Tracy-Widom Limit for the Wishart Unitary Ensemble

A thesis submitted in partial fulfillment of the requirements for the degree of
Master of Science in Mathematics

Author:
Andreas Schulthess

Supervisor:
Dr. Gaultier Lambert

University of Zurich[UZH]

Department of Mathematics
University of Zurich
Switzerland
March, 2022

Abstract

The Wishart distribution arises in multivariate statistics in the estimation of covariance matrices when one samples from a multivariate normal distribution. Estimation of eigenvalues of such random matrices is a practical problem appearing in various fields such as finance, machine learning, physics, biology and many more.

Nowadays data and the covariance matrix is often high-dimensional leading to several statistical challenges like the curse of dimensionality and the inconsistency of classical estimators. Methods for dimensionality reduction such as the principal component analysis and the canonical correlation analysis require the understanding of the statistical behaviour of the eigenvalues. The theory which explains how the spectrum behaves is random matrix theory.

This thesis gives an insight into the application of random matrix theory to high-dimensional statistics and the underlying probabilistic framework. We investigate the distributions of the β-Wishart ensembles for the cases of uncorrelated and correlated entries. With the theory of orthogonal polynomials we show that the eigenvalues of the Wishart unitary ensemble form a determinantal point process and derive useful eigenvalue statistics. In the last two chapters we explore the theory of Fredholm determinants and derive by the use of asymptotic techniques such as the method of steepest descent the limiting spectral distributions of the Wishart unitary ensemble. For the global regime this is the Marchenko-Pastur distribution and for the biggest eigenvalue statistics this is the Tracy-Widom distribution.

Acknowledgedments

Foremost I would like to thank my supervisor Dr. Gaultier Lambert for his tremendous help, for his patience and his time spent explaining the topics. I am very grateful that he gave me the opportunity to write a thesis on such a fascinating topic covering both theoretical and practical use.

On a personal note I also want to thank Dr. Gaultier Lambert for showing me some very beautiful mathematics, helping me to acquire new techniques and knowledge. During the writing of the thesis I got to see connections between various different fields of mathematics that gave me also a new perspective on mathematics itself.

I thank my beloved parents and brother for all the love and support they gave me and for teaching me to appreciate art and education.

Contents

1 Introduction **9**

1.1 Problems in high-dimensional statistical inference 10

1.1.1 Preliminaries 10

1.1.2 High-dimensional phenomenas 12

1.2 Application of random matrix theory 16

1.2.1 Solving the curse of dimensionality 16

1.2.2 Artificial neural networks 21

1.3 Where high-dimensional data occurs 23

2 Orthogonal Polynomials **25**

2.1 Orthogonal polynomials 26

2.1.1 The integral representation of polynomials 26

2.1.2 Three-term recurrence relation: 30

2.2 Laguerre polynomials 31

2.2.1 The three-term recurrence relation and the differential equation 33

2.3 Special functions 36

2.3.1 Bessel function 36

2.3.2 Airy function 37

3 Wishart Ensemble **41**

3.1 The Wishart distribution 42

3.1.1 The Wishart distribution in multivariate statistics 45

3.2 The β-Wishart ensembles 47

3.3 Spectral densities 48

3.3.1 White β-Wishart eigendistribution 48

3.3.2 Correlated β-Wishart eigendistribution 54

3.4 Real and complex correlated Wishart ensembles 58

3.4.1 Jack polynomials 59

3.4.2 Harish-Chandra-Itzykson-Zuber integral 62

3.5 Wishart unitary ensemble 65

3.5.1 Orthogonal polynomial method 65

3.5.2 The n-point correlation functions and some important statistics 69

3.6 Determinantal point processes 74

4 Fredholm Determinants **77**
4.1 Trace and determinants . . . 78
4.1.1 The exterior power of a linear operator . . . 79
4.1.2 Preliminaries from functional analysis . . . 83
4.2 Traces and determinants in infinite dimension . . . 84
4.2.1 Trace class operators . . . 86
4.2.2 Lidskii's trace theorem . . . 89
4.2.3 Fredholm Determinants . . . 90
4.2.4 The Fredholm formula . . . 90
4.2.5 Numerical computation of Fredholm determinants . . . 93

5 Limit Theorems for the Wishart Unitary Ensemble **95**
5.1 The limit theorems . . . 96
5.2 The method of steepest descent . . . 98
5.2.1 Method of steepest descent . . . 99
5.3 The saddle points of the Laguerre polynomial . . . 102
5.3.1 Saddle points . . . 103
5.4 The Marchenko-Pastur law . . . 104
5.4.1 The Laguerre asymptotics in the bulk . . . 106
5.4.2 The Marchenko-Pastur law . . . 110
5.5 Tracy-Widom limit . . . 114
5.5.1 Airy asymptotics . . . 114
5.5.2 Airy kernel . . . 118
5.5.3 Tracy-Widom limit . . . 120

A Appendix **123**
A.1 List of symbols . . . 123
A.2 Combinatorics . . . 124
A.3 Code examples in Matlab . . . 126
A.3.1 Marchenko-Pastur example . . . 126
A.3.2 Screeplot example . . . 127

Chapter 1

Introduction

Data is usually represented as a $n \times p$ matrix X where n is the *observation* size and p the *dimension* of the parameter space. Various fields have different names for the parameters for example the word *feature* is common in machine learning while *covariates* or *predictors* are used in statistics.

Typically the data matrix X is modelled as random which allows us to use random matrix theory to analyse it. It turns out that the theory is very appropriate in the modern statistical setting "*large p, large n*" dealing with *high-dimensional* data.

The traditional multivariate statistics theory assumes "*small p, large n*" dealing with *low-dimensional* data. Typically p is fix or grows of much smaller order than n leading to

$$\frac{p(n)}{n} \ll 1.$$

The progress of technology and the rise of *big data* (the latter refers to data where n is huge) made it necessary to also consider high-dimensional data where p is no longer small and *comparable to n*. The dimension p can still be fix but typically p grows with n and with similar rate

$$p(n) \approx n \qquad \text{or} \qquad p(n) > n.$$

Several problems can occur in this setting such as the *curse of dimensionality* and the *inconsistency of estimators*. The first problem refers to various statistical and computational phenomena that occur in high dimension. In statistics it means loosely speaking that if we increase the dimension of the parameters p we exponentially increase the possible configurations of a statistical model.

The second problem refers to the phenomena that functions of random variables might not converge to the same limit in high dimension as in low dimension. This is a consequence of the new asymptotic regime that one has to consider in the high-dimensional setting because "*p comparable to n*" means that as $n \to \infty$ we also have $p \to \infty$. Classical limit theorems from probability work under the assumption that p is fix as $n \to \infty$ which is the low-dimensional statistics setting.

$p \to \infty$				
n	$X_1^{(1)}$	$X_1^{(2)}$	$X_1^{(3)}$	$\cdots$
$\downarrow$	$X_2^{(1)}$	$X_2^{(2)}$	$X_2^{(3)}$	$\cdots$
∞	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Fig. 1.1: The asymptotic regime in high-dimensional statistics.

1.1 Problems in high-dimensional statistical inference

In order to understand the high-dimensional phenomena we briefly recap some essential definitions and concepts from statistics and random matrix theory.

1.1.1 Preliminaries

The definitions and concepts can be found in [HB14, ch. 3] and [LNV18, ch. 3].

Statistical inference

In statistical inference we assume that the data X was produced by a certain probability distribution but typically we don't know some of its parameters. The goal is to estimate these parameters based on the data.

Statistical terminology

- (Estimator) Let θ be an unknown parameter. We call any function

$$\hat{\theta} = h(X_1, \ldots, X_n)$$

to estimate θ based on some *random sample* an *estimator*.

Example: Assume we have an i.i.d. *random sample* $X_1, \ldots, X_n$ and each $X_i \sim \mathcal{N}_p(0, \Sigma_p)$ where Σ_p is unknown. Σ_p is called the *population covariance matrix*. The canonical estimator is the *sample covariance matrix* denoted as $\widehat{\Sigma}_p$ and defined as

$$\widehat{\Sigma}_p := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T = \frac{1}{n}\left(\sum_{i=1}^{n} X_i X_i^T\right) - \overline{X}\overline{X}^T. \tag{1.1}$$

where $\overline{X}$ denotes the *sample mean* defined as

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.2}$$

- (Regression) An algorithm to predict a continuous quantity is called a *regressor*. The corresponding task is referred to as *regression*.
- (Randomness) Data is considered as a realization of some distribution. We model it as a random sample $X_1, \ldots, X_p$ and denote its realization as non-random $x_1, \ldots, x_p$.
- (Non-parametric model) If we do not fix the dimension of the parameter p of a statistical model we have a *non-parametric model*. A model where the number of parameters grow with the sample size n is a non-parametric model.

Random matrix theory

Emperical spectral measure: For the sample covariance matrix $\widehat{\Sigma}_p$ and $p_n := p(n)$ we denote the eigenvalues as

$$\lambda_1 := \lambda_1(\widehat{\Sigma}), \ldots, \ \lambda_p := \lambda_p(\widehat{\Sigma})$$

and we introduce the normalized counting measure of the eigenvalues called the *empirical spectral measure*

$$\mu_n := \frac{1}{p_n} \sum_{i=1}^{p_n} \delta_{\lambda_i}. \tag{1.3}$$

For any test function $f : \mathbb{R} \to \mathbb{C}$ we have

$$\frac{1}{p_n} \operatorname{Tr}\left(f\left(\widehat{\Sigma}_p\right)\right) = \frac{1}{p_n} \sum_{i=1}^{p_n} f(\lambda_i) =: \mu_n[f] \tag{1.4}$$

which is a linear statistic. It implies that for any set $A \subseteq \mathbb{R}$ by choosing $f := \mathbf{1}_A$ we get

$$\frac{1}{p_n} \sum_{i=1}^{p_n} \mathbf{1}_A(\lambda_i) = \int_A \mathrm{d}\mu_n = \mu_n(A). \tag{1.5}$$

Convergence: As $n \to \infty$ and $p(n)/n \to \gamma \in (0, +\infty)$ the random measure μ_n converges in *distribution almost surely* (or *weakly almost surely*) to some deterministic distribution μ_{MP} meaning

$$\mathbb{P}\left(\mu_n \to \mu_{MP} \text{ in distribution}\right) = 1 \tag{1.6}$$

and by convergence in distribution we mean for any test function f

$$\lim_{n\to\infty} \frac{1}{p_n} \sum_{i=1}^{p_n} f(\lambda_i) = \lim_{n\to\infty} \int f \ \mathrm{d}\mu_n = \int f \ \mathrm{d}\mu_{MP}. \tag{1.7}$$

In particular this means for any set $A = (-\infty, x]$ we have for $x \in \mathbb{R} \setminus \{0\}$ that

$$\lim_{n\to\infty} \mu_n(A) = \mu_{MP}(A). \tag{1.8}$$

Notice also that $\widehat{\Sigma}_p$ is a scaled matrix because it is of the form

$$\widehat{\Sigma}_p = \frac{1}{n} X^T X \tag{1.9}$$

for some Gaussian matrices X.

Remark 1: *(More types of convergence) Convergence in distribution almost surely implies convergence in distribution in probability*

$$\mathbb{P}\left(\left|\int f \, \mathrm{d}\mu_n - \int f \, \mathrm{d}\mu_{MP}\right| > \varepsilon\right) \xrightarrow{n\to\infty} 0. \tag{1.10}$$

There exist also a weaker form of convergence the convergence of the expectation of μ_n. Define for a bounded continuous test function f

$$\int f \mathrm{d}\mathbb{E}\left[\mu_n\right] := \mathbb{E}\left[\int f \mathrm{d}\mu_n\right] \tag{1.11}$$

then convergence in expectation is

$$\int f \mathrm{d}\mathbb{E}\left[\mu_n\right] \to \int f \mathrm{d}\mu_n. \tag{1.12}$$

Average spectral density: Given the joint probability density function for the unordered eigenvalues $f_p(\lambda_1, \ldots, \lambda_p)$ of a random matrix. Let use define $\lambda := \lambda_1$ then the average spectral density is defined as

$$f_1(\lambda) := \int \cdots \int f_p(\lambda, \ldots, \lambda_p) \, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_p. \tag{1.13}$$

The reason we call it "average" is because if we take the expectation of the empirical spectral measure (1.3) for some particle λ we get

$$\begin{aligned}\mathbb{E}[\mu_n(\lambda)]) &= \int \cdots \int f_p(\lambda_1, \ldots, \lambda_p) \frac{1}{p} \sum_{i=1} \delta_{\lambda_i}(\lambda) \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_p \\ &= \int \cdots \int f_p(\lambda, \ldots, \lambda_p) \, \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_p = f_1(\lambda)\end{aligned}$$

since f_p is a symmetric function. As we will later see in the case of the unitary ensemble we can write the density in terms of a kernel function

$$f_1(\lambda) = \frac{1}{p} K(\lambda, \lambda). \tag{1.14}$$

1.1.2 High-dimensional phenomenas

These examples are based on the examples in [Wai19, pp. 5-9].

Example 1 (The curse of dimensionality in machine learning / statistical learning theory)**:**

A non-parametric regression model

Given some data set $(x_1, y_1), \ldots, (x_n, y_n) \in [0,1]^p \times \mathbb{R}$ consisting of i.i.d. features $x_i = (x_i^{(1)}, \ldots, x_i^{(p)})$ and some response

$$y_i = g(x_i)$$

that was produced by some unknown function g. When we model (X, Y) as a random variable we could search for a function f that minimizes the *least-squares prediction error*

$$\underset{f}{\operatorname{argmin}}\, \mathbb{E}\left[(Y - f(X))^2\right] \tag{1.15}$$

yielding $g = f + \varepsilon$ for some random error ε. It can be shown that the optimal solution to this problem is the *regression function*

$$f(x) = \mathbb{E}\left[Y|X = x\right]. \tag{1.16}$$

But to construct f we would need to know the underlying probability measure $\mathbb{P}_{X,Y}$ which we usually don't know in practice. So we construct an estimator $\hat{f}$ based on our data points and try to minimize the *mean-square error (MSE)*

$$\|\hat{f} - f\|_{L^2}^2 := \mathbb{E}_X\left[(\hat{f}(X) - f(X))^2\right]. \tag{1.17}$$

Remark 2: *(Bias-variance tradeoff) Notice the right-hand side of (1.17) can be brought into the bias-variance decomposition*

$$\begin{aligned}\mathbb{E}_X\left[(\hat{f}(X) - f(X))^2\right] &= \left(\mathbb{E}[\hat{f}(X)] - f(X)\right)^2 + \mathbb{E}[\hat{f}(X)^2] - \mathbb{E}[\hat{f}(X)]^2 \\ &= \mathrm{Bias}(\hat{f}(X))^2 + \mathrm{Var}(\hat{f}(X)).\end{aligned}$$

The curse of dimensionality

The question we are interested in is:

"How big should the data size n be for any given feature dimension p?"

Assume that the unknown probability distribution is a uniform distribution on $[0, 1]^p$ so $X \sim \mathrm{Uni}([0, 1]^p)$ and suppose we have already found a good estimate $\hat{f}$ based on our data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

Since $\hat{f}$ is already a good estimate we would expect that if we draw another random variable $X' \sim \mathrm{Uni}([0, 1]^p)$ any realisation of $X' = x'$ falls close to any of the data points $\{x_1, \ldots, x_n\}$. So the quantity we are interested in to study is

$$\psi(n, p) := \mathbb{E}_{X',X}\left[\min_{i=1,\ldots,n} \|X' - X_i\|_\infty\right] \tag{1.18}$$

for a given n and p. We can reformulate our question to "how big should n be for given p such that $\psi(n, p)$ falls below some threshold δ?". By solving this optimization problem one can prove the following lower bound

$$\log(\psi(n, p)) \geq \log\left(\frac{p}{2(p+1)}\right) - \frac{\log(n)}{p}. \tag{1.19}$$

This implies that we need $n > (1/\delta)^p$ to get $\psi(n, p) \leq \delta$. Therefore the necessary sample size grows exponentially with the feature dimension. Let us compute some examples to get a more intuitive understanding. For the thresholds $\delta = \{0.1,\ 0.01\}$ we have

$\delta = 0.1$	p	n	$\delta = 0.01$	p	n
	2	100		2	$10'000$
	5	$100'000$		5	10^{10}
	10	10^{10}		10	10^{20}
	50	10^{50}		50	10^{100}

On a fixed sample size increasing the feature space dimension leads eventually to sparse data at some point. This shows that the *curse of dimensionality* is closely tied to the geometry of the data and therefore unavoidable in high dimension without preventive measures.

Remark 3: *(The right amount of features) The curse of dimensionality does not automatically mean "the less features the better the model performs". This depends on the complexity of the task and the sample size. However it does mean that if a model has "too many" features they will capture a lot of "statistical noise" which can lead to overfitting on too small data sets. The model will then perform poorly on new data.*

Example 2 (Inconsistency of the estimators and the Marchenko-Pastur law)**:**

Consider a random sample $X_1, \dots, X_n$ that are i.i.d. and drawn from $\mathcal{N}_p(\mu, \Sigma_p)$ where the true mean vector μ and covariance matrix Σ_p are unknown. The task is to search for the most "*probable*" probability measure $\mathbb{P}_\theta$ where $\theta = (\mu, \Sigma_p)$ based on our sample with good estimators $\widehat{\mu}$ and $\widehat{\Sigma}_p$.

By "*probable*" we mean that any estimator $\hat{\theta}_n$ should have good statistical qualities. A typical quality an estimator should have is *consistency* (convergence in probability) i.e. for $\varepsilon > 0$ and sample size $n \to \infty$ that

$$\mathbb{E}\left[\hat{\theta}_n\right] = \theta, \ \mathrm{Var}\left[\hat{\theta}_n\right] \to 0 \quad \Longrightarrow \quad \mathbb{P}\left(|\hat{\theta}_n - \theta| > \varepsilon\right) \to 0$$

While it is theoretically possible to construct an estimator that is consistent but not useful for a finite sample size n, in practice such estimators are not used.

Traditional setting: $p \ll n$

The traditional solutions of this statistical inference problem are the *maximum likelihood estimators (MLE)* gained from maximizing the likelihood. In the case of the normal distribution these are the sample mean (1.2) and the sample covariance matrix (1.1) defined previously

$$\widehat{\mu}_{MLE} := \overline{X} \qquad \text{and} \qquad \widehat{\Sigma}_{MLE} := \widehat{\Sigma}.$$

It is well-known that they are consistent estimators. Consistency describes the asymptotic behaviour of an estimator and therefore relies on the *law of large numbers (LLN)* which works like the *central limit theorem (CLT)* under the assumption that the parameter size p is fixed while $n \to \infty$, which is in this setting no problem.

High-dimensional setting: $p(n) \approx n$

The question we want to study is whether the maximum likelihood estimator is still a consistent estimator in high dimension. Let us assume that $\mu = (0, \dots, 0)$ and the true covariance matrix is the identity matrix so $\Sigma_p := I_p$. Then we expect as $n \to \infty$ under some moment conditions that we have

$$\mathbb{P}\left(\left\|\widehat{\Sigma}_{MLE} - \Sigma_p\right\| > \varepsilon\right) = 0, \qquad \text{for some } \varepsilon > 0.$$

for the spectral norm defined with the euclidean norm

$$\left\|\widehat{\Sigma}_{MLE} - \Sigma_p\right\| := \sup_{x \neq 0} \frac{\left\|(\widehat{\Sigma}_{MLE} - \Sigma_p)x\right\|_2}{\|x\|_2}. \tag{1.20}$$

Let us investigate this experimentally with different (n, p).

p	n	p/n	p	n	p/n
100	500	0.2	2000	5000	0.4

In both cases p is still smaller than n but $p/n = \gamma$ is not that small. We expect the histogram of the eigenvalues of $\widehat{\Sigma}_{MLE}$ to concentrate around the eigenvalue 1.

Experimental results

Running the experiment 10 times and plotting the normalized results in fig. 1.2 and fig. 1.3.

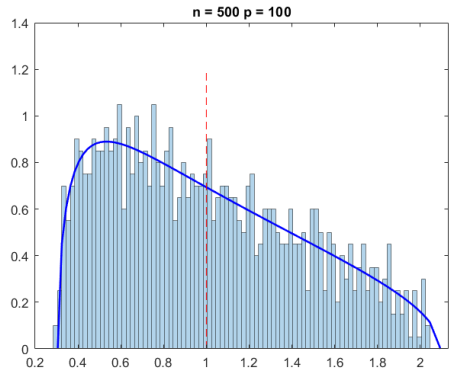

Fig. 1.2: $n = 500,\ p = 100.$

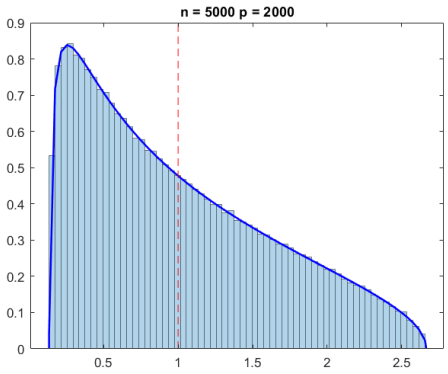

Fig. 1.3: $n = 5000,\ p = 2000.$

These plots are not centred around 1 suggesting that as n increases with p the eigenvalues will not converge to 1. This means as $n \to \infty$ with

$$\frac{p(n)}{n} = \gamma \neq 0 \qquad \Longrightarrow \qquad \widehat{\Sigma}_{MLE} \text{ will not converge to } \Sigma_p \tag{1.21}$$

and therefore the maximum likelihood estimator is no longer consistent in high dimension. The distribution that the histogram approaches is the *Marchenko-Pastur distribution*. We have as $n \to \infty$ and $p(n)/n \to \gamma \in (0, +\infty)$ that

$$\mathbb{P}\left(\mu_n \to \mu_{MP} \text{ in distribution}\right) = 1 \tag{1.22}$$

where μ_{MP} is the deterministic distribution with density

$$\left(1-\gamma^{-1}\right)^{+}\delta_0(x)+\frac{1}{2\pi\gamma x}\sqrt{(B_+-x)(x-B_-)}\mathbf{1}_{[B_-,B_+]}(x) \tag{1.23}$$

with

$$B_+ := (1+\sqrt{\gamma})^2 \qquad B_- := (1-\sqrt{\gamma})^2$$

and $(x)^+ := \max(0,x)$. A more detailed description of this distribution follows in the last chapter 5.

Surprisingly this experiment shows that the shape of the limiting distribution is already visible for reasonably small n and p suggesting that this distribution can be used already as an estimating distribution for the empirical spectral distribution of our estimator. In fact this is one of the strong points advocating the use of random matrix theory for statistical analysis.

1.2 Application of random matrix theory

We now show two examples of how random matrix theory can be applied to statistics and machine learning. The scree plot part in the first example is inspired of [EK18] and the second example follows [Pas20].

1.2.1 Solving the curse of dimensionality

The dominant strategy used to deal with high-dimensional data is *dimensionality reduction* which reduces the dimension p of the features. Some popular methods that are used are *principal component analysis (PCA)*, the *linear discriminant analysis (LDA)* and the *canonical correlation analysis (CCA)*. All these techniques are based on the spectrum of the data matrix.

Example 3 (PCA and random matrix theory)**:**

Let $D \in \mathbb{R}^{n\times p}$ be a data matrix with p features $X = \{X_1,\dots,X_p\}$ and covariance matrix Σ_p. To reduce the dimension

$$D \in \mathbb{R}^{n\times p} \to \widetilde{D} \in \mathbb{R}^{n\times d}$$

for some $d < p$ we either remove some features (*feature selection*) or create a new set of features $\{Z_1,\dots,Z_d\}$ based on some function

$$Z_i = f(X_1,\dots,X_p) \tag{1.24}$$

evaluating all the features (*feature extraction*), the latter is what PCA is doing.

Principal component analysis (PCA)

PCA makes a *principal component decomposition* of D

$$Z^{(d)} = DW^{(d)}, \qquad W^{(d)} \in \mathbb{R}^{p\times d} \tag{1.25}$$

where $Z^{(d)}$ consists of the *principal components*. These are orthonormal vectors $z_1, \dots, z_p$ that *maximize the variance*. Heuristically this means if two data points p_1, p_2 are far from each other in D they will stay far from each other in $Z^{(d)}$ and vice versa for points close to each other thus capturing most of the original information.

To compute the principal components one needs to solve the following optimization problem

$$\max_z z^T \Sigma_p z, \qquad \text{with} \qquad zz^T = \mathrm{Id} \tag{1.26}$$

which can be reduced to an eigenvalue problem

$$\Sigma_p z_i = \lambda_i z_i \tag{1.27}$$

and therefore

principal components = maximum eigenvectors of the covariance matrix.

Since there are in total p eigenvectors we have to choose a smaller subset of d principal components.

Scree plot method

The task is now to find out which eigenvectors to choose such that the new data matrix $Z^{(d)}$ is still meaningful. In practice one uses often the *scree plot method* (also called *elbow criterion*) where the ordered eigenvalues of the sample covariance matrix $\widehat{\Sigma}_p$ are plotted. Then one looks for the biggest change in steepness ("*elbow*") in the plot and all the eigenvalues above this point will be kept. But this heuristic method is not reliable in high dimension as we will show experimentally.

Spiked models

Let $X \sim \mathcal{N}_p(0, \Sigma_p)$. If we can split up Σ_p into

$$\Sigma_p = I_p + P \quad \text{ with } \operatorname{rank}(P) = \operatorname{diag}(s_1, \dots, s_d)$$

where P is some perturbation then we call it a *spiked model* with d *spikes*. A model without spikes meaning $\Sigma_p = I_p$ is referred to as a *null model*. We now want to try the scree plot method for the spiked model.

Consider a spiked model where $\Sigma_p^{(S)}$ is a diagonal matrix with $d = 7$ eigenvalues $\{5,\ 4.5,\ 4,\ 3.5,\ 3,\ 2.5,\ 2\}$ and $p - d$ eigenvalues equal to 1 so. Therefore we have d spikes

$$(s_1, \dots, s_d) = (4,\ 3.5,\ 3,\ 2.5,\ 2,\ 1.5,\ 1).$$

Let us choose a low- and high-dimensional setting. We plot the eigenvalues of the sample covariance matrix $\widehat{\Sigma}_p^{(S)}$ in blue and the true eigenvalues of $\Sigma_p^{(S)}$ as a dotted red line in fig. 1.4 and fig. 1.5.

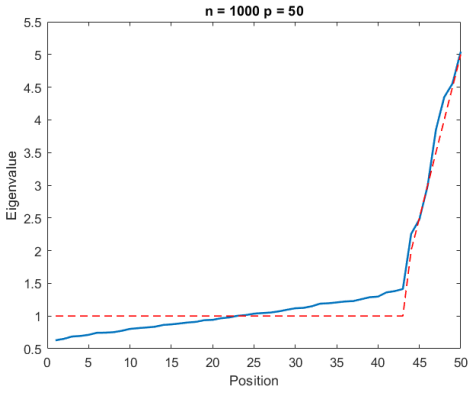

Fig. 1.4: $n = 1000,\ p = 50.$

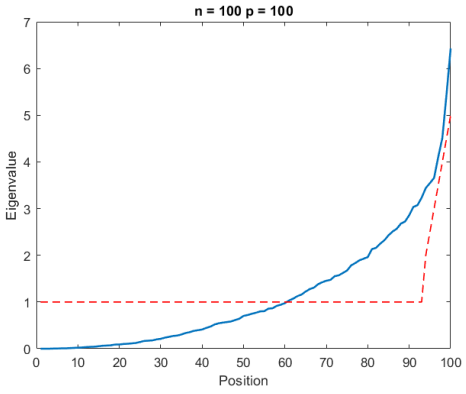

Fig. 1.5: $n = 100,\ p = 100.$

Experimental results

In fig. 1.4 we have $p/n = 0.05$ and the elbow in the blue line is clearly visible. This means any perturbation in the covariance matrix $\Sigma_p^{(S)}$ is still visible in the eigenvalue plot of the sample covariance matrix $\widehat{\Sigma}_p^{(S)}$. One can now proceed to only use the principal components for those eigenvalues above the elbow point.

In fig. 1.5 when $p/n = 1$ the elbow vanishes in the blue line. In this case the scree plot method fails.

Remark 4: *(Level repulsion in statistics) The experiment captures an important phenomena of high-dimensional statistics that the eigenvalues of $\widehat{\Sigma}_p$ are more spread out than the actual eigenvalues of Σ_p are. As an example a simple simulation for $n = 10,\ p = 10$ and $\Sigma_p = I_p$ shows that the largest and smallest eigenvalue of $\widehat{\Sigma}_p$ are*

$$\lambda_1 = 3.61 \quad \textit{and} \quad \lambda_{10} = 0$$

and $\lambda_2, \ldots, \lambda_9$ are almost equally distributed in between. This is a consequence of the "level repulsion" of the eigenvalues of the Wishart ensemble due to the Vandermonde determinant in its density. Without random matrix theory it is hard to understand why they are not close to 1.

Comparing the spiked with the null model

Assume we have no knowledge about the $d = 7$ spikes. The goal is to estimate the numbers of spikes. If we want to proceed in the high-dimensional setting we can compare the spiked model with a null model of dimension $p - d$. Let "cov" stand for covariance matrix and "eig" for its eigenvalues. The setting is:

A spiked model

$$\text{cov} = \Sigma_p^{(S)} \qquad \text{eig} = \{1 + s_1, 1 + s_2, \ldots, 1 + s_{d-1}, 1 + s_d, 1, 1 \ldots, 1\}$$

its sample covariance matrix

$$\text{cov} = \widehat{\Sigma}_p^{(S)} \qquad \text{eig} = \{l_1 \geq \cdots \geq l_d \geq l_{d+1} \geq \cdots \geq l_p\}.$$

A null model

$$\text{cov} = I_{p-d} \qquad \text{eig} = \{1, 1, \ldots, 1, 1\}$$

its sample covariance matrix

$$\text{cov} = \widehat{\Sigma}_{p-d}^{(0)} \qquad \text{eig} = \{\tau_1 \geq \cdots \geq \tau_{p-d}\}.$$

Random matrix theory tells us that as $n, p \to \infty$ the biggest eigenvalues of $\widehat{\Sigma}_{p-d}^{(0)}$ will fluctuate around the upper edge of the Marchenko-Pastur distribution according to the *Tracy-Widom* distribution. What about $\widehat{\Sigma}_p^{(S)}$? There are a couple of different approaches.

For example we can study the behaviour of the smallest spiked eigenvalue l_d or we can study the behaviour of the largest non-spiked eigenvalue l_{d+1}.

If a spike s_i is small we would expect that the eigenvalue $1+s_i$ behaves approximately like 1. So another thing to study is how big can s_i be so that l_i behaves still the same as it would have in the case $s_i = 0$.

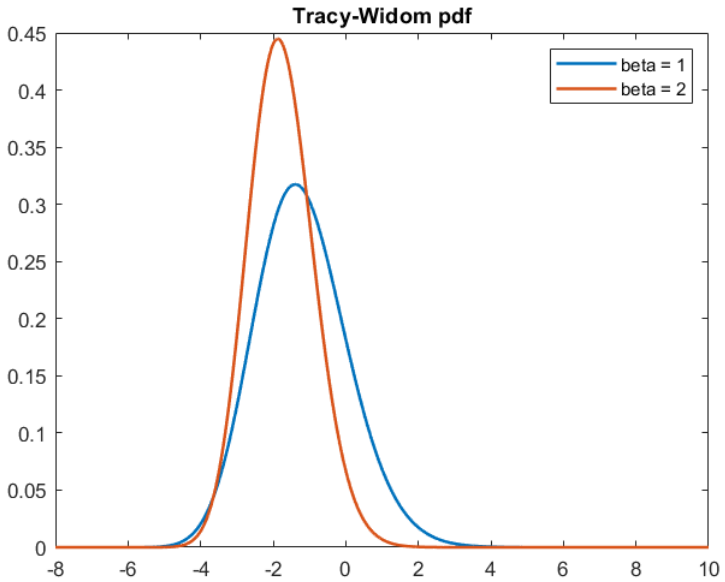

Fig. 1.6: Tracy-Widom density for $\beta = 1,\ \beta = 2$.

Spiked model limit theorems

In [CHP17, pp.10-11] it was shown for the real case that under a couple of mild assumptions as $p, n \to \infty$ the largest non-spiked eigenvalue l_{d+1} of $\widehat{\Sigma}_p^{(S)}$ will follow

the Tracy-Widom distribution.

Further in [CHP17, p. 12] based on this Tracy-Widom result the following bound for estimating the first non-spiked eigenvalue was derived:

$$l_{d+1} \leq l_{n^{1/6}} + \log(n \times n^{-5/9})$$

In [BS06, pp. 1385-1387] it was shown for the real case under some finite 4-th moment assumption that as $p, n \to \infty$ and $p/n \to \gamma$

- if $s_i > \sqrt{\gamma}$ then l_i will converge outside of the Marchenko-Pastur support to $1 + \sqrt{\gamma} + \gamma\frac{(1+s_i)}{s_i}$ almost surely.
- If $0 \leq s_i \leq \sqrt{\gamma}$ then l_i lies inside the support of the Marchenko-Pastur distribution.

In [BV12] it was shown (for general β) that for the biggest eigenvalue l_1 we have a so called *BPP phase transition* where l_i converges in distribution either to the Tracy-Widom or the normal distribution:

- if $0 \leq s_i < \sqrt{\gamma}$ then
$$n^{2/3}\frac{l_1 - \mu(\gamma)}{\sigma(\gamma)} \Rightarrow TW_\beta.$$
- if $s_i > \sqrt{\gamma}$ then
$$n^{1/2}\frac{l_1 - \rho(s_i, \gamma)}{\tau(s_i, \gamma)} \Rightarrow \mathcal{N}(0, 1).$$

Finally we note that in [Joh01b, pp. 301-303] it was shown for the real case that when $d < p$ the largest non-spiked eigenvalue l_{d+1} is *stochastically smaller* than the biggest eigenvalue τ_1 of $\widehat{\Sigma}_p^{(0)}$. This is a consequence of the *interlacing property* that says the smallest real eigenvalue of the matrix $\Sigma_p^{(S)}$ is less than the smallest real eigenvalue of its principal submatrix I_{p-d}.

Again as with the Marchenko-Pastur limit the Tracy-Widom limit is already a good approximation for the null model for a reasonably small setting. A good investigation on how well Tracy-Widom performs under different n, p can be found in [Ma08, p. 5] and in [Joh01b, p. 302].

Tracy-Widom hypothesis testing

We can use the Tracy-Widom distribution also for hypothesis testing.

$$\begin{aligned} H_0 &: \Sigma_p = I_p \qquad \text{(null hypothesis)} \\ H_1 &: \Sigma_p \neq I_p \qquad \text{(alternative)} \end{aligned}$$

For the largest eigenvalue l_1 of $\widehat{\Sigma}_p$ we compute the *TW test statistics* (valid for $n \geq p$):

$$L_{TW} := \frac{nl_1 - \mu_{n,p}}{\sigma_{n,p}}$$

$$\mu_{n,p} = \left(\sqrt{n} + \sqrt{p}\right)^2, \quad \sigma_{n,p} = (\sqrt{n} + \sqrt{p}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}\right)^{1/3}.$$

Then we compute the p-value

$$\mathbb{P}(L_{TW} \geq t | H_0). \tag{1.28}$$

1.2.2 Artificial neural networks

Despite the popularity and success of neural networks in deep learning there are still unresolved theoretical questions. Random matrix theory gives a new approach to study these models.

We briefly explain mathematically the simplest neural network a *fully-connected, feed-forward neural network (FFNN)*. Fully-connected means every neuron on layer $l-1$ is connected to every neuron on layer l.

Definition 1 (fully-connected, feed-forward neural network) [Pas20, p. 1-2]**.** *A FFNN is a function*

$$\mathcal{N} : \mathbb{R}^{n_0} \to \mathbb{R}^{n_L} \tag{1.29}$$

with $\{n_0, n_1, \ldots, n_L\}$ representing the amount of neurons per layer. We have $L-1$ hidden layers and $\mathcal{N}$ is a composition of the form

$$\mathcal{N}(x) = \left(\Psi_L \circ \mathcal{A}^{(L)}_{n_L,n_{L-1}} \circ \Psi_{L-1} \circ \mathcal{A}^{(L-1)}_{n_{L-1},n_{L-2}} \circ \cdots \circ \Psi_1 \circ \mathcal{A}^{(1)}_{n_1,n_0}\right)(x) \tag{1.30}$$

where $\Psi_1, \ldots, \Psi_L$ are some activation functions and $\mathcal{A}^{(1)}_{n_1,n_0}, \ldots, \mathcal{A}^{(L)}_{n_L,n_{L-1}}$ are affine transformations.

These affine transformation are of the form

$$\mathcal{A}^{(l)}_{n_l,n_{l-1}}(x) = W^{(l)}x^{(l-1)} + b^{(l)} \tag{1.31}$$

where

$$W^{(l)} \in \mathbb{R}^{n_l \times n_{l-1}}, \; x^{(l-1)} \in \mathbb{R}^{n_{l-1}}, \; b^{(l)} \in \mathbb{R}^{n_l} \tag{1.32}$$

for $1 \leq l \leq L$ were the dimension correspond to the neurons on the previous and next layer. $W^{(l)}$ is a random matrix called the *weights* and $b^{(l)}$ is a random vector called the *bias*. The activation functions are of the form

$$\Psi_l(y) = (\psi_1(y_1^{(l)}), \ldots, \psi_{n_l}(y_{n_l}^{(l)})) \qquad \text{for} \quad y^{(l)} = W^{(l)}x^{(l-1)} + b^{(l)} \tag{1.33}$$

and they measure whether a certain threshold was achieved for each entry $y^{(l)} = (y_1^{(l)} \ldots, y_{n_l}^{(l)})$ after the affine transformation. This means

$$x^{(l+1)} = \Psi_{l+1}\left(W^{(l+1)}x^{(l)} + b^{(l+1)}\right) \tag{1.34}$$

respectively

$$x^{(L)} = \Psi_L\left(W^{(L)}\Psi_{L-1}\left(\ldots \Psi_1\left(W^{(1)}x^{(0)} + b^{(1)}\right) \cdots + b^{L-1}\right) + b^L\right). \tag{1.35}$$

The input-output Jacobian

Assume that $x^{(0)} = (x_1^{(0)}, \dots, x_{n_0}^{(0)})$ are the neurons of the input layer and $x^{(L)} = (x_1^{(L)}, \dots, x_{n_L}^{(L)})$ the neurons of the output layer. The *input-output Jacobian* is defined as the $n_L \times n_0$ matrix

$$J^L := J^L_{n_1,\dots,n_L} = \left\{ \frac{\partial x_{j_L}^{(L)}}{\partial x_{j_0}^{(0)}} \right\}_{j_0, j_L = 1}^{n_0, n_L} = \begin{pmatrix} \frac{\partial x_1^{(L)}}{\partial x_1^{(0)}} & \cdots & \frac{\partial x_1^{(L)}}{\partial x_{n_0}^{(0)}} \\ \vdots & & \vdots \\ \frac{\partial x_{n_L}^{(L)}}{\partial x_1^{(0)}} & \cdots & \frac{\partial x_{n_L}^{(L)}}{\partial x_{n_0}^{(0)}} \end{pmatrix}. \tag{1.36}$$

The input-output Jacobian measures the effect that changes of the input values have on the output. It turns out that the spectrum of the *singular values*, i.e. the square roots of the eigenvalues of the $n^L \times n^L$ matrix

$$M^L := J^L (J^L)^T \tag{1.37}$$

has an influence on the learning speed of a neural network. For some concentration one can speed up the learning process of deep neural network (see [PSG18]). Therefore one is interested in studying the spectral distribution of this matrix.

