) Use the result from part (1) to show that the asymptotic averaged eigenvalue distribution of $W_N := X_N X_N^*$ is the same as the square of the semicircle distribution, i.e. the distribution of Y^2 if Y has a semicircular distribution.

(3) Calculate the explicit form of the asymptotic averaged eigenvalue distribution of W_N.

(4) Again, the convergence is here also in probability and almost surely. Produce histograms of samples of the random matrix W_N for large N and compare it with the analytic result from (3).

Exercise 18. We consider now random matrices $W_N = X_N X_N^*$ as in Exercise 17, but now we allow the X_N to be rectangular matrices, i.e., of the form

$$X_N = \frac{1}{\sqrt{p}}(x_{ij})_{\substack{1 \le i \le N \\ 1 \le j \le p}},$$

where again all x_{ij} are independent and identically distributed. We allow now real or complex entries. (In case the entries are real, X_N^* is of course just the transpose X_N^T.) Such matrices are called *Wishart matrices*. Note that we can now not multiply X_N and X_N^* in arbitrary order, but alternating products as in W_N make sense.

(1) What is the general relation between the eigenvalues of $X_N X_N^*$ and the eigenvalues of $X_N^* X_N$. Note that the first is an $N \times N$ matrix, whereas the second is a $p \times p$ matrix.

(2) Produce histograms for the eigenvalues of $W_N := X_N X_N^*$ for $N = 50$, $p = 100$ as well as for $N = 500$, $p = 1000$, for different distributions of the x_{ij};

- standard real Gaussian random variables
- standard complex Gaussian random variables
- Bernoulli random variables, i.e., x_{ij} takes on values $+1$ and -1, each with probability $1/2$.

(3) Compare your histograms with the density, for $c = 0.5 = N/p$, of the *Marchenko–Pastur* distribution which is given by

$$\frac{\sqrt{(\lambda^+ - x)(x - \lambda^-)}}{2\pi c x} \mathbb{1}_{[\lambda^-, \lambda^+]}(x), \qquad \text{where} \qquad \lambda^\pm := \left(1 \pm \sqrt{c}\right)^2.$$

13.7 Assignment 7

Exercise 19. Prove – by adapting the proof for the GOE case and parametrizing a unitary matrix in the form $U = e^{-iH}$, where H is a selfajoint matrix – Theorem 7.6: The joint eigenvalue distribution of the eigenvalues of a GUE(N) is given by a density

$$\hat{c}_N e^{-\frac{N}{2}(\lambda_1^2 + \cdots + \lambda_N^2)} \prod_{k<l} (\lambda_l - \lambda_k)^2,$$

restricted on $\lambda_1 < \lambda_2 < \cdots < \lambda_N$,

Exercise 20. In order to get a feeling for the repulsion of the eigenvalues of GOE and GUE compare histograms for the following situations:

- the eigenvalues of a GUE(N) matrix for one realization
- the eigenvalues of a GOE(N) matrix for one realization
- N independently chosen realizations of a random variable with semicircular distribution

for a few suitable values of N (for example, take $N = 50$ or $N = 500$).

Exercise 21. For small values of N (like $N = 2, 3, 4, 5, 10$) plot the histogram of averaged versions of GUE(N) and of GOE(N) and notice the fine structure in the GUE case. In the next assignment we will compare this with the analytic expression for the GUE(N) density from class.

13.8 Assignment 8

Exercise 22. In this exercise we define the Hermite polynomials H_n by

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

and want to show that they are the same polynomials we defined in Definition 7.10 and that they satisfy the recursion relation. So, starting from the above definition show the following.

(1) For any $n \geq 1$,
$$xH_n(x) = H_{n+1}(x) + nH_{n-1}(x).$$

(2) H_n is a monic polynomial of degree n. Furthermore, it is an even function if n is even and an odd function if n is odd.

(3) The H_n are orthogonal with respect to the Gaussian measure
$$d\gamma(x) = (2\pi)^{-1/2} e^{-x^2/2} dx.$$
More precisely, show the following:
$$\int_{\mathbb{R}} H_n(x) H_m(x) d\gamma(x) = \delta_{nm} n!$$

Exercise 23. Produce histograms for the averaged eigenvalue distribution of a GUE(N) and compare this with the exact analytic density from Theorem 7.21.