By using the chain rule the input-output Jacobian can be rewritten as

$$J^L = \prod_{l=1}^{L} D^{(l)} W^{(l)} \tag{1.38}$$

where $D^{(l)}$ is a diagonal matrix given as

$$D^{(l)} = \left\{ D_{j_l}^{(l)} \delta_{j_l, k_l} \right\}_{j_l, k_l = 1}^{n_l} \qquad D_{j_l}^{(l)} = \psi'_{j_l} \left(\sum_{j_{l-1}}^{n_{l-1}} W_{j_l, j_{l-1}}^{(l)} x_{j_{l-1}}^{l-1} + b_{j_l}^l \right), \; l = 1, \dots, L. \tag{1.39}$$

Convergence of the empirical spectral measure

In [Pas20] it was shown with the use of random matrix theory that under certain conditions such as that the weights $W^{(l)}$ and the bias $b^{(l)}$ are Gaussian and for appropriate scaling, the *empirical spectral measure* $\nu_{M_n^L}$ of (1.37) of a FFNN converges as $n \to \infty$ to a non-random measure ν_{M^L}. The input $x^{(0)}$ can be deterministic or random. The limiting measure is of the form

$$\nu_{M^L} = \nu_{K^1} \diamond \nu_{K^2} \diamond \cdots \diamond \nu_{K^L} \diamond \delta_1 \tag{1.40}$$

where $\{\nu_{K^l}\}$ for $l = 1, \dots, L$ are the probability distributions of the random variables $\left\{ \psi' \left(\gamma \sqrt{q^l} \right)^2 \right\}$ with $\gamma \sim \mathcal{N}(0,1)$ and q^l is some deterministic quantity defined recursively, δ_1 is the unit measure concentrated at 1. The operation $\diamond$ is defined as follow: Suppose you have a random matrix of the form

$$M_n = n^{-1} R_n^{1/2} X_n^T K_n X_n R_n^{1/2} \tag{1.41}$$

where X_n is a $n \times n$ Gaussian matrix and K_n and R_n are positive definite matrices (random or not). Consider the emperical spectral measure ν_{M_n}. The expression

$$\nu_M = \nu_K \diamond \nu_R$$

means that as $n \to \infty$ and $\nu_{M_n} \to \nu_M$ weakly almost surely, the limiting measure ν_M is uniquely determined by the limits ν_K and ν_R via the *Stieltjes transform*. The details can be read in [Pas20].

1.3 Where high-dimensional data occurs

Some examples where high-dimensional data occurs are:

Deep learning (image and speech recognition), finance (portfolio optimization theory), internet of things (smart grid), genetics (microarray analysis techniques), astrophysics, satellite imagery (which has a broad application from meteorology, online mapping to intelligence assessment).

Chapter 2

Orthogonal Polynomials

The *classical orthogonal polynomials* play an important role in random matrix theory. Their weight functions ω_β appear in the joint eigenvalue density of the classical ensembles

$$\widehat{\mathbb{P}}_p(\mathrm{d}\lambda_1, \ldots, \mathrm{d}\lambda_p) = \frac{1}{\widehat{Z}_L^{\beta,p}} \prod_{1 \le i < j \le p} |\lambda_j - \lambda_i|^\beta \prod_{1 \le j \le p} \omega_\beta(\lambda_j) \mathrm{d}\lambda_j \tag{2.1}$$

on the space of ordered eigenvalues

$$\mathbb{R}^p_{\ge} := \mathbb{R}^p \cap W_p, \qquad W_p := \{(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p : \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p\}$$

where $\widehat{Z}_L^{\beta,p}$ is the normalisation constant for ordered eigenvalues. In the case of the unitary ensembles ($\beta = 2$) one can exploit the properties of the Vandermonde determinant and rewrite the density as a determinant in terms of orthogonal polynomials. Thus we call this ensemble the *orthogonal polynomial ensemble*.

It turns out that for *Wishart matrices* $W = X^*X$ the corresponding orthogonal polynomials are the *generalized Laguerre polynomials* with weights

$$\omega_{\beta,\alpha}(\lambda) = \lambda^\alpha e^{-\beta\lambda/2}, \qquad \alpha = \frac{\beta}{2}(n - p + 1) - 1 \tag{2.2}$$

where $n \times p$ is the dimension of the Gaussian matrix X. We will come back later in the chapter on the Wishart distribution to this.

The goal of this chapter is to state some important results from the theory of orthogonal polynomials and about Laguerre polynomials. We introduce the Airy and Bessel functions.

This chapter is mostly based on [Dei00],[Sze75], [Koo13] and [Olv97].

Notation and assumptions: Throughout this chapter we use the following notation

$p_n = \kappa_n x^n +$ lower order terms	for orthonormal polynomial
$\pi_n = x^n +$ lower order terms	for monic polynomial

2.1 Orthogonal polynomials

Definition 2 (Orthogonal polynomials) [Dei00, Ch. 3, p. 37]**.** *Given a positive Borel measure μ on $S \subseteq \mathbb{R}$ with finite moments, i.e.*

$$\mu_n := \int_S |x|^n \mathrm{d}\mu < \infty, \quad \text{for } n = 0, 1, 2, \ldots.$$

We call the set of real polynomials $\{p_n(x)\}$ orthogonal polynomials if $\deg(p_n) = n$ *for $n = 0, 1, 2, \ldots$ and*

$$\int_S p_n(x) p_s(x) \mathrm{d}\mu = c_n \delta_{ns} \tag{2.3}$$

where δ_{ns} is the Kronecker delta. If μ is absolutely continuous we call the density $\omega(x)$ a weight function. If $c_n = 1$ for all n we call $\{p_n\}$ orthonormal.

Theorem 1 (Construction) [Gau04, Ch. 1, p. 3]**.** *Let $S \subseteq \mathbb{R}$ and the inner product (2.3) is positive definite on the space of polynomials $\mathbb{P}$, then there exist a unique infinite sequence $\{\pi_k\}$ of monic orthogonal polynomials.*

Proof. We can construct such polynomials by orthogonalizing the monomials of x

$$1, x, x^2, \ldots, x^n, \ldots.$$

We define the Gram-Schmidt process recursively with $\pi_0 = 1$ and

$$\pi_k = x^k - \sum_{l=0}^{k-1} a_l \pi_l, \qquad a_l = \frac{\langle x^k, \pi_l \rangle}{\|\pi_l\|^2}. \tag{2.4}$$

By the positivity of the scalar product $\langle \pi_l, \pi_l \rangle > 0$ the polynomial π_k is uniquely defined and orthogonal by construction to all polynomials π_j for $j \neq k$. □

2.1.1 The integral representation of polynomials

Let M_n be the *matrix of moments* for a measure μ and denote it's determinant as D_n, so

$$D_n := \det(M_n) = \det\left(\int x^{i+j} \mathrm{d}\mu\right)_{0 \leq i,j \leq n} = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ldots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ \mu_n & \mu_{n-1} & \cdots & \mu_{2n} \end{vmatrix}. \tag{2.5}$$

Remark 5: *(Positive-definiteness) M_n is positive-definite. We notice that for vector $(t_0, \ldots, t_n) \in \mathbb{R}^n$ that is note the zero vector we have*

$$0 < \left\| \sum_{i=0}^{n} x^i t_i \right\|_{L^2(S,\mu)} = \int_S \left(\sum_{i=0}^{n} x^i t_i \right)^2 \mathrm{d}\mu = \sum_{i=0}^{n} \sum_{j=0}^{n} t_i t_j \int_S x^{i+j} \mathrm{d}\mu = \langle t, M_n t \rangle.$$

Let M_n^0 be the matrix of moments where the last row is replaced by $(1, x, x^2, \ldots, x^n)$, then define the determinant function $D_n^0(x)$ as

$$D_n^0(x) := \det(M_n^0) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \vdots & \vdots & \cdots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ 1 & x & \cdots & x^n \end{vmatrix}. \tag{2.6}$$

so that $D_n^0(x)$ is a function where the rows $0, \ldots, n-1$ are fix. The ^{0} in the notation is to distinguish M_n from M_n^0 resp. D_n from $D_n^0(x)$.

Lemma 1 [Dei00, Ch. 3, p. 38]**.** *Let $\{p_n\}$ be orthonormal polynomials with corresponding measure μ. Then*

$$p_n(x) = \frac{1}{\sqrt{D_{n-1}D_n}} D_n^0(x). \tag{2.7}$$

Proof. Define $\tilde{p}_n(x) := \frac{1}{\sqrt{D_{n-1}D_n}} D_n^0(x)$ and observe that for $j < n$ by Laplace expansion

$$\sqrt{D_{n-1}D_n} \int s^j \tilde{p}_n(s) \mathrm{d}\mu = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \cdots & \vdots \\ \mu_{n-1} & \mu_n & \cdots & \mu_{2n-1} \\ \int s^j \mathrm{d}\mu & \int s^{j+1} \mathrm{d}\mu & \cdots & \int s^{j+n} \mathrm{d}\mu \end{vmatrix} \tag{2.8}$$

Because $j < n$ we have $j + n \leq 2n - 1$ and so the last row is must be equal to one of the rows above and therefore the determinant is 0 which implies that $\int s^j \tilde{p_n}(s) \mathrm{d}\mu = 0$ for all $j < n$. Thus we have the orthogonality condition and only have to show that $\tilde{p}_n$ are orthonormal polynomials. By (2.8) and using Laplace expansion on the last row we see that

$$\begin{aligned} \sqrt{D_{n-1}D_n} \int \tilde{p}_n(s) \mathrm{d}\mu &= s^n \det\left(\int x^{i+j} \mathrm{d}\mu \right)_{0 \leq i,j \leq n-1} + \ldots \\ &= s^n D_{n-1} + \cdots \end{aligned}$$

which proves that the left-hand side is a polynomial since $D_{n-1} > 0$. On the other hand by the definition of $\tilde{p}_n$

$$\begin{aligned} \sqrt{D_{n-1}D_n} \int \left(\tilde{p}_n(s)\right)^2 \mathrm{d}\mu &= \int \left(D_n^0(s)\right)^2 \mathrm{d}\mu \\ &= \int \left(s^n D_{n-1} + \cdots\right) D_n^0(s) \mathrm{d}\mu \\ &= D_{n-1} \int s^n D_n^0(s) \mathrm{d}\mu \qquad \left(\text{as } \int s^j D_n \mathrm{d}\mu = 0,\ j < n \right) \\ &= D_{n-1} D_n \end{aligned}$$

□

Theorem 2 (Heine's integral representation) [Dei00, Ch. 3, p. 39]**.** *Let* $\{p_n\}$ *be orthonormal polynomials with corresponding measure* μ *and* D_n *denote the determinant of the matrix of moments, then Heine's integral representation is true for* $n \geq 1$*:*

$$D_n = \frac{1}{(n+1)!} \underbrace{\int \cdots \int}_{n+1} \prod_{0 \leq i < j \leq n} (x_i - x_j)^2 \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_n) \tag{2.9}$$

$$p_n(x) = \frac{(D_{n-1} D_n)^{-1/2}}{n!} \underbrace{\int \cdots \int}_{n} \prod_{i=0}^{n-1} (x - x_j)^2 \prod_{0 \leq i < j \leq n} (x_i - x_j)^2 \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_{n-1}). \tag{2.10}$$

Proof. Using Leibniz's determinant formula and the multilinearity of the determinant

$$\begin{aligned}
D_n = \det \left(\int x_j^{i+j} \mathrm{d}\mu \right)_{0 \leq i,j \leq n} &= \sum_{\tau \in \mathcal{S}_n} \operatorname{sgn}(\tau) \prod_{i=0}^{n} \int x_i^{i+\tau(i)} \mathrm{d}\mu(x_i) \\
&= \underbrace{\int \cdots \int}_{n+1} \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_n) \sum_{\tau \in \mathcal{S}_n} \operatorname{sgn}(\tau) \prod_{i=0}^{n} x_i^{i+\tau(i)} \\
&= \int \cdots \int \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_n) \det \left(x_j^{i+j} \right)_{0 \leq i,j \leq n} \\
&= \int \cdots \int \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_n) \det \left(x_j^{i} \right)_{0 \leq i,j \leq n} x_0^0 \cdots x_n^n
\end{aligned}$$

so we have reduced the degree inside the determinant then

$$\det \left(x_j^i \right)_{0 \leq i,j \leq n} = \begin{vmatrix} x_0^0 & x_1^0 & \cdots & x_1^0 \\ x_0^1 & x_1^1 & \cdots & x_1^1 \\ \vdots & \vdots & \cdots & \vdots \\ x_0^n & x_1^n & \cdots & x_n^n \end{vmatrix}$$

Now observe that for any of the $(n+1)!$ permutations $\sigma \in \mathcal{S}_{n+1}$

$$\begin{aligned}
D_n &= \underbrace{\int \cdots \int}_{n+1} \mathrm{d}\mu(x_{\sigma_{(0)}}) \cdots \mathrm{d}\mu(x_{\sigma_{(n)}}) \det \left(x_{\sigma(j)}^i \right)_{0 \leq i,j \leq n} x_{\sigma(0)}^0 \cdots x_{\sigma(n)}^n \\
&= \int \cdots \int \mathrm{d}\mu(x_0) \cdots \mathrm{d}\mu(x_n) \det \left(x_j^i \right)_{0 \leq i,j \leq n} \operatorname{sgn}(\sigma) x_{\sigma(0)}^0 \cdots x_{\sigma(n)}^n
\end{aligned}$$

Therefore we take the sum

$$\begin{aligned}
D_n &= \frac{1}{(n+1)!}\sum_{\sigma\in\mathcal{S}_{n+1}}\int\cdots\int \mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_n)\det\left(x_j^i\right)_{0\le i,j\le n}\operatorname{sgn}(\sigma)x_{\sigma(0)}^0\cdots x_{\sigma(n)}^n\\
&= \frac{1}{(n+1)!}\int\cdots\int \mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_n)\det\left(x_j^i\right)_{0\le i,j\le n}\sum_{\sigma\in\mathcal{S}_{n+1}}\operatorname{sgn}(\sigma)x_{\sigma(0)}^0\cdots x_{\sigma(n)}^n\\
&= \frac{1}{(n+1)!}\int\cdots\int \mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_n)\left[\det\left(x_j^i\right)_{0\le i,j\le n}\right]^2\\
&= \frac{1}{(n+1)!}\int\cdots\int \mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_n)\prod_{0\le i<j\le n}(x_i-x_j)^2
\end{aligned}$$

where we have used the Vandermonde determinant formula in the last line which proves (2.9). For (2.10) we apply the exact same reasoning as above to $D_n^0(x)$ but this time by x_i^n we denote the x^n in row i. and observe that for any of the $n!$ permutations $\sigma\in\mathcal{S}_n$

$$\begin{aligned}
D_n^0(x) &= \underbrace{\int\cdots\int}_{n}\mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_{n-1})\begin{vmatrix} x_0^0 & \cdots & x_0^n\\ \vdots & \vdots & \vdots\\ x_{n-1}^0 & \cdots & x_{n-1}^n\\ 1 & \cdots & x^n\end{vmatrix}\operatorname{sgn}(\sigma)x_{\sigma(0)}^0\cdots x_{\sigma(n-1)}^{n-1}\\
&= \frac{1}{n!}\underbrace{\int\cdots\int}_{n}\mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_{n-1})\begin{vmatrix} x_0^0 & \cdots & x_0^n\\ \vdots & \vdots & \vdots\\ x_{n-1}^0 & \cdots & x_{n-1}^n\\ 1 & \cdots & x^n\end{vmatrix}\det\left(x_i^j\right)_{0\le i,j\le n-1}\\
&= \frac{1}{n!}\underbrace{\int\cdots\int}_{n}\mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_{n-1})\underbrace{\begin{vmatrix} 1 & x_0 & \cdots & x_0^n\\ \vdots & \vdots & \vdots & \vdots\\ 1 & x_{n-1} & \cdots & x_{n-1}^n\\ 1 & x_n & \cdots & x^n\end{vmatrix}}_{:=\det(A)}\det\left(x_i^j\right)_{0\le i,j\le n-1}
\end{aligned}$$

Observe that $\det(A)$ is a Vandermonde determinant with variables $x_0, x_1, \ldots, x_{n-1}, x$ and since $\det(A)=\det(A^T)$ we have $\det(A)=\prod_{i=0}^{n-1}(x-x_i)\prod_{0\le i<j\le n-1}(x_i-x_j)$ and therefore

$$D_n^0(x) = \frac{1}{n!}\underbrace{\int\cdots\int}_{n}\mathrm{d}\mu(x_0)\cdots\mathrm{d}\mu(x_{n-1})\prod_{i=0}^{n-1}(x-x_i)\prod_{0\le i,j\le n-1}(x_i-x_j)^2$$

Then by (2.7)

$$p_n = \frac{1}{\sqrt{D_nD_{n-1}}}D_n^0(x)$$

which proves (2.9). □

Remark 6: *(Use in RMT) Heine's integral representation is a useful tool for computing integrals involving the Vandermonde determinant. It can be used to compute the normalisation constant of the orthogonal polynomial ensembles. For general β one needs the more sophisticated Selberg integral formula.*

2.1.2 Three-term recurrence relation:

Orthogonal polynomials satisfy a three-term recurrence which is an important tool for construction and computation.

Theorem 3 (Three-term recurrence relation) [Koo13, p. 3]**.** *A sequence of orthonormal polynomials $\{p_n\}$ satisfy the three-term recurrence for $n > 0$*

$$\begin{aligned} xp_n(x) &= \alpha_n p_{n+1}(x) + \beta_n p_n(x) + \alpha_{n-1} p_{n-1}(x) \\ xp_0(x) &= a_0 p_0(x) + b_0 p_0(x) \end{aligned} \tag{2.11}$$

with real constants α_n and β_n and

$$\alpha_n = \frac{\kappa_n}{\kappa_{n+1}} \tag{2.12}$$

Proof. We can rewrite xp_n with an orthogonal polynomial basis

$$xp_n = \sum_{k=0}^{n+1} \mu_k p_k(x)$$

and since for $j < n-2$ by orthogonality

$$0 = \langle xp_n, p_j \rangle = \sum_{k=0}^{n+1} \mu_k \langle p_k, p_j \rangle = \mu_j \langle p_j, p_j \rangle$$

it must follow that $\mu_j = 0$ for $j < n-2$ and hence

$$xp_n = \alpha_n p_{n+1}(x) + \beta_n p_n(x) + \gamma_n p_{n-1}(x)$$

for some constants $\alpha_n,\ \beta_n,\ \gamma_n$. By orthogonality

$$\langle xp_n, p_n \rangle = \sum_{k=0}^{n+1} \mu_k \langle p_k, p_n \rangle = \mu_n \langle p_n, p_n \rangle = \gamma_{n+1} \langle p_n, p_n \rangle$$

and so

$$\gamma_{n+1} = \frac{\langle xp_{n+1}, p_n \rangle}{\langle p_n, p_n \rangle} = \frac{\langle xp_{n+1}, p_n \rangle}{\langle p_{n+1}, p_{n+1} \rangle} = \alpha_n.$$

□

Remark 7: *(Favard's theorem) The inverse of theorem 3 for general orthogonal polynomials:*

$\{p_n\}$ *satisfy the three-term recurrence* $\implies$ $\{p_n\}$ *are orthogonal polynomials.*

is also true and called Favard's theorem (see for example [Chi78]).

The Christoffel-Darboux formula

The Christoffel-Darboux formula will play a crucial when we rewrite the density of the eigenvalues for the unitary ensembles.

Theorem 4 (Christoffel-Darboux formula) [Koo13, p. 3]**.** *Let $\{p_k\}$ be orthonormal polynomials with respect to the positive measure μ with density $\omega(x)$. Then*

$$\sum_{k=0}^{n} p_k(x)p_k(y) = \frac{\kappa_n}{\kappa_{n+1}} \frac{p_{n+1}(x)p_n(y) - p_{n+1}(y)p_n(x)}{x-y} \tag{2.13}$$

and

$$\sum_{k=0}^{n} p_k(x)p_k(x) = \frac{\kappa_n}{\kappa_{n+1}} \left(p'_{n+1}(x)p_n(x) - p_{n+1}(x)p'_n(x)\right) \tag{2.14}$$

Proof. Using theorem 3 we multiply $p_n(y)$ to the three-term recurrence relation with the variable x and rearrange it

$$xp_n(y)p_n(x) - \beta_n p_n(y)p_n(x) - \alpha_{n-1}p_n(y)p_{n-1}(x) = \alpha_n p_n(y)p_{n+1}(x)$$

Now we subtract the same thing but with x and y interchanged

$$\begin{aligned}(x-y)p_n(y)p_n(x) = \alpha_n &\left(p_n(y)p_{n+1}(x) - p_n(x)p_{n+1}(y)\right) \\ &- \alpha_{n-1}\left(p_n(x)p_{n-1}(y) - p_n(y)p_{n-1}(x)\right)\end{aligned}$$

and sum it

$$\begin{aligned}(x-y)\sum_{k=0}^{n} p_k(y)p_k(x) = \sum_{k=0}^{n} &\alpha_k \left(p_{k+1}(x)p_k(y) - p_{k+1}(y)p_k(x)\right) \\ &- \alpha_{k-1}\left(p_k(x)p_{k-1}(y) - p_k(y)p_{k-1}(x)\right)\end{aligned}$$

Since $p_{-1} = 0$ we realize that the right-hand side is a telescope sum and

$$\begin{aligned}\sum_{k=0}^{n} p_k(y)p_k(x) &= \alpha_n \frac{\left(p_n(y)p_{n+1}(x) - p_n(x)p_{n+1}(y)\right)}{(x-y)} \\ &= \frac{\kappa_n}{\kappa_{n+1}} \frac{\left(p_n(y)p_{n+1}(x) - p_n(x)p_{n+1}(y)\right)}{(x-y)}\end{aligned}$$

where we have used that $\alpha_n = \frac{\kappa_n}{\kappa_{n+1}}$. Now taking the limit $x \to y$ we get the second identity. □

2.2 Laguerre polynomials

As we have briefly mentioned the relevant orthogonal polynomials for the Wishart ensemble are the *generalized Laguerre polynomials*. We will generally just called them *Laguerre polynomials*. The goal of this section is to derive some important properties of the Laguerre polynomials and their differential equations.

Definition 3 (Laguerre polynomials) [Olv97, Chapter 6, p. 48]**.** *The generalized Laguerre polynomials* $\{L_n^{(\alpha)}(x)\}$ *for* $\alpha > -1$ *are the orthogonal polynomials defined through*

$$S = [0, \infty], \qquad \omega(x) = e^{-x}x^{\alpha}, \qquad a_{n,n} = (-1)^n/n! \tag{2.15}$$

Proposition 1 [Olv97, Chapter 2, p. 49]**.** *The following holds for the Laguerre polynomials:*

1. *Rodrigues formula:*

$$L_n^{(\alpha)}(x) = \frac{e^x x^{-\alpha}}{n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}(e^{-x}x^{n+\alpha}) \tag{2.16}$$

2. *Normalisation:*

$$c_{n,n} = \langle L_n^{(\alpha)}, L_n^{(\alpha)}\rangle = \frac{\Gamma(n+\alpha+1)}{n!} \tag{2.17}$$

Proof. Let use abbreviate the differential operator $\mathrm{D}^n := \frac{\mathrm{d}^n}{\mathrm{d}x^n}$ and define

$$\phi_n(x) := \frac{e^x x^{-\alpha}}{n!}\,\mathrm{D}^n(e^{-x}x^{n+\alpha})$$

and $p(x)$ be any polynomial. Then we use repeated integration by parts on the following

$$\begin{aligned}
&\int_0^\infty e^{-x}x^{\alpha}\phi_n(x)p(x)\mathrm{d}x \\
&\qquad = \frac{1}{n!}\int_0^\infty \mathrm{D}^n(e^{-x}x^{n+\alpha})p(x)\mathrm{d}x \\
&\qquad = \frac{1}{n!}\left(\left[\mathrm{D}^{n-1}(e^{-x}x^{n+\alpha})p(x)\right]_0^\infty - \int_0^\infty \mathrm{D}^{n-1}(e^{-x}x^{n+\alpha})p^{(1)}(x)\mathrm{d}x\right) \\
&\qquad = \frac{1}{n!}\Bigg(\left[\mathrm{D}^{n-1}(e^{-x}x^{n+\alpha})p(x)\right]_0^\infty - \left[\mathrm{D}^{n-2}(e^{-x}x^{n+\alpha})p^{(1)}(x)\right]_0^\infty \\
&\qquad\qquad + \int_0^\infty \mathrm{D}^{n-2}((e^{-x}x^{n+\alpha})p^{(2))}\mathrm{d}x\Bigg)
\end{aligned}$$

and continue and notice that

$$\left[\mathrm{D}^{n-1}(e^{-x}x^{n+\alpha})p(x)\right]_0^\infty = \left[\mathrm{D}^{n-2}(e^{-x}x^{n+\alpha})p^{(1)}(x)\right]_0^\infty = \cdots = 0$$

so

$$\int_0^\infty e^{-x}x^{\alpha}\phi_n(x)p(x)\mathrm{d}x = \frac{(-1)^n}{n!}\int_0^\infty e^{-x}x^{n+\alpha}p^{(n)}(x)\mathrm{d}x \tag{2.18}$$

and $p^{(n)} = 0$ for any polynomial with degree $< n$, thus ϕ_n must be orthogonal with respect to the Laguerre weight $\omega(x) = e^{-x}x^{\alpha}$. And if we expand ϕ_n in descending

powers of x it must be clear that the x^n coefficient must be $(-1)^n/n!$ which proves the Rodrigues formula. Next choosing $p(x) = \phi_n(x)$ and using (2.18)

$$\begin{aligned} c_{n,n} = \int_0^\infty w(x) L_n^{(\alpha)} L_n^{(\alpha)} \mathrm{d}x &= \frac{(-1)^n}{n!} \int_0^\infty e^{-x} x^{n+\alpha} p^{(n)}(x) \mathrm{d}x \\ &= \frac{1}{n!} \int_0^\infty e^{-x} x^{n+\alpha} \mathrm{d}x \\ &= \frac{\Gamma(n+\alpha+1)}{n!} \end{aligned}$$

□

2.2.1 The three-term recurrence relation and the differential equation

Our goal is to derive a three-term recurrence formula for the Laguerre polynomials $\{L_n^\alpha\}$ and to show that they are the solution of the *Laguerre differential equation*

$$xy'' + (\alpha + 1 - x)y' + ny = 0. \tag{2.19}$$

We first derive a series representation and a useful way to deal with derivatives. The series representation gives us the possibility to derive the generating function which we will need for the three-term recurrence relation. With all the tools collected we show that the Laguerre polynomials solve the ODE.

Lemma 2 (Serie representation) [Sze75, p.101]**.** *We have the following series representation*

$$L_n^\alpha(x) = \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} \tag{2.20}$$

and for derivatives and $k \leq n$

$$L_n^{(\alpha)}(x) = (-1)^k L_{n-k}^{(\alpha+k)}(x). \tag{2.21}$$

Proof. By using the general Leibniz rule on (2.16)

$$\begin{aligned} L_n^\alpha(x) &= \frac{x^{-\alpha} e^x}{n!} \sum_{k=0}^n (-1)^k e^{-x} \binom{n}{k} (n-k)! \binom{n+\alpha}{n-k} x^{n+\alpha-(n-k)} \\ &= \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{(n-k)!}{n!} \binom{n+\alpha}{n-k} x^k \\ &= \sum_{k=0}^n (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} \end{aligned}$$

which gives (2.20). For (2.21) notice the following

$$L_n'^\alpha(x) = \sum_{k=1}^n (-1)^k \binom{n+\alpha}{n-k} \frac{k x^{k-1}}{k!} = \sum_{k=0}^n (-1)^{k+1} \binom{n-1+\alpha+1}{n-1-k} \frac{x^k}{k!} = -L_{n-1}^{\alpha+1}(x)$$

and repeat $k-1$ times. □

Theorem 5 (Generating function) [Sze75, p.101]**.** *The generating function of the Laguerre polynomial $L_n^{(\alpha)}$ is*

$$G(t,x) = \sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x) = \frac{1}{(1-t)^{\alpha+1}} e^{-tx/(1-t)}. \tag{2.22}$$

Proof. We have

$$\sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} t^n$$

and

$$\frac{1}{(1-t)^{\alpha+1}} e^{-tx/(1-t)} = \sum_{k=0}^{\infty} \frac{(-tx)^k}{k!} \frac{1}{(1-t)^{k+\alpha+1}}$$

We want to rewrite the last factor into a series. We can do this by taking the derivatives of the geometric series

$$\frac{\mathrm{d}^{k+\alpha}}{\mathrm{d}t^{k+\alpha}} \sum_{n=0}^{\infty} t^n = \frac{\mathrm{d}^{k+\alpha}}{\mathrm{d}t^{k+\alpha}} \frac{1}{(1-t)} = \frac{(-1)^{k+\alpha}(-1)^{k+\alpha}(k+\alpha)!}{(1-t)^{k+\alpha+1}} = \frac{(k+\alpha)!}{(1-t)^{k+\alpha+1}}$$

Now we compute the left-hand side

$$\begin{aligned} \frac{\mathrm{d}^{k+\alpha}}{\mathrm{d}t^{k+\alpha}} \sum_{n=0}^{\infty} t^n &= \sum_{n=k+\alpha}^{\infty} n(n-1)\cdots(n-k-\alpha+1) t^{n-k-\alpha} \\ &= \sum_{n=k+\alpha}^{\infty} \frac{n!}{(n-k-\alpha)!} t^{n-k-\alpha} = \sum_{n=k}^{\infty} \frac{(n+\alpha)!}{(n-k)!} t^{n-k} \end{aligned}$$

Combining both we get

$$\frac{1}{(1-t)^{k+\alpha+1}} = \frac{1}{(k+\alpha)!} \sum_{n=k}^{\infty} \frac{(n+\alpha)!}{(n-k)!} t^{n-k}$$

and hence

$$\begin{aligned} \frac{1}{(1-t)^{\alpha+1}} e^{-tx/(1-t)} &= \sum_{k=0}^{\infty} \frac{(-tx)^k}{k!} \frac{1}{(k+\alpha)!} \sum_{n=k}^{\infty} \frac{(n+\alpha)!}{(n-k)!} t^{n-k} \\ &= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} (-1)^k \frac{(n+\alpha)!}{(n-k)!(k+\alpha)!} \frac{x^k}{k!} t^n \\ &= \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} t^n \end{aligned}$$

And by symmetry of summation we have

$$\sum_{k=0}^{\infty} \sum_{n=k}^{\infty} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} t^n = \sum_{n=0}^{\infty} \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{n-k} \frac{x^k}{k!} t^n = \sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x)$$

□

Corollary 1 (Contour integral representation)**.** *The Laguerre polynomial $L_n^{(\alpha)}$ can be expressed as a contour integral*

$$L_n^{(\alpha)}(x) = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{e^{-xt/(1-t)}}{(1-t)^{\alpha+1}\, t^{n+1}} \mathrm{d}t \tag{2.23}$$

where the contour circle $\mathcal{C}$ is counterclockwise oriented around 0 in $\mathbb{C}$.

Proof. By Cauchy's integral formula and Taylors theorem we have for an analytic function $g(t) = \sum_{n=0}^{\infty} a_n (t-t_0)^n$ that $a_n = (n!)^{-1} g^{(n)}(t_0)$ and

$$\frac{1}{2\pi \mathrm{i}} \oint_C \frac{g(t)}{(t-t_0)^{n+1}}\, dt = \frac{1}{2\pi \mathrm{i}} \oint_C \frac{a_n}{t-t_0}\, dt = a_n,$$

So choose $g(t) = G(t,x)$ then $t_0 = 0$ and $a_n = L_n^{(\alpha)}(x)$. □

Laguerre three-term recurrence relation

Theorem 6 (Recurrence relation) [Sze75, p.101]**.** *The Laguerre polynomials $\{L_n^{(\alpha)}\}$ satisfy the following three-term recurrence for $n \geq 1$*

$$(n+1)L_{n+1}^{(\alpha)}(x) = (\alpha + 1 - x + 2n)L_n^{(\alpha)}(x) - (\alpha + n)L_{n-1}^{(\alpha)}(x) \tag{2.24}$$

Proof. We take the derivative of the generating function (2.22) with respect to t

$$\begin{aligned}\sum_{n=0}^{\infty} n t^{n-1} L_n^{(\alpha)}(x) &= \left(\frac{\alpha+1}{(1-t)} + \frac{-x}{(1-t)^2}\right) \frac{1}{(1-t)^{\alpha+1}} e^{-tx/(1-t)} \\ &= \left(\frac{\alpha+1}{(1-t)} + \frac{-x}{(1-t)^2}\right) \sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x).\end{aligned}$$

Multiply both sides with $(1-t)^2$

$$(1-2t+t^2)\sum_{n=0}^{\infty} n t^{n-1} L_n^{(\alpha)}(x) = (\alpha - \alpha t + 1 - t - x)\sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x) \tag{2.25}$$

The left-hand side of (2.25) can be written in form of t^n as

$$\begin{aligned}&(1-2t+t^2)\sum_{n=0}^{\infty} n t^{n-1} L_n^{(\alpha)}(x) \\ &\qquad = \sum_{n=0}^{\infty}(n+1)t^n L_{n+1}^{(\alpha)}(x) - 2\sum_{n=0}^{\infty} n t^n L_n^{(\alpha)}(x) + \sum_{n=2}^{\infty}(n-1)t^n L_{n-1}^{(\alpha)}(x)\end{aligned}$$

and the right-hand side of (2.25) can be written in form of t^n as

$$\begin{aligned}&(\alpha - t(\alpha+1) + 1 - x)\sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x) \\ &\qquad = (\alpha + 1 - x)\sum_{n=0}^{\infty} t^n L_n^{(\alpha)}(x) - (\alpha+1)\sum_{n=1}^{\infty} t^n L_{n-1}^{(\alpha)}(x)\end{aligned}$$

We put everything together and drop the sum and divide by t^n to get

$$(n+1)L_{n+1}^{(\alpha)}(x) - 2nL_n^{(\alpha)}(x) + (n-1)L_{n-1}^{(\alpha)}(x) \\ = (\alpha+1-x)L_n^{(\alpha)}(x) - (\alpha+1)L_{n-1}^{(\alpha)}(x)$$

and simplify

$$(n+1)L_{n+1}^{(\alpha)}(x) = (\alpha+1-x+2n)L_n^{(\alpha)}(x) - (\alpha+n)L_{n-1}^{(\alpha)}(x).$$

□

Laguerre differential equation

Theorem 7 (Differential equation) [Koo13, p. 9]**.** *The Laguerre polynomials* $\{L_n^{(\alpha)}\}$ *are the regular solutions of the second order differential equation*

$$xy'' + (\alpha+1-x)y' + ny = 0 \tag{2.26}$$

with $x \in \mathbb{R}_+$.

Proof. Using the identity (2.21) for derivatives we can rewrite the equation as

$$-x\frac{\mathrm{d}}{\mathrm{d}x}L_{n-1}^{(\alpha+1)} - (\alpha+1-x)L_{n-1}^{(\alpha+1)} + nL_n^{(\alpha)} = 0$$

Then notice that

$$x^{-\alpha}e^x\frac{\mathrm{d}}{\mathrm{d}x}\left(x^{\alpha+1}e^{-x}L_{n-1}^{\alpha+1}(x)\right) = (\alpha+1)L_{n-1}^{\alpha+1}(x) - xL_{n-1}^{\alpha+1}(x) + x\frac{\mathrm{d}}{\mathrm{d}x}L_{n-1}^{\alpha+1}(x) \\ = (\alpha+1-x)L_{n-1}^{\alpha+1}(x) + x\frac{\mathrm{d}}{\mathrm{d}x}L_{n-1}^{\alpha+1}(x)$$

and also that

$$x^{-\alpha}e^x\frac{\mathrm{d}}{\mathrm{d}x}\left(x^{\alpha+1}e^{-x}L_{n-1}^{\alpha+1}(x)\right) = x^{-\alpha}e^x\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x^{\alpha+1}e^{-x}e^x x^{-\alpha+1}}{(n-1)!}\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}}(e^{-x}x^{n+\alpha})\right) \\ = nL_n^{(\alpha)}$$

which proves the equation. □

2.3 Special functions

2.3.1 Bessel function

Definition 4 (Complex Bessel function) [Olv97, Chapter 2, p. 55]**.** *Let* $v \in \mathbb{Z}$ *then the complex Bessel function is defined as*

$$J_v(z) = \frac{1}{\pi}\int_0^\pi \cos(v\theta - z\sin(\theta))\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^\pi \exp(-\mathrm{i}v\theta + \mathrm{i}z\sin(\theta))\mathrm{d}\theta. \tag{2.27}$$

Let $w := J_v(z)$ *then* w *satisfies the Bessel differential equation*

$$z^2w'' + zw' + (z^2 - v^2)w = 0. \tag{2.28}$$

Definition 5 (General complex Bessel function) [Olv97, Chapter 2, p. 57]**.** *Let* $v \in \mathbb{R}$ *we then define the general complex Bessel function as*

$$J_v(z) = \left(\frac{x}{2}\right)^v \sum_{m=0}^\infty \frac{(-1)^m}{m!\Gamma(m+v+1)}\left(\frac{x}{2}\right)^{2m}. \tag{2.29}$$

2.3.2 Airy function

Definition 6 (Complex Airy function) [Olv97, Chapter 2, p. 53]**.** *The complex Airy function is defined as*

$$\mathrm{Ai}(x) = \frac{1}{2\pi i}\int_C \exp\left(\frac{s^3}{3} - xs\right)\mathrm{d}s \tag{2.30}$$

where C is any contour that begins at a point in infinity in the sector S and ends at a point in infinity in the conjugate sector T

$$S := -\tfrac{1}{2}\pi \le \mathrm{Arg}(x) \le -\tfrac{1}{6}\pi, \qquad T := \tfrac{\pi}{6} \le \mathrm{Arg}(x) \le \tfrac{\pi}{2}.$$

Remark 8: *(Derivative) It's easy to see that*

$$\mathrm{Ai}'(x) = \frac{1}{2\pi i}\int_C \frac{\mathrm{d}}{\mathrm{d}x}\exp\left(\frac{s^3}{3} - xs\right)\mathrm{d}s = -\frac{1}{2\pi i}\int_C s\exp\left(\frac{s^3}{3} - xs\right)\mathrm{d}s.$$

One of the most important features of the Airy function is that it solves a second-order differential equation with basic form

$$y'' - xy = 0. \tag{2.31}$$

Then we notice that by Cauchy's integral theorem

$$\begin{aligned}\mathrm{Ai}''(x) - x\,\mathrm{Ai}(x) &= \frac{1}{2\pi i}\int_C \left(s^2 - x\right)\exp\left(\frac{s^3}{3} - xs\right)\mathrm{d}s \\ &= \frac{1}{2\pi i}\left[\exp\left(\frac{s^3}{3} - xs\right)\right]_C \\ &= 0.\end{aligned}$$

Other solutions are $\mathrm{Ai}(wx)$ and $\mathrm{Ai}(w^2x)$ for $w = \exp(2\pi\mathrm{i}/3)$ and we have

$$\mathrm{Ai}(x) + w\,\mathrm{Ai}(wx) + w^2\,\mathrm{Ai}(w^2x) = 0.$$

The function

$$\mathrm{Bi}(x) = \mathrm{i}w^2\,\mathrm{Ai}(w^2x) - \mathrm{i}w\,\mathrm{Ai}(wx) \tag{2.32}$$

is also a solution and named the *Airy function of second kind* and a real solution when x is real. The general solutions are of the form $y = \alpha\,\mathrm{Ai}(x) + \beta\,\mathrm{Bi}(x)$.

If we look at the figures 2.1 and 2.2 we can see that $\mathrm{Ai}(x)$ has an oscillatory region for negative x and a region with exponential-like decay for all positive x. The transitional point at $x = 0$ is called a turning point.

The connection to the Laguerre polynomials:

Comparing the Airy differential equation with the Laguerre differential equation

$$\begin{aligned} y'' - zy &= 0 \qquad \text{(Airy)} \\ xy'' + (\alpha + 1 - x)y' + ny &= 0 \qquad \text{(Laguerre)}\end{aligned}$$

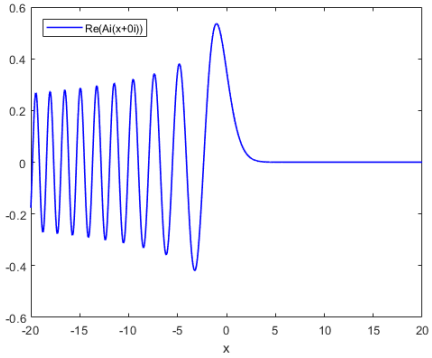

Fig. 2.1: Ai(x) at $x + 0$i.

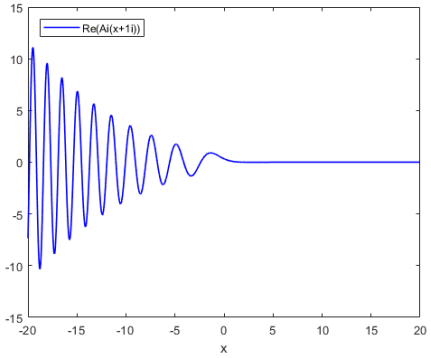

Fig. 2.2: Ai(x) at $x + 1$i.

it is clear that by eliminating the first derivative term of the Laguerre equation we get almost the same Airy equation. So make the substitution:

$$y = uf, \qquad y' = u'f + uf', \qquad y'' = u''f + 2u'f' + uf''$$

for some functions u and f, then

$$x\left(u''f + 2u'f' + uf''\right) + (\alpha + 1 - x)\left(u'f + uf'\right) + nuf = 0$$

group by order of u

$$xu''f + u'\left(2xf' + (\alpha + 1 - x)f\right) + u\left(xf'' + (\alpha + 1 - x)f' + nf\right) = 0 \tag{2.33}$$

In order for u' to disappear we must choose f in such a way that

$$2xf' + (\alpha + 1 - x)f = 0. \tag{2.34}$$

This is a simple homogeneous ODE that we can solve by separation of variables

$$2xf' = -(\alpha + 1 - x)f$$
$$\int \frac{f'}{f}\mathrm{d}x = -\frac{1}{2}\int \frac{(\alpha + 1 - x)}{x}\mathrm{d}x$$

The solution are the inverse of the squared weights of the Laguerre polynomial with a shift in the α parameter

$$f(x) = Ce^{-\frac{1}{2}\int \frac{(\alpha+1-x)}{x}\mathrm{d}x} = Cx^{-\frac{1}{2}(\alpha+1)}e^{\frac{1}{2}x}.$$

Since u is not defined we choose $C = 1$. We abbreviate $g(x) := \frac{\alpha+1-x}{x}$ then

$$f = \exp\left(-\frac{1}{2}\int g(x)\mathrm{d}x\right)$$
$$f' = -\frac{1}{2}g(x)\exp\left(-\frac{1}{2}\int g(x)\mathrm{d}x\right) = -\frac{1}{2}g(x)f(x)$$
$$f'' = -\frac{1}{2}g'(x)f(x) + \frac{1}{4}g^2(x)f(x).$$

Now divide (2.33) by x and replace the derivatives of f to get

$$u''f + u\left(f'' + \frac{(\alpha+1-x)}{x}f' + \frac{n}{x}f\right) = u''f - u\left(\frac{1}{2}g'f - \frac{1}{4}g^2f + \frac{1}{2}g^2f - \frac{n}{x}f\right)$$

and divide by f to get an Airy-type equation

$$u'' - u\left(\frac{1}{2}g' + \frac{1}{4}g^2 - \frac{n}{x}\right) = 0. \tag{2.35}$$

We can transform the Laguerre equation into an Airy-type equation

$$xy'' + (\alpha+1-x)y' + ny \qquad \rightsquigarrow \qquad u'' = h(x)u \tag{2.36}$$

with

$$\begin{aligned} h(x) :&= \frac{1}{2}g'(x) + \frac{1}{4}g^2(x) - \frac{n}{x} \\ &= -\frac{\alpha+1}{2x^2} + \frac{1}{4}\frac{(\alpha+1-x)^2}{x^2} - \frac{n}{x} \end{aligned} \tag{2.37}$$

so the general formula to transform such a differential equation into an Airy-type equation is

$$\begin{aligned} y'' + g(x)y + q(x)y \qquad &\rightsquigarrow \qquad u'' = h(x)u \\ h(x) &:= \frac{1}{2}g'(x) + \frac{1}{4}g^2(x) - q(x). \end{aligned}$$

In Fig. 2.3 and Fig. 2.4 we plot the product $L_n^{(\alpha)}(x)f^{-1}(x)$ for $n = 10, \alpha = 0$ and $n = 20, \alpha = 0$. We indeed obtain an Airy-type function but shifted to the positive axis with turning point t near 40 respectively 80 so this suggests that t is close to $4n$. The factor $4n$ is the neighbourhood of the largest zero and used to derive Airy asymptotics for a scaling $x_0 =: (4n + \alpha + 2)$ and $L_n^{(\alpha)}(x)$.

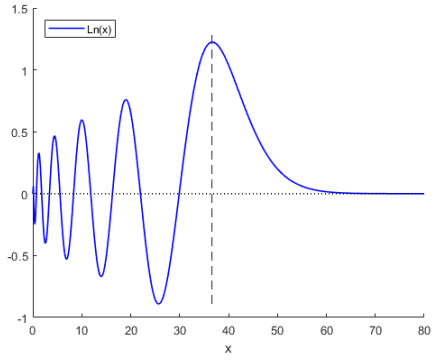

Fig. 2.3: $L_n^{(\alpha)}f^{-1}$ with $n = 10$.

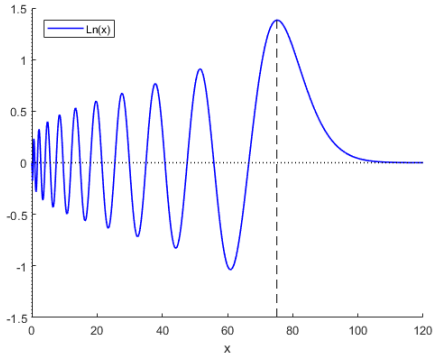

Fig. 2.4: $L_n^{(\alpha)}f^{-1}$ with $n = 20$.

Chapter 3

Wishart Ensemble

In this chapter we introduce the β-Wishart ensemble and learn how to compute the eigenvalue density. The first goal is to give an overview over the different Wishart cases and see different techniques that can be applied. Then we look at the special case of the Wishart unitary ensemble and use the orthogonal polynomial method to bring the spectral density into a determinantal form that will be needed to derive the asymptotic spectral density for an infinite matrix. This will lead us to the concept of determinantal point processes. We will make a case study and follow this map:

White β-Wishart	Correlated β-Wishart
Multivariate statistics	Broken-arrow models
Tridiagonal models	Jack polynomials
	HCIZ-Integral

Wishart Unitary Ensemble

Orthogonal polynomial method
Determinantal point processes

The theory regarding the Wishart distribution and the matrix models is mostly based on [GN00], [DE02], [DEKV13] and [Jam64]. The results for the WUE and DPP are mostly from [Dei00],[BDT16], [LNV18], [Joh05].

Notation and assumptions: Throughout this chapter we are going to assume that $n, p \in \mathbb{N}$. We follow the statistical convention that data vectors X_i of size p are rows of the data matrix X and so $\dim(X^T X) = p \times p$. In random matrix literature one sometimes uses the variable N or m instead of p (e.g. Dumitriu-Edelman [DE02]) therefore we summarize the changes:

	Random Matrix Theory	Multivariate Statistics
Sample size:	n	n
Covariates:	N	p
Wishart dimension:	$N \times N$	$p \times p$
Data matrix:	$n \times N$	$n \times p$

3.1 The Wishart distribution

There are many different ways to approach the Wishart distribution but it is a good introduction to talk first about some of the characteristics that matrix variate distributions can have by analysing the family of Wishart distributions. But first two important examples where the Wishart distribution appears in statistics:

Multivariate statistics

It arises as the distribution of the sample covariance matrix $\widehat{\Sigma}_p$ when one samples from a multivariate normal distribution. We will prove this later.

Bayesian inference

If Σ_p follows a Wishart distribution then Σ_p^{-1} follows a *inverse-Wishart* distribution. In Bayesian statistics if the *posterior distribution* is a matrix normal distribution with density $f(\Sigma_p|X)$ then the *conjugate prior* follows the inverse-Wishart distribution

$$f(\Sigma_p|X) = \frac{f(X|\Sigma_p)f(\Sigma_p)}{f(X)} \qquad \text{(posterior)} = \frac{\text{(likelihood)(prior)}}{\text{(marginal likelihood)}}.$$

The parametrization of the Wishart distribution family

Matrix variate distributions depend not only on the distribution of the entries (a_{ij}) but also on the structure of the matrix $W = (a_{ij})_{ij}$. The Wishart distribution family has three important characteristics that subdivide the family:

- (a_{ij}) in $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$ determined through β.
- Uncorrelatedness of rows and columns

  $$\begin{pmatrix}\text{uncorrelatedness} & \rightarrow \\ & \downarrow\end{pmatrix} \quad \text{which we denote by} \quad \Phi_n \otimes \Sigma_p$$

  where Σ_p is the covariance matrix for rows and Φ_n is the covariance matrix for columns. In general we have uncorrelated rows and therefore $\Phi_n := I_n$.
- Centrality of W determined by $\mathbb{E}[a_{ij}] = 0$ for all i, j.

The two simplest cases

$$a_{ij} \in \mathbb{R} \text{ and } \mathbb{E}[a_{ij}] = 0 \quad \text{for all } i,j \quad \text{and} \quad I_n \otimes I_p$$
$$a_{ij} \in \mathbb{C} \text{ and } \mathbb{E}[a_{ij}] = 0 \quad \text{for all } i,j \quad \text{and} \quad I_n \otimes I_p$$

are called *real* and *complex (central) white Wishart* distributions where *white* refers to the fact that their entries are uncorrelated $I_n \otimes I_p$. Many textbooks reserve the term *Wishart distribution* for the particular case where only rows are uncorrelated and

$$a_{ij} \in \mathbb{R} \text{ and } \mathbb{E}[a_{ij}] = 0 \quad \text{for all } i,j \quad \text{and} \quad I_n \otimes \Sigma_p \qquad \text{(Wishart)}$$

and we are going to stick to that. Before we give an explicit formal definition of the distribution we try to give an intuitive approach by showing the connection to the normal distribution and the χ^2-distribution.

The matrix variate normal distribution

Definition 7 (Matrix variate normal distribution) [GN00, Ch. 2, p. 55]**.** *Consider* $X_1, \dots, X_n \sim \mathcal{N}_p(\mu, \Sigma_p)$ *then the* $n \times p$ *matrix*

$$X = \begin{pmatrix} X_1^{(1)} & X_1^{(2)} & \cdots & X_1^{(p)} \\ \vdots & \vdots & \vdots & \vdots \\ X_n^{(1)} & X_n^{(2)} & \cdots & X_n^{(p)} \end{pmatrix}$$

follows a matrix variate normal distribution written $X \sim \mathcal{N}_{n,p}(M, \Phi_n \otimes \Sigma_p)$ *where*

M *is the* $n \times p$ *mean matrix* $(\mu, \mu, \dots, \mu)^T$.
Σ_p *is a* $p \times p$ *covariance matrix for the rows.*
Φ_n *is a* $n \times n$ *covariance matrix for the columns.*

with probability density function

$$\frac{1}{(2\pi)^{np/2} \det(\Sigma_p)^{n/2} \det(\Phi_n)^{p/2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left\{\Phi_n^{-1}(X-M)\Sigma_p^{-1}(X-M)^T\right\}\right). \quad (3.1)$$

We also have to introduce the multivariate generalization of the gamma function $\Gamma^{(\beta)}(a)$.

Definition 8 (Generalized multivariate gamma function) [DEKV13, p. 18]**.** *The generalized multivariate gamma function is*

$$\Gamma_p^{(\beta)}(a) = \pi^{\beta \frac{1}{4} p(p-1)} \prod_{i=1}^{p} \Gamma\left(a - \frac{\beta}{2}(i-1)\right) \quad (3.2)$$

for $\operatorname{Re}(a) > (p-1)\beta/2$.

Remark 9: *Sometimes in the literature one encounters* $\Gamma_p^1(a)$ *or* $\Gamma_p^2(a)$ *only in this numerical representation. This form doesn't really explain why there exist two multivariate gamma function. The gamma functions also have a matrix integral representation and because we integrate over two different matrix spaces we have two different functions*

$$\Gamma_p^{(1)}(a) = \int_{\mathbb{S}_p^+} \exp(\operatorname{tr}(-A)) \det(A)^{a-\frac{1}{2}(p+1)} \mathrm{d}A$$
$$\Gamma_p^{(2)}(a) = \int_{\mathbb{H}_p^+} \exp(\operatorname{tr}(-A)) \det(A)^{a-p} \mathrm{d}A$$

where $\mathbb{S}_p^+$ *is the space of* $p \times p$ *positive symmetric matrices,* $\mathbb{H}_p^+$ *the space of* $p \times p$ *positive Hermitian matrices and* $\mathrm{d}A$ *the product Lebesgue measure of the distinct elements. A proof of the first identity can be found in [GN00, Ch. 1, p. 19].*

The Wishart distribution

The Wishart distribution occurs in the following setting: let X be a $n \times p$ matrix consisting of i.i.d. zero mean vectors $X_1, \ldots, X_n$ following a p-variate normal distribution with covariance Σ_p. Then we have

$$X \sim \mathcal{N}_{n,p}(0, I_n \otimes \Sigma_p) \qquad \text{and} \qquad X^T X = \sum_{i=1}^{n} X_i X_i^T \sim W_p(n, \Sigma_p) \tag{3.3}$$

where $W_p(n, \Sigma_p)$ means the $p \times p$ matrix $X^T X$ follows a Wishart distribution with n *degrees of freedom*, i.e. the number of varying variables.

Remark 10: *We want to emphasize that the ^{T}-operator changes the position depending on whether we look at the $n \times p$ data matrix X or the $p \times 1$ vectors X_i. If we do not use the statistical convention and instead choose a $p \times n$ data matrix $Y \sim \mathcal{N}_{p,n}(0, \Sigma_p \otimes I_n)$) then the ^{T}-operator stays in the same position*

$$YY^T = \sum_{i=1}^{n} Y_i Y_i^T \sim W_p(n, \Sigma_p).$$

Definition 9 (Wishart distribution) [GN00, Ch. 3, p. 87]**.** *A $p \times p$ random symmetric positive definite Matrix $S := X^T X > 0$ follows the Wishart distribution with parameters p, n such that $n \geq p$ and $\Sigma_p > 0$ if it's probability density function is given as*

$$\left\{ 2^{\frac{np}{2}} \Gamma_p^{(1)} \left(\frac{n}{2} \right) \det\left(\Sigma_p\right)^{\frac{n}{2}} \right\}^{-1} \det\left(S\right)^{\frac{1}{2}(n-p-1)} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\Sigma_p^{-1} S \right) \right) \tag{3.4}$$

where $\Gamma_p^{(1)}$ is the generalized multivariate Gamma function

$$\Gamma_p^{(1)} \left(\frac{n}{2} \right) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left(\frac{n}{2} - \frac{j-1}{2} \right)$$

defined for $\operatorname{Re}(n/2) > (p-1)/2$.

Remark 11: *(Singularity) In the case when $n < p$ and $S \sim W_p(n, \Sigma_p)$ one obtains the singular Wishart distribution since* $\operatorname{rank}(S) < \dim(S)$ *the distribution becomes singular.*

Remark 12: *(Noncentrality) If the $X_1, \ldots, X_n$ have a non-zero mean vector we call the distribution of $X^T X$ a noncentral Wishart distribution with noncentrality parameter Θ_p*

<u>central Wishart</u>	<u>noncentral Wishart</u>
$\mu^T = (0, \ldots, 0)$	$\mu^T \neq (0, \ldots, 0)$
$\Theta_p = 0$	$\Theta_p \neq 0$

We assume by default that a Wishart distribution is central and refer to it just as Wishart. The probability density function for a noncentral Wishart matrix is similar to the central case but more complicated to evaluate since it involves the hypergeometric function of matrix argument ${}_0F_1$ (see [GN00, Ch. 3, p. 113-114])

$$\left\{ \textit{central Wishart density} \right\} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\Theta_p\right) \right) {}_0F_1\left(\frac{n}{2}, \frac{1}{4}\Theta_p \Sigma_p^{-1} X^T X \right) \tag{3.5}$$

we will later elaborate on this function in the subchapter on eigendistribution for correlated Wishart matrixes.

Definition 10 (Complex Wishart distribution) [Goo63, p. 162]**.** *A $p \times p$ random Hermitian positive definite Matrix $S = X^*X > 0$ follows a complex Wishart distribution with parameters p, n such that $n \geq p$ and $\Sigma_p > 0$ if its probability density function is given as*

$$\left\{ \Gamma_p^{(2)}(n) \det\left(\Sigma_p\right)^n \right\}^{-1} \det\left(S\right)^{n-p} \exp\left(-\operatorname{tr}\left(\Sigma_p^{-1} S \right) \right). \tag{3.6}$$

where $\Gamma_p^{(2)}$ is the generalized multivariate gamma function for $\operatorname{Re}(a) > (p-1)$.

3.1.1 The Wishart distribution in multivariate statistics

Recall the *sample covariance estimator* and *sample mean*

$$\widehat{\Sigma}_p = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})^T \quad \text{and} \quad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.7}$$

for independent $X_1, \ldots, X_n \sim \mathcal{N}_p(\mu, \Sigma_p)$ and define $S := (n-1)\widehat{\Sigma}_p$. It turns out that this estimator follows a central Wishart distribution

$$\widehat{\Sigma}_p \sim W_p\left(n-1, \frac{1}{n-1}\Sigma_p \right) \qquad \text{and} \quad S \sim W_p(n-1, \Sigma_p). \tag{3.8}$$

We have $n-1$ degrees of freedom because if we know $\overline{X}$ and $X_1, \ldots, X_{n-1}$ we automatically know the last X_n.

Remark 13: *(Degrees of freedom) If we instead know the true mean $\mu = 0$ a priori we can replace $\overline{X}$ with μ in the estimator and have again n degrees of freedom (since we don't know $\overline{X}$ we need all $X_1, \ldots, X_n$) and so*

$$S = \sum_{i=1}^{n} X_i X_i^T \sim W_p(n, \Sigma_p)$$

Remark 14: *(Wishart and χ^2) The Wishart distribution can be interpreted as a matrix variate distribution of the χ^2-distribution then if in addition $p = 1$ and $S \sim W_1(N, \sigma^2)$ that is identical to $S \sim \sigma^2\chi^2$ with N degrees of freedom.*

We now come to the arguably most important theorem for statistics regarding the Wishart distribution that establishes the connection of the Wishart distribution with statistical inference theory.

Theorem 8 (Sample covariance estimator follows Wishart) [GN00, Ch. 3, p. 92]**.** *Let $X_1, \dots, X_n \in \mathcal{N}_p(\mu, \Sigma_p)$ be independent with $\Sigma_p > 0$ and $n > p$. Let $S := (n-1)\hat{\Sigma}_p$, then*

$$\begin{aligned} &S \text{ and } \overline{X} \text{ are independent} \\ &\overline{X}_p \sim \mathcal{N}_p\left(\mu, \frac{1}{n}\Sigma_p\right) \\ &S \sim W_p\left(n-1, \Sigma_p\right) \end{aligned} \tag{3.9}$$

where W_p denotes the (central) Wishart distribution.

Proof. We ignore the statistical convention for this proof and look at a $p \times n$ matrix $X = [X_1, X_2, \dots, X_n]$ then $X \sim \mathcal{N}_{p,n}(\mu e^T, \Sigma_p \otimes I_n)$ where $e^T := (1, 1, \dots, 1)$ has $1 \times n$ dimension. By (3.1) The probability density function of X is

$$(2\pi)^{-\frac{1}{2}np} \det(\Sigma_p)^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}\operatorname{tr}\left\{(X-\mu e^T)^T \Sigma_p^{-1}(X-\mu e^T)\right\}\right) \tag{3.10}$$

Since $\operatorname{tr}(ABC) = \operatorname{tr}(CAB)$ we are free to change the order in cyclic permutation. Consider an orthogonal $n \times n$-matrix $H = \begin{pmatrix} \frac{1}{\sqrt{n}}e & B \end{pmatrix}$ where B has dimension $n \times n-1$ and perform an orthogonal transformation $X_O := XH$ then clearly $X_O X_O^T = XHH^T X^T = XX^T$ and notice that the i-th entry of the first column of X_O will be $\sqrt{n}(X_1^{(i)} + \cdots + X_n^{(i)})$ so

$$X_O = \begin{pmatrix} \sqrt{n}\,\overline{X} & XB \end{pmatrix}, \qquad \text{and} \qquad XX^T = X_O X_O^T = n\overline{X}\overline{X}^T + XB(XB)^T. \tag{3.11}$$

Then we want to compute

$$(X-\mu e^T)(X-\mu e^T)^T = XX^T - \mu e^T X^T - Xe\mu^T + \mu e^T e \mu^T \tag{3.12}$$

and replace each term on the right-hand side. The first term can be replaced by (3.11). For the second term notice

$$\mu e^T X^T = \mu e^T H X_O^T = \mu e^T \begin{pmatrix} \frac{1}{\sqrt{n}}e & B \end{pmatrix} \begin{pmatrix} \sqrt{n}\overline{X}^T \\ B^T X^T \end{pmatrix} = n\mu\overline{X}^T$$

and for the third term $Xe\mu^T = n\overline{X}\mu^T$. Replacing everything in (3.12) and using the fact that $e^T e = n$ we get

$$\begin{aligned} (X-\mu e^T)(X-\mu e^T)^T &= n\overline{X}\overline{X}^T + XB(XB)^T - n\mu\overline{X}^T - n\overline{X}\mu^T + n\mu\mu^T \\ &= n(\overline{X}-\mu)(\overline{X}-\mu)^T + XB(XB)^T \end{aligned} \tag{3.13}$$

then $(\overline{X}-\mu)^T = (\overline{X}^T - \mu^T)$. Now compute the Jacobian of the orthogonal transformation X_O which is $J(X \to XH) = \det(H) = 1$ and use (3.13) in the probability density function (3.10) to get the density for $X_O = \left(\sqrt{n}\ \overline{X}, XB\right)$

$$f\left(\sqrt{n}\ \overline{X}, XB\right) = (2\pi)^{-\frac{1}{2}p} \det(\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{n}{2} \operatorname{tr}\left\{(\overline{X}-\mu)^T \Sigma_p^{-1} (\overline{X}-\mu)\right\}\right)$$
$$\cdot (2\pi)^{-\frac{1}{2}(n-1)p} \det(\Sigma)^{-\frac{1}{2}(n-1)} \exp\left(-\frac{1}{2} \operatorname{tr}\left\{\left((XB)^T \Sigma_p^{-1} XB\right)\right\}\right)$$

which shows that $\overline{X}$ and XB are independent and $\overline{X} \sim \mathcal{N}_p(\mu, \frac{1}{n}\Sigma_p)$ and $XB \sim \mathcal{N}_{p,n}(0, \Sigma_p \otimes I_n)$. Since $XX^T = XBB^TX^T$ we have $XX^T \sim W_p(n-1, \Sigma_p)$. Further from (3.13) if we substitute $\overline{X}$ for μ we get

$$(n-1)\hat{\Sigma}_p = (X - \overline{X}e^T)(X - \overline{X}e^T)^T = XB(XB)^T$$

which proves $(n-1)\hat{\Sigma}_p \sim W_p(n-1, \Sigma_p)$. □

3.2 The β-Wishart ensembles

The β-Wishart ensembles $X^\dagger X \sim W_p^\beta(n, \Sigma_p)$ are defined as

$\beta = 1:$	$X_i \sim \mathcal{N}_p(0, \Sigma_p)$	WOE/LOE	(real)
$\beta = 2:$	$X_i \sim \mathbb{C}\mathcal{N}_p(0, \Sigma_p)$	WUE/LUE	(complex)
$\beta = 4:$	$X_i \sim \mathbb{H}\mathcal{N}_p(0, \Sigma_p)$	WSE/LSE	(quaternion)

where

- $\mathbb{C}\mathcal{N}$ is the complex normal distribution and $\mathbb{H}\mathcal{N}$ denotes that X_i is a quaternion and each element follows a real normal distribution.
- WOE/WUE/WSE resp. LOE/LUE/LSE stands for *Wishart* resp. *Laguerre orthogonal/unitary/symplectic ensemble.*

Summary of the spectral densities

We briefly summarize the eigenvalue densities before deriving them, "S" stands for "integration space".

Unordered eigenvalues

$$\text{density: } \widehat{f}_p, \qquad \text{S: } \mathbb{R}_+^p$$

Ordered eigenvalues

$$\text{density: } f_p = p! \cdot \widehat{f}_p, \qquad \text{S: } \mathbb{R}_{\geq,+}^p := \mathbb{R}_+^p \cap W_p$$

$$W_p := \{(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p\}.$$

White β-Wishart density

$$f_p(\lambda_1, \ldots, \lambda_p) = \frac{1}{Z_L^{\beta,a}} \prod_{1 \le i < j \le p} |\lambda_j - \lambda_i|^\beta \prod_{j=1}^{p} \lambda_j^{\frac{(n-p+1)\beta}{2}-1} e^{-\beta\lambda_j/2} \tag{3.14}$$

Correlated β-Wishart density

$$f_p(\lambda_1, \ldots, \lambda_p) = \frac{1}{C_L^\beta} {}_0F_0^{(\beta)}\left(-\frac{\beta}{2}\Sigma_p^{-1}, \Lambda\right) \prod_{1 \le i < j \le p} |\lambda_j - \lambda_i|^\beta \prod_{j=1}^{p} \lambda_j^{\frac{n-p+1}{2}\beta-1} \tag{3.15}$$

Normalization constants:

$$Z_L^{\beta,a} = 2^{pa} \prod_{j=1}^{p} \frac{\Gamma(1+\frac{\beta}{2}j)\Gamma(a-\frac{\beta}{2}(p-j))}{\Gamma(1+\frac{\beta}{2})}, \qquad a := \frac{\beta}{2}n$$

$$C_L^\beta = \left(\frac{2}{\beta}\right)^{np\beta/2} \det(\Sigma_p)^{n\beta/2} \frac{\Gamma_p^{(\beta)}(n\beta/2)\Gamma_p^{(\beta)}(p\beta/2)}{\pi^{p(p-1)\beta/2}}$$

$$\widehat{Z}_L := \frac{1}{p!} Z_L, \qquad \widehat{C}_L := \frac{1}{p!} C_L$$

Scaling of the Wishart ensemble for deriving limiting distributions

$$W = \frac{1}{n} X^* X \sim W_p^\beta(n, I_p).$$

3.3 Spectral densities

For the spectral density we have two cases to distinguish:

	White Wishart	Correlated Wishart
Covariance:	$I_n \otimes I_p$	$I_n \otimes \Sigma_p$
Model:	Tridiagonal	Broken-Arrow
β:	$\beta > 0$	$\beta > 0$

The tridiagonal matrix models are introduced in ([DE02]). For the correlated case the tridiagonal matrix model work for $\beta = 1, 2, 4$ however a better model for general $\beta > 0$ called the *broken-arrow model* was later introduced in ([DEKV13]).

The eigendistribution for correlated Wishart ensembles involves the hypergeometric function of matrix arguments ${}_0F_0$ therefore it is more difficult to evaluate and needs additional theory to derive.

3.3.1 White β-Wishart eigendistribution

We study the case for $I_n \otimes I_p$.

Derivation of the matrix models

For a white Wishart matrix $S_\beta = X^T X \sim W_p^\beta(n, I_p)$ it is possible to find a decomposition

$$S_\beta = B_\beta^T B_\beta \qquad \text{with bidiagonal matrices } B_\beta, B_\beta^T.$$

One achieves this by repeatedly make a Housholder transformation on X^T and X. As an example for the procedure for X^T we start with $X_{(1)}^T := X^T$ and then do repeatedly

$$X_{(i+1)}^T = \widetilde{P}_{(i)}^T X_{(i)}^T P_{(i)}^T$$

to obtain the desired result

$$B_\beta^T = \widetilde{P}_{(p-2)}^T \cdots \widetilde{P}_{(1)}^T X^T P_{(1)}^T \cdots P_{(p-1)}^T$$

where $\widetilde{P}_{(i)}^T$ is a $p \times p$ Householder matrix for eliminating row entries above the diagonal and $P_{(i)}^T$ is a $n \times n$ Householder matrix for eliminating column entries below the subdiagonal.

The tridiagonal matrix model

It turns out that for $a = \frac{\beta}{2} n$ such that $a > \frac{\beta}{2}(p-1)$ we have the following tridiagonal matrix model

$$S_\beta = B_\beta^T B_\beta, \qquad B_\beta^T = \begin{pmatrix} \chi_{2\alpha} & & & \\ \chi_{\beta(p-1)} & \chi_{2a-\beta} & & \\ & \ddots & \ddots & \\ & & \chi_\beta & \chi_{2a-\beta(p-1)} \end{pmatrix} \tag{3.16}$$

where B_β^T is a *sparse* $p \times p$ matrix with $2p-1$ non-zero, mutually independent elements.