(1) Rewrite first the averaged eigenvalue density
$$p_N(\mu) = \frac{1}{N} K_N(\mu, \mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{k!} H_k(\mu)^2 e^{-\mu^2/2}$$
for the unnormalized GUE(N) to the density $q_N(\lambda)$ for the normalized GUE(N) (with second moment normalized to 1).

(2) Then average over sufficiently many normalized GUE(N), plot their histograms, and compare this to the analytic density $q_N(\lambda)$. Do this at least for $N = 1, 2, 3, 5, 10, 20, 50$.

(3) Check also numerically that q_N converges, for $N \to \infty$, to the semicircle.

(4) For comparison, also average over GOE(N) and over Wigner ensembles with non-Gaussian distribution for the entries, for some small N.

Exercise 24. In this exercise we will approximate the Dyson Brownian motions from Section 8.3 by their discretized random walk versions and plot the corresponding walks of the eigenvalues.

(1) Approximate the Dyson Brownian motion by its discretized random walk version
$$A_N(k) := \sum_{i=1}^{k} \Delta \cdot A_N^{(i)}, \qquad \text{for } 1 \leq k \leq K$$

where $A_N^{(1)}, \ldots, A_N^{(K)}$ are K independent normalized GUE(N) random matrices. Δ is a time increment. Generate a random realization of such a Dyson random walk $A_N(k)$ and plot the N eigenvalues $\lambda_1(k), \ldots, \lambda_N(k)$ of $A_N(k)$ versus k in the same plot to see the time evolution of the N eigenvalues. Produce at least plots for three different values of N.
Hint: Start with $N = 15$, $\Delta = 0.01$, $K = 1500$, but also play around with those parameters.

(2) For the same parameters as in part (i) consider the situation where you replace GUE by GOE and produce corresponding plots. What is the effect of this on the behaviour of the eigenvalues?

(3) For the three considered cases of N in parts (1) and 2i), plot also N independent random walks in one plot, i.e.,

$$\tilde{\lambda}_N(k) := \sum_{i=1}^{k} \Delta \cdot x^{(i)}, \qquad \text{for } 1 \leq k \leq K$$

where $x^{(1)}, \ldots, x^{(K)}$ are K independent real standard Gaussian random variables.

You should get some plots like in Section 8.3.

13.9 Assignment 9

Exercise 25. Produce histograms for the Tracy–Widom distribution by plotting $(\lambda_{\max} - 2)N^{2/3}$.

(1) Produce histograms for the largest eigenvalue of GUE(N), for $N = 50$, $N = 100$, $N = 200$, with at least 5000 trials in each case.

(2) Produce histograms for the largest eigenvalue of GOE(N), for $N = 50$, $N = 100$, $N = 200$, with at least 5000 trials in each case.

(3) Consider also real and complex Wigner matrices with non-Gaussian distribution for the entries.

(4) Check numerically whether putting the diagonal equal to zero (in GUE or Wigner) has an effect on the statistics of the largest eigenvalue.

(5) Bonus: Take a situation where we do not have convergence to semicircle, e.g., Wigner matrices with Cauchy distribution for the entries. Is there a reasonable guess for the asymptotics of the distribution of the largest eigenvalue?

(6) Superbonus: Compare the situation of repelling eigenvalues with "independent" eigenvalues. Produce N independent copies $x_1, \ldots, x_N$ of variables distributed according to the semicircle distribution and take then the maximal

value $x_{\max}$ of these. Produce a histogram of the statistics of $x_{\max}$. Is there a limit of this for $N \to \infty$; how does one have to scale with N?

Exercise 26. Prove the estimate for the Catalan numbers

$$C_k \leq \frac{4^k}{k^{3/2}\sqrt{\pi}} \qquad \forall k \in \mathbb{N}.$$

Show that this gives the right asymptotics, i.e., prove that

$$\lim_{k\to\infty} \frac{4^k}{k^{3/2}C_k} = \sqrt{\pi}.$$

Exercise 27. Let $H_n(x)$ be the Hermite polynomials. The Christoffel-Darboux identity says that

$$\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(x-y)\ (n-1)!}.$$

(1) Check this identity for $n = 1$ and $n = 2$.
(2) Prove the identity for general n.