Vandermonde determinant formula for tridiagonal matrix models

Lemma 3 [DE02, p. 8]**.** *Given a symmetric tridiagonal matrix S with positive subdiagonal $b = (b_{p-1}, \ldots, b_1)$ and eigendecomposition $S = Q\Lambda Q^T$ where the first row of the eigenvector matrix Q is $q = (q_1, \ldots, q_p)$. Let $\lambda = \operatorname{diag}(\Lambda)$ then the Vandermonde determinant for the ordered eigenvalues of S is given by*

$$\triangle(\lambda) = \prod_{i<j} (\lambda_i - \lambda_j) = \frac{\prod_{i=1}^{p-1} b_i^i}{\prod_{i=1}^{p} q_i} \tag{3.17}$$

Proof. Let $\lambda_i^{(k)}$ for $i = 1, \ldots, k$ be the eigenvalues of the $k \times k$ lower right-corner submatrix of T. The associated characteristic polynomial of this submatrix is $P_k(x) = \prod_{i=1}^{k} (x - \lambda_i^{(k)})$ and satisfies for $k = 1, \ldots, p$ the three-term recurrence

$$P_k(x) = (x - a_k) P_{k-1}(x) - b_{k-1}^2 P_{k-2}(x). \tag{3.18}$$

and the two-term relation

$$\prod_{1\le i\le k}\prod_{1\le j\le k-1}|\lambda_i^{(k)}-\lambda_j^{(k-1)}| = \left|\prod_{i=1}^{k}P_{k-1}(\lambda_i^{(k)})\right| = \left|\prod_{i=1}^{k-1}P_k(\lambda_j^{(k-1)})\right| \tag{3.19}$$

from (3.18) we get

$$\left|\prod_{i=1}^{k-1}P_k(\lambda_i^{(k-1)})\right| = b_{k-1}^{2(k-1)}\left|\prod_{i=1}^{k-1}P_{k-2}(\lambda_i^{(k-1)})\right|. \tag{3.20}$$

By repeatedly applying (3.18) and (3.19) we get

$$\prod_{i=1}^{p-1}|P_p(\lambda_i^{(p-1)})| = b_{p-1}^{2(p-1)}\left|\prod_{i=1}^{p-2}P_{p-1}(\lambda_i^{(p-2)})\right| \tag{3.21}$$

$$= b_{p-1}^{2(p-1)}b_{p-2}^{2(p-2)}\left|\prod_{i=1}^{p-2}P_{p-3}(\lambda_i^{(p-2)})\right| \tag{3.22}$$

$$= \dots \tag{3.23}$$

$$= \prod_{i=1}^{p-1}b_i^{2i}. \tag{3.24}$$

Now we apply *theorem 7.9.2* from [Par98] which says

$$q_i^2 = \left|\frac{P_{p-1}(\lambda_i)}{P_p'(\lambda_i)}\right| = \left|\frac{P_{p-1}(\lambda_i^{(p)})}{P_p'(\lambda_i^{(p)})}\right| \tag{3.25}$$

so it follows

$$\prod_{i=1}^{p}q_i^2 = \frac{\prod_{i=1}^{p}|P_{p-1}(\lambda_i^{(p)})|}{\triangle(\lambda)^2} = \frac{\prod_{i=1}^{p-1}b_i^{2i}}{\triangle(\lambda)^2}.$$

which proves the lemma. □

Jacobian for tridiagonal matrix models

Lemma 4 [DE02, p. 11]**.** *Let B^T be a bidiagonal matrix with positive diagonal $x = (x_p, \dots, x_1)$ and positive subdiagonal $y = (y_{p-1}, \dots, y_1)$. Consider $S = B^TB$ and let $a = (a_p, \dots, a_1)$ be the diagonal part and $b = (b_{p-1}, \dots, b_1)$ the subdiagonal part of S. Then we have for the Jacobian $J_{(B^T\to S)}$ that*

$$J_{(B^T\to S)} = \left(2^p x_1\prod_{i=2}^{p}x_i^2\right)^{-1}. \tag{3.26}$$

Proof. Differentials are defined in decreasing order

$$\mathrm{d}\vec{x} = \bigwedge_{i:1\le i\le p}\mathrm{d}x_i := \mathrm{d}x_p\wedge\dots\wedge\mathrm{d}x_1 \text{ and } \mathrm{d}\vec{y} = \bigwedge_{i:1\le i\le p-1}\mathrm{d}y_i := \mathrm{d}y_{p-1}\wedge\dots\wedge\mathrm{d}y_1$$

then we have

$$\bigwedge_{i:1\le i\le p} \mathrm{d}x_i \wedge \bigwedge_{j:1\le j\le p-1} y_j = J_{(B^T\to S)} \bigwedge_{i:1\le i\le p} \mathrm{d}a_i \wedge \bigwedge_{j:1\le j\le p-1} \mathrm{d}b_j.$$

We notice that S is a symmetric matrix and computing the product B^TB we get

$$\begin{aligned} a_p &= x_p^2 \\ a_i &= y_i^2 + x_i^2, && i = p-1, p-2, \ldots, 1 \\ b_i &= y_i x_{i+1}, && i = p-1, p-2, \ldots, 1 \end{aligned}$$

To compute the Jacobian $J_{(B\to S)}$ for the change of variable $\{x_i, y_i\} \to \{a_i, b_i\}$ we compute the differentials

$$\begin{aligned} \mathrm{d}a_p &= 2x_p\mathrm{d}x_p \\ \mathrm{d}a_i &= 2(x_i\mathrm{d}x_i + y_i\mathrm{d}y_i), && i = p-1, p-2, \ldots, 1 \\ \mathrm{d}b_i &= x_{i+1}\mathrm{d}y_i + y_i\mathrm{d}x_{i+1}, && i = p-1, p-2, \ldots, 1 \end{aligned}$$

and then take the wedge product on both sides

$$\mathrm{d}a_p \wedge \mathrm{d}a_i \wedge \mathrm{d}b_i = (2x_p\mathrm{d}x_p) \wedge 2(x_i\mathrm{d}x_i + y_i\mathrm{d}y_i) \wedge (x_{i+1}\mathrm{d}y_i + y_i\mathrm{d}x_{i+1}).$$

Proceed by induction and claim the factor of the right-hand side is $2^p x_1 \prod_{i=2}^{p} x_i^2$. Take $p = 2$ as initial case

$$\begin{aligned} I_2 &= (2x_2\mathrm{d}x_2) \wedge 2(x_1\mathrm{d}x_1 + y_1\mathrm{d}y_1) \wedge (x_2\mathrm{d}y_1 + y_1\mathrm{d}x_2) \\ &= 4x_2\mathrm{d}x_2 \wedge (x_1x_2\mathrm{d}x_1 \wedge \mathrm{d}y_1 + x_1y_1\mathrm{d}x_1 + y_1^2\mathrm{d}y_1 \wedge \mathrm{d}x_2) \\ &= 4(x_1x_2^2\mathrm{d}x_2 \wedge \mathrm{d}x_1 \wedge \mathrm{d}y_1) \\ &= 2^2x_1x_2^2\ \mathrm{d}x_2 \wedge \mathrm{d}x_1 \wedge \mathrm{d}y_1. \end{aligned}$$

Assume it's true for general p

$$\begin{aligned} I_p &= (2x_p\mathrm{d}x_p) \wedge \bigwedge_{i:1\le i\le p-1} 2(x_i\mathrm{d}x_i + y_i\mathrm{d}y_i) \wedge \bigwedge_{i:1\le i\le p-1} (x_{i+1}\mathrm{d}y_i + y_i\mathrm{d}x_{i+1}) \\ &= (2^p x_1 \prod_{i=2}^{p} x_i^2) \bigwedge_{i:1\le i\le p} \mathrm{d}x_i \wedge \bigwedge_{i:1\le i\le p-1} \mathrm{d}y_i \end{aligned}$$

then for the case $p+1$

$$\begin{aligned} I_{p+1} &= (2x_{p+1}\mathrm{d}x_{p+1}) \wedge \bigwedge_{i:1\le i\le p} 2(x_i\mathrm{d}x_i + y_i\mathrm{d}y_i) \wedge \bigwedge_{i:1\le i\le p} (x_{i+1}\mathrm{d}y_i + y_i\mathrm{d}x_{i+1}) \\ &= C \wedge \left(\bigwedge_{i:1\le i\le p-1} 2(x_i\mathrm{d}x_i + y_i\mathrm{d}y_i) \wedge \bigwedge_{i:1\le i\le p-1} (x_{i+1}\mathrm{d}y_i + y_i\mathrm{d}x_{i+1}) \right) \end{aligned}$$

where C is similar to the case I_2 but shifted

$$\begin{aligned} C &= (2x_{p+1}\mathrm{d}x_{p+1}) \wedge 2(x_p\mathrm{d}x_p + y_p\mathrm{d}y_p) \wedge (-1)^{p-1}(x_{p+1}\mathrm{d}y_p + y_i\mathrm{d}x_{p+1}) \\ &= (-1)2^2x_px_{p+1}^2\,(\mathrm{d}x_{p+1} \wedge \mathrm{d}x_p \wedge \mathrm{d}y_p) \\ &= 2x_{p+1}^2\left(\mathrm{d}x_{p+1} \wedge 2x_p\mathrm{d}x_p \wedge (-1)^{p-1}\mathrm{d}y_p\right) \end{aligned}$$

Now use the identity for I_p

$$\begin{aligned} I_{p+1} &= \left(2x_{p+1}^2 2^p x_1 \prod_{i=2}^{p} x_i^2\right) \mathrm{d}x_{p+1} \wedge \mathrm{d}x_p \wedge \bigwedge_{i:1\le i\le p-1} \mathrm{d}x_i \wedge (-1)^{2p-2}\mathrm{d}y_p \wedge \bigwedge_{i:1\le i\le p-1} \mathrm{d}y_i \\ &= \left(2^{p+1} x_1 \prod_{i=2}^{p+1} x_i^2\right) \bigwedge_{i:1\le i\le p+1} \mathrm{d}x_i \bigwedge_{i:1\le i\le p} \mathrm{d}y_i \end{aligned}$$

therefore we have $(J_{B^T\to S}) = \left(2^p x_1 \prod_{i=2}^{p} x_i^2\right)^{-1}$ which proves (3.26). □

Eigendistribution for white β-Wishart

Theorem 9 (Eigendistribution for white β-Wishart ensembles) [DE02, p. 14]**.** *Assume $S_\beta \sim W_p^\beta(n, I_p)$ and $S_\beta = B_\beta^T B_\beta$ be the corresponding tridiagonal matrices. Let $S_\beta = Q\Lambda Q^T$ be the eigendecomposition such that the sign of the first row q of Q is non-negative and the eigenvalues in Λ are increasingly ordered. Then q and Λ are independent and the eigenvalues have the following joint probability density function*

$$f_p(\lambda_1, \ldots, \lambda_p) = \frac{1}{Z_L^{\beta,a}} \prod_{1\le i<j\le p} |\lambda_j - \lambda_i|^\beta \prod_{j=1}^{p} \lambda_j^{a-r} e^{-\lambda_j/2} \tag{3.27}$$

where

$$a = \frac{\beta}{2}n, \qquad r = 1 + \frac{\beta}{2}(p-1), \qquad a - r = -1 + \frac{\beta}{2}(n-p+1)$$

and

$$Z_L^{\beta,a} = 2^{pa} \prod_{j=1}^{p} \frac{\Gamma(1+\frac{\beta}{2}j)\Gamma(a-\frac{\beta}{2}(p-j))}{\Gamma(1+\frac{\beta}{2})}$$

Further $q = (q_1, \ldots, q_p)$ is normalized to unit length and has joint law:

$$c_q^\beta \prod_{i=1}^{p-1} q_i^{\beta-1} \mathrm{d}q, \qquad c_q^\beta := \frac{2^{p-1}\Gamma\left(\frac{\beta}{2}p\right)}{\Gamma\left(\frac{\beta}{2}\right)} \tag{3.28}$$

Proof. Let $\mathrm{d}B_\beta^T$ be the joint element distribution of B_β^T

$$\mathrm{d}B_\beta^T \equiv \mu(x,y)\mathrm{d}x\ \mathrm{d}y = \prod_{i=0}^{p-1} x_{p-i}^{a-\beta i-1} e^{-x_i^2/2} \prod_{i=1}^{p-1} y_i^{\beta i-1} e^{-y_i^2/2} \mathrm{d}x\ \mathrm{d}y$$

By (3.26) we have

$$\begin{aligned} \mathrm{d}S_\beta &\equiv J_{(B^T\to T)}\mu(x,y)\mathrm{d}x\ \mathrm{d}y \\ &\equiv 2^{-p} c_{x,y} x_1^{2a-\beta(p-1)-2} e^{-x_1^2/2} \prod_{i=0}^{p-2} x_{p-i}^{a-\beta i-3} e^{-x_i^2/2} \prod_{i=1}^{p-1} y_i^{\beta i-1} e^{-y_i^2/2} \mathrm{d}x\ \mathrm{d}y \end{aligned}$$

where

$$c_{x,y} = \frac{\prod_{i=1}^{p-1} \Gamma(i\frac{\beta}{2}) \prod_{i=1}^{p} \Gamma(a - \frac{\beta}{2}(i-1))}{2^{2p-1}}$$

rewrite everything in terms of x, y, λ and $q = (q_1, \ldots, q_p)$

$$\begin{aligned} \mathrm{d}S_\beta &\equiv 2^{-p} c_{x,y} e^{-\sum_{i=1}^{p} x_i^2/2} e^{-\sum_{i=1}^{p-1} y_i^2/2} \frac{\prod_{i=1}^{p-1} (x_{i+1} y_i)}{\prod_{i=1}^{p} q_i} x_1^{2a-\beta(p-1)-2} \\ &\quad \times \prod_{i=0}^{p-2} x_{p-i}^{2a-\beta(p-i)-3} \prod_{i=1}^{p-1} y_i^{\beta i - 1} \mathrm{d}q \mathrm{d}\lambda \\ &\equiv 2^{-p} c_{x,y} e^{-\sum_{i=1}^{p} x_i^2/2} e^{-\sum_{i=1}^{p-1} y_i^2/2} \frac{\prod_{i=0}^{p-1} x_{p-1}^{2a-\beta(p-i)-2} \prod_{i=1}^{p-1} y_i^{\beta i}}{\prod_{i=1}^{p} q_i} \mathrm{d}q \mathrm{d}\lambda \end{aligned}$$

Using lemma 3 we can write the Vandermonde with respect to the subdiagonal b and q and the ordered eigenvalues λ as

$$\triangle(\lambda) = \frac{\prod_{i=1}^{p-1} b_i^i}{\prod_{i=1}^{p} q_i} \qquad \text{and so it follows that} \qquad \triangle(\lambda) = \frac{\prod_{i=1}^{p-1} (x_{i+1} y_i)^i}{\prod_{i=1}^{p} q_i}$$

This means we can rewrite

$$\begin{aligned} \mathrm{d}S_\beta &\equiv 2^{-p} c_{x,y} e^{-\sum_{i=1}^{p-1} x_{n-i}^2/2} e^{-\sum_{i=1}^{p-1} y_i^2/2} \frac{\prod_{i=1}^{p-1} (x_{i+1} y_i)^{\beta i}}{\prod_{i=1}^{p} q_i^{\beta}} \prod_{i=1}^{p-1} q_i^{\beta-1} \prod_{i=0}^{p-1} x_{p-i}^{2a-\beta(p-1)-2} \mathrm{d}q \, \mathrm{d}\lambda \\ &\equiv 2^{-p} c_{x,y} e^{-\sum_{i=1}^{p-1} x_{n-i}^2/2} e^{-\sum_{i=1}^{p-1} y_i^2/2} \triangle(\lambda)^\beta \prod_{i=1}^{p-1} q_i^{\beta-1} \left(\prod_{i=1}^{p-1} x_{p-i} \right)^{2a-\beta(p-1)-2} \mathrm{d}q \, \mathrm{d}\lambda \end{aligned}$$

By invariance $\mathrm{tr}(S_\beta) = \mathrm{tr}(Q\Lambda Q^T) = \mathrm{tr}(\Lambda)$ and $\det(S_\beta) = \det(Q\Lambda Q^T) = \det(\Lambda)$ and so

$$\begin{aligned} \mathrm{tr}(S_\beta) &= \sum_{i=0}^{p-1} x_{p-i}^2 + \sum_{i=1}^{p-1} y_i^2 = \sum_{i=1}^{p} \lambda_i \\ \det(S_\beta) &= \prod_{i=0}^{p-1} x_{p-i}^2 = \prod_{i=1}^{p} \lambda_i \end{aligned}$$

using this and substituting r for $1 + \frac{\beta}{2}(p-1)$ we obtain

$$\mathrm{d}S_\beta \equiv \left(c_q^\beta \prod_{i=1}^{p-1} q_i^{\beta-1} \mathrm{d}q \right) \left(p! \left(Z_L^{\beta,a} \right)^{-1} e^{-\sum_{i=1}^{p} \lambda_i/2} \triangle(\lambda)^\beta \prod_{i=1}^{p} \lambda_i^{a-r} \mathrm{d}\lambda \right)$$

where $c_{x,y} = c_q^\beta (Z_L^{\beta,a})^{-1}$ for

$$Z_L^{\beta,a} = 2^{pa} \prod_{j=1}^{p} \frac{\Gamma(1+\frac{\beta}{2}j)\Gamma(a-\frac{\beta}{2}(p-j))}{\Gamma(1+\frac{\beta}{2})}$$

we see that q and λ are independent. If we drop the ordering the $m!$ vanishes and the joint eigenvalue density is given while q is distributed like a normalized vector of χ^β random variables. □

Remark 15: *(Different parametrizations) It convenient to work with a slightly different parametrization particularly with a $e^{-\beta\lambda_j/2}$ instead a $e^{-\lambda_j/2}$:*

$$\mathbb{P}_p(\mathrm{d}\lambda_1, \ldots, \mathrm{d}\lambda_p) = \frac{1}{Z_L^{\beta,a}} \prod_{1\le i<j\le p} |\lambda_j - \lambda_i|^\beta \prod_{j=1}^{p} \lambda_j^{\beta w/2} e^{-\beta\lambda_j/2} \mathrm{d}\lambda_j \tag{3.29}$$

where $w = n - p + 1 - \frac{2}{\beta}$ and

$$Z_L^{\beta,a} = 2^{p(w\beta/2+1+(p-1)\beta/2)} \prod_{j=1}^{p} \frac{\Gamma(1+\frac{\beta}{2}j)\Gamma(1+\beta w/2+\beta(j-1)/2)}{\Gamma(1+\frac{\beta}{2})}.$$

Notice that they are the same

I. $\beta\frac{w}{2} = \frac{\beta}{2}(n-p+1) - 1 = a - r$

II. $p\left(w\frac{\beta}{2} + 1 + (p-1)\frac{\beta}{2}\right) = p\left((n-p+1-\frac{2}{\beta})\frac{\beta}{2} + 1 + (p-1)\frac{\beta}{2}\right) = p\left(\frac{\beta}{2}n\right) = pa.$

III. $1 + \frac{\beta}{2}w + \frac{\beta}{2}(j-1) = 1 + \frac{\beta}{2}\left(n-p+1-\frac{2}{\beta}\right) + \frac{\beta}{2}(j-1) = a - \frac{\beta}{2}(p-j).$

3.3.2 Correlated β-Wishart eigendistribution

We introduce the broken-arrow model for Wishart matrices with $I_n \otimes \Sigma_p$ covariance matrix. This matrix model resembles a broken $\searrow$ and is defined iteratively.

Notation: We change the use of notation for this subchapter to follow closely ([DEKV13]):

Σ_p	is w.l.o.g. a positive-definite diagonal covariance matrix
$\Sigma_{p,p}$	stands for the (p,p) entry.
$\Sigma_{1:p-1;1:p-1}$	stand for the minor from $(1,1)$ to $(p-1,p-1)$.
$\triangle^2(\sigma)$	is defined as $\prod_{i<j} \lvert\sigma_i^2 - \sigma_j^2\rvert$.

Broken-Arrow matrix model

For $S_\beta \sim W_p^\beta(n, \Sigma_p)$ we introduce a $p-1 \times p-1$ Wishart matrix $Q_\beta \sim W_{p-1}^\beta(n, \Sigma_{p-1})$ and get from there the $p \times p$ case.

	dimension		dimension	
$Q_\beta = X_\beta^T X_\beta$	$p-1 \times p-1$	X_β	$n \times p-1$	
$S_\beta = Z_\beta^T Z_\beta$	$p \times p$	Z_β	$p \times p$	(square matrix!)

The broken-arrow model is defined

for $p = 1$

$$Z = [\chi_{n\beta}\Sigma_{1,1}^{1/2}, 0, \ldots, 0]^T \qquad \text{with } \; n-1 \; \text{ zeros.}$$

for $p > 1$ and $n \geq p$ as

$$S_\beta = Z_\beta^T Z_\beta, \qquad Z_\beta = \begin{pmatrix} \tau_1 & & & \chi_\beta \Sigma_{p,p}^{1/2} \\ & \ddots & & \vdots \\ & & \tau_{p-1} & \chi_\beta \Sigma_{p,p}^{1/2} \\ & & & \chi_{(n-p+1)\beta}\Sigma_{p,p}^{1/2} \end{pmatrix} \tag{3.30}$$

where

- $\tau_1, \ldots, \tau_{p-1}$ are the singular values of Q_β.
- Σ_p is the positive-definite diagonal covariance matrix.
- χ_β are independent random variables.

Then if $\sigma_1, \ldots, \sigma_n$ are the singular values of Z_β the $\sigma_1^2, \ldots, \sigma_n^2$ are the eigenvalues of S_β.

Eigendistribution for correlated β-Wishart:

The main result of this subchapter involves the hypgergeometric function ${}_0F_0^{(\beta)}$ and a couple of identities for Jack polynomials. We first state the result and postpone the definitions.

Theorem 10 (Eigendistribution for correlated β-Wishart ensembles) [DEKV13, p.19/22]**.** *Let $S \sim W_p^\beta(n, \Sigma_p)$ be a correlated Wishart matrix with eigenvalues $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$ such that $\lambda_1 > \cdots > \lambda_p$. Then the eigenvalues have the following probability density function*

$$f_p(\lambda_1, \ldots, \lambda_p) = \frac{1}{C_L^\beta} {}_0F_0^{(\beta)}\left(-\frac{\beta}{2}\Sigma_p^{-1}, \Lambda\right) \prod_{j=1}^p \lambda_j^{\frac{n-p+1}{2}\beta-1} \prod_{1 \leq i < j \leq p} |\lambda_j - \lambda_i|^\beta \tag{3.31}$$

with constant:

$$C_L^\beta = \left(\frac{2}{\beta}\right)^{np\beta/2} \det(\Sigma_p)^{n\beta/2} \frac{\Gamma_p^{(\beta)}(n\beta/2)\Gamma_p^{(\beta)}(p\beta/2)}{\pi^{p(p-1)\beta/2}} \tag{3.32}$$

where ${}_0F_0^{(\beta)}$ denotes the hypergeometric function for two matrix arguments.

Remark 16: *The eigenvalue density for correlated Wishart can be found for $\beta = 1$ in [Jam64, p.483] (with typo!), for $\beta = 2$ in [RVA04, p. 1] and for $\beta = 4$ in [LX09, p. 19]. The proof of the theorem is very long and technical. It requires many preliminary results about Jack polynomials and the Jacobian of arrow/broken-arrow matrix models. Therefore we only sketch it and the details of the proof can be looked up in ([DEKV13]) and the additional references mentioned there.*

Proof. (Sketch) Full proof in ([DEKV13]) and derivation of the normalisation constant too. The proof works like this: compute the proportional density for the singular values via induction. Then perform a change of variable to get the proportional eigenvalue density. For the singular values $\sigma_1 > \cdots > \sigma_p$ denote the diagonal matrix $\Phi := \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$.

First case $p = 1$: we have only one singular value σ_1 and $\sigma_1 \sim \chi_{n\beta}\Sigma_{1,1}^{1/2}$ having density f_{σ_1} then using $\exp\left(-\frac{\sigma^2}{2\Sigma_{1,1}}\right) = {}_0F_0^{(\beta)}\left(-\frac{1}{2}\sigma_1^2, \Sigma_{1,1}^{-1}\right)$

$$f_{\sigma_1} \propto \Sigma_{1,1}^{-n\beta/2}\sigma_1^{n\beta-1}\,{}_0F_0^{(\beta)}\left(-\frac{1}{2}\sigma_1^2, \Sigma_{1,1}^{-1}\right)$$

where the identity used can be found in ([For10]) and ${}_0F_0^{(\beta)}(S, I_p) = {}_0F_0^{(\beta)}(S)$ by definition. Then make the change of variable $\sigma_1 = \sqrt{\lambda_1}$ with $\mathrm{d}\sigma = \frac{1}{2}\lambda^{-1/2}\mathrm{d}\lambda$ to get

$$f_{\lambda_1} \propto \Sigma_{1,1}^{-n\beta/2}\lambda_1^{\frac{n}{2}\beta-1}\,{}_0F_0^{(\beta)}\left(-\frac{1}{2}\lambda_1, \Sigma_{1,1}^{-1}\right)$$

Case $p > 1$: Let

$$Z_\beta = \begin{pmatrix} \tau_1 & & & \chi_\beta\Sigma_{p,p}^{1/2} \\ & \ddots & & \vdots \\ & & \tau_{p-1} & \chi_\beta\Sigma_{p,p}^{1/2} \\ & & & \chi_{(n-p+1)\beta}\Sigma_{p,p}^{1/2} \end{pmatrix} = \begin{pmatrix} \tau_1 & & & a_1 \\ & \ddots & & \vdots \\ & & \tau_{p-1} & a_{p-1} \\ & & & a_p \end{pmatrix}$$

then (a_i) are χ-distributed and by hypothesis, assume (τ_i) come from the β-Wishart distribution. Let $f_{\tau_1,\ldots,\tau_{p-1},a_1,\ldots,a_n}$ be their joint density which is proportional to

$$\det(\Sigma)^{-n\beta/2}\prod_{i=1}^{p-1}\tau_i^{(n-p+2)/\beta-1}\triangle^2(\tau)^\beta\,{}_0F_0^{(\beta)}\left(-\frac{1}{2}T^2, \Sigma_{1:p-1;1:p-1}^{-1}\right)$$
$$\times\left(\prod_{i=1}^{p-1}a_i\right)^{\beta-1}a_p^{(n-p+1)\beta-1}\exp\left(-\frac{1}{2\Sigma_{p,p}}\sum_{i=1}^{p}a_i^2\right)\mathrm{d}a\mathrm{d}\tau$$

where $T = \mathrm{diag}(\tau_1, \ldots, \tau_{p-1})$. One can then apply lemma 2 and 3 from ([DEKV13]) to get the following transformation

$$\mathrm{d}a\mathrm{d}\tau = \left(\prod_{i=1}^{p-1} a_i\right) \prod_{i=1}^{p} q_i^{-1} \mathrm{d}q\mathrm{d}\sigma.$$

By using further transformation lemmas for the Jacobian of arrow/broken-arrow from ([DEKV13]) one eventually gets

$$f_{\tau_1,\ldots,\tau_{p-1},a_1,\ldots,a_n} \propto \det(\Sigma)^{-n\beta/2} \prod_{i=1}^{p-1} \sigma_i^{(n-p+1)\beta-1} \triangle^2(\sigma)^\beta {}_0F_0^{(\beta)}\left(-\frac{1}{2}T^2, \Sigma^{-1}_{1:p-1,1:p-1}\right) \exp\left(-\frac{1}{2\Sigma_{p,p}} \sum_{i=1}^{n} a_i^2\right) \prod_{i=1}^{p} q_i^{\beta-1} \mathrm{d}\sigma\mathrm{d}q$$

To proceed one needs the result that

$${}_0F_0^{(\beta)}\left(-\frac{1}{2}\Phi^2, \Sigma^{-1}\right) \propto \int \prod_{i=1}^{p} q_i^{\beta-1} e^{-\|a\|^2/(2\Sigma_{p,p})} {}_0F_0^{(\beta)}\left(-\frac{1}{2}T^2, \Sigma^{-1}_{1:(p-1);1:(p-1)}\right) \mathrm{d}q.$$

where we can use $\|a\|^2 + \sum_{i=1}^{p-1} \tau_i^2 = \sum_{i=1}^{p} \sigma_i^2$ to reduce this expression to a simpler form. By rearranging things and applying lemma 10 of ([DEKV13]) the problem reduces to

$$\begin{aligned} &{}_0F_0^{(\beta)}\left(-\frac{1}{2}\Phi^2, \Sigma^{-1} - \frac{1}{\Sigma_{p,p}} I_p\right) \\ &\quad \propto \int \prod_{i=1}^{p} q_i^{\beta-1} {}_0F_0^{(\beta)}\left(-\frac{1}{2}T^2, \Sigma^{-1}_{1:(p-1);1:(p-1)} - \frac{1}{\Sigma_{p,p}} I_{p-1}\right) \mathrm{d}q. \end{aligned}$$

Then both ${}_0F_0^{(\beta)}$ must be expressed in infinitely many Jack polynomials

$$\begin{aligned} &\sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_k} C_\kappa^{(\beta)}\left(-\frac{1}{2}\Phi^2\right) C_\kappa^{(\beta)}\left(\Sigma^{-1} - \frac{1}{\Sigma_{p,p}} I_p\right) \\ \propto \int \prod_{i=1}^{p} q_i^{\beta-1} &\sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_k} C_\kappa^{(\beta)}\left(-\frac{1}{2}T^2\right) C_\kappa^{(\beta)}\left(\Sigma^{-1}_{1:(p-1);1:(p-1)} - \frac{1}{\Sigma_{p,p}} I_{p-1}\right) \mathrm{d}q \end{aligned}$$

One needs now to compare if the Jack polynomials are proportional which can be shown by applying some more lemmas and theorems from ([DEKV13]). Finally one arrives at the proportional density for the singular values where one has to perform a last change of variable to get the proportional eigenvalue density. A scaling of Σ_p^{-1} to $\beta\Sigma_p^{-1}$ gives the $\beta^{-np\beta/2}$ coefficient. □

3.4 Real and complex correlated Wishart ensembles

Hypergeometric functions of matrix arguments

By an ordered partition $\kappa = (k_1, k_2, \ldots, k_j)$ of $k \in \mathbb{N}$ for $j \leq p$ we mean that

$$k_1 \geq \cdots \geq k_j \geq 0 \quad \text{and} \quad \sum_{i=1}^{j} k_i = k \quad \text{and} \quad l(\kappa) = \#\text{ non-zeros entries}$$

$$\kappa = (k_1, k_2, \ldots, k_{l(\kappa)}, 0, \ldots, 0), \qquad l(\kappa) \leq j$$

For a p we denote by $\mathcal{P}_k$ the space of all partitions $\kappa = (k_1, k_2, \ldots, k_j)$ of k of size j. We also assume that they are *lexicographically ordered* meaning $(4,3,2,0) > (4,3,1,1) > (3,3,3,3)$. Example:

$$\mathcal{P}_4 = \{(4,0,0,0), (3,1,0,0), (2,2,0,0), (2,1,1,0), (1,1,1,1)\}.$$

We usually just write (4) instead of $(4,0,0,0)$ so

$$\mathcal{P}_4 = \{(4), (3,1), (2,2), (2,1,1), (1,1,1,1)\}.$$

Definition 11 (hypergeometric function of matrix argument) [DEKV13, p.17]**.** *Let S, T be a $p \times p$ matrix then*

$${}_tF_q^{(\beta)}\left[\begin{matrix} a_1 & \ldots & a_t \\ b_1 & \ldots & b_q \end{matrix}; S, T\right] := \sum_{k=0}^{\infty} \sum_{\kappa \in \mathcal{P}_k} \frac{(a_1)_\kappa^{(\beta)} \cdots (a_t)_\kappa^{(\beta)}}{(b_1)_\kappa^{(\beta)} \cdots (b_q)_\kappa^{(\beta)}} \frac{C_\kappa^{(\beta)}(S) C_\kappa^{(\beta)}(T)}{C_\kappa^{(\beta)}(I_p) k!} \tag{3.33}$$

where β is the Dyson-index and

- *$\mathcal{P}_k$ is the space of all integer partitions of k.*
- *$(a)^{(\beta)}$ are the generalized Pochhammer symbols*

$$(a)_\kappa^{(\beta)} = \prod_{i=1}^{l(\kappa)} \left(a - \frac{\beta}{2}(i-1)\right)_{k_i} = \prod_{i=1}^{l(\kappa)} \prod_{j=1}^{k_i} \left(a - \frac{\beta}{2}(i-1) + j - 1\right).$$

  for the partition $\kappa = (k_1, k_2, \ldots, k_i)$ of k.

- *$C_k^{(\beta)}(S)$ are Jack polynomials under the C normalization. If $k > p$ then they are 0.*
- *I_p is the $p \times p$ identity matrix.*

Remark 17: *We define also the single matrix case*

$${}_tF_q^{(\beta)}\left[\begin{matrix} a_1 & \ldots & a_t \\ b_1 & \ldots & b_q \end{matrix}; S\right] := {}_tF_q^{(\beta)}\left[\begin{matrix} a_1 & \ldots & a_t \\ b_1 & \ldots & b_q \end{matrix}; S, I_p\right]. \tag{3.34}$$

A list of sufficient conditions for converges of (3.34) can be found in ([DEKV13, p.17]). One of them is that $t \leq q$.

Notation warning: *Some authors prefer to use the parameter $\alpha := \frac{2}{\beta}$ so*

$$\beta = 1 \leftrightarrow \alpha = 2 \quad | \quad \beta = 2 \leftrightarrow \alpha = 1 \quad | \quad \beta = 4 \leftrightarrow \alpha = 1/2$$

Remark 18: *(Hypgeometric function as integral) The following identities can be found in [Jam64, p. 483/489].*

$$ {}_0F_0^{(1)}\left(-\frac{1}{2}\Lambda, \Sigma_p^{-1}\right) = \int_{\mathbb{O}(p)} e^{-\frac{1}{2}\operatorname{tr}(\Sigma_p^{-1} H\Lambda H^T)} \mathrm{d}H \qquad \text{if } \beta = 1 $$
$$ {}_0F_0^{(2)}\left(-\frac{1}{2}\Lambda, \Sigma_p^{-1}\right) = \int_{\mathbb{U}(p)} e^{-\frac{1}{2}\operatorname{tr}(\Sigma_p^{-1} U\Lambda U^*)} \mathrm{d}U \qquad \text{if } \beta = 2 $$

with corresponding normalized Haar measure $\mathrm{d}H$ *for* $\mathbb{O}(p)$ *resp.* $\mathrm{d}U$ *for* $\mathbb{U}(p)$*. For the integrals there are two different ways to evaluate.*

1. $\beta = 1$*: can be solved with the theory of zonal polynomials to find an infinite series expansion.*
2. $\beta = 2$*: can be solved with the Harish-Chandra-Itzykson-Zuber integral (HCIZ Integral).*

3.4.1 Jack polynomials

We call a polynomial $f(x_1, \ldots, x_n)$ *homogeneous* if all the nonzeros terms have the same degree. Example:

$$ x^5 + xy^4 + x^3y^2 \qquad \text{then} \qquad \deg(x^5) = \deg(xy^4) = \deg(x^3y^2) = 5. $$

We call a polynomial $f(x_1, \ldots, x_n)$ *symmetric* if $f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for any permutation σ.

Definition 12 (Jack polynomial) [DEKV13, p.11]**.** *Let* S *be a symmetric* $p \times p$ *matrix and* $\lambda_1, \ldots, \lambda_p$ *be its eigenvalues and* $\kappa = (k_1, \ldots, k_p)$ *are partition of* k*. The Jack polynomial* $C_\kappa^{(\beta)}$ *is the only homogenous polynomial that is an eigenfunction of the Laplace-Beltrami-type operator:*

$$ \Delta_S = \sum_{i=1}^{p} \lambda_i^2 \frac{\partial^2}{\partial \lambda_i^2} + \beta \cdot \sum_{1 \le i \ne j \le p} \frac{\lambda_i^2}{\lambda_i - \lambda_j} \frac{\partial}{\partial \lambda_i} \tag{3.35} $$

with eigenvalue $\rho_k^\beta + k(p-1)$ *where*

$$ \rho_k^\beta := \sum_{i=1}^{l(\kappa)} k_i(k_i - 1 - \beta(i-1)). $$

Meaning $C_\kappa^{(\beta)}$ *satisfy the ODE:*

$$ \Delta_S C_\kappa^{(\beta)}(S) = \left[\rho_k^\beta + k(p-1)\right] C_\kappa^{(\beta)}(S). $$

For a partition $\kappa = (k_1, k_2, \ldots, k_p)$ *we have*

$$ C_\kappa^{(\beta)} = \sum_{\gamma \le \kappa} c_{\kappa,\gamma}^{(\beta)} m_\gamma(S) = c_\kappa^{(\beta)} \lambda_1^{k_1} \cdots \lambda_p^{k_p} + \textit{lower weight} $$

where $m_\gamma(S)$ denotes the monomial symmetric functions of the eigenvalues of S. In addition

$$(\operatorname{tr} S)^k = \sum_{\kappa\in\mathcal{P}_k, l(\kappa)\le p} C_\kappa^{(\beta)}(S) = (\lambda_1 + \cdots + \lambda_p)^k. \tag{3.36}$$

Remark 19: *For a partition $\kappa = (k_1, \ldots, k_p)$ of k and matrix S with eigenvalues $\lambda_1, \ldots, \lambda_p$ the monomial symmetric functions are*

$$m_1(S) = \lambda_1 + \cdots + \lambda_p, \qquad m_2(S) = \lambda_1^2 + \cdots + \lambda_p^2 \qquad m_{1,1}(S) = \sum_{i<j} \lambda_i\lambda_j$$

$$m_\kappa(S) = \lambda_1^{k_1} \cdots \lambda_p^{k_p} + \textit{symmetric terms}$$

Example 4 : Let us use an easier notation for visibility $c_{\kappa,\gamma} := c_{\kappa,\gamma}^{(\beta)}$ and

$$(4, 3^2, 2, 1^4) := (4, 3, 3, 2, 1, 1, 1, 1).$$

Let $k = 4$ then the Jack polynomials have the following forms

$$\begin{aligned}
C_4^{(\beta)} &= c_4 m_4(S) + c_{4,(3,1)} m_{(3,1)}(S) + c_{4,(2^2)} m_{(2^2)}(S) \\
&\quad + c_{4,(2,1^2)} m_{(2,1^2)}(S) + c_{4,(1^4)} m_{(1^4)}(S) \\
C_{(3,1)}^{(\beta)} &= c_{(3,1),(3,1)} m_{(3,1)}(S) + c_{(3,1),(2,1^2)} m_{(2,1^2)}(S) + c_{(3,1),(1^4)} m_{(1^4)}(S) \\
C_{(2,2)}^{(\beta)} &= c_{(2,2),(2,2)} m_{(2,2)}(S) + c_{(2,2),(2,1^2)} m_{(2,1^2)}(S) + c_{(2,2),(1^4)} m_{(1^4)}(S) \\
C_{(2,1,1)}^{(\beta)} &= c_{(2,1^2 1),(2,1^2)} m_{(2,1^2)}(S) + c_{(2,1^2),(1^4)} m_{(1^4)}(S) \\
C_{(1,1,1,1)}^{(\beta)} &= c_{(1^4)} m_{(1^4)}(S)
\end{aligned}$$

One can define the table of coefficients

		λ				
		(4)	$(3,1)$	$(2,2)$	$(2,1,1)$	$(1,1,1,1)$
	(4)	1	$c_{4,(3,1)}$	$c_{4,(2,1)}$	$c_{4,(2,1^2)}$	$c_{4,(1^4)}$
	$(3,1)$	0	$c_{(3,1),(3,1)}$	$c_{(3,1),(2,2)}$	$c_{(3,1),(2,1^2)}$	$c_{(3,1),(1^4)}$
κ	$(2,2)$	0	0	$c_{(2^2),(2^2)}$	$c_{(2^2),(2,1^2)}$	$c_{(2^2),(1^4)}$
	$(2,1,1)$	0	0	0	$c_{(2,1^2),(2,1^2)}$	$c_{(2,1^2),(1^4)}$
	$(1,1,1,1)$	0	0	0	0	$c_{(1^4),(1^4)}$

where the first coefficient $c_{(4),(4)}$ must be one because of

$$(\operatorname{tr} S)^k = \sum_{\kappa\in\mathcal{P}_k, l(\kappa)\le p} C_\kappa^{(\beta)}(S) = (\lambda_1 + \cdots + \lambda_p)^k$$

The coefficients can be computed with the following recurrence ([DES04, p. 11]):

$$c_{\kappa,\lambda}^{(\beta)} = \frac{\beta}{\rho_\kappa^{(\beta)} - \rho_\lambda^{(\beta)}} \sum_{\lambda<\mu\le\kappa} \left((l_i + t) - (l_j - t)\right) c_{\kappa,\mu}^{(\beta)} \tag{3.37}$$

where $\lambda = (l_1, \ldots, l_i, \ldots, l_j, \ldots, l_p)$ and $\lambda = (l_1, \ldots, l_i + t, \ldots, l_j - t, \ldots, l_p)$ and ρ_κ^β as before

$$\rho_\kappa^\beta := \sum_{i=1}^{l(\kappa)} k_i(k_i - 1 - \beta(i-1)).$$

Example 5 : It follows straight from the definitions that

$$ {}_0F_0^{(\beta)}(S,I) = {}_0F_0^{(\beta)}(S) = \sum_{k=0}^{\infty} \sum_{\kappa\in\mathcal{P}_k, l(k)\le p} \frac{C_\kappa^{(\beta)}(S)}{k!} = \sum_{k=0}^{\infty} \frac{(\mathrm{tr}(S))^k}{k!} = \exp\left(\mathrm{tr}(S)\right). $$

Further it can be shown that

$$ C_\kappa^{(\beta)}(I_p) = \left(\frac{2}{\beta}\right)^{2k} k! \left(\frac{\beta}{2}p\right)_\kappa^{(\beta)} \frac{\prod_{1\le j<r\le l(\kappa)} (2k_j - 2k_r - j + r)}{\prod_{j=1}^{l(\kappa)} (2k_j + l(\kappa) - j)!} $$

For the case $\beta = 1$ see [Jam64].

Remark 20: *The Jack polynomials are a generalization of the zonal polynomials $C_\kappa(S) := C_k^{(1)}(S)$ ($\beta = 1$,$\alpha = 2$). Some authors ([Jam64] and [Tak84]) also use the name complex zonal polynomials for $C_k^{(2)}(S)$. The expression "under the C normalization" refers to the fact that there are different normalizations of Jack polynomials and we use the C-normalization.*

Zonal polynomials

Let S be a $p \times p$ symmetric matrix then we denote by V_k the vector space of homogeneous polynomials $\phi(S)$ of degree k in $q := \frac{1}{2}p(p+1)$ different elements of S. So if we number the elements in the upper triangular matrix of S from $1, \ldots, q$ then the dimension of V_k is the number of monomials

$$ \prod_{i\le j}^{p} s_{ij}^{k_{ij}} = s_1^{k_1} s_2^{k_2} \ldots s_q^{k_q} \qquad \text{with} \qquad \sum_{i=1}^{q} k_i = k $$

which is

$$ \dim(V_k) = \binom{k+q-1}{q-1} = \binom{k+\frac{1}{2}p(p+1)-1}{\frac{1}{2}p(p+1)-1} $$

It can be shown ([Jam64, p. 478]) that V_k can be decomposed into a direct sum of irreducible π-invariant subspaces

$$ V_k = \bigoplus_{\kappa\in\mathcal{P}_k} V_\kappa $$

where by π-invariant we mean invariant under the linear transformation of $\phi \in V_k(S)$ by the map

$$ L : \phi \to \pi\phi \qquad \text{with} \qquad (\pi\phi)(S) = \phi(L^{-1}S(L^{-1})^T) $$

induced by the congruence transformation $S \to L^{-1}S(L^{-1})^T$ for any invertible $p \times p$ matrix L. So a subspace $V_i \subset V_k$ is invariant if we have $\pi V_i \subset V_i$. By irreducible we mean that V_i has no proper invariant subspace.

3.4.2 Harish-Chandra-Itzykson-Zuber integral

In the beginning of this subchapter we have briefly mentioned that in the case $\beta = 2$ the integral representation of the function ${}_0F_0^{(2)}$ can be solved with the Harish-Chandra-Itzykson-Zuber integral formula.

Theorem 11 (HCIZ-Formula) [McS21, p.2]**.** *Let A and B hermitian $p \times p$ matrices with corresponding eigenvalues $a_1 < \cdots < a_p$ and $b_1 < \cdots < b_p$. The Harish-Chandra-Itzykson-Zuber integral formula is defined as*

$$\int_{\mathbb{U}(p)} e^{t \operatorname{tr} AUBU^*} \mathrm{d}U = \left(\prod_{j=1}^{p-1} j! \right) \frac{\det\left[\exp(t a_i b_j)\right]_{i,j=1}^{p}}{t^{(n^2-n)/2} \triangle(A) \triangle(B)} \tag{3.38}$$

for some complex constant $t \neq 0$ where $\triangle(A) := \prod_{i<j} (a_j - a_i)$.

Proof. This proof can be found in [Tao13]. We follow the *Dyson Brownian motion* approach. Let us denote the space of hermitian $p \times p$ matrices as V_p with corresponding Haar measure $\mathrm{d}M_p$. Let $f : V_p \to \mathbb{C}$ be a class function, that is $f(UAU^*) = f(A)$ for all $U \in \mathbb{U}(p)$ and all $A \in V_p$. Since Hermitian matrices are normal, we can diagonalize them by unitary matrices, therefore f depends only on the eigenvalues of $(\lambda_i(A))$ so we will use the notation

$$f(A) = f(\lambda(A)). \tag{3.39}$$

since we can see f as a function $f : \mathbb{R}^p_{\geq} \to \mathbb{C}$ where

$$\mathbb{R}^p_{\geq} := \{\lambda = (\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p : \lambda_1 \geq \cdots \geq \lambda_p\}$$

Let f be a smooth class function exponentially decaying, then by *Riesz representation theorem* ([Bre11, p. 97]) there exist a function $w : \mathbb{R}^p_{\geq} \to \mathbb{C}$ such that

$$\int_{V_p} f(M_p) \mathrm{d}M_p = T(f) = \langle f, w \rangle = \int_{\mathbb{R}^p_{\geq}} f(\lambda) w(\lambda) \mathrm{d}\lambda \tag{3.40}$$

for the Lebesgue measure $\mathrm{d}\lambda$ on $\mathbb{R}^p_{\geq}$ such that $\|T\|_{\mathrm{Op}} = \|w\|$. Now we want to compute explicitly $w(\lambda)$ and consider the following probability measure on V_p (i.e. the GUE)

$$\mathbb{P}_p(\mathrm{d}M_p) = C_p \exp(-\operatorname{tr}(M_p^2)/2) \mathrm{d}M_p$$

then we know that the eigenvalue density for $M_p \in V_p$ will be

$$\mathbb{P}_p(\lambda_1, \ldots, \lambda_p) = \frac{1}{(2\pi)^{p/2} c_p} \triangle_p(\lambda)^2 \exp(-|\lambda|^2/2) \mathrm{d}\lambda$$

and computing the expectation for the class function f

$$\begin{aligned} \mathbb{E}(f(M_p)) &= \int_{V_p} f(M_p) C_p \exp(-\operatorname{tr}(M_p^2)/2) \mathrm{d}M_p \\ &= \int_{\mathbb{R}^p_{\geq}} f(\lambda) \frac{1}{(2\pi)^{p/2} c_p} \triangle_p(\lambda)^2 \exp(-|\lambda|^2/2) \mathrm{d}\lambda. \end{aligned}$$

Because $\mathrm{tr}(M_p^2) = \|\lambda(M_p)\|^2$ we can use (3.40) to identify

$$w(\lambda) := \frac{1}{C_p(2\pi)^{p/2}c_p}\triangle_p(\lambda)^2$$

and get

$$\int_{V_p} f(M_p)\mathrm{d}M_p = \frac{1}{C_p(2\pi)^{p/2}c_p}\int_{\mathbb{R}^p_{\geq}} f(\lambda)\triangle_p(\lambda)^2\mathrm{d}\lambda. \tag{3.41}$$

Now consider two class functions f, g that are smooth and exponentially decaying, then for $t \in \mathbb{R}_{\geq 0}$ consider the integral

$$\begin{aligned}\int_{V_p}\int_{V_p} f(A)g(B)\frac{C_p}{t^{p^2/2}}\exp\left(-\mathrm{tr}\left(\frac{(A-B)^2}{2t}\right)\right)\mathrm{d}A\mathrm{d}B \\ = \frac{C_p}{t^{p^2/2}}\int_{V_p}\int_{V_p} f_t(A)g_t(B)\exp\left(\frac{1}{t}\mathrm{tr}\,(AB)\right)\mathrm{d}A\mathrm{d}B\end{aligned} \tag{3.42}$$

where $f_t(A) := \exp\left(-\frac{1}{2t}\mathrm{tr}\,(A^2)\right)f(A)$ and $g_t(B) := \exp\left(-\frac{1}{2t}\mathrm{tr}\,(B^2)\right)g(B)$. Now we conjugate $B \to UBU^*$ with some unitary matrices, then by invariance of tr by cyclic permutation $\mathrm{tr}((UBU^*)^2) = \mathrm{tr}(UUB^2U^*U^*) = \mathrm{tr}(U^*U^*UUB^2) = \mathrm{tr}(B^2)$ and $g(UBU^*) = g(B)$, so when we integrate the right-hand side of (3.42) with respect to the Haar measure $\mathrm{d}U$ on $\mathbb{U}(n)$ we get

$$\begin{aligned}\frac{C_p}{t^{p^2/2}}\int_{\mathbb{U}_p}\int_{V_p}\int_{V_p} f_t(A)g_t(B)\exp\left(\frac{1}{t}\mathrm{tr}\,(AUBU^*)\right)\mathrm{d}A\mathrm{d}B\mathrm{d}U \\ = \frac{C_p}{t^{p^2/2}}\int_{V_p}\int_{V_p} f_t(A)g_t(B)K_t(A,B)\mathrm{d}A\mathrm{d}B\end{aligned}$$

where

$$K_t(A,B) = \int_{\mathbb{U}(p)}\exp\left(\frac{1}{t}\mathrm{tr}\,(AUBU^*)\right)\mathrm{d}U$$

Now $K_t(A,B)$ is invariant in both A and B and therefore only depend on the eigenvalues of A and B and can be seen as a function $K_t : \mathbb{R}^p_{\geq} \times \mathbb{R}^p_{\geq} \to \mathbb{C}$ so we write $K_t(\lambda,\nu)$. We apply twice (3.41) and get thus

$$\begin{aligned}\int_{V_p}\int_{V_p} f(A)g(B)\frac{C_p}{t^{p^2/2}}\exp\left(-\mathrm{tr}\left(\frac{(A-B)^2}{2t}\right)\right)\mathrm{d}A\mathrm{d}B \\ = \frac{1}{C_p(2\pi)^p c_p^2 t^{p^2/2}}\int_{\mathbb{R}^p_{\geq}}\int_{\mathbb{R}^p_{\geq}} f_t(\lambda)g_t(\nu)K_t(\lambda,\nu)\triangle_p(\lambda)^2\triangle_p(\nu)^2\mathrm{d}\lambda\mathrm{d}\nu\end{aligned} \tag{3.43}$$

where λ denote the eigenvalues of A and ν the eigenvalues of B. Now if we look at

$$\int_{V_p} f(A)g\frac{C_p}{t^{p^2/2}}\exp\left(-\mathrm{tr}\left(\frac{(A-B)^2}{2t}\right)\right)\mathrm{d}A \tag{3.44}$$

for a fixed $B := B_0$ this can be understood as the expectation of $f(B+t^{1/2}G)$ for the Brownian motion $B_t = B_0 + t^{1/2}G$ where G is drawn from the GUE. Let $\lambda_1, \ldots, \lambda_p$

be the eigenvalues of $B_t = B + t^{1/2}G$ then by *Johansson's formula* ([Joh01a, p. 3]) the eigenvalue density is

$$\mathbb{P}_p(\lambda_1, \dots, \lambda_p) = \frac{1}{(2\pi t)^{p/2}} \frac{\triangle_p(\lambda)}{\triangle_p(\lambda(B))} \det\left(\exp\left(-\frac{(\lambda_i - \lambda_j(B))^2}{2t}\right)\right)_{1\le i,j\le p} \mathrm{d}\lambda$$

and so (3.44) becomes

$$\int_{\mathbb{R}^p_\geq} f(\lambda) \frac{1}{(2\pi t)^{p/2}} \frac{\triangle_p(\lambda)}{\triangle_p(\lambda(B))} \det\left(\exp\left(-\frac{(\lambda_i - \lambda_j(B))^2}{2t}\right)\right)_{1\le i,j\le p} \mathrm{d}\lambda.$$

Now we integrating against $g(B)$ and use (3.41) and write it in terms of eigenvalues $\nu_1, \dots \nu_p$ of B

$$\int_{\mathbb{R}^p_\geq} f(\lambda) g(\nu) \frac{1}{(2\pi t)^{p/2}} \frac{\triangle_p(\lambda)}{\triangle_p(\nu)} \det\left(\exp\left(-\frac{(\lambda_i - \nu_j)^2}{2t}\right)\right)_{1\le i,j\le p} \frac{1}{C_p (2\pi)^{p/2} c_p} \triangle_p(\nu)^2 \mathrm{d}\lambda \mathrm{d}\nu$$

and compare this result with (3.43). We conclude that

$$\exp(-\lambda^2/2t - \nu^2/2t) K_t(\lambda, \nu) = c_p t^{(p^2-p)/2} \frac{\det\left(\exp\left(-\frac{(\lambda_i-\nu_j)^2}{2t}\right)\right)_{1\le i,j\le p}}{\triangle_p(\nu)\triangle_p(\lambda)}$$

and

$$\int_{\mathbb{O}(p)} e^{\mathrm{tr}(AUBU^*)/t} \mathrm{d}U = c_p t^{(p^2-p)/2} \frac{\det\left(\exp\left(-\frac{(\lambda_i-\nu_j)^2}{2t}\right)\right)_{1\le i,j\le p}}{\triangle_p(\nu)\triangle_n(\lambda)}$$

and by analytic continuation we can replace t with $1/t \in \mathbb{C} \setminus \{0\}$ to obtain the Harish-Chandra formula where $c_p = \prod_{j=1}^{p-1} j!$. □

Remark 21: *(Validity of the use of Johansson's formula) There exist two proofs for the Johansson's formula in [Joh01a]. One is based on the HCIZ formula and the other is based on an analysis of the Dyson Brownian motion without the HCIZ formula. Therefore the proof is not circular.*

Remark 22: *(Harish-Chandra integral) This formula is a special case of the Harish-Chandra integral formula for the integral*

$$\int_G e^{\langle \mathrm{Ad}_g\, x, y\rangle} \mathrm{d}g$$

defined for any compact Lie group G that is connected and semisimple. We briefly explain the notation: $x, y \in \mathfrak{t} \subset \mathfrak{g} = \mathrm{Lie}(G)$ in the Cartan subalgebra, i.e. a nilpotent subalgebra of the Lie algebra $\mathfrak{g}$. The $\langle \cdot, \cdot \rangle$ denotes a Ad*-invariant scalar product, meaning Ad_g is an isometry $\langle \mathrm{Ad}_g\, x, , \mathrm{Ad}_g\, y\rangle = \langle x, y\rangle$ for all $x, y \in \mathfrak{g}$, $g \in G$. And $\mathrm{Ad}_g(\cdot) := D_e c_g(\cdot)$ denotes the derivative at some point e of the map $c_g(h) := ghg^{-1}$ for $g, h \in G$. And $\mathrm{d}g$ is the normalized Haar measure.*

A lengthy proof of the general version can be found in [McS21, p.2].

Remark 23: *(Stochastic calculus for the Wishart ensemble) Since we have used Dyson's Brownian motion we want to point out that there exist a similar process for the Wishart ensembles. Let $B(t)$ be a $n \times p$ matrix with independent standard Brownian motion entries. Let D be some deterministic matrix $n \times p$. Then*

$$W(t) = B(t)^T B(t) \tag{3.45}$$

defines the Wishart process with initial state $W_0 := D^T D$. The process for complex $B(t) = (b_{ij}(t))$ is usually called the Laguerre process. If $p = 1$ then $W(t) = B(t)^T B(t) = \|B(t)\|^2$ is the squared Bessel process of order $2n$.

In [MF89, pp. 127-129] it was now shown that the eigenvalue process $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ of the real Wishart process satisfies the stochastic differential equation

$$\mathrm{d}\lambda_i(t) = 2\sqrt{\lambda_i(t)}(\mathrm{d}v_i(t)) + n \ \mathrm{d}t + \sum_{k \neq i} \frac{\lambda_i(t) + \lambda_k(t)}{\lambda_i(t) - \lambda_k(t)} \mathrm{d}t. \tag{3.46}$$

where $v_1, v_2, \ldots, v_p$ are independent Brownian motions.

3.5 Wishart unitary ensemble

Let $S \in W_p^2(n, I_p)$ then we know by theorem 9 that the corresponding probability measure is

$$\mathbb{P}_p(\mathrm{d}\lambda_1, \ldots, \mathrm{d}\lambda_p) = \frac{1}{Z_L^{2,n}} |\triangle(\lambda)|^2 \prod_{j=1}^{p} \omega_\alpha(\lambda_j) \mathrm{d}\lambda_j \tag{3.47}$$

where

$$|\triangle(\lambda)|^2 := \prod_{1 \leq i < j \leq p} |\lambda_j - \lambda_i|^2, \qquad \omega_\alpha(\lambda_j) := \lambda_j^\alpha e^{-\lambda_j}, \qquad \alpha = n - p.$$

$$Z_L^{2,n} = 2^{pn} \prod_{j=1}^{p} \Gamma(1 + j)\Gamma(n - p + j)).$$

3.5.1 Orthogonal polynomial method

We now apply a classical technique to derive a determinantal form of the density f_p for the unitary ensembles. To summarize the procedure we follow:

I Transform $|\triangle(\lambda)|^2 \prod_{j=1}^{p} \omega_\alpha(\lambda_j)$ into $\frac{1}{\left(\prod_{j=0}^{p-1} \kappa_j\right)^2} \det\left(K_p(\lambda_i, \lambda_k)\right)$ where κ_j is the leading coefficient of a polynomial and $K_p(\lambda, \lambda)$ is a kernel.

II Integrating the kernel

III Derive some basic statistics and the n-point correlation function.

We follow closely [Dei00, Ch. 5.4], [LNV18, Ch. 10] and [BDT16, Ch. 6.2-6.3].

Rewriting the Vandermonde determinant

Let $\{p_k\}$ be any polynomials and denote the corresponding monic polynomials $\{\pi_k\}$ such that $p_k = \kappa_k \pi_k$. Then by properties of the Vandermonde determinant

$$|\triangle(\lambda)|^2 = \det \begin{pmatrix} \pi_0(\lambda_1) & \dots & \pi_0(\lambda_p) \\ \pi_1(\lambda_1) & \dots & \pi_1(\lambda_p) \\ \vdots & \vdots & \vdots \\ \pi_{p-1}(\lambda_1) & \dots & \pi_{p-1}(\lambda_p) \end{pmatrix}^2$$
$$= \frac{1}{(\kappa_0 \kappa_1 \dots \kappa_{p-1})^2} \det \underbrace{\begin{pmatrix} p_0(\lambda_1) & \dots & p_0(\lambda_p) \\ p_1(\lambda_1) & \dots & p_1(\lambda_p) \\ \vdots & \vdots & \vdots \\ p_{p-1}(\lambda_1) & \dots & p_{p-1}(\lambda_p) \end{pmatrix}^2}_{=:B}.$$

For any matrix $A = [a_{ij}]_{1 \le i,j \le p}$ we have

$$\det(A)^2 = \det(A^T A) = \det\left(\sum_{j=1}^{p} a_{ji} a_{jk} \right)$$

therefore when applied on B we get that

$$|\triangle(\lambda)|^2 = \frac{1}{(\kappa_0 \kappa_1 \dots \kappa_{p-1})^2} \det\left(\sum_{j=1}^{p} p_{j-1}(\lambda_i) p_{j-1}(\lambda_k) \right). \tag{3.48}$$

We replace the Vandermonde determinant in the density (3.47) with (3.48) and pull the Laguerre polynomial weights $\omega_\alpha(\lambda_j)$ inside the determinant

$$\begin{aligned} f_p(\lambda_1, \dots, \lambda_p) &= \frac{1}{Z_L^{\beta,a}} \frac{1}{(\kappa_0 \kappa_1 \dots \kappa_{p-1})^2} \det\left(\sum_{j=1}^{p} p_{j-1}(\lambda_i) p_{j-1}(\lambda_k) \right) \prod_{j=1}^{p} \omega_\alpha(\lambda_j) \\ &= \frac{1}{Z_L^{\beta,a}} \frac{1}{(\kappa_0 \kappa_1 \dots \kappa_{p-1})^2} \det\left(\sum_{j=1}^{p} p_{j-1}(\lambda_i) \sqrt{\omega_\alpha(\lambda_i)} p_{j-1}(\lambda_k) \sqrt{\omega_\alpha(\lambda_k)} \right) \\ &= \frac{1}{Z_L^{\beta,a}} \frac{1}{(\kappa_0 \kappa_1 \dots \kappa_{p-1})^2} \det\left(\sum_{j=1}^{p} \phi_j(\lambda_i) \phi_j(\lambda_k) \right) \\ &= \frac{1}{Z_L^{\beta,a}} \left(\prod_{j=0}^{p-1} \kappa_j \right)^{-2} \det\left(K_p(\lambda_i, \lambda_k) \right) \end{aligned}$$

where we introduced the *kernel* for $1 \le i, j \le p$

$$K_p(\lambda_i, \lambda_k) = \sum_{j=1}^{p} \phi_j(\lambda_i) \phi_j(\lambda_k), \qquad \phi_j(\lambda_i) = p_{j-1}(\lambda_i) \sqrt{\omega_\alpha(\lambda_i)}.$$

Integrating the kernel

The trick is now to pick orthonormal polynomials $\{p_j\}$ with respect to $\omega_\alpha(x)$ because then we get the *reproducing property*

$$\int K_p(x,y)K_p(y,z)\mathrm{d}y = \sum_{l=1}^{p}\sum_{j=1}^{p}\phi_l(x)\phi_j(z)\underbrace{\int \phi_l(y)\phi_j(y)\mathrm{d}y}_{0 \text{ when } l\neq j \text{ otherwise } 1}$$
$$= \sum_{j=1}^{p}\phi_j(x)\phi_j(z) = K_p(x,z).$$

which we need for the following lemma.

Lemma 5 (Integrating-out lemma) [Dei00, Ch. 5, p.103]**.** *Let $J_p = J_p(x) = (J_{ij})_{1\leq i,j\leq p}$, $x \in \mathbb{R}^p$ be a $p \times p$ matrix such that*

1. *$J_{ij} = f(x_i, x_j)$ for some measurable function $f : \mathbb{R}^2 \to \mathbb{C}$.*
2. *$\int f(x,y)f(y,z)\mathrm{d}\mu(y) = f(x,z)$ (reproducing property).*

then

$$\int \det(J_p(x))\mathrm{d}\mu(x_p) = (d-p+1)\det(J_{p-1}) \tag{3.49}$$

where $J_{p-1} = (J_{ij})_{1\leq i,j\leq p-1}$ and $d := \int f(x,x)\mathrm{d}\mu(x)$.

Proof.

$$\int \det(J_p(x))\mathrm{d}\mu(x_p) = \sum_{\tau\in\mathcal{S}_p}\mathrm{sgn}(\tau)\int\prod_{i=1}^{p} f(x_i, x_{\tau(i)})\mathrm{d}\mu(x_p)$$
$$= \sum_{k=1}^{p}\sum_{\tau:\tau(p)=k}\mathrm{sgn}(\tau)\int\prod_{i=1}^{p} f(x_i, x_{\tau(i)})\mathrm{d}\mu(x_p)$$

where the inner sum goes over all permutations such that the last index $\tau(p) = k$. In the case $\tau(p) = p$ we only integrate the last one

$$\sum_{\tau:\tau(p)=p}\int\prod_{i=1}^{p} f(x_i, x_{\tau(i)})\mathrm{d}\mu(x_p) = \sum_{\tau:\tau(p)=p}\prod_{i=1}^{p-1} f(x_i, x_{\tau(i)})\int f(x_p,x_p)\mathrm{d}\mu(x_p)$$
$$= d\sum_{\tau:\tau(p)=p}\prod_{i=1}^{p-1} f(x_i, x_{\tau(i)}) = d\cdot\det(J_{p-1}).$$

In the case $\tau(p) = k < p$ we have for some $j < p$ that $\tau(j) = p$ and by the reproducing property

$$\int f(x_j, x_{\tau(j)})f(x_p, x_{\tau(p)})\mu(x_p) = \int f(x_j,x_p)f(x_p,x_k)\mu(x_p) = f(x_j,x_k)$$

removing again the last x_p. Hence combining everything together we get

$$\int \det(J_p(x))\mathrm{d}\mu(x_p) = d \cdot \det(J_{p-1}) + \sum_{k=1}^{p-1} \sum_{\tau:\tau(p)=k} \operatorname{sgn}(\tau) f(x_1, x_\tau(1)) \cdots f(x_j, x_k) \cdots f(x_{p-1}, x_{\tau(p-1)}).$$

Now let $\widehat{\tau}$ be a permutation of $\{1, \ldots, p-1\}$ defined by τ and $\sigma(p) = k < p$, $\sigma(j) = p$ such that

$$\{\widehat{\tau}(1), \ldots, \widehat{\tau}(p-1)\} = \{\tau(1), \ldots, \tau(j-1), k, \tau(j+1), \ldots \tau(p-1)\}$$

then $\operatorname{sgn}(\widehat{\tau}) = -\operatorname{sgn}(\tau)$ so we get a sign change and it follows that

$$\int \det(J_p(x))\mathrm{d}\mu(x_p) = (d - (p-1)) \cdot \det(J_{p-1}) = (d - p + 1) \cdot \det(J_{p-1}).$$

□

Applying the *integrating-out lemma* (lemma 5) on the kernel $K_p(\lambda, \lambda)$

$$J_{ik} = K_p(\lambda_i, \lambda_k), \qquad d = \int K_p(\lambda, \lambda)\, \mathrm{d}\lambda = \sum_{j=1}^{p} \int \phi_j^2(\lambda)\mathrm{d}\lambda = p$$

we get

$$\begin{aligned} \int \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p}\, \mathrm{d}x_p &= (p - p + 1) \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p-1} \\ &= \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p-1}. \end{aligned} \tag{3.50}$$

Repeat the procedure

$$\int \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p}\, \mathrm{d}x_{p-1}\mathrm{d}x_p = 2 \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p-2}.$$

so by induction

$$\begin{aligned} \int \cdots \int \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p}\, \mathrm{d}x_2 \cdots \mathrm{d}x_p &= (p-1)! \det(K_p(\lambda_i, \lambda_k)) \\ \int \cdots \int \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p}\, \mathrm{d}x_1 \cdots \mathrm{d}x_p &= p! \end{aligned}$$

and for $1 \le n < p$

$$\underbrace{\int \cdots \int}_{p-n} \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le p}\, \mathrm{d}x_{n+1} \cdots \mathrm{d}x_p = (p-n)! \det(K_p(\lambda_i, \lambda_k))_{1\le i,k\le n}. \tag{3.51}$$

Determinantal form of the eigenvalue density

As a consequence we get the *determinantal form* of the density of the unordered eigenvalues

$$f_p(\lambda_1, \dots, \lambda_p) = \frac{1}{p!} \det\left(K_p(\lambda_i, \lambda_k)\right)_{1 \le i,k \le p} \tag{3.52}$$

respectively for the ordered eigenvalues $\lambda_1 > \cdots > \lambda_p$

$$\widehat{f}_p(\lambda_1, \dots, \lambda_p) = \det\left(K_p(\lambda_i, \lambda_k)\right)_{1 \le i,k \le p}. \tag{3.53}$$

Apply now the *Christoffel-Darboux Identity* (4) to the kernel

$$\begin{aligned} K_p(\lambda_i, \lambda_k) &= \sum_{j=1}^{p} \phi_j(\lambda_i)\phi_j(\lambda_k) \\ &= \sum_{j=0}^{p-1} p_j(\lambda_i)p_j(\lambda_k)\sqrt{\omega_\alpha(\lambda_i)}\sqrt{\omega_\alpha(\lambda_k)} \\ &= \frac{\kappa_{p-1}}{\kappa_p} \cdot \frac{p_p(\lambda_i)p_{p-1}(\lambda_k) - p_p(\lambda_k)p_{p-1}(\lambda_i)}{\lambda_i - \lambda_k} \sqrt{\omega_\alpha(\lambda_i)}\sqrt{\omega_\alpha(\lambda_k)}. \end{aligned}$$

Notice the change $j \to j-1$ because $\phi_j(u) = p_{j-1}(u)\sqrt{\omega_\alpha(u)}$.

3.5.2 The n-point correlation functions and some important statistics

For a bounded measurable function $h : \mathbb{R} \to \mathbb{R}$ we have

$$\begin{aligned} \mathbb{E}\left(\prod_{j=1}^{p}(1+h(\lambda_i))\right) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_p(\lambda_1, \dots, \lambda_p) \prod_{j=1}^{p}(1+h(\lambda_i))\mathrm{d}\lambda_j \\ &= \frac{1}{p!} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K_p(\lambda_i, \lambda_k)\right)_{1 \le i,k \le p} \prod_{j=1}^{p}(1+h(\lambda_i))\mathrm{d}\lambda_j \\ &= \int_{\lambda_1 > \cdots > \lambda_p} \det\left(K_p(\lambda_i, \lambda_k)\right)_{1 \le i,k \le p} \prod_{j=1}^{p}(1+h(\lambda_i))\mathrm{d}\lambda_j. \end{aligned} \tag{3.54}$$

Another useful tool is the following identity.

Lemma 6 (Andréief's formula) [BDT16, p.143]**.** *Let (X, μ) be a measure space and $f_i(x), g_i(x) \in L^2(X, \mu)$ for $i = 1, \dots, p$. Then*

$$\begin{aligned} &\int_X \cdots \int_X \det(f_i(x_j))_{1 \le i,j \le p} \det(g_i(x_j))_{1 \le i,j \le p} \prod_{k=1}^{p} \mathrm{d}\mu(x_k) \\ &\qquad = p!\, \det\left(\int_X f_i(x)g_i(x)\mathrm{d}\mu(x)\right)_{1 \le i,j \le p} \end{aligned} \tag{3.55}$$

Proof. Use Leibniz determinant formula on $\det(g_i(x_j))_{1\leq i,j\leq p}$ and then use the multilinearity of the determinant

$$\begin{aligned}
&\sum_{\pi\in S_p}\operatorname{sgn}(\tau)\int_X\cdots\int_X\det(f_i(x_j))\prod_{j=1}^p g_{\tau(j)}(x_j)\prod_{k=1}^p \mathrm{d}\mu(x_k)\\
&=\sum_{\pi\in S_p}\operatorname{sgn}(\tau)\int_X\cdots\int_X\\
&\quad\cdot\det\begin{pmatrix} f_1(x_1)g_{\tau(1)}(x_1) & f_1(x_2)g_{\tau(2)}(x_2) & \cdots & f_1(x_p)g_{\tau(p)}(x_p)\\ \vdots & \vdots & \ddots & \vdots\\ f_p(x_1)g_{\tau(1)}(x_1) & f_p(x_2)g_{\tau(2)}(x_2) & \cdots & f_p(x_p)g_{\tau(p)}(x_p)\end{pmatrix}\prod_{k=1}^p\mathrm{d}\mu(x_k)\\
&=\sum_{\pi\in S_p}\operatorname{sgn}(\tau)\det\left(\int_X f_i(x)g_{\tau(j)}(x)\mathrm{d}\mu(x)\right)_{1\leq i,j\leq p}\\
&=p!\ \det\left(\int_X f_i(x)g_j(x)\mathrm{d}\mu(x)\right)_{1\leq i,j\leq p}
\end{aligned}$$

□

Combining all together we derive the following important theorem.

Theorem 12 [BDT16, p.145]**.** *Let $h:\mathbb{R}\to\mathbb{R}$ be a bounded measurable function and $\{p_k\}$ are orthonormal polynomials with respect to the measure $\mathrm{d}\mu=\omega_\alpha(x)\mathrm{d}x$. Let $\{\kappa_k\}$ denote the leading coefficient of the $\{p_k\}$. Then*

$$\begin{aligned}
\mathbb{E}\left(\prod_{j=1}^p(1+h(\lambda_i))\right)&=\int_{\lambda_1>\cdots>\lambda_p}\det\left(K_p(\lambda_i,\lambda_k)\right)_{1\leq i,k\leq p}\prod_{j=1}^p(1+h(\lambda_i))\mathrm{d}\lambda_j\\
&=\det\left(I_p+K_pM_h\right)
\end{aligned}\tag{3.56}$$

where $K_p:L^2(\mathbb{R})\to L^2(\mathbb{R})$ is defined to be

$$K_p(u,v)=\frac{\kappa_{p-1}}{\kappa_p}\cdot\frac{p_p(u)p_{p-1}(v)-p_p(v)p_{p-1}(u)}{u-v}\sqrt{\omega_\alpha(u)}\sqrt{\omega_\alpha(v)}\tag{3.57}$$

where $M_h:L^2(\mathbb{R})\to L^2(\mathbb{R})$ is the multiplication by h.

Proof. Follows from the calculations above and the fact that K_p is bounded and therefore K_pM_h is a *trace class operator* (see definition 25 in chapter 4 *Fredholm Determinants*). □

The n-point correlation function

We are now ready to define the k-point correlation function.

Definition 13 (n-point correlation function) [Dei00, p.108]**.** *We define the k-point correlation function:*

$$R_k(x_1,\ldots,x_k)=\frac{p!}{(p-k)!}\underbrace{\int\cdots\int}_{p-k}f_p(x_1,\ldots,x_p)\mathrm{d}x_{k+1}\cdots\mathrm{d}x_p.$$

Using (3.52) and then using (3.51) we get for $k \leq p$

$$\underbrace{\int \cdots \int}_{p-k} f_p(\lambda_1, \ldots, \lambda_p) \prod_{j=k+1}^{p} \mathrm{d}\lambda_j = \frac{1}{p!} \underbrace{\int \cdots \int}_{p-k} \det\left(K_p(\lambda_i, \lambda_j)\right)_{1 \leq i,j \leq p} \prod_{j=k+1}^{p} \mathrm{d}\lambda_j$$
$$= \frac{(k-p)!}{p!} \det\left(K_p(\lambda_i, \lambda_j)\right)_{1 \leq i,j \leq k}.$$

and so we get the *determinantal form* for the correlation function

$$R_k(\lambda_1, \ldots, \lambda_k) = \det\left(K_p(\lambda_i, \lambda_j)\right)_{1 \leq i,j \leq k} \tag{3.58}$$

Zero eigenvalues in an interval

For a subset $I \subset \mathbb{R}$ and eigenvalues $\lambda_1, \ldots, \lambda_p$ we call the probability that there is no eigenvalue in I a *gap probability* and it is defined as

$$\mathbb{P}\left(\{\text{No eigenvalues in } I\}\right) = \int_{\lambda_1 > \cdots > \lambda_p} \prod_{j=1}^{p} (1 - \chi_I(\lambda_j)) \widehat{f}_p(\lambda_1, \ldots, \lambda_p) \prod_{j=1}^{p} \mathrm{d}\lambda_j$$
$$= \mathbb{E}\left[\prod_{j=1}^{p} (1 - \chi_I(\lambda_j))\right]$$

k intervals with exactly one eigenvalue

We follow closely [BDT16, p.149-152]. Let $\Delta_1, \ldots, \Delta_k$ be small positive numbers such that $I_j = [x_j, x_j + \Delta_j]$ for $1 \leq j \leq k$ are disjoint sets for some particles $x_1 > \cdots > x_k$ meaning we have particles in sets like this

$$\text{particles:} \quad \underbrace{x_k} \quad < \quad \cdots \quad < \quad \underbrace{x_2} \quad < \quad \underbrace{x_1}$$

$$\text{sets:} \qquad I_k \quad | \quad \cdots \quad | \quad I_2 \quad | \quad I_1$$

Let $\lambda_1, \ldots \lambda_p$ be eigenvalues with $p \geq k$ and look at the subset

$$\{\lambda_{i_1}, \ldots, \lambda_{i_k}\} \subseteq \{\lambda_1, \ldots, \lambda_p\} \qquad i_1 < \cdots < i_k.$$

We are interested in the probability that all of the intervals $I_1, \ldots, I_k$ have exactly one eigenvalue and we don't care about the order so all $I_1, \ldots, I_k$ can possibly contain any of the $\lambda_{i_1}, \ldots, \lambda_{i_k}$.

$$\mathbb{P}_p\left(\bigcap_{1 \leq j \leq k} \{I_k \text{ contains exactly 1 eigenvalue}\}\right)$$
$$= \mathbb{P}_p\left(\bigcup_{i_1 < \cdots < i_k} \bigcap_{1 \leq j \leq k} \{I_j \text{ contains the } \lambda_{i_j} \text{ eigenvalue}\}\right)$$
$$= \mathbb{E}\left[\sum_{i_1 < \cdots < i_k} \prod_{j=1}^{k} \chi_{I_j}(\lambda_{i_j})\right] =: \mathbb{E}\left[g_{sym}(\lambda_1, \ldots, \lambda_p)\right]$$

So this probability is actually the expectation of a symmetric function of the eigenvalues. When we integrate

$$\begin{aligned}
&\int\cdots\int g_{sym}(\lambda_1,\ldots,\lambda_p)\mathbb{P}_p(\mathrm{d}\lambda_1,\ldots,\mathrm{d}\lambda_p)\\
&\quad=\sum_{i_1<\cdots<i_k}\int\cdots\int\chi_{I_1}(\lambda_{i_1})\cdots\chi_{I_k}(\lambda_{i_k})f_p(\lambda_1,\ldots,\lambda_p)\prod_{j=1}^{p}\mathrm{d}\lambda_j\\
&\quad=\sum_{i_1<\cdots<i_k}\int\cdots\int\chi_{I_1}(\eta_1)\cdots\chi_{I_k}(\eta_k)f_p(\eta_1,\ldots,\eta_p)\prod_{j=1}^{p}\mathrm{d}\eta_j\\
&\quad=\binom{p}{k}\underbrace{\int\cdots\int}_{p}\chi_{I_1}(\eta_1)\cdots\chi_{I_k}(\eta_k)f_p(\eta_1,\ldots,\eta_p)\prod_{j=1}^{p}\mathrm{d}\eta_j
\end{aligned}$$

where we did a change of variable $\lambda_{i_j}\to\eta_j$ in the second last line and used the fact that the density f_p is a symmetric function in $(\lambda_1,\ldots,\lambda_p)$. This gives us the same integral for all $i_1<\cdots<i_k$. Therefore we get

$$\begin{aligned}
&\lim_{\Delta_1\downarrow 0\cdots\Delta_k\downarrow 0}\frac{\mathbb{P}_p\left(\bigcap\limits_{1\le j\le k}\{I_k\text{ contains exactly 1 eigenvalue}\}\right)}{\Delta_1\cdots\Delta_k}\\
&\quad=\binom{p}{k}\underbrace{\int\cdots\int}_{p-k}f_p(y_1,\ldots,y_k,\eta_{k+1},\ldots,\eta_p)\prod_{j=k+1}^{p}\mathrm{d}\eta_j\\
&\quad=\binom{p}{k}\underbrace{\int\cdots\int}_{p-k}f_p(x_1,\ldots,x_k,\eta_{k+1},\ldots,\eta_p)\prod_{j=k+1}^{p}\mathrm{d}\eta_j\\
&\quad=p!\,R_k(x_1,\ldots,x_k)
\end{aligned}\tag{3.59}$$

Remark 24: *It's important to notice that R_k is not a probability distribution then*

$$\int\cdots\int R_k(x_1,\ldots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k=\frac{p!}{(p-k)!}.$$

Lemma 7 [BDT16, p.150/151]**.** *For any subset $I\subset\mathbb{R}$ we have*

$$\int_I\cdots\int_I R_n(x_1,\ldots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k=\mathbb{E}[\#\text{ of }k\text{-tuples of eigenvalues in }I\,]\tag{3.60}$$

and in particular

$$\int_I R_1(x_1)\mathrm{d}x_1=\mathbb{E}[\#\text{ of eigenvalues in }I\,].\tag{3.61}$$

Proof. Let $T_k=\{(i_1,\ldots,i_k)\in\{1,\ldots,p\}^k : i_1,\ldots,i_k\text{ are all distinct}\}$. Then

$$\int_I \cdots \int_I R_k(x_1, \ldots, x_k) \mathrm{d}x_1 \cdots \mathrm{d}x_k \\ = \frac{p!}{(p-k)!} \int_{\mathbb{R}^p} \left(\prod_{j=1}^{k} \chi_I(\lambda_j) \right) f_p(\lambda_1, \ldots, \lambda_p) \prod_{j=1}^{p} \mathrm{d}\lambda_j \tag{3.62}$$

notice that $|T_k| = \frac{p!}{(p-k)!}$ and so the right-hand side is the same as

$$\int_{\mathbb{R}^p} \sum_{(i_1,\ldots,i_k)\in T_k} \left(\prod_{j=1}^{k} \chi_I(\lambda_{i_j}) \right) f_p(\lambda_1, \ldots, \lambda_p) \prod_{j=1}^{p} \mathrm{d}\lambda_j = \mathbb{E}\left[\sum_{(i_1,\ldots,i_k)\in T_k} \prod_{j=1}^{k} \chi_I(\lambda_{i_j}) \right]$$

which is (3.60). □

Fredholm expansion of the gap probability

For a subset $I \subset \mathbb{R}$ and eigenvalues $\lambda_1, \ldots, \lambda_p$ we can generalize the gap probability. Let t be some parameter then by using some combinatorics (appendix A.1 lemma 15) we get that

$$\mathbb{E}\left[\prod_{j=1}^{p} (1 - t\chi_I(\lambda_j) \right] \\ = \sum_{k=0}^{p} (-t)^k \sum_{i_1<\cdots<i_k} \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}_{p} \left(\prod_{j=1}^{k} \chi_I(\lambda_{i_j}) \right) f_p(\lambda_1, \ldots, \lambda_p) \prod_{j=1}^{p} \mathrm{d}\lambda_j$$

using (3.62) the right-hand side becomes

$$\sum_{k=0}^{p} (-t)^k \sum_{i_1<\cdots<i_k} \frac{(p-k)!}{p!} \underbrace{\int_I \cdots \int_I}_{k} R_k(x_1, \ldots, x_n) \prod_{j=1}^{k} \mathrm{d}x_j \\ = \sum_{k=0}^{p} \frac{(-t)^k}{k!} \int_I \cdots \int_I R_k(x_1, \ldots, x_k) \prod_{j=1}^{k} \mathrm{d}x_j \\ = \sum_{k=0}^{p} \frac{(-t)^k}{k!} \int_I \cdots \int_I \det\left(K_p(x_i, x_j) \right)_{1\le i,j\le k} \prod_{j=1}^{k} \mathrm{d}x_j$$

where we used (3.58) in the last equality. Futher let $K_{p,I}$ be the restriction of K_p to $L^2(I)$ then by theorem 12 we know that

$$\begin{aligned}\mathbb{E}\left[\prod_{j=1}^{p}(1-t\chi_I(\lambda_j)\right] &= \det(I_p - tK_{p,I}) \\ &= \sum_{k=0}^{p}\frac{(-t)^k}{k!}\int_I\cdots\int_I \det\left(K_p(x_i,x_j)\right)_{1\le i,j\le k}\prod_{j=1}^{k}\mathrm{d}x_j \\ &= \sum_{k=0}^{\infty}\frac{(-t)^k}{k!}\int_I\cdots\int_I \det\left(K_p(x_i,x_j)\right)_{1\le i,j\le k}\prod_{j=1}^{k}\mathrm{d}x_j\end{aligned}$$

We were able to expand the series to infinity because the series terminates at $k = p$ since $\mathrm{Rank}(K_p) = p$.