13.10 Assignment 10

Exercise 28. Work out the details for the "almost sure" part of Corollary 9.6, i.e., prove that almost surely the largest eigenvalue of GUE(N) converges, for $N \to \infty$, to 2.

Exercise 29. Consider the rescaled Hermite functions

$$\tilde{\Psi}(x) := N^{1/12}\Psi_N(2\sqrt{N} + xN^{-1/6}).$$

(1) Check numerically that the rescaled Hermite functions have a limit for $N \to \infty$ by plotting them for different values of N.
(2) Familarize yourself with the Airy function. Compare the above plots of $\tilde{\Psi}_N$ for large N with a plot of the Airy function.

Hint: MATLAB has an implementation of the Airy function, see

https://de.mathworks.com/help/symbolic/airy.html

Exercise 30. Prove that the Hermite functions satisfy the following differential equations:

$$\Psi_n'(x) = -\frac{x}{2}\Psi_n(x) + \sqrt{n}\Psi_{n-1}(x)$$

and

$$\Psi_n''(x) + (n + \frac{1}{2} - \frac{x^2}{4})\Psi_n(x) = 0.$$

13.11 Assignment 11

Exercise 31. Read the notes "Random Matrix Theory and its Innovative Applications" by A. Edelman and Y. Wang,

http://math.mit.edu/~edelman/publications/random_matrix_theory_innovative.pdf

and implement its "Code 7" for calculating the Tracy–Widom distribution (via solving the Painlevé II equation) and compare the output with the histogram for the rescaled largest eigenvalue for the GUE from Exercise 25. You should get a plot like after Theorem 9.10.

Exercise 32. For $N = 100, 1000, 5000$ plot in the complex plane the eigenvalues of one $N \times N$ random matrix $\frac{1}{\sqrt{N}} A_N$, where all entries (without symmetry condition) are independent and identically distributed according to the

(i) standard Gaussian distribution;
(ii) symmetric Bernoulli distribution;
(iii) Cauchy distribution.

14 Literature

Books

(1) Gernot Akemann, Jinho Baik, Philippe Di Francesco: *The Oxford Handbook of Random Matrix Theory*, Oxford Handbooks in Mathematics, 2011.

(2) Greg Anderson, Alice Guionnet, Ofer Zeitouni: *An Introduction to Random Matrices*, Cambridge University Press, 2010.

(3) Zhidong Bai, Jack Silverstein: *Spectral Analysis of Large Dimensional Random Matrices*, Springer-Verlag 2010.

(4) Patrick Billingsley: *Probability and Measure*, John Wiley & Sons, 3rd edition, 1995.

(5) Stéphane Boucheron, Gábor Lugosi, Pascal Massart: *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, Oxford, 2013.

(6) Alice Guionnet: *Large Random Matrices: Lectures on Macroscopic Asymptotics*, Springer-Verlag 2009.

(7) Madan Lal Mehta: *Random Matices*, Elsevier Academic Press, 3rd edition, 2004.

(8) James Mingo, Roland Speicher: *Free Probability and Random Matrices*, Springer-Verlag, 2017.

(9) Alexandru Nica, Roland Speicher: *Lectures on the Combinatorics of Free Probability*, Cambridge University Press 2006.

Lecture Notes and Surveys

(10) Nathanaël Berstycki: *Notes on Tracy–Widom Fluctuation Theory*, 2007.

(11) Charles Bordenave, Djalil Chafaï: *Around the circular law*, Probability Surveys 9 (2012) 1-89.

(12) Alan Edelman, Raj Rao: *Random matrix theory*, Acta Numer. 14 (2005), 233-297.

(13) Alan Edelman, Raj Rao: *The polynomial method for random matrices*, Found. Comput. Math. 8 (2008), 649-702.

(14) Todd Kemp: *MATH 247A: Introduction to Random Matrix Theory*, lecture notes, UC San Diego, fall 2013.