Computing it explicitly for the WUE

We have to pick orthonormal polynomials with respect to $\omega_\alpha(x_j) = x_j^\alpha e^{-x_j}$.
So we introduce the *Laguerre functions*

$$p_k(x) = L_k^{(\alpha)}(x)\frac{\sqrt{k!}}{\sqrt{(k+\alpha)!}} \tag{3.63}$$

$$\kappa_k = \frac{\sqrt{k!}}{\sqrt{(k+\alpha)!}}\frac{(-1)^k}{k!} = \frac{(-1)^k}{\sqrt{(k+\alpha)!k!}} \tag{3.64}$$

$$\pi_k(x) = L_k^{(\alpha)}(x)\frac{\sqrt{k!}}{\sqrt{(k+\alpha)!}}\sqrt{x_j^\alpha e^{-x}} \tag{3.65}$$

3.6 Determinantal point processes

In the last chapter we have seen that the k-point correlation function has the following determinantal form

$$R_k(\lambda_1,\ldots,\lambda_k) = \det\left(K_p(\lambda_i,\lambda_j)\right)_{1\le i,j\le k}. \tag{3.66}$$

The eigenvalues of the Wishart unitary ensemble form a *determinantal point processes (DPP)*. The setting for the theory of DPP (see [Joh05, p.2-5]):

- Λ is a complete separable metric space.
- $\xi = \sum_{x\in\Lambda}\delta_x$ is a point process on Λ and we abbreviate $\xi(x_1) := \xi(\{x_1\})$. ξ is a *simple* point process meaning $\mathbb{P}(\xi(x_i)) \le 1$ for all $x \in \Lambda$.
- Let $A_1,\ldots,A_m$ be bounded, disjoints Borel sets in Λ and k_i, $1 \le i \le m$ such that $1 \le k_i \le k$ and $k_1 + \cdots + k_m = k$. We define the factorial moment measure M_k on Λ^k

$$M_k(A_1^{k_1}\times\cdots\times A_k^{k_m}) = \mathbb{E}\left[\prod_{i=1}^{m}\frac{(\xi(A_i))!}{(\xi(A_i)-k_i)!}\right] \tag{3.67}$$

so $M_k(A) = \mathbb{E}[\xi(A)]$ for each bounded $A \subseteq \Lambda^k$. Further we have for all Borel sets A_i in Λ that $M_k(A_1, \ldots, A_k) = \mathbb{E}[\xi(A_1) \cdots \xi(A_k)]$.

- If M_k is absolutely continuous with respect to the reference measure λ^k (usually this is the Lebesgue measure) on Λ^k then

$$\begin{aligned}\mathbb{E}[\xi(A_1) \cdots \xi(A_k)] &= M_k(A_1, \ldots, A_k) \\ &= \int_{A_1 \times \cdots \times A_k} \rho_k(x_1, \ldots, x_k)\, \mathrm{d}\lambda(x_1) \cdots \mathrm{d}\lambda(x_k)\end{aligned} \tag{3.68}$$

for all Borel sets A_i in Λ. We call the Radon–Nikodym density ρ_k the k-point correlation function. Heuristically this is

$$\rho_k(\mathrm{d}x_1, \cdots, \mathrm{d}x_k) \approx \mathbb{P}(\xi(x_1) \cdots \xi(x_k)) = \mathbb{P}(x_1, \ldots, x_k \in \Lambda).$$

Definition 14 (Determinantal point process) [Joh05, p.7]**.** *Let Λ be a complete separable metric space and ξ be a point process for some reference measure λ. Assume the correlation function ρ_k exist. If there exist a function $K : \Lambda \times \Lambda \to \mathbb{C}$ such that*

$$\rho_k(x_1, \ldots, x_k) = \det\left(K(x_i, x_j)\right)_{1 \le i,j \le k} \tag{3.69}$$

for all $x_1, \ldots, x_k \in \Lambda$, $k \ge 1$, then ξ is called a determinantal point process (DPP). The function K is called the correlation kernel of the process.

Some useful statistics for a DPP

Consider a DPP ξ on Λ with kernel K. Let $\phi : \Lambda \to \mathbb{C}$ be a bounded Borel function with support in a bounded set B. By using again A.1 lemma 15 we have

$$\begin{aligned}\mathbb{E}\left[\prod_i (1 + \phi(x_i))\right] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{1}{k!} \sum_{x_{i_1} \neq \cdots \neq x_{i_k}} \phi(x_{i_1}) \cdots \phi(x_{i_k})\right] \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \int \left(\prod_{i=1}^{k} \phi(x_i)\right) \rho_k(x_1, \ldots, x_k) \mathrm{d}\lambda(x_1) \cdots \lambda(x_k) \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} \int \left(\prod_{i=1}^{k} \phi(x_i)\right) \det\left(K(x_i, x_j)\right)_{1 \le i,j \le k} \mathrm{d}\lambda(x_1) \cdots \lambda(x_k) \\ &= \det(1 + K)\end{aligned} \tag{3.70}$$

and for $\phi := -\chi_B$ we get the useful gap probability

$$\begin{aligned}\mathbb{P}(\text{ no particles in B }) &= \mathbb{E}\left[\prod_i (1 - \chi_B(x_i))\right] \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{B^k} \det\left(K(x_i, x_j)\right)_{1 \le i,j \le k}\, \mathrm{d}\lambda(x_1) \cdots \lambda(x_k).\end{aligned} \tag{3.71}$$

For $x_{\max} := x_1 > x_2 > \cdots > x_k$ of ξ we have

$$\begin{aligned}\mathbb{P}\left(x_{\max} \leq t\right) &= \mathbb{P}\left(\text{ no particles in } (t,\infty)\right) \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{n!} \int_{(t,\infty)^k} \det\left(K(x_i,x_j)\right)_{1\leq i,j\leq k} \, \mathrm{d}\lambda(x_1)\cdots\lambda(x_k) \qquad (3.72) \\ &= \det(1-\chi_{(t,\infty)}K).\end{aligned}$$

Chapter 4

Fredholm Determinants

In random matrix theory some of the objects of interests are infinite matrices because their spectral density becomes deterministic. In the case of the scaled white Wishart matrices

$$S_n = \frac{1}{n} X^T X$$

it means that for the eigenvalues $\lambda_1(S_n) \geq \cdots \geq \lambda_p(S_n)$ the empirical spectral distribution

$$\mu_S := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(S)} \tag{4.1}$$

converges almost surely to the Marchenko-Pastur distribution μ_{MP} as $n \to \infty$. Since we are now in infinite dimension we need to be careful when we talk about functions of matrices such as the determinant or the trace which are introduced on finite-dimensional vector spaces.

As we have seen in the last chapter the determinant plays a crucial role in random matrix theory and some interesting probabilities appear in the form

$$\mathbb{P}\left(\{\dots\}\right) = \det\left(I + K\right)$$

where I is the identity operator and K an integral operator. These determinants are called *Fredholm determinants* and they are defined for *trace class operators* in a separable Hilbert space. The fundamental result that guarantees the existence of such determinants is *Lidskiis's trace theorem*

The goal of this chapter is to explain what Fredholm determinants are and to derive the *Fredholm formula*. First we study the determinant and its connection to the *exterior product* and recall some functional analysis. This part is mostly based on [Spi99],[Sim77], [Ger14] and [Bre11]. Then we study trace class operators and determinants in infinite dimension which is mostly based on [GGK00], [Sim05],[Sim77] and [Bor09].

Notation and assumptions: We assume some simplifications such as that the infinite topological vector spaces are all *Banach spaces*. And by *dual space* V^* we refer to the *continuous dual* so that the functionals are bounded. We use the

notation $\|\cdot\|$ for a norm on V instead of $\|\cdot\|_V$ with the exception of the operator norm. $\mathbb{K}$ will denote that the space is either over $\mathbb{R}$ or $\mathbb{C}$.

4.1 Trace and determinants

It is a classical fact that in a finite dimensional setting every matrix corresponds to a linear operator and vice versa. In infinite dimension every operator on a separable Hilbert space corresponds to a matrix but not every matrix corresponds to an operator.

Let V be a vector space over $\mathbb{K}$. The trace and the determinant of a linear operator $A : V \to V$ are defined as

$$\operatorname{tr}(A) := d_1\lambda_1 + \cdots + d_n\lambda_n \qquad \text{and} \qquad \det(A) := \lambda_1^{d_1} \cdots \lambda_n^{d_n} \tag{4.2}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues with algebraic multiplicities of A.

Some facts about the trace and determinant

The connection between the determinant and the trace is through

$$\log\det A = d_1 \log(\lambda_1) + \cdots + d_n \log(\lambda_n)$$

and since the eigenvalues of the matrix exponential $\exp(A)$ are exactly $\exp(\lambda_i)$ we have

$$\log(\det(\exp(A))) = \operatorname{tr}(A). \tag{4.3}$$

The determinant and the trace are closely tied to the characteristic polynomial

$$\det(A - xI) = (\lambda_1 - x)^{d_1}(\lambda_2 - x)^{d_2} \cdots (\lambda_n - x)^{d_n}. \tag{4.4}$$

Letting $x = 0$ gives the determinant. For general x by expanding the polynomial one can show that the the sum of the eigenvalues must be the coefficient of the second highest term

$$\det(A - xI) = (-1)^n(x^{d_1+\cdots+d_n} - \operatorname{tr}(A)x^{d_1+\cdots+d_n-1} + \cdots + \det(A)). \tag{4.5}$$

Remark 25: *(Trace as the diagonal entries) The trace can also be defined as the sum of diagonal elements (a_{ii}) of the matrix representation of A. If the vector space V is endowed with a scalar product then we have for any orthonormal basis $e_1, \ldots, e_n$ in V that*

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} \langle e_j, e_i \rangle \right) = \sum_{i=1}^{n} \left\langle \sum_{j=1}^{n} a_{ij} e_j, e_i \right\rangle = \sum_{i=1}^{n} \langle Ae_i, e_i \rangle.$$

The second equality is true because of orthogonality. Choosing any eigenbasis gives the same definition.

4.1.1 The exterior power of a linear operator

Definition 15 [Ger14, Ch. 6, p. 368]. *Given a vector space V with basis $\phi_1, \dots, \phi_n$. The exterior power $\Lambda^k(V)$ is defined to be the vector space*

$$\Lambda^k(V) = \otimes^k(V)/J^k(V) \tag{4.6}$$

where $\otimes^k(V)$ is the k-fold tensor product space of V and

$$J^k(V) := \text{span}\{v_1 \otimes \cdots \otimes v_n \mid \textit{where } v_i = v_j \textit{ for } (i,j) \textit{ with } i \neq j\}.$$

The dimension of $\Lambda^k(V)$ is $\binom{n}{k}$ and a basis is given by the set

$$\{\phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_k} \mid 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}. \tag{4.7}$$

Given any linear operator $A : V \to V$ this induces a multilinear and alternating map $\Lambda^k(A) : \Lambda^k(V) \to \Lambda^k(V)$ such that for any $v_1, \dots, v_k \in V$ one has

$$\Lambda^k(A)(v_1 \wedge \cdots \wedge v_k) = Av_1 \wedge \cdots \wedge Av_k \tag{4.8}$$

Remark 26: *(Exterior power properties) Let V be a $\mathbb{K}$-vector space then*

- *$\Lambda^1(V) = V$ and $\Lambda^0(V) = \mathbb{K}$.*
- *If $\dim(V) = n$ and $k > n$ then $\Lambda^k(V) = 0$.*
- *If $k = n$ then $\Lambda^n(A) : \Lambda^n(V) \to \Lambda^n(V)$ and since $\dim(\Lambda^n(V) = \binom{n}{n} = 1$ it means $\Lambda^n(A)$ can be represented as a 1×1 matrix which is just a multiplication by a scalar.*

By the definition of $J^k(V)$ any elment in $\Lambda^k(V)$ is an alternating tensor because

$$\begin{aligned}
0 &= v_1 \wedge \cdots \wedge (v_i + v_j) \wedge \cdots \wedge (v_i + v_j) \wedge \cdots \wedge v_k \\
&= (v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k) + (v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k) \\
&+ (v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k) + (v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k) \\
&= 0 + (v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_i \wedge \cdots \wedge v_k) + (v_1 \wedge \cdots \wedge v_i \wedge \cdots \wedge v_j \wedge \cdots \wedge v_k) + 0.
\end{aligned}$$

The projection from $\otimes^n(V)$ onto $\Lambda^n(V)$

For any tensor $T \in \otimes^n(V)$ we can define the projection onto the subspace $\Lambda^n(V)$ called the *alternation of T* and denoted as $\text{Alt}\, T$.

Definition 16 [Spi99, Ch. 7, p. 202]. *We define the map $\text{Alt} : \otimes^n(V) \to \Lambda^n(V)$ for any tensor $v_1 \otimes \cdots \otimes v_n$ as*

$$\begin{aligned}
\text{Alt}(v_1 \otimes \cdots \otimes v_n) &= \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \text{sgn}(\sigma)(v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}) \\
&= \frac{1}{n!}\, v_1 \wedge \cdots \wedge v_n
\end{aligned} \tag{4.9}$$

where $\text{sgn}(\sigma) = 1$ if σ is an even permutation and $\text{sgn}(\sigma) = -1$ if σ is odd.

The *wedge product* of $\omega_k \in \Lambda^k(H)$ and $\omega_l \in \Lambda^l(H)$ can now be defined as

$$\omega_k \wedge \omega_l = \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega_k \otimes \omega_l) \tag{4.10}$$

Remark 27: *(Dual basis / dual set) It's a classical result from linear algebra that (see [Ger14, Ch. 6, p. 332]) that for any finite-dimensional vector space V with basis $v_1, \dots, v_n$ there exist for every $i = 1, \dots, n$ exactly one linear functional $f_i^* : V \to \mathbb{K}$ such that they are biorthogonal meaning*

$$f_i^*(v_j) = \delta_{ij} \tag{4.11}$$

and $f_1^*, \dots, f_n^*$ form a basis of the dual space V^*. This is not true in infinite-dimensions. In general we call any set $f_1^*, f_2^*, \dots \in W^*$ that is biorthogonal (but not necessarily a basis) a dual set.

Let $v_1, \dots, v_n$ be a basis of a vector space H and $f_1^*, \dots, f_n^*$ be it's dual basis, then

$$\begin{aligned}(f_1^* \wedge \cdots \wedge f_n^*)(v_1, \dots, v_n) &= \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)(f_{\sigma(1)}^* \otimes \cdots \otimes f_{\sigma(n)}^*)(v_1, \dots, v_n) \\ &= (f_1^* \otimes \cdots \otimes f_n^*)(v_1, \dots, v_n) = 1\end{aligned}$$

The determinant and the exterior power

Let $v_1, \dots, v_n$ be a basis for V and define

$$w_i := \sum_{j=1}^{n} a_{ji} v_j, \qquad i = 1, \dots, n$$

then a consequence of the definition of the $\wedge$ product is that

$$w_1 \wedge \cdots \wedge w_n = (\det(a_{ij}))\, v_1 \wedge \cdots \wedge v_n. \tag{4.12}$$

According to (4.8) for a linear operator $A : V \to V$ with $\dim(V) = n$ the operator $\Lambda^n(A) : \Lambda^n(V) \to \Lambda^n(V)$ is precisely the map

$$\omega \mapsto \det(A)\, \omega$$

for any $\omega \in \Lambda^n(V)$ which is just multiplication by $\det(A)$.

This gives a new abstract definition of the determinant for the linear operator $A : V \to V$ independent of the basis of V.

Definition 17 (Determinant of a linear operator)**.** *The determinant of a linear operator $A : V \to V$ is the unique scalar* $\det(A)$ *such that for a basis $\phi_1, \dots, \phi_n \in V$*

$$\Lambda^n(A)(\phi_1 \wedge \cdots \wedge \phi_n) = \det(A)\, \phi_1 \wedge \cdots \wedge \phi_n. \tag{4.13}$$

Example 6 : Let $A : V \to V$ be linear operator and v_1, v_2 be a basis of V. Suppose the matrix representation with respect to this basis is

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then

$$\begin{aligned} \Lambda^2(A)(v_1 \wedge v_2) = Av_1 \wedge Av_2 &= (av_1 + cve_1) \wedge (bv_2 + dv_2) \\ &= ad\ v_1 \wedge v_2 + bc\ v_2 \wedge v_1 \\ &= (ad - bc)\ v_1 \wedge v_2. \end{aligned}$$

The eigenvalues of $\Lambda^k(A)$

The map $\Lambda^n(\cdot)$ is a functor ([Sim77, p. 248]) from and into the category of vector spaces so it holds that

$$\Lambda^n(AB) = \Lambda^n(A)\Lambda^n(B) \qquad \text{and} \qquad \Lambda^n(A^*) = \Lambda^n(A)^*$$

and thus $|\Lambda^n(A)| = \Lambda^n(|A|)$.

Let $A : V \to V$ be a linear operator with an *eigenbasis* $v_1, \ldots, v_n$ and eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$. Then one has for $i_1 < \cdots < i_k$ that

$$\begin{aligned} \Lambda^k(A)(v_{i_1} \wedge \cdots \wedge v_{i_k}) &= (Av_{i_1}) \wedge \cdots \wedge (Av_{i_k}) \\ &= (\lambda_{i_1}(A) \cdots \lambda_{i_k}(A))\ v_{i_1} \wedge \cdots \wedge v_{i_n} \end{aligned}$$

This means the eigenvalues of $\Lambda^k(A)$ are

$$(\lambda_{i_1}(A) \cdots \lambda_{i_k}(A)) \qquad \text{for } i_1 < \cdots < i_k. \tag{4.14}$$

This also shows that both definitions of the determinants that we have seen are equivalent since

$$\det(A) = (\lambda_{i_1}(A) \cdots \lambda_{i_n}(A)). \tag{4.15}$$

The trace of $\Lambda^k(A)$

Combining the definition (4.2) that the trace is the sum of its eigenvalues with (4.14) we thus have

$$\operatorname{tr}(\Lambda^k(A)) = \sum_{i_1 < \cdots < i_k} \lambda_{i_1}(A) \cdots \lambda_{i_k}(A). \tag{4.16}$$

Proposition 2 [Bor09, p. 10]**.** *Let $A : V \to V$ be a linear operator on a complex finite dimensional vector space V and let I be the identity operator. Then for $z \in \mathbb{C}$ we have*

$$\det(I + zA) = \sum_{k=0}^{n} z^k \operatorname{tr}(\Lambda^k(A)) \tag{4.17}$$

Proof. Without loss of generality assume we have eigenvalues with multiplicity 1. Notice that by (4.4) we get

$$z^{-n}\det(I+zA)=\det\left(z^{-1}I+A\right)=(\lambda_1+z^{-1})\cdots(\lambda_n+z^{-1})$$

therefore

$$\det(I+zA)=(z\lambda_1+1)\cdots(z\lambda_n+1). \tag{4.18}$$

Expanding (4.18) we get

$$\det(I+zA)=(1+z\lambda_1)\ldots(1+z\lambda_n)=\sum_{k=0}^{n}z^k\sum_{i_1<\cdots<i_k}\lambda_{i_1}\cdots\lambda_{i_k}. \tag{4.19}$$

Then apply (4.16) to (4.19). □

The Cauchy–Binet formula of the determinant

From (4.16) if we evaluate the trace of the matrix representation of A with respect to an orthonormal eigenbasis $\{\eta_i\}_{i=1}^n$ then we have for a scalar product

$$A_{i,i}:=\langle\eta_i,A\eta_i\rangle=\lambda_i(A) \tag{4.20}$$

and so

$$\begin{aligned}\operatorname{tr}(\Lambda^k(A))&=\sum_{i_1<\cdots<i_k}\lambda_{i_1}(A)\cdots\lambda_{i_k}(A)\\&=\sum_{i_1<\cdots<i_k}\langle\eta_{i_1},A\eta_{i_1}\rangle\cdots\langle\eta_{i_k},A\eta_{i_k}\rangle\\&=\sum_{i_1<\cdots<i_k}\det(\langle\eta_{i_p}A,\eta_{i_q}\rangle)_{p,q=1}^k\\&=\sum_{i_1<\cdots<i_k}\det(A_{i_p,i_q})_{p,q=1}^k\\&=\frac{1}{k!}\sum_{i_1,\ldots,i_k=1}^{k}\det(A_{i_p,i_q})_{p,q=1}^k\end{aligned}$$

and using proposition (2) we get the so called *Cauchy–Binet formula* of the determinant. The factor $(k!)^{-1}$ comes from the $k!$ duplicates that we get. The next definition uses the concept of a Schauder basis which we introduce in the next subchapter.

Definition 18 (Cauchy–Binet formula) [GGK00, p. 22]**.** *Let $\{\omega_i\}_1^\infty$ be a Schauder basis with biorthogonal system $\{\eta_i\}_1^\infty$. Then*

$$\det(I+zA)=\sum_{k=0}^{n}\frac{z^k}{k!}\sum_{i_1,\ldots,i_n=1}^{n}\det(A_{i_p,i_q})_{p,q=1}^k. \tag{4.21}$$

where $A_{i_p,i_q}:=\langle\omega_{i_p},A\eta_{i_p}\rangle$.

4.1.2 Preliminaries from functional analysis

It's time to introduce some definitions and basic theorems for operators: *the uniform boundedness principle* and the *spectral theorem for compact self-adjoint operators.* However in order to not blow-up this chapter we just state the theorems and refer for the proofs to [Bre11] (or any other standard book on functional analysis).

Definition 19 [Bre11, Ch. 2, p. 43]**.** *Let E and F be two Banach spaces. We say a linear operator $T : E \to F$ is bounded if there exists a constant $M \geq 0$ such that*

$$\|Tx\| \leq M\|x\| \qquad \forall x \in E \tag{4.22}$$

We denote the space of all linear bounded operators from E to V with $\mathcal{L}(E, V)$ and $\mathcal{L}(E) := \mathcal{L}(E, E)$. For a bounded operator we can then define the operator norm

$$\|T\|_{\mathcal{L}(E,F)} = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} \tag{4.23}$$

Theorem 13 (Uniform Boundedness Principle) [Bre11, Ch. 2, p. 32]**.** *Let E and F be two Banach spaces and let $(T_i)_{i_I}$ be a family of bounded linear operators from E into F. Assume that*

$$\sup_{i \in I} \|T_i x\| < \infty \quad \forall x \in E,\ \|x\| = 1.$$

Then

$$\sup_{i \in I} \|T_i\|_{\mathcal{L}(E,F)} < \infty \tag{4.24}$$

Proof. See [Bre11, Ch. 2, p. 32]. □

Definition 20 [Bre11, Ch. 6, p. 157]**.** *An operator $T \in \mathcal{L}(E, F)$ is said to be*

- *a finite-rank operator if the range $R(T)$ is of finite dimension. We denote the space of all finite-rank operators as $\mathcal{F}(E, F)$.*
- *a compact operator if the closed unit ball $B = \{x \in E : \|x\| \leq 1\}$ has compact closure in F in the usual topology, i.e.*

  $$\overline{T(B)} \in F$$

  is compact.

Definition 21 [Bre11, Ch. 6, p. 165]**.** *Let H, K be two Hilbert spaces and $T : H \to K$ a linear bounded operator. The adjoint operator $T^* : K \to H$ is the unique operator defined through*

$$\langle Tu, v\rangle = \langle u, T^* v\rangle. \tag{4.25}$$

where $u \in H$ and $v \in K$. If $H = K$ and $T = T^*$ then T is called self-adjoint.

Definition 22 [Bre11, Ch. 3, p. 72]**.** *We call a metric space E separable if there exist a subset $D \subset E$ that is countable and dense.*

Definition 23 [Bre11, Ch. 5, p. 146]**.** *Let B be a Banach space over $\mathbb{K}$. A sequence $(e_n)_{n\geq 1}$ is called a (Schauder) basis if for ever $v \in B$ there exist a unique sequence $(a_n)_{n\geq 1}$ in $\mathbb{K}$ such that $v = \sum_{i=1}^{\infty} a_n e_n$.*

Theorem 14 (Spectral theorem for compact self-adjoint operators) [Bre11, Ch. 6, p. 167]**.** *Let H be a separable Hilbert space and T a compact self-adjoint operator on it. Then there exist a basis composed of eigenvectors of T.*

Proof. See [Bre11, Ch. 6, p. 167]. □

Remark 28: *Not all separable Banach spaces have a Schauder basis which is a consequence of a result shown by Enflo [Enf73].*

4.2 Traces and determinants in infinite dimension

This chapter follows closely [GGK00]. Given any finite-rank operator T our first results shows that we can decompose a Hilbert space B according to the kernel structure of the operator T so that we have a finite-dimensional subspace.

Lemma 8 (Decomposing the Hilbert space) [GGK00, Ch. 1, pp. 5-6]**.** *Any finite-rank operator $T : B \to B$ on a Hilbert space B can be represented as a direct sum*

$$B = M_T \oplus N_T \tag{4.26}$$

where

$$T(M_T) \subset M_T \quad \text{and} \quad N_T \subset \ker T. \tag{4.27}$$

Proof. Consider a basis $\varphi_1, \ldots, \varphi_n \in \operatorname{Im} T$ and some vectors $v_1, \ldots, v_n \in B$ such that $Tv_j = \varphi_j$ for $j = 1, \ldots, n$. Let $f_1^*, \ldots, f_n^* \in B^*$ be a biorthogonal set of functionals for $\{\varphi_i\}$ meaning they are orthonormal

$$f_k^*(\varphi_i) = \delta_{ik}.$$

Because of this and since $\{\varphi_i\}$ is a basis we have for each $x \in B$ the decomposition

$$Tx = \sum_{j=1}^{n} \langle Tx, f_j \rangle \varphi_j$$

and so

$$T = \sum_{j=1}^{n} \varphi_j \otimes T^* f_j^* = \sum_{j=1}^{n} \varphi_j \otimes g_j$$

for $g_j := T^* f_j^*$. Set

$$M_T = \operatorname{span}\{\varphi_1, \ldots, \varphi_n, v_1, \ldots, v_n\}, \quad \text{and} \quad R_T = \bigcap_{j=1}^{n} \ker g_j$$

We can write any vector $x \in B$ as $x = u + v$ for

$$u = x - \sum_{j=1}^{n} \langle x, g_j \rangle v_j \in R_T, \qquad v = \sum_{j=1}^{n} \langle x, g_j \rangle v_j \in M_T$$

therefore we have the equality

$$B = M_T + R_T \tag{4.28}$$

and since

$$g_j(v_i) = T^* f_j^*(v_i) = f_j^*(Tv_i) = f_j^*(\varphi_i) = \delta_{ij}$$

we have

$$T(M_T) \subset M_T, \qquad \text{and} \qquad R_T \subset \ker T.$$

For the intersection $S_T := M_T \cap R_T$ we have $\dim(S_T) < \infty$ so R_T can be represented as $S_T \oplus N_T$. And so $B = M_T \oplus N_T$ is a direct sum and the subset M_T and N_T satisfy the conditions of the lemma. □

Determinants and traces of finite-rank operators

By looking at the finite-dimensional subspace M_T of the decomposition we can now define the trace of a finite-rank operator T and the determinant of $I + T$ where I denotes the identity operators. We define

$$T_1 := T \mid_{M_T}, \qquad I_1 := I \mid_{M_T}, \qquad I_2 := I \mid_{N_T} .$$

Clearly the functionals $\mathrm{tr}(T_1)$ and $\det(I_1 + T_1)$ are well-defined since M_T is finite-dimensional. Then for the operators T and $I + T$ we can decompose them with lemma (8)

$$T = \begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad I + T = \begin{pmatrix} I_1 + T_1 & 0 \\ 0 & I_2 \end{pmatrix} \tag{4.29}$$

and simply define

$$\mathrm{tr}(T) := \mathrm{tr}(T_1), \qquad \det(I + T) := \det(I + T_1). \tag{4.30}$$

This definition does not depend on the choice of M_T since the Jordan block that correspond to the non-zero eigenvalues do not depend on M_T

$$\mathrm{tr}(T) = \sum_j \lambda_j(T), \qquad \det(I + T) := \prod_j (1 + \lambda_j(T)). \tag{4.31}$$

The question now is: "*can we extend the definition of the trace and determinant to any bounded linear operators?*" or in other words can we assume for a sequence of bounded linear operators $\{T_n\}$ that the following is true

$$T_n \to T \text{ in } \mathcal{L}(\mathcal{B}) \implies \det(I + T) = \lim_{n \to \infty} \det(I + T_n)$$

The answer is no as the following example given in [GGK00, Ch. 2, p. 25] shows.

Example 7 : Let B be a Banach space with Schauder basis $(b_i)_{i\geq 1}$ and a corresponding dual set of bounded functionals $(e_i^*)_{i\geq 1} \in B^*$ (such that they are orthonormal). Consider the following operator

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b_i \otimes e_i^* \tag{4.32}$$

For any vector $v \in B$ by linearity of the functional e_i^* and the condition $e_i^*(b_k) = \delta_{ik}$ we have

$$b_i \otimes e_i^*(v) = b_i \otimes e_i^*(v_1 b_1 + v_2 b_2 + \dots) = b_i v_i \otimes e_i^*(b_i) = b_i v_i$$

Then we have $\|T_n\| = \frac{1}{\sqrt{n}}\|b_i v_i\| = \frac{1}{\sqrt{n}}\|v_i\|$ and so by the uniform boundedness principle it follows that $\lim_{n\to\infty} \|T_n\| = 0$. However $I + T_n$ is a diagonal matrix and

$$\det(I + T_n) = \det\left(\left(\delta_{jk} + \frac{1}{\sqrt{n}} e_j^*(b_k)\right)_{1\leq j,k\leq n}\right) = \left(1 + \frac{1}{\sqrt{n}}\right)^n \to \infty$$

and

$$\operatorname{tr}(I + T_N) = \frac{1}{\sqrt{n}} \sum_{j}^{n} e_j^*(b_j) = \frac{n}{\sqrt{n}} \to \infty$$

hence we have $\lim_{n\to\infty} \|T_n\| = 0$ but $\lim_{n\to\infty} \det(I + T_n) = \infty$.

So this example showed that it is impossible to just simply extend the definition to the space $\mathcal{L}(B)$. What can we do to still expand the determinant to infinite dimension? The answer is to restrict the class of operators.

4.2.1 Trace class operators

Let A be a bounded operator on a Hilbert space H. The operator A^*A is self-adjoint and has non-negative eigenvalues. First notice because of the anti-distributivity of the adjoint operator, i.e. $(AB)^* = B^*A^*$ we have

$$(A^*A)^* = (A^*A^{**}) = (A^*A).$$

Second from $A^*Ax = \lambda x$ and the positivity of the scalar product it follows that for an eigenvector $x \neq 0$ we have

$$\langle A^*Ax, x\rangle = \langle \lambda x, x\rangle = \lambda\langle x, x\rangle = \langle Ax, Ax\rangle \geq 0.$$

We recall the definition of singular values for a compact operator A.

Definition 24 (Singular values) [GGK00, Ch. 4, p. 49]**.** *Consider the eigenvalues of A^*A*

$$\lambda_1(A^*A) \geq \lambda_2(A^*A) \geq \dots$$

then the singular values or s-number of A are defined as

$$s_j(A) := \sqrt{(\lambda_j(A^*A))}, \qquad j = 1, 2, \dots$$

Since the eigenvalues of A^*A are non-negative numbers the same is true for the singular values. Notice that if A is a matrix

$$\prod_j s_j(A)^2 = \det(A^*A) = \det(A^*)\det(A) = |\det(A)|^2$$

and so

$$\prod_j s_j(A)^2 = \prod_j |\lambda_j(A)|^2. \tag{4.33}$$

A decomposition of compact operators

Theorem 15 (**Schmidt representation of compact operators**)[GGK00, Ch. 4, p. 49]**.** *Let A be a compact operator on a Hilbert space H. Then A admits a representation of the form*

$$Ax = \sum_{j=1}^{\nu(A)} s_j(A)\langle x, \phi_j\rangle\psi_j \tag{4.34}$$

for any $x \in H$ where $\nu(A)$ denotes the number of non-zero singular values of A and $\phi_1, \phi_2, \ldots$ and $\psi_1, \psi_2, \ldots$ are orthonormal systems.

Proof. The spectral theorem for compact self-adjoint operators on Hilbert spaces ensures that there exist an orthonormal basis of eigenvectors $\phi_1, \phi_2, \ldots$ such that

$$A^*Ax = \sum_{j=1}^{\nu(A)} \lambda(A^*A)\langle x, \phi_j\rangle\phi_j = \sum_{j=1}^{\nu(A)} (s_j(A))^2\langle x, \phi_j\rangle\phi_j$$

Now choose $\psi_j := s_j^{-1}A\phi_j$ for each j. Then we have another orthonormal system

$$\begin{aligned}\langle\psi_j, \psi_k\rangle = s_j^{-1}s_k^{-1}\langle A\phi_j, A\phi_k\rangle &= s_j^{-1}s_k^{-1}\langle A^*A\phi_j, \phi_k\rangle \\ &= s_j^{-1}s_k^{-1}\langle s_j^2\phi_j, \phi_k\rangle = \delta_{jk}\end{aligned}$$

Any $x \in H$ can be written as

$$x = \sum_{j=1}^{\nu(A)} \langle x, \phi_j\rangle\phi_j + u$$

where $u \in \ker(A^*A)$ since $\nu(A)$ only counts non-zero elements. But for a $u \in \ker(A^*A)$ we have $\langle A^*Au, u\rangle = \langle Au, Au\rangle$ so $\ker(A^*A) = \ker(A)$. And by definition $\psi_j s_j = A\phi_j$ so

$$Ax = \sum_{j=1}^{\nu(A)} s_j(A)\langle x, \phi_j\rangle\psi_j.$$

□

Trace class operators

Definition 25 (Trace class operator) [GGK00, Ch. 4, p. 59]**.** *Let A be a compact linear operator acting on a separable Hilbert space H such that the trace class norm is finite, i.e.*

$$\|A\|_1 := \sum_j s_j(A) < \infty \tag{4.35}$$

then we call this operator a trace class operator. We denote the space of all trace class operators with $\mathcal{S}_1$ and we define it's absolute value $|A| := \sqrt{A^*A}$.

In short

$$\sqrt{A^*A} \text{ has a finite trace} \quad \Longleftrightarrow \quad A \text{ is trace class.}$$

Remark 29: *(Schatten class) The $\mathcal{S}_1$ stands for Schatten class 1 which is the class of trace class operators. The Schatten class $\mathcal{S}_p$ for $p > 0$ are the set of linear compact operators A such that*

$$\|A\|_p := \sum_j s_j^p(A) < \infty. \tag{4.36}$$

The class $\mathcal{S}_2$ is very important. An operator $H \in \mathcal{S}_2$ is called a Hilbert-Schmidt operator. These are compact linear operators such that H^*H is a trace class operators.

Fredholm determinants and trace of a trace class operator

Definition 26 (Fredholm determinant) [Sim77, p. 245]**.** *Let $A \in \mathcal{S}_1$ then we define the Fredholm determinant as*

$$\det(I + zA) = \sum_{n=0}^{\infty} z^n \operatorname{tr}(\Lambda^n(A)). \tag{4.37}$$

The following lemma shows that if A is a trace class operator then $\Lambda^n(A)$ is also a trace class operator and hence the construction makes sense.

Lemma 9 [Sim77, Ch. 3, p. 253]**.** *Let $A \in \mathcal{S}_1(H)$ then $\Lambda^k(A) \in \mathcal{S}_1(\Lambda^k(H))$ and*

$$\|\Lambda^k(A)\|_1 \leq \frac{\|A\|_1^k}{k!} \tag{4.38}$$

Proof. We have $|\Lambda^k(A)| = \Lambda^k(|A|)$ so $\Lambda^k(A)$ has singular values $s_{i_1}(A) \cdots s_{i_k}(A)$ with $i_1 < \cdots < i_k$ so

$$\|\Lambda^k(A)\|_1 = \sum_{i_1 < \cdots < i_k} s_{i_1}(A) \cdots s_{i_k}(A) \leq \frac{1}{k!} \left(\sum_{n=1}^{\infty} s_{i_n}(A) \right)^k = \frac{1}{k!} \|A\|_1^k \tag{4.39}$$

□

Definition 27 (Trace of a trace class operator) [Sim77, p. 244]**.** *Let $A \in \mathcal{S}_1$ on a Hilbert space H and $e_1, e_2, \ldots$ be any orthonormal basis of H. Then we define the trace as*

$$\operatorname{tr}(A) = \sum_{i=1}^{\infty} \langle Ae_i, e_i \rangle. \tag{4.40}$$

4.2.2 Lidskii's trace theorem

Theorem 16 (Lidskii's Trace Theorem) [GGK00, Ch. 4, p. 63]**.** *Let A be a trace class operator and let $\lambda_1(A), \lambda_2(A), \dots$ be the sequence of eigenvalues of A (multiplicites taken into account). Then*

$$\operatorname{tr}(A) = \sum_{j=1}^{\infty} \lambda_j(A) \tag{4.41}$$

and

$$\det(I + A) = \prod_{j=1}^{\infty} (1 + \lambda_j(A)) . \tag{4.42}$$

Proof. We have

$$\det(I + zA) = \sum_{n=0}^{\infty} z^n \operatorname{tr}(\Lambda^n(A))$$

and the series converges uniformly by lemma 9. A has a basis of eigenvectors $(\phi_j)_{j=1}^{\infty}$ with eigenvalues $(\lambda_j)_{j=1}^{\infty}$ by the spectral theorem. By definition 27

$$\operatorname{tr}(A) = \sum_{i=1}^{\infty} \langle A\phi_i, \phi_i \rangle = \sum_j \lambda_j. \tag{4.43}$$

Similarly by (4.16)

$$\operatorname{tr}\left(\Lambda^n(A)\right) = \sum_{j_1 < \cdots < j_n} \lambda_{j_1} \cdots \lambda_{j_n}.$$

Now $g(z) := \prod_{j=1}^{\infty} (1 + z\lambda_j(A))$ is an entire function since

$$\sum_j |\lambda_j| \leq \sum_j |s_j| \leq \operatorname{tr}(A)$$

and it's coefficient are

$$\sum_{j_1 < \cdots < j_n} \lambda_{j_1} \cdots \lambda_{j_n} = \operatorname{tr}\left(\Lambda^n(A)\right)$$

thus

$$g(z) = \sum_{n=0}^{\infty} z^n \operatorname{tr}(\Lambda^n(A)) = \det(I + zA).$$

□

Remark 30: *It's a classical result (see for example [Ahl79, Ch. 5, p. 192]) that for $\lambda_j \in \mathbb{C}$*

$$\begin{aligned} &\prod_j (1 + \lambda_j) \text{ converges absolutely} \\ &\qquad \iff \sum_j |\lambda_j| \text{ converges.} \end{aligned} \tag{4.44}$$

If we let A to be the zero-matrix we have $\det(I - A) = 1$. It should be clear that for $z \in \mathbb{C}$ we have $\det(I - zA) = \prod_j (1 - z\lambda_j(A))$.

4.2.3 Fredholm Determinants

We have seen in definition 26 that we can define for a trace class operator A the *Fredholm determinant* $\det(I+zA)$ through the exterior algebra. However there exist several equivalent constructions of the Fredholm determinant. The central result that proves the equivalence of all these definitions is *Lidksii's theorem*. Detailed proofs that all construction are equivalent can be found in [GGK00, Ch. 2] and [Sim05, Ch. 3].

Definition 28 (Fredholm determinant) [Bor09, p .10]**.** *Let A be a trace class operater. The Fredholm determinant of A has three more constructions*

I Directly from Lidskii's trace theorem for a $z \in \mathbb{K}$

$$\det(I - zA) = \prod_j (1 - z\lambda_j(A)) \,. \tag{4.45}$$

II For any sequence of finite-rank operators A_n with $A_n \to A$ in the trace class norm, we have the following convergence for a $z \in \mathbb{K}$

$$\det(I + zA_n) \to \det(I + zA). \tag{4.46}$$

III For a $z \in \mathbb{K}$ such that $|z| < 1/|\lambda_1(A)|$ the following converges

$$\det(I + zA) = \exp(\mathrm{tr}(\log(I + zA))) = \exp\left(-\sum_{n=1}^{\infty} \frac{(-z)^n}{n} \mathrm{tr}(A^n)\right). \tag{4.47}$$

4.2.4 The Fredholm formula

The central object of Fredholm's original work was the *inhomogeneous Fredholm integral equation* of the second kind

$$u(x) + z \int_a^b K(x, y)u(y)\mathrm{d}y = f(x), \quad x \in [a, b] \tag{4.48}$$

where f and K are assumed to be continuous on $[a, b]$. Fredholm's approach was to discretize the integral diving the interval $[a, b]$ into n intervals of length h which leads to the system of equations in matrix form

$$(I + zhK)u = f \qquad \text{where} \quad u, f \in \mathbb{R}^n. \tag{4.49}$$

Its determinant is given by

$$D(h) := \det(I + zhK_{ij}) \tag{4.50}$$

Making now a Taylor expansion and taking the limit $n \to \infty$ so that $h \to 0$ leads to Fredholms famous determinant formula which we call the *Fredholm formula*.

A fully construction can be found in [Lax02, Ch. 24, pp 260]. Our approach will be different by making use of the definition 26 through the exterior power.

Integral operators

Definition 29 (Integral operator) [Lax02, Ch. 16, p.173]**.** *Let T,S be Hausdorff spaces equipped with measures μ and ν. A linear operator K on $L^2(T,\mu)$ is called an integral operator if it has the following form for $f \in L^2(T,\mu)$*

$$(Kf)(s) = \int_T k(s,t) f(t) \mathrm{d}\mu(t) \tag{4.51}$$

where $k : S \times T \to \mathbb{C}$ is a measurable function called kernel.

Remark 31: *(Condition to be trace class and Mercer's theorem) A sufficient condition for an integral operator K to be a trace class operator is when the kernel is a continous Hermitian non-negative kernel on a compact set T. A proof can be found in ([GGK00, p. 74]) but it is based on Mercer's theorem which is a corollary of the spectral theorem so we omit it here. The spectral theorem guarantees the existence of a eigendecomposition of the kernel $k(s,t)$*

$$k(s,t) = \sum_{j=1}^{\infty} \lambda_j e_j(s) e_j(t) \tag{4.52}$$

where $\lambda_j \geq 0$ are the eigenvalues with eigenfunctions e_j. Mercer's theorem is also used in statistical learning theory for kernel methods.

The trace of an integral operator

Let K be an integral operator with kernel $k : T \times T \to \mathbb{C}$. We now investigate the question what the meaning of $\otimes^n(K)$ and $\Lambda^n(K)$ is. Consider two functions $f_1 \in L^2(T)$ and $f_2 \in L^2(T)$. Then

$$K \otimes K(f)(s,t) = \int_T \int_T k_1(x_1,t) k_2(x_2,s) f_1(x_1) f_2(x_2) \mathrm{d}x_1 \mathrm{d}x_2$$

so if we define $f := f_1(s_1) \cdots f_n(s_n)$ we have $f \in L^2(T^n)$ and

$$\otimes^n Kf\ (s_1,\ldots,s_n) = \int_T \cdots \int_T k(x_1,s_1) \cdots k(x_n,s_n) f(x_1,\ldots,x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

Now we make use of $\mathrm{Alt}(\otimes^n Kf)$ and get

$$\begin{aligned}
&\Lambda^n Kf\ (s_1,\ldots,s_n) \\
&\quad = \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \mathrm{sgn}(\sigma) \int_T \cdots \int_T k(x_{\sigma(1)},s_1) \cdots k(x_{\sigma(n)},s_n) f(x_1,\ldots,x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&\quad = \frac{1}{n!} \int_T \cdots \int_T \det\left(k(x_i,s_j)\right)_{1 \leq i,j \leq n} f(x_1,\ldots,x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n
\end{aligned} \tag{4.53}$$

This means $\Lambda^n Kf\ (s_1,\ldots,s_n)$ is an integral operator with kernel

$$K(s_1,\ldots,s_n) := \frac{1}{n!} \det\left(k(x_i,s_j)\right)_{1 \leq i,j \leq n}. \tag{4.54}$$

Theorem 17 (Trace of an integral operator with continuous kernel) [Sim05, Ch. 3, p. 35]**.** *Let $A \in \mathcal{S}_1$ be an integral operator with continuous kernel k on $T \times T$ where T is compact. Then*

$$\operatorname{tr}(A) = \int_T k(x,x)\mathrm{d}x. \tag{4.55}$$

Proof. Without loss of generality $T = [0,1]$. For each n, let $\{\phi_{n,1}, \phi_{n,2}, \dots, \phi_{n,2^n}\}$ be the 2^n functions

$$\phi_{n,m} = \begin{cases} 2^{n/2} & \frac{m-1}{2^n} \le x \le \frac{m}{2^n} \\ 0 & \text{otherwise} \end{cases}$$

and denote with P_n the projection onto the span $\{\phi_{n,1}, \phi_{n,2}, \dots, \phi_{n,2^n}\}$. We have $\mathrm{A} = \lim_{n\to\infty} \operatorname{tr}(P_n A P_n)$ since one can easily construct a basis $\varphi_1, \varphi_2 \dots$ in $L^2(T)$ with $\varphi_1, \dots, \varphi_{2^n} \in \operatorname{Ran} P_n$. Now

$$\operatorname{tr}(P_n A P_n) = \sum_m \langle \phi_{n,m}, A\phi_{n,m}\rangle = 2^n \sum_{m=1}^{n} \int\int_{(m-1)/2^n \le x,y \le m/2^n} [k(x,y)]\mathrm{d}x\mathrm{d}y \tag{4.56}$$

Using the uniform continuity of $k(x,y)$ this integral converges to $\int_T k(x,x)\mathrm{d}x$ as $n \to \infty$ □

It is important to stress out that A has to have a continuous kernel for theorem 17 to be true.

The Fredholm formula

We now come to our main result.

Theorem 18 (Fredholm formula) [Sim05, Ch. 3, p. 36]**.** *Let $A \in \mathcal{S}_1$ be an integral operator with continuous kernel k on $T \times T$. Then*

$$\det(I + A) = \sum_{n=0}^{\infty} \frac{\alpha_n}{n!} \tag{4.57}$$

where

$$\alpha_n = \int_T \cdots \int_T \det\left[(k(x_i, s_j))_{1\le i,j\le n}\right] \mathrm{d}x_1 \dots \mathrm{d}x_n \tag{4.58}$$

Proof. From theorem 17 we know

$$\operatorname{tr}(A) = \int_T k(x,x)\mathrm{d}x$$

so combing it with (4.53) we see that $\operatorname{tr}(\Lambda^n(A)))$ is just an integral operator (Kf) with kernel $(n!)^{-1}\det(k(x_i,s_j))_{1\le i,j\le n}$ applied to $f = 1$ and hence

$$\operatorname{tr}(\Lambda^n(A)) = \frac{1}{n!}\int_T \cdots \int_T \det\left[(k(x_i,s_j))_{1\le i,j\le n}\right]\mathrm{d}x_1 \cdots \mathrm{d}x_n$$

and so

$$\begin{aligned}\det(I+A) &= \sum_{n=0}^{\infty} \operatorname{tr}(\Lambda^n(A)) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_T \cdots \int_T \det\left[(k(x_i,s_j))_{1\le i,j\le n}\right] \mathrm{d}x_1 \cdots \mathrm{d}x_n\end{aligned}$$

□

Remark 32: *Since the determinant is a homogeneous function it is easy to see that*

$$\begin{aligned}\det(I+zA) &= \sum_{n=0}^{\infty} \frac{z^n \alpha_n}{n!} \\ &= 1 + \sum_{n=1}^{\infty} \frac{z^n}{n!} \int_T \cdots \int_T \det\left[(k(x_i,s_j))_{1\le i,j\le n}\right] \mathrm{d}x_1 \ldots \mathrm{d}x_n.\end{aligned} \tag{4.59}$$

4.2.5 Numerical computation of Fredholm determinants

The numerical methods to compute Fredholm determinants can be divided into two classes.

1. Methods whose approximations are expressible as integrals (Quadrature).
2. Projective methods.

Which method is used depends on which construction $I - IV$ from (28) is used for the Fredholm determinant.

In ([Bor09]) two methods are explained: the *Nyström-type quadrature method* and the *projection method of Galerkin type*. The advoacted method is the Nyström-type quadrature method with *positive weights*.

Nyström-type quadrature method

For a quadrature rule

$$Q(f) = \sum_{j=1}^{m} w_j f(x_j) \approx \int_a^b f(x)\mathrm{d}x \tag{4.60}$$

the Nyström-type approximation of the Fredholm formula (4.59) with kernel $k(x,x)$ is

$$d_Q(z) = \det\left(\delta_{ij} + z w_i k(x_i,x_j)\right)_{i,j=1}^{n}. \tag{4.61}$$

If the weights w_j of the chosen quadrature rule (4.60) are positive then

$$d_Q(z) = \det(\delta_{ij} + z A_Q), \qquad A_Q := \left(w_i^{1/2} K(x_i,x_j) w_j^{1/2}\right)_{i,j=1}^{m}. \tag{4.62}$$

For the *von Koch form*: If we introduce the function

$$K_n(t_1, \ldots, t_n) := \det(K(t_p, t_q))_{p,q}^n \tag{4.63}$$

the Nyström-type approximation in *von Koch form* (4.21)

$$\begin{aligned} d_Q(z) &= 1 + \sum_{n=1}^{m} \frac{z^n}{n!} \sum_{j_1,\ldots,j_n=1}^{m} \det(w_{j_p} K(x_{j_p}, x_{j_q})_{p,q=1}^n \\ &= 1 + \sum_{n=1}^{m} \frac{z^n}{n!} \sum_{j_1,\ldots,j_n=1}^{m} w_{j_1} \cdots w_{j_1} K_n(x_{j_1}, \ldots, x_{j_n}) \\ &= 1 + \sum_{n=1}^{m} \frac{z^n}{n!} Q^n(K_n) \end{aligned} \tag{4.64}$$

Convergence of the Nyström-type quadratur method

Let $\{Q_m\}$ be a family of quadratur rules, $\{d_{Q_m}(z)\}$ be the corresponding Nyström-type approximations and $d(z)$ the Fredholm determinant. In [Bor09, Ch. 6, p. 21] it was proved that if

$$Q_m(f) \to \int f(x)\mathrm{d}x$$

for a continuous f then as $m \to \infty$

$$d_{Q_m}(z) \to d(z) \tag{4.65}$$

uniformly for bounded z.

Projection method of Galerkin type

The projection method is based on construction II from definition 28 and uses a sequence of m-dimensional subspaces $V_m \subset H$ with their projection

$$P_m : H \to V_m$$

and the Galerkin projection $P_m A P_m$. One can then show that as $m \to \infty$ that P_m converges to the identity operator I and so

$$\det(I + zP_m A P_m) \to \det(I + zA).$$

Comparison of Nyström method with projection method

In ([Bor09]) the two methods are compared when applied to the integral operator A that solves the *Poisson equation* with given $f \in L^2(0,1)$ such that

$$-\Delta u(x) = f(x), \qquad u(0) = u(1) = 1. \tag{4.66}$$

The error estimate is defined with resepect to the trace class norm as

$$\|P_m A P_m - A\|_1. \tag{4.67}$$

For the projection method a convergence rate of $\mathcal{O}(m^{-1})$ was achieved while for the quadrature method on has $\mathcal{O}(m^{-2})$. Thus the projection method has a worse error bound than the quadrature method.

Chapter 5

Limit Theorems for the Wishart Unitary Ensemble

In the introduction chapter we have seen the importance of the spectral distribution and the extreme eigenvalue distribution of Wishart matrices with their corresponding *asymptotic limits* for multivariate statistics. In this chapter we will study both limits for the Wishart unitary ensemble.

Notation and assumptions: In order to be consistent with the random matrix theory literature we make a small notational change and use N instead of p. The observation size will still be n. In summary we have

$$X \text{ is } n \times N \qquad \text{and} \qquad X^*X \text{ is } N \times N.$$

with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$. We define the scaled and unscaled Wishart matrix S_n and $W \sim W_N(0, I_N)$ and the corresponding empirical eigenvalue measure for S_n

$$S_n = \frac{1}{n}W, \qquad W = X^*X, \qquad \mu_{S_n} = \frac{1}{N}\sum_{k=1}^{N} \delta_{\lambda_k(S_n)}$$

Throughout the chapter we assume

$$\mathbb{E}[X_{ij}] = 0, \qquad \mathbb{E}[X_{ij}^2] = 0, \qquad \mathbb{E}|X_{ij}|^2 = 1. \tag{5.1}$$

where X_{ij} are i.i.d. complex Gaussian entries of X. We further assume that as $N, n \to \infty$ we have $N/n = \gamma \in (0, \infty)$.

The identities (3.52) and (3.58) for the joint law of eigenvalues f_N and the k-point correlation function R_k for the Wishart unitary ensemble are:

$$f_N(\lambda_1, \ldots, \lambda_N) = \frac{1}{N!} \det(K_N(\lambda_i, \lambda_j))_{1 \leq i,j \leq N}$$
$$R_k(\lambda_1, \ldots, \lambda_k) = \det(K_N(\lambda_i, \lambda_j))_{1 \leq i,j \leq k}$$

and the average spectral density f_1

$$f_1(\lambda_1) = \frac{1}{N}K_N(\lambda_1, \lambda_1), \qquad R_1(\lambda_1) = K_N(\lambda_1, \lambda_1)$$

We are interested in what happens to the kernel $K_N(x, x)$ under appropriate scaling in the limit.

5.1 The limit theorems

We briefly summarize the basic limit theorems for the Wishart unitary ensemble without proofs.

Global regime

Let μ_{S_n} be the empirical spectral distribution of S_n.

Theorem 19 (Marchenko-Pastur law for the WUE)**.** *Under the assumptions (5.1) for the scaled Wishart matrix S_n as $N, m \to \infty$ with $N/n = \gamma \in (0, \infty)$ we have*

$$\mathbb{P}\left(\mu_{S_n} \to \mu_{MP} \text{ in distribution}\right) = 1 \tag{5.2}$$

where μ_{MP} denotes the Marchenko-Pastur distribution with density

$$\hat{\rho}(x) = \left(1 - \gamma^{-1}\right)^+ \delta_0(x) + \frac{\sqrt{(B_+ - x)(x - B_-)}}{2\pi x \gamma} \mathbf{1}_{[B_-, B_+]}(x) \tag{5.3}$$

where

$$B_+ := (1 + \sqrt{\gamma})^2, \qquad B_- := (1 - \sqrt{\gamma})^2 \tag{5.4}$$

Observe that the distributions spreads over

$$B_+ - B_- = (1 + \sqrt{\gamma})^2 - (1 - \sqrt{\gamma})^2 = 4\sqrt{\gamma}.$$

Remark 33: *(The atomic part) The atomic part of the Marchenko-Pastur density vanishes if $\gamma \leq 1$. This is the case when S_n is a non-singular matrix meaning $n > N$. Assume $\gamma > 1$ then we have $N > n$ and so $\operatorname{rank}(S_n) = n$ which implies that S_n is a singular Wishart-Matrix with $N - n$ zero eigenvalues. In this case we can split-up μ_{S_n} into the form*

$$\mu_{S_n} = \frac{1}{N} \sum_{k=1}^{N} \delta_k = \frac{N - n}{N} \delta_0 + \frac{1}{N} \sum_{k=1}^{n} \delta_k.$$

Now taking the limit yields for the left term

$$\lim_{N,n \to \infty} \frac{N - n}{N} \delta_0 = \lim_{N,n \to \infty} \left(1 + \frac{n}{N}\right) \delta_0 = (1 + \gamma^{-1}) \delta_0.$$

Remark 34: *(General variance case) One can extend the theorem to a general variance assumption $\mathbb{E}[X_{ij}^2] = \sigma^2$ then the density of the Marchenko-Pastur distribution becomes*

$$\left(1 - \gamma^{-1}\right)^+ \delta_0(x) + \frac{\sqrt{(B_+ - x)(x - B_-)}}{\sigma^2 2\pi x \gamma} \mathbf{1}_{[B_-, B_+]}(x) \tag{5.5}$$

with

$$B_+ := \sigma^2 (1 + \sqrt{\gamma})^2, \qquad B_- := \sigma^2 (1 - \sqrt{\gamma})^2. \tag{5.6}$$

Remark 35: *(The hard edge) In the case that $\gamma = 1$ we have $B_+ = 4$ and $B_- = 0$ the density becomes*

$$\hat{\rho}(x) = \left(\frac{1}{2\pi} \sqrt{\frac{4}{x} - 1}\right) \mathbf{1}_{[0,4]}(x) \tag{5.7}$$

where at the lower edge $x = 0$ we have a singularity.

The limit kernels

A central object of study in the local regime are the correlation functions

$$R_k(\lambda_1, \dots, \lambda_k) = \det(K_N(\lambda_i, \lambda_j))_{1 \le i,j \le k}$$

For the Wishart ensemble $K_N(\lambda_i, \lambda_j)$ converges either to the sine, Airy or Bessel kernel.

Theorem 20 (Sine and Airy kernel)**.** *Under the assumptions (5.1). Let $\hat{\rho}(x)$ be the Marchenko-Pastur density. For a bulk point $x_0 \in (B_-, B_+)$ we have*

$$\lim_{N\to\infty} \left(\frac{1}{N\hat{\rho}(x_0)}\right) K_N\left(x_0 + \frac{x}{N\hat{\rho}(x_0)}, x_0 + \frac{y}{N\hat{\rho}(x_0)},\right) = \frac{\sin(\pi(x-y))}{\pi(x-y)}. \tag{5.8}$$

Let $\gamma = 1$, $B_+ = 4$, then at the upper edge we find convergence to the Airy kernel

$$\begin{aligned} \lim_{n\to\infty} \frac{4^{2/3}}{N^{2/3}} K_N\left(4 + \frac{4^{2/3}}{N^{2/3}}x, 4 + \frac{4^{2/3}}{N^{2/3}}y\right) \\ = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}. \end{aligned} \tag{5.9}$$

Theorem 21 (Bessel kernel) [For10, p. 290]**.** *Under the assumptions (5.1). At the lower edge we find convergence to the Airy kernel except in the case when $\gamma = 1$. Then we have a hard wall and*

$$\lim_{n\to\infty} \frac{1}{4N} K_N\left(\frac{x}{4N}, \frac{y}{4N}\right) = \frac{J_a(\sqrt{x})\sqrt{y}J_a'(\sqrt{y}) - J_a'(\sqrt{x})\sqrt{x}J_a(\sqrt{y})}{2(x-y)}. \tag{5.10}$$

where $a = n - N$ is fixed and J_a denotes the Bessel function.

Largest eigenvalue distribution

Let $\lambda_1 := \lambda_1(W)$ be the biggest eigenvalue for the unscaled complex Wishart matrix W then λ_1 converges under appropriate scaling to the Tracy-Widom distribution. We have the following limit theorem which is a consequence of theorem 20.

Theorem 22 (Tracy-Widom law for the WUE)**.** *Under the assumptions (5.1) as $N, n \to \infty$ and $N/n = \gamma \in (0, \infty)$ we have*

$$\mathbb{P}\left(\frac{\lambda_1 - \mu_{n,N}}{\sigma_{n,N}} \le s\right) = F_2(s) \tag{5.11}$$

where F_2 is the complex Tracy-Widom distribution and the centering and scaling are

$$\mu_{n,N} = \left(\sqrt{n} + \sqrt{N}\right)^2, \tag{5.12}$$

$$\sigma_{n,N} = (\sqrt{n} + \sqrt{N})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{N}}\right)^{1/3}. \tag{5.13}$$

Remark 36: *(Evaluation of the TW distribution) The Tracy-Widom distribution $F_2(s)$ can be computed explicitly either as a Fredholm determinant with Airy kernel*

$$F_2(s) = \det\left(1 - K_{\mathrm{Ai}}\right) \tag{5.14}$$

$$= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_s^{\infty} \cdots \int_s^{\infty} \det\left(\mathrm{K}_{\mathrm{Ai}}(x_i, x_j)\right)_{i,j=1}^{k} \prod_{j=1}^{k} \mathrm{d}x_j \tag{5.15}$$

where

$$K_{\mathrm{Ai}}(x, y) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x - y} \tag{5.16}$$

or by solving (see [Sos01, p. 4])

$$F_2(s) = \exp\left(-\int_s^{\infty} (t - s) q^2(t) \mathrm{d}t\right) \tag{5.17}$$

where q satisfies the Painlevé II equation

$$q''(s) = sq(s) + 2q^3(s) \tag{5.18}$$

with boundary condition

$$q(s) \sim \mathrm{Ai}(s) \quad \text{as } s \to \infty. \tag{5.19}$$

Remark 37: *(Airy process) Theorem 20 implies that also the second, third, etc biggest eigenvalues $\lambda_2(W), \lambda_3(W), \ldots$ follow the Airy point process law.*

Remark 38: *(Tracy-Widom universality) In [Sos01] theorem 22 was relaxed so that it is true for non-Gaussian $X = \{X_{ij}\}$ matrices such that $W = X^*X$ with independent entries under some conditions:*

1. $\mathbb{E}[X_{ij}] = 0,\ \mathbb{E}[X_{ij}] = 0$ *and* $\mathbb{E}|X_{ij}|^2 = 1$.
2. X_{ij} *have symmetric laws of distribution.*
3. $\mathbb{E}\left[X_{ij}^{2m}\right] \leq (c \cdot m)^m$ *for* $m = 1, 2, \ldots$ *and some constant* c.
4. $n - N = \mathcal{O}(N^{1/3})$.

5.2 The method of steepest descent

This subchapter about the method of steepest descent is based on [Won01, pp. 84-90].

5.2.1 Method of steepest descent

Consider two analytic functions $f(z)$ and $g(z)$ where for $z = x + \mathrm{i}y$

$$f(z) = u(x, y) + \mathrm{i}v(x, y)$$

and the following integral problem over a contour C

$$F(\lambda) = \int_C g(z)e^{-\lambda f(z)}\mathrm{d}z. \tag{5.20}$$

By Cauchy's theorem we can change the contour of the integral in such a way that we don't pass any singularity and that $v(x, y) = V_0$ is a constant. Then

$$\begin{aligned}\int_C g(z)e^{-\lambda f(z)}\mathrm{d}z &= \int_{\widetilde{C}} g(z)e^{-\lambda \mathrm{i}\mathfrak{Im}(f(z))-\lambda\mathfrak{Re}(f(z))}\mathrm{d}z \\ &= e^{-\mathrm{i}\lambda V_0}\int_{\widetilde{C}} g(z)e^{-\lambda\mathfrak{Re}(f(z))}\mathrm{d}z\end{aligned}$$

Sometimes the path $\widetilde{C}$ must be further divided into subpaths $\widetilde{C}_1, \ldots, \widetilde{C}_n$ where each path has a corresponding $V_0^{(1)}, \ldots, V_0^{(n)}$ leading to a sum of integrals

$$\sum_i e^{-\mathrm{i}\lambda V_0^i}\int_{\widetilde{C}_i} g(z)e^{-\lambda\mathfrak{Re}(f(z))}\mathrm{d}z.$$

The main strategy is to find a new integration contour $\widetilde{C}$ such that

- $\widetilde{C}$ passes through one or more saddle points z_0.
- the imaginary part $v(x, y)$ is constant.
- $\widetilde{C}$ avoids any singularity.

It turns out that to achieve $v(x, y) = V_0$ for some constant V_0 we have to integrate along the path of *steepest descent*.

Saddle points

Suppose $z_0 = x_0 + \mathrm{i}y_0$ is a point such that $f'(z_0) = 0$ then by the Cauchy-Riemann equations

$$f'(z_0) = u_x(x_0, y_0) + \mathrm{i}v_x(x_0, y_0) = u_x(x_0, y_0) - \mathrm{i}u_y(x_0, y_0) = 0$$

implying

$$u_x(x_0, y_0) = u_y(x_0, y_0) = 0 \quad \Longrightarrow \quad (x_0, y_0) \text{ is a critical point of } u(x, y).$$

Further since $u(x, y)$ is a harmonic function by the *maximum modulus principle* the maximum/minimum can not be on the interior and therefore (x_0, y_0) is a saddle point of $u(x, y)$. In short:

$$f'(z_0) = 0 \implies z_0 \text{ is a saddle point of } u(x, y).$$

Geometric reasoning for steepest descent paths

Consider the set of parametric equations of a differentiable curve $\gamma(s) = (x(s), y(s))$ and for simplicity let s be the *arc length* of γ. Then

$$x = x(s), \qquad y = y(s), \qquad u = u(s) = u(x(s), y(s)) \tag{5.21}$$

forms a set of parametric equations on a surface S in the space with x-axis, y-axis and u-axis. By the arc length formula

$$\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2 = 1 \tag{5.22}$$

and we define the parameter θ by

$$\cos(\theta) = \frac{\mathrm{d}x}{\mathrm{d}s}, \qquad \sin(\theta) = \frac{\mathrm{d}y}{\mathrm{d}s}. \tag{5.23}$$

The change of $u(s)$ along γ is

$$\frac{\mathrm{d}u}{\mathrm{d}s} = \frac{\mathrm{d}u}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}s} + \frac{\mathrm{d}u}{\mathrm{d}y}\frac{\mathrm{d}y}{\mathrm{d}s} = u_x\cos(\theta) + u_y\sin(\theta). \tag{5.24}$$

To measure the steepness of the curve on S we need the angle α formed by the u-axis and the curve. The formula for the angle of two lines yields

$$\cos(\alpha) = \frac{\frac{\mathrm{d}u}{\mathrm{d}s}}{\sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}s}\right)^2 + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2}} = \frac{\frac{\mathrm{d}u}{\mathrm{d}s}}{\sqrt{1 + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2}}$$

For a point on S we consider all differentiable curves that pass through that point. We are interested in the extrema of $\cos(\alpha)$ and differentiate with respect to θ. By using

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\mathrm{d}u}{\mathrm{d}s} = -u_x\sin(\theta) + u_y\cos(\theta)$$

and the fact that for any function $f(x)$

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{f(x)}{\sqrt{1+f(x)^2}} = \frac{f'(x)}{(1+f(x)^2)^{3/2}}$$

we get

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\cos(\alpha) = \frac{1}{\left[1 + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2\right]^{3/2}}(-u_x\sin(\theta) + u_y\cos(\theta)).$$

The stationary points are given by

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\cos(\alpha) = 0 \implies (-u_x\sin(\theta) + u_y\cos(\theta)) = 0$$

and by the Cauchy-Riemann equations this is equivalent to

$$\frac{\mathrm{d}v}{\mathrm{d}s} = v_x \cos(\theta) + v_y \sin(\theta) = 0$$

so these are the points where $v(x, y) = V_0$ is constant. If we compute the second derivative at a stationary point

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\cos(\alpha) = -\frac{1}{\left[1 + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2\right]^{3/2}}\frac{\mathrm{d}u}{\mathrm{d}s}$$

we see that

$$\begin{aligned}
&\frac{\mathrm{d}v}{\mathrm{d}s} = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}s} > 0 \implies \text{minimum for } \cos(\alpha)\\
&\frac{\mathrm{d}v}{\mathrm{d}s} = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}s} < 0 \implies \text{maximum for } \cos(\alpha)\\
&\frac{\mathrm{d}u}{\mathrm{d}s} = 0 \implies \text{critical point or saddle point.}
\end{aligned}$$

This means if we change the point on the surface along a curve where $v(x, y)$ is constant, then the tangent vector of the curve γ will have a maximum or minimum inclination with the u-axis as long as we don't pass through a singularity or points where $\frac{\mathrm{d}u}{\mathrm{d}s} = 0$. Therefore when $\frac{\mathrm{d}v}{\mathrm{d}s} = 0$ we have either curves of steepest ascent or steepest descents on S.

Determining the steepest descent paths

Suppose z_0 is a saddle point such that for $m \geq 2$

$$f'(z_0) = f''(z_0) = \cdots = f^{m-1}(z_0) = 0, \quad \text{but} \quad f^m(z_0) = ae^{\mathrm{i}\phi}$$

where $a > 0$ with real argument ϕ, then we call z_0 a saddle point of *order* $m - 1$. Suppose

$$z = z_0 + re^{\mathrm{i}\theta}, \qquad z - z_0 = re^{\mathrm{i}\theta}$$

then when we do a Taylor expansion of the phase $f(z)$ at z_0 we get

$$\begin{aligned}
f(z) &= f(z_0) + \frac{f^m(z_0)}{m!}(z - z_0)^m + \mathcal{O}(z^{m+1})\\
&= f(z_0) + ae^{\mathrm{i}(m\theta+\phi)}\frac{r^m}{m!} + \mathcal{O}(z^{m+1})
\end{aligned}$$

and for the real and imaginary part of $f(z)$

$$\begin{aligned}
\mathfrak{Re}(f(z)) = u(x, y) &= u(x_0, y_0) + \frac{r^m}{m!}a\cos(m\theta + \phi) + \cdots\\
\mathfrak{Im}(f(z)) = v(x, y) &= v(x_0, y_0) + \frac{r^m}{m!}a\sin(m\theta + \phi) + \cdots -
\end{aligned}$$

To get $v(x, y)$ constant we need to have $\sin(m\theta + \phi) = 0$ which happens when

$$\theta = -\frac{\phi}{m} + \frac{k\pi}{m}, \qquad k = 0, 1, \ldots, 2m - 1. \tag{5.25}$$

with constant $V_0 := v(x_0, y_0)$. Analogously if we want to have the level curves we need $u(x, y)$ to be constant and so $\cos(m\theta + \phi) = 0$ which happens at

$$\theta = -\frac{\phi}{m} + \left(k + \frac{1}{2}\right)\frac{\pi}{m}, \qquad k = 0, 1, \ldots, 2m-1. \tag{5.26}$$

This gives us a total of $2m$ steepest directions

- m direction of steepest ascent up the *hills*.
- m direction of steepest descent down into the *valleys*.
- The transition between the hills and valleys are the $2m$ level curves when $u(x, y)$ is constant.

The direction of steepest descents are chosen because this guarantees us that the integral (5.20) doesn't diverge when we make the transformation

$$g(z)e^{-\lambda f(z)} = g(z)e^{-\lambda f(z_0)-\lambda\tau^2}$$

because $\tau \in \mathbb{R}$ will be a local coordinate. Finally we are left with the problem to find an asymptotic expansion for

$$e^{-\lambda f(z_0)} \int_{\mathbb{R}} g(z)e^{-\lambda\tau^2}\frac{\mathrm{d}z}{\mathrm{d}\tau}\mathrm{d}\tau. \tag{5.27}$$

From here one proceeds using classical asymptotic methods such as *Watson's lemma* or *Laplace's method*.

5.3 The saddle points of the Laguerre polynomial

Consider the contour integral for a closed $\gamma \in D$ circle around 0 in a counterclockwise direction without enclosing the essential singularity at $t = 1$

$$L_m^{(a)}(z) = \frac{1}{2\pi i}\oint_C \frac{e^{-zt/(1-t)}}{(1-t)^{\alpha+1}\,t^{m+1}}\,dt$$

where $a = n - N$. We are interested in the asymptotics of the form $N \to \infty$ on the regime

$$\begin{cases} m = N - k & \text{for} \quad k \in \mathbb{N}_0 \quad \text{fixed.} \\ z = Nx & \text{with} \quad x > 0. \\ a = N\alpha \end{cases}$$

Rewrite the integrand

$$\begin{aligned} \partial_z^l L_{N-k}^{(N\alpha)}(Nx) &= \frac{1}{2\pi i}\oint_\gamma e^{-Nx\frac{t}{1-t}+N\alpha\log(1-t)+N\log(t)}\frac{t^{l+k}}{(1-t)^{l+1}t^1}\mathrm{dt} \\ &= \frac{1}{2\pi i}\oint_\gamma e^{-N\phi(t)}\frac{f_{k,l}(t)}{(1-t)t}\mathrm{dt}. \end{aligned} \tag{5.28}$$

we get the following phase ϕ and f:

$$\begin{cases} \phi(t,x) & = x\frac{t}{1-t} + \alpha\log(1-t) + \log(t) \\ f_{k,l}(t) & = \frac{t^{l+k}}{(1-t)^l} \\ F_{k,l}(t) & = \frac{t^{l+k-1}}{(1-t)^{l+1}}. \end{cases}$$

By $\curvearrowleft$ we define the upper half curve where the contours circles the origin in counterclockwise direction but not enclosing the essential singularity point $t = 1$ till it reaches the x-axis. Both $\phi, f_{k,l}$ are real-analytic functions so by symmetry

$$\partial_z^l L_{N-k}^{(N\alpha)}(Nx) = \frac{1}{\pi}\mathfrak{Im}\left\{ \int_{\curvearrowleft} e^{-N\phi(t)} \frac{f_{k,l}(t)}{(1-t)t}\mathrm{dt} \right\}.$$

Derivatives

We first compute the derivatives

$$\begin{aligned} \phi_t(t,x;u,\alpha) &= \frac{x}{(1-t)^2} - \frac{\alpha}{1-t} + \frac{1}{t} = \frac{tx - \alpha t(1-t) + (1-t)^2}{(1-t)^2 t} \\ \phi_x(t,x;u,\alpha) &= \frac{t}{(1-t)} \\ \phi_{tt}(t;u,\alpha) &= \frac{2x}{(1-t)^3} - \frac{\alpha}{(1-t)^2} - \frac{1}{t^2} = \frac{2xt^2 - \alpha t^2(1-t) - (1-t)^3}{(1-t)^2 t^2} \\ \phi_{tx}(t,x;u,\alpha) &= \phi_{xt}(t,x;u,\alpha) = \frac{1}{(1-t)^2} \\ \phi_{xx}(t,x;u,\alpha) &= 0 \\ \phi_{ttt}(t;u,\alpha) &= \frac{6x}{(1-t)^4} - \frac{2\alpha}{(1-t)^3} + \frac{2}{t^3} \\ \phi_{txt}(t,x;u,\alpha) &= \phi_{xtt}(t,x;u,\alpha) = \phi_{ttx}(t,x;u,\alpha) = \frac{2}{(1-t)^3} \end{aligned}$$

we perform the change of variable

$$x = \alpha + 2 + 2\sqrt{\alpha+1}u = (\sqrt{\alpha+1} + u)^2 - u^2 + 1, \qquad u \in [-1,1] \tag{5.29}$$

then for $\phi_t(t,x) = 0$

$$\iff (\alpha + 2 + 2\sqrt{\alpha+1}u)t - \alpha t(1-t) + (1-t)^2 = 2\sqrt{\alpha+1}tu + t^2(\alpha+1) + 1$$

$$\iff (t\sqrt{1+\alpha} + u)^2 = u^2 - 1.$$

5.3.1 Saddle points

For $u \in [-1,1]$ the saddle points are

$$t = -\frac{u \pm \mathrm{i}\sqrt{1-u^2}}{\sqrt{1+\alpha}} \tag{5.30}$$

This means there are two types of saddle points depending on the $\pm$ and u:

- two complex conjugate saddle points for $u \in (-1,1)$ (bulk)
- one real saddle point for $u \in \{-1,1\}$ (edge/boundary)

Edge region

At the lower boundary called soft/hard edge for $u=-1$:

$$t_{-1} = \frac{1}{\sqrt{1+\alpha}} \tag{5.31}$$

At the upper boundary called soft edge for $u=1$:

$$t_1 = -\frac{1}{\sqrt{1+\alpha}} \tag{5.32}$$

Bulk region

In the bulk region for $u \in (-1,1)$ we make a change of variable

$$u = \cos(\theta), \qquad \theta = \arccos(u), \qquad \theta \in (0,\pi)$$

then

$$t = -\frac{u \pm \mathrm{i}\sqrt{1-u^2}}{\sqrt{1+\alpha}} = -\frac{\cos(\theta) \pm \mathrm{i}\sin(\theta)}{\sqrt{1+\alpha}} = \frac{-e^{\pm \mathrm{i}\theta}}{\sqrt{1+\alpha}}.$$

Since $-e^{-\mathrm{i}\theta}$ runs counterclockwise and $[-e^{-\mathrm{i}\cdot 0}, -e^{-\mathrm{i}\cdot\pi}] = [-1,1]$ the root in the upper plane is

$$t_u = \frac{-e^{-\mathrm{i}\theta(u)}}{\sqrt{1+\alpha}}, \qquad \theta \in (0,\pi). \tag{5.33}$$

5.4 The Marchenko-Pastur law

We apply the steepest descent method in the bulk region. Our goal is to make a Taylor expansion of the phase at (5.33).

Lemma 10 .

$$\partial_{tt}\phi(t_u) = 2\mathrm{i}\sqrt{\alpha+1}\frac{\sqrt{1-u^2}}{(1-t_u)^2 t_u}. \tag{5.34}$$

and its absolute value is

$$|\partial_{tt}\phi(t_u)| = 2(\alpha+1)^2\frac{\sqrt{1-u^2}}{x}. \tag{5.35}$$

Proof. We first compute (5.34). We have

$$\begin{aligned}\partial_{tt}\phi(t;u,\alpha)\Big|_{t=t_u} &= \frac{\partial_t(xt-\alpha t(1-t)+(1-t)^2)}{(1-t)^2 t}\Big|_{t=t_u} \\ &\quad + \left(xt-\alpha t(1-t)+(1-t)^2\right)\partial_t\frac{1}{(1-t)^2 t}\Big|_{t=t_u} \\ &= 2\frac{\sqrt{\alpha+1}u+(\alpha+1)t_u}{(1-t_u)^2 t_u}.\end{aligned}$$

Using that

$$\sqrt{\alpha+1}\, t_u + u = \mathrm{i}\sqrt{1-u^2} \tag{5.36}$$

we multiply (5.36) by $\sqrt{\alpha+1}$ so

$$(\alpha+1)t_u + \sqrt{\alpha+1}u = \mathrm{i}\sqrt{\alpha+1}\sqrt{1-u^2}$$

and therefore

$$\partial_{tt}\phi(t_u) = 2\mathrm{i}\sqrt{\alpha+1}\frac{\sqrt{1-u^2}}{(1-t_u)^2 t_u}. \tag{5.37}$$

Now we compute the second part (5.35). By using (5.29) we compute

$$\begin{aligned}|1-t_u|^2 = (1-t_u-\overline{t_u}+\overline{t_u}t_u) &= \frac{\alpha+1}{\alpha+1} + \frac{2u\sqrt{\alpha+1}}{\alpha+1} + \frac{1}{\alpha+1} \\ &= \frac{(\sqrt{1+\alpha}+u)^2+1-u^2}{\alpha+1} \\ &= \frac{\alpha+2+2\sqrt{1+\alpha}u}{\alpha+1}\end{aligned}$$

hence

$$|1-t_u|^2 = \frac{x}{\alpha+1} \tag{5.38}$$

and since

$$|t_u| = (\sqrt{1+\alpha})^{-1}$$

we get

$$|\partial_{tt}\phi(t_u)| = 2(\alpha+1)^2\frac{\sqrt{1-u^2}}{x}. \tag{5.39}$$

□

Since $u \in (-1,1)$ by lemma 10 we have $\phi''(t_u) \neq 0$ and therefore the critical point t_u is not degenerated. Now we make a Taylor expansion of the phase at t_u

$$N\phi\left(t_u + \frac{se^{\mathrm{i}\sigma}}{\sqrt{N}}\right) = N\phi(t_u) + \frac{1}{2}\phi''(t_u)s^2e^{\mathrm{i}2\sigma} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{5.40}$$

We choose the direction

$$\sigma = \frac{\pi}{4} - \frac{\theta}{2} \tag{5.41}$$

and make a change of variable

$$t = t_u + \frac{(1-t_u)e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}}}{\sqrt{2N(1+\alpha)}}\tau$$

and the Taylor expansion becomes

$$N\phi\left(t_u + \frac{(1-t_u)e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}}}{\sqrt{2N(1+\alpha)}}\tau\right) = \phi(t_u) + \frac{\tau^2}{2}\sqrt{1-u^2} + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{5.42}$$

5.4.1 The Laguerre asymptotics in the bulk

By the change of variable $x = \alpha + 2 + 2\sqrt{\alpha+1}u$

$$\begin{aligned}
&\partial_z^l L_{n-k}^{(\alpha)}\left(N\left(\alpha+2+2\sqrt{\alpha+1}u\right)\right)\\
&= \frac{1}{\pi}\mathfrak{Im}\left\{e^{-N\phi(t_u)}\int_{\mathbb{R}} e^{-\frac{\tau^2}{2}\sqrt{1-u^2}}\left(\frac{f_{k,l}(t_u)}{t_u}+\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)\right)\frac{e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}}}{\sqrt{2N(1+\alpha)}}\mathrm{d}\tau\right\}\\
&= \mathfrak{Im}\left\{e^{-N\phi(t_u)}\frac{e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}}}{\sqrt{2N(1+\alpha)}}\frac{1}{\pi}\int_{\mathbb{R}} e^{-\frac{\tau^2}{2}\sqrt{1-u^2}}\left(\frac{f_{k,l}(t_u)}{t_u}+\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)\right)\mathrm{d}\tau\right\}.
\end{aligned}$$

By the change of variable $s = \frac{\tau}{\sqrt{2}}(1-u^2)^{1/4}$ we get a Gaussian integral yielding

$$\int_{\mathbb{R}} e^{-\frac{\tau^2}{2}\sqrt{1-u^2}}\mathrm{d}\tau = \sqrt{\frac{2\pi}{(1-u^2)^{1/2}}}$$

and ignoring the error terms we obtain

$$\begin{aligned}
&\partial_z^l L_{n-k}^{(\alpha)}\left(N(\alpha+2+2\sqrt{\alpha+1}u)\right)\\
&\qquad\simeq \frac{1}{(1-u^2)^{1/4}}\mathfrak{Im}\left\{e^{-N\phi(t_u)}\frac{e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}}}{t_u\sqrt{\pi N(1+\alpha)}}f_{k,l}(t_u)\right\} \qquad (5.43)\\
&\qquad= \frac{-1}{(1-u^2)^{1/4}\sqrt{\pi N}}\mathfrak{Im}\left\{e^{-N\phi(t_u)}e^{-\mathrm{i}\frac{\theta}{2}+\mathrm{i}\frac{\pi}{4}+\mathrm{i}\theta}f_{k,l}(t_u)\right\}.
\end{aligned}$$

Define the function

$$\psi(u) := \phi(t_u; u, \alpha) \qquad (5.44)$$

and split it into the real and imaginary part in (5.43) and we obtain

$$\begin{aligned}
&\partial_z^l L_{n-k}^{(\alpha)}(N(\alpha+2+2\sqrt{\alpha+1}u))\\
&\qquad\simeq \frac{-e^{-N\mathfrak{Re}\psi(u)}}{(1-u^2)^{1/4}\sqrt{\pi N}}\mathfrak{Im}\left\{e^{-\mathrm{i}\left(N\mathfrak{Im}\psi(u)-\frac{\theta}{2}-\frac{\pi}{4}\right)}f_{k,l}(t_u)\right\}. \qquad (5.45)
\end{aligned}$$

The asymptotics with the weights

Lemma 11 . *Consider the constant*

$$c_\alpha := \psi(-1) + \frac{x(-1)-\alpha\log(x(-1))}{2} \qquad (5.46)$$

then

$$c_\alpha = \frac{\alpha-(\alpha+1)\log(1+\alpha)}{2}. \qquad (5.47)$$

Proof. For $u=-1$ and $x = x(-1)$ we have

$$x(-1) = (\sqrt{\alpha+1}-1)^2 - 1 + 1 = (\sqrt{\alpha+1}-1)^2, \qquad \sqrt{x(-1)} = (\sqrt{\alpha+1}-1)$$
$$t_u = \frac{1}{\sqrt{1+\alpha}}$$

and

$$\begin{aligned}\psi(-1) &= \frac{x}{\sqrt{\alpha+1}-1} + \alpha\log\left(\sqrt{1+\alpha}-1\right) - (\alpha+1)\log\left(\sqrt{1+\alpha}\right)\\ &= \frac{x}{\sqrt{x}} + \alpha\log\left(\sqrt{x}\right) - (\alpha+1)\log\left(\sqrt{1+\alpha}\right)\\ \frac{x(-1)}{2} &= \frac{\alpha-2\sqrt{x}}{2} = -\sqrt{x}+\frac{\alpha}{2}\end{aligned}$$

and since $\alpha\log(\sqrt{x}) = \frac{1}{2}\alpha\log(x)$

$$\begin{aligned}c_\alpha &= \sqrt{x} + \alpha\log(\sqrt{x}) - (\alpha+1)\log\left(\sqrt{(1+\alpha)}\right) - \sqrt{x} + \frac{\alpha}{2} + \frac{\alpha\log(x)}{2}\\ &= \frac{\alpha-(\alpha+1)\log(1+\alpha)}{2}.\end{aligned}$$

□

Lemma 12 . *For $u \in (-1,1)$ we have*

$$\psi(u) = -\frac{x(u)-\alpha\log(x(u))}{2} + c_\alpha + \mathrm{i}\pi\int_{-1}^{u}\rho(s)\mathrm{d}s \tag{5.48}$$

where c_α is as lemma 11 and $\rho(s)$ is the following probability density

$$\rho(s) = \frac{2(\alpha+1)}{\pi}\frac{\sqrt{1-s^2}}{x(s)}. \tag{5.49}$$

Proof. The proof idea is we take the derivative of $\psi(u) = \phi(t_u; u, \alpha)$ and then integrate it from $[-1, u]$.

$$\partial_u\psi(u) = \underbrace{\partial_u t_u\ \partial_t\phi\Big|_{t=t_u}}_{=0} + \partial_u\phi\Big|_{t=t_u} = 2\sqrt{\alpha+1}\frac{t_u}{1-t_u} \tag{5.50}$$

and

$$2\sqrt{\alpha+1}\frac{t_u}{1-t_u} = \frac{2\sqrt{1+\alpha}}{|1-t_u|^2}t_u(1-\overline{t_u}) \tag{5.51}$$

where we used that $|1-t_u|^2 = (1-t_u)(1-\overline{t_u})$. By (5.38) we get

$$\begin{aligned}\partial_u\psi(u) &= \frac{-2\sqrt{\alpha+1}}{x}e^{-\mathrm{i}\theta}\left(\sqrt{1+\alpha}+e^{\mathrm{i}\theta}\right)\\ &= \frac{-2\sqrt{\alpha+1}}{x}(u-\mathrm{i}\sqrt{1-u^2})\left(\sqrt{1+\alpha}+(u+\mathrm{i}\sqrt{1-u^2})\right)\\ &= \frac{-2\sqrt{\alpha+1}}{x}\left(\sqrt{1+\alpha}u+1-\mathrm{i}\sqrt{1+\alpha}\sqrt{1-u^2}\right)\\ &= \underbrace{-\sqrt{\alpha+1}\left(1-\frac{\alpha}{x(u)}\right)}_{=:A} + \mathrm{i}\underbrace{\frac{2(\alpha+1)}{x(u)}\sqrt{1-u^2}}_{=\rho(u)\pi}\end{aligned}$$

where we used $x(u) = \alpha + 2 + 2\sqrt{1+\alpha}u$ and $u = \cos(\theta)$. Then

$$\frac{1}{2}\partial_u\Big(x(u) - \alpha\log\left(x(u)\right)\Big) = \underbrace{\sqrt{\alpha+1}\left(1 - \frac{\alpha}{x(u)}\right)}_{=A}.$$

Choosing $u = -1$ implies that $\theta = \arccos(-1) = \pi$ and $t_u = (1+\alpha)^{-1/2}$ therefore

$$\begin{aligned}\psi(-1) &= \frac{x(-1)}{\sqrt{\alpha+1}-1} + \alpha\log\left(1 - \frac{1}{\sqrt{1+\alpha}}\right) - \frac{1}{2}\log(1+\alpha)\\ &= \frac{x(-1)}{\sqrt{\alpha+1}-1} + \alpha\log\left(\sqrt{1+\alpha}-1\right) - \alpha\log\left(\sqrt{1+\alpha}\right) - \log(\sqrt{1+\alpha})\\ &= \frac{x(-1)}{\sqrt{\alpha+1}-1} + \alpha\log\left(\sqrt{1+\alpha}-1\right) - (\alpha+1)\log\left(\sqrt{1+\alpha}\right) \in \mathbb{R}\end{aligned}$$

$$\begin{aligned}\mathfrak{Re}(\psi(u)) &= -\frac{x(u) - \alpha\log(x(u))}{2} + \psi(-1) + \frac{x(-1) - (\alpha+1)\log(x(1))}{2}\\ &= -\frac{x(u) - \alpha\log(x(u))}{2} + c_\alpha\end{aligned}$$

we conclude

$$\begin{aligned}\mathfrak{Re}(\psi(u)) &= -\frac{x(u) - \alpha\log(x(u))}{2} + c_\alpha\\ \mathfrak{Im}(\psi(u)) &= 2(\alpha+1)\int_{-1}^{u}\frac{\sqrt{1-s^2}}{x(s)}\mathrm{d}s = \pi\int_{-1}^{u}\rho(s)\mathrm{d}s.\end{aligned}$$

□

By applying lemma 12 and lemma 11 to (5.45) we get the following proposition.

Proposition 3 .

$$\begin{aligned}&\partial_z^l L_{n-k}^{(N\alpha)}(Nx)e^{-N\frac{x-\alpha\log(x)}{2}}\\ &\qquad= \frac{-e^{-Nc_\alpha}}{\sqrt{\pi N}(1-u^2)^{1/4}}\mathfrak{Im}\left\{e^{-\mathrm{i}\left(N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s - \frac{\theta}{2} - \frac{\pi}{4}\right)}f_{k,l}\left(\frac{-e^{-\mathrm{i}\theta}}{\sqrt{1+\alpha}}\right)\right\}\end{aligned} \tag{5.52}$$

where $f_{k,l}$ is

$$f_{k,l}(t) = \frac{t^{k+l}}{(1-t)^l} \tag{5.53}$$

and c_α as in lemma 11 and $\rho(s)$, u, $\theta(u)$ as in lemma 12.

Asymptotics of the Laguerre functions

We normalize the Laguerre polynomials

$$\begin{aligned}\int_0^\infty L_m^{(a)}(z)^2 z^a e^{-z}\mathrm{d}z &= \frac{\Gamma(m+a+1)}{m!}\\ \implies \int_0^\infty L_{n-k}^{(N\alpha)}(Nx)^2 x^{N\alpha}e^{-Nx}\mathrm{d}x &= \frac{\Gamma(N(\alpha+1)+1-k)}{(N-k)!N^{N\alpha+1}}\end{aligned} \tag{5.54}$$

Define the Laguerre Functions with weights

$$\begin{aligned}\phi_{N-k}^{(N,\alpha)}(x) :&= \left(\frac{\Gamma(N(1+\alpha)+1-k)}{(N-k)!N^{N\alpha+1}}\right)^{-1/2} L_{N-k}^{(N\alpha)}(Nx)\sqrt{x^{N\alpha}e^{-Nx}} \\ &= \kappa_{N-k}L_{N-k}^{(N\alpha)}(Nx)\sqrt{w_{N\alpha}(Nx)}\end{aligned} \tag{5.55}$$

then we compute

$$c_N := \frac{\kappa_{N-1}}{\kappa_N} = \left(\frac{\Gamma(N(1+\alpha)+1)}{N!}\right)^{1/2}\left(\frac{(N-1)!}{\Gamma(N(1+\alpha))}\right)^{1/2} = \sqrt{1+\alpha}. \tag{5.56}$$

Theorem 23 . *For the Laguerre functions (5.55) we have the following asymptotics*

$$\begin{aligned}&\partial_x^l\phi_{N-k}^{(N,\alpha)}(x) \\ &\quad\simeq \frac{N^l}{(\pi^2(1-u^2))^{1/4}}\frac{1}{(1+\alpha)^{\frac{1}{4}-k/2}}\mathfrak{Im}\left\{e^{+\mathrm{i}\left(N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s-\frac{\theta}{2}-\frac{\pi}{4}\right)}\overline{f_{k,l}\left(\frac{-e^{-\mathrm{i}\theta}}{\sqrt{1+\alpha}}\right)}\right\}.\end{aligned} \tag{5.57}$$

Proof. Apply Stirling's formula on the normalization coefficient

$$\begin{aligned}&\frac{\Gamma(N(1+\alpha)+1-k)}{(N-k)!N^{N\alpha+1}} \\ &\approx \sqrt{\frac{(N-k+N\alpha)}{(N-k-1)}}\left(\frac{1}{e}\right)^{N\alpha}\left(\frac{(N(1+\alpha)-k)^{N-k+N\alpha}}{(N-k)^{N-k}N^{N\alpha+1}}\right) \\ &= \frac{1}{N}\sqrt{\frac{(1-\frac{N}{k}+\alpha)}{1-\frac{k}{N}-\frac{1}{N}}}e^{-N\alpha}\left(\frac{N(1+\alpha)-k}{(N-k)}\right)^{N-k}\left(\frac{N(1+\alpha)-k}{N}\right)^{N\alpha}.\end{aligned}$$

Notice that

$$\frac{(N(1+\alpha)-k)}{(N-k)} = \frac{N\alpha}{N-k}+1 = \frac{N\alpha-\alpha k+\alpha k}{N-k}+1 = \alpha+\frac{\alpha k}{N-k}+1$$

so the approximation becomes

$$\frac{\Gamma(N(1+\alpha)+1-k)}{(N-k)!N^{N\alpha+1}} \simeq \frac{\sqrt{1+\alpha}e^{-\alpha N}}{N}\left(1+\alpha-\frac{k}{N}\right)^{N\alpha}\left(1+\alpha+\frac{\alpha k}{N-k}\right)^{N-k} \tag{5.58}$$

factor out $(1+\alpha)$ so the right-hand side of (5.58) becomes

$$\begin{aligned}&\frac{\sqrt{1+\alpha}e^{-\alpha N}}{N}(1+\alpha)^{N\alpha+N-k}\underbrace{\left(1-\frac{k}{(1+\alpha)N}\right)^{N\alpha}}_{\to e^{-\frac{k\alpha}{1+\alpha}}}\underbrace{\left(1+\frac{\alpha k}{(1+\alpha)(N-k)}\right)^{N-k}}_{\to e^{\frac{k\alpha}{1+\alpha}}} \\ &\simeq \frac{\sqrt{1+\alpha}e^{-\alpha N}}{N}e^{\log(1+\alpha)(N(\alpha+1)-k)} \\ &\simeq e^{-\alpha N+\log(1+\alpha)(N(\alpha+1))}\frac{N^{-1}}{(1+\alpha)^{k-1/2}}\end{aligned}$$

hence

$$\left(\frac{(N-k)!N^{N\alpha+1}}{\Gamma(N(1+\alpha)+1-k)}\right)^{1/2} \simeq \left(\frac{N}{(1+\alpha)^{1/2-k}}\right)^{1/2} e^{N\frac{(\alpha-(\alpha+1)\log(1+\alpha)}{2}}. \tag{5.59}$$

By lemma 11

$$\left(\frac{(N-k)!N^{N\alpha+1}}{\Gamma(N(1+\alpha)+1-k)}\right)^{1/2} \simeq \left(\frac{N}{(1+\alpha)^{1/2-k}}\right)^{1/2} e^{Nc_\alpha} \tag{5.60}$$

and using (5.60) on the Laguerre functions with weights (5.55) we get

$$\phi_{N-k}^{(N,\alpha)}(x) \simeq \left(\frac{N}{(1+\alpha)^{1/2-k}}\right)^{1/2} e^{Nc_\alpha} L_{N-k}^{(N\alpha)}(Nx) e^{-N\frac{(x+\alpha\log(x))}{2}}. \tag{5.61}$$

Applying proposition 3 to (5.61) we get the following asymptotics for the Laguerre functions with weights

$$\begin{aligned}&\partial_x^l \phi_{N-k}^{(N,\alpha)}(x)\\ &\quad\simeq \frac{N^l}{\left(\pi^2(1-u^2)\right)^{1/4}} \frac{1}{(1+\alpha)^{\frac{1}{4}-k/2}} \mathfrak{Im}\left\{ e^{+\mathrm{i}\left(N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s-\frac{\theta}{2}-\frac{\pi}{4}\right)} \overline{f_{k,l}\left(\frac{-e^{-\mathrm{i}\theta}}{\sqrt{1+\alpha}}\right)} \right\}.\end{aligned} \tag{5.62}$$

□

5.4.2 The Marchenko-Pastur law

By theorem 12 the kernel is

$$K_N(x,y) = c_N \frac{\phi_N(x)\phi_{N-1}(y) - \phi_N(y)\phi_N(x)}{x-y}$$

where c_N is defined in (5.56). Taking the limit by Bernoulli-L'Hôpital rule

$$K_N(x,x) = c_N \left(\phi_N'(x)\phi_{N-1}(x) - \phi_{N-1}'(x)\phi_N(x)\right).$$

We compute the asymptotics.

Lemma 13 . *Define*

$$\Gamma_N(u) = N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s - \frac{\theta(u)}{2} - \frac{\pi}{4} \tag{5.63}$$

where $\rho(s)$, u and $\theta(u)$ as in lemma 12. Then

$$\phi_{N-k}(x) \simeq \frac{(-1)^k}{\sqrt{\pi}} \frac{(1+\alpha)^{-1/4}}{(1-u^2)^{1/4}} \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N+\mathrm{i}k\theta}\right) \tag{5.64}$$

and

$$\phi_{N-k}'(x) \simeq \frac{(-1)^k N}{2\sqrt{\pi}} \frac{(1+\alpha)^{-3/4}}{(1-u^2)^{1/4}} \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N+\mathrm{i}k\theta}\overline{\psi}'\right) \tag{5.65}$$

where $\psi(u)$ is defined in (5.44).

Proof. Recall

$$f_{k,l}(t) = \frac{t^{k+l}}{(1-t)^l}$$

First we compute (5.64). Let $l = 0$ then

$$f_{k,0} = t^k \qquad \text{and} \qquad \overline{f_{k,0}(t_u)} = (-1)^k \frac{e^{\mathrm{i}k\theta}}{(1+\alpha)^{k/2}}.$$

By theorem 23

$$\phi_{N-k}(x) \simeq \frac{(1+\alpha)^{-\frac{1}{4}+k/2}}{(\pi^2(1-u^2))^{1/4}} \mathfrak{Im}\left\{ e^{+\mathrm{i}\left(N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s - \frac{\theta}{2} - \frac{\pi}{4}\right)} \overline{f_{k,l}\left(\frac{-e^{-\mathrm{i}\theta}}{\sqrt{1+\alpha}}\right)} \right\}. \tag{5.66}$$

and so

$$\phi_{N-k}(x) \simeq (-1)^k \left(\frac{(1+\alpha)^{-1}}{\pi^2(1-u^2)}\right)^{1/4} \mathfrak{Im}\left\{ e^{+\mathrm{i}\left(N\pi\int_{-1}^{u}\rho(s)\mathrm{d}s - \frac{\theta}{2} - \frac{\pi}{4}\right)} e^{\mathrm{i}k\theta} \right\}. \tag{5.67}$$

Then by (5.63)

$$\phi_{N-k}(x) \simeq \frac{(-1)^k}{\sqrt{\pi}} \left(\frac{(1+\alpha)^{-1}}{(1-u^2)}\right)^{1/4} \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N + \mathrm{i}k\theta}\right)$$

which proves (5.64) and now we compute (5.65). Let $l = 1$ then we know by (5.50)

$$\frac{\partial_u \psi(u)}{2\sqrt{\alpha+1}} = \frac{t_u}{1-t_u}$$

where $\psi(u)$ is defined in (5.44) and so

$$f_{k,1} = \left.\frac{t^{k+1}}{1-t}\right|_{t=t_u} = \frac{\partial_u \psi(u)}{2(\alpha+1)^{\frac{k+1}{2}}} e^{-\mathrm{i}k\theta}(-1)^k$$

then by theorem 23

$$\phi'_{N-k}(x) \simeq \frac{(-1)^k N}{\sqrt{\pi}} \frac{(1+\alpha)^{-3/4}}{2(1-u^2)^{1/4}} \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N + \mathrm{i}k\theta}\overline{\psi}'\right).$$

□

Theorem 24 (Marchenko-Pastur law)**.** *The scaled kernel*

$$\widetilde{K_N}(x) := \frac{K_N(x,x)}{N} \tag{5.68}$$

converges for $x \in [B_-, B_+]$ *to*

$$\widetilde{K_N}(x) \to \hat{\rho}(x)$$

where $\hat{\rho}(x)$ *is the Marchenko-Pastur density*

$$\hat{\rho}(x) = \frac{\sqrt{(B_+ - x)(x - B_-)}}{2\pi x}. \tag{5.69}$$

with the boundary points at

$$B_\pm = (\sqrt{\alpha+1} \pm 1)^2. \tag{5.70}$$

Proof. By lemma 13

$$\phi_{N-k_1}(x)\phi'_{N-k_2}(y) \simeq \frac{(-1)^{k_1+k_2}N}{2\pi}\frac{(1+\alpha)^{-1}}{\sqrt{(1-u^2)}}\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N+\mathrm{i}k_1\theta}\right)\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N+\mathrm{i}k_2\theta}\overline{\psi}'\right) \quad (5.71)$$

The N vanishes because of the scaling of the kernel. Now we do the algebra

$$\begin{aligned}\widetilde{K_N}(x) &\simeq \frac{1}{2\pi}\frac{(1+\alpha)^{-1}c_N}{\sqrt{1-u^2}}\left(\underbrace{\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N+\mathrm{i}\theta}\overline{\psi'}\right)}_{:=A}\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\right) - \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi'}\right)\underbrace{\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}e^{\mathrm{i}\theta}\right)}_{:=B}\right)\end{aligned}$$

and we split $e^{\mathrm{i}\theta}$ into its real and imaginary part

$$\begin{aligned}A &= \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi'}\right)\mathfrak{Re}\left(e^{\mathrm{i}\theta}\right) + \mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi'}\right)\underbrace{\mathfrak{Im}\left(e^{\mathrm{i}\theta}\right)}_{\sqrt{1-u^2}}\\ &= \left(u\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi'}\right) + \mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi'}\right)\sqrt{1-u^2}\right)\end{aligned}$$

$$\begin{aligned}B &= \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\right)\mathfrak{Re}\left(e^{\mathrm{i}\theta}\right) + \mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\right)\mathfrak{Im}\left(e^{\mathrm{i}\theta}\right)\\ &= \left(u\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\right) + \mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\right)\sqrt{1-u^2}\right)\end{aligned}$$

giving

$$\begin{aligned}\widetilde{K_N}(x) &\simeq \frac{1}{2\pi(1+\alpha)^{\frac{1}{2}}}\left(\mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\right)\mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi}'\right) - \mathfrak{Im}\left(e^{\mathrm{i}\Gamma_N}\overline{\psi}'\right)\mathfrak{Re}\left(e^{\mathrm{i}\Gamma_N}\right)\right)\\ &= \frac{1}{2\pi(1+\alpha)^{\frac{1}{2}}}\mathfrak{Im}\left(e^{+\mathrm{i}\Gamma_N}e^{-\mathrm{i}\Gamma_N}\psi'\right)\\ &= \frac{1}{2\pi(1+\alpha)^{\frac{1}{2}}}\mathfrak{Im}\left(\psi'\right)\end{aligned}$$

and so

$$\widetilde{K_N}(x) \simeq \frac{1}{2(1+\alpha)^{\frac{1}{2}}}\rho(u). \quad (5.72)$$

By lemma 12 we have

$$\rho(u) = \frac{2(\alpha+1)}{\pi}\frac{\sqrt{1-u^2}}{x(u)} \quad (5.73)$$

where $u = \frac{x-(\alpha+2)}{2\sqrt{1+\alpha}} \in (-1,1)$ and so

$$\widetilde{K_N}(x) \simeq \frac{\sqrt{\alpha+1}}{\pi}\frac{\sqrt{1-u^2}}{x(u)} = \frac{\sqrt{\alpha+1}}{\pi}\frac{\sqrt{1-u}\sqrt{u+1}}{x(u)}. \quad (5.74)$$

We compute

$$\sqrt{1-u} = \sqrt{\frac{2\sqrt{1+\alpha}-x+(\alpha+2)}{2\sqrt{1+\alpha}}} = \sqrt{\frac{(B_+-x)}{2\sqrt{1+\alpha}}} \quad (5.75)$$

and

$$\sqrt{1+u} = \sqrt{\frac{2\sqrt{1+\alpha}+x-(\alpha+2)}{2\sqrt{1+\alpha}}} = \sqrt{\frac{(x-B_-)}{2\sqrt{1+\alpha}}} \tag{5.76}$$

where

$$B_\pm = \alpha + 2 \pm 2\sqrt{1+\alpha} = \left(\sqrt{\alpha+1}\pm 1\right)^2 \tag{5.77}$$

so

$$\widetilde{K_N}(x) \approx \frac{\sqrt{(B_+-x)(x-B_-)}}{2\pi x} =: \hat{\rho}(x). \tag{5.78}$$

We conclude for $x \in [B_-, B_+]$ the scaled Kernel converges to the Marchenko-Pastur density. □

The sine-kernel

Theorem 25 . *Under appropriate scaling of the kernel we have as $N \to \infty$*

$$K_N\left(x_0 + \frac{\lambda}{N\hat{\rho}(x_0)}, x_0 + \frac{\mu}{N\hat{\rho}(x_0)}\right)\frac{1}{N\hat{\rho}(x_0)} \simeq \frac{\sin(\pi(\lambda-\mu))}{\pi(\lambda-\mu)}.$$

Proof. For the sine-kernel we scale at a point

$$\begin{cases} x = x_0 + \frac{\lambda}{N} \\ y = y_0 + \frac{\mu}{N} \end{cases}$$

and we evaluate (5.63) for $x_0 + \frac{\lambda}{N}$ giving approximately

$$\begin{aligned}\Gamma_N\left(x_0+\frac{\lambda}{N}\right) &\simeq \Gamma_N(x_0) + \frac{\Gamma_N'(x_0)}{N}\lambda + \mathcal{O}(N^{-1}) \\ &\simeq \Gamma_N(x_0) + \pi\hat{\rho}(x_0)\lambda + \mathcal{O}(N^{-1})\end{aligned}$$

where the last line follows from

$$\frac{\mathrm{d}}{\mathrm{d}x}\Gamma_N(x) = \frac{\mathrm{d}u}{\mathrm{d}x}\Gamma_N'(u) \simeq \frac{N}{2\sqrt{1+\alpha}}\rho(u) = N\pi\hat{\rho}(x)$$

and by (5.72) from the previous proof we know

$$\hat{\rho}(x) = \frac{\rho(u)}{2\sqrt{1+\alpha}}.$$

Then

$$\begin{aligned}K_N(x,y) &\simeq \left(\frac{\mathfrak{Im}(e^{\mathrm{i}\Gamma_N(x)})\mathfrak{Im}(e^{\mathrm{i}\Gamma_N(y)}e^{\mathrm{i}\theta(x_0)}) - \mathfrak{Im}(e^{\mathrm{i}\Gamma_N(y)})\mathfrak{Im}(e^{\mathrm{i}\theta(x_0)}e^{\mathrm{i}\Gamma_N(x)})}{\pi(y-x)\sqrt{\sin(\theta(x_0))}\sqrt{\sin(\theta(x_0))}}\right) \\ &\simeq \frac{\mathfrak{Im}(e^{\mathrm{i}\Gamma_N(x)-\mathrm{i}\Gamma_N(y)})}{\pi(\lambda-\mu)\sin(\theta(x_0))}N \\ &\simeq N\frac{\sin(\pi\hat{\rho}(x_0)(\lambda-\mu))}{\pi(\lambda-\mu)}\end{aligned}$$

so it follows as $N \to \infty$

$$K_N\left(x_0 + \frac{\lambda}{N\hat{\rho}(x_0)}, x_0 + \frac{\mu}{N\hat{\rho}(x_0)}\right)\frac{1}{N\hat{\rho}(x_0)} \simeq \frac{\sin(\pi(\lambda-\mu))}{\pi(\lambda-\mu)}.$$

□

5.5 Tracy-Widom limit

We continue the analysis of (5.28). For the Tracy-Widom limit we need the saddle point in the upper edge region which we already have computed in (5.32). The change of variable

$$x = \alpha + 2 + 2\sqrt{\alpha+1}u$$

and choosing $u = 1$ gives the saddle point at upper boundary

$$t_1 = -\frac{1}{\sqrt{1+\alpha}} \qquad \text{for } u = 1 \tag{5.79}$$

5.5.1 Airy asymptotics

For simplicity we assume $\alpha = 0$.

Lemma 14 . *For the scaling*

$$x = 4 + \frac{2^{(4/3)}\lambda}{N^{2/3}}, \qquad t = -1 + \frac{2^{(2/3)}s}{N^{1/3}} \tag{5.80}$$

we get the weights

$$\sqrt{\omega_\alpha(Nx)} = e^{-2N - N^{1/3}2^{1/3}\lambda} \tag{5.81}$$

and have the following asymptotics

$$\partial_z^l L_{N-k}^{(N\alpha)}(Nx)\sqrt{\omega_\alpha(Nx)} \approx \frac{(-1)^{-N+1}2^{(2/3)}}{N^{1/3}}\frac{1}{2\pi\mathrm{i}}\int_C \exp\left(\frac{s^3}{3} - \lambda s\right) F_{k,l}\left(-1 + \frac{2^{2/3}s}{N^{1/3}}\right)\mathrm{ds}$$

where C is the Airy contour as in definition 6 that starts at $\infty e^{-\frac{1}{3}\pi\mathrm{i}}$ encircles the origin in negative direction and ends in the conjugate sector $\infty e^{\frac{1}{3}\pi\mathrm{i}}$.

Proof. We make a Taylor expansion of $\phi(t,x)$ so we compute the values for $(t_0, x_0) = (-1, 4)$ are

$$\begin{aligned}
\phi(-1,4) &= -2 + \log(-1)\\
\phi_t(-1,4) &= \frac{4}{4} - 1 = 0, \qquad \phi_x(-1,4) = -\frac{1}{2}\\
\phi_{tt}(-1,4) &= \frac{8}{8} - 1 = 0, \quad \phi_{tx}(-1,4) = \frac{1}{4}, \quad \phi_{xt}(-1,4) = \frac{1}{4}, \quad \phi_{xx}(-1,4) = 0\\
\phi_{ttt}(-1,4) &= \frac{24}{16} - 2 = -\frac{1}{2}\\
\phi_{txt}(t,x;u,\alpha) &= \phi_{xtt}(t,x;u,\alpha) = \phi_{ttx}(t,x;u,\alpha) = \frac{1}{4}
\end{aligned}$$

We take the ansatz

$$x = 4 + \frac{a\lambda}{N^{2/3}} \qquad \text{and} \qquad t = -1 + \frac{cs}{N^{1/3}} \tag{5.82}$$

and make a Taylor expansion at $(t_0, x_0) = (-1, 4)$. The only derivative that contribute are $\phi_x, \phi_{xt}, \phi_{tx}, \phi_{ttt}$. The derivatives $\phi_{txt} = \phi_{xtt} = \phi_{ttx}$ go to the error term.

$$\begin{aligned}
\phi(-1,4) &= -2 + \log(-1) \\
\phi_x(-1,4)(x - x_0) &= -\frac{1}{2}\frac{a\lambda}{N^{2/3}} \\
\frac{\phi_{ttt}(-1,4)}{3!}(t - t_0)^3 &= -\frac{1}{12}\frac{(cs)^3}{N} \\
2\frac{\phi_{xt}(-1,4)}{2!}(x - x_0)(t - t_0) &= \frac{1}{4}\frac{a\lambda cs}{N} \\
\frac{\phi_{txt}(-1,4)}{2!}(x - x_0)(t - t_0)^2 &= \frac{1}{12}\frac{a\lambda (cs)^2}{N^{\frac{4}{3}}}
\end{aligned}$$

So the Taylor-expansion becomes

$$\begin{aligned}
\phi(x,t) &= (-2 + \log(-1)) - \frac{1}{2}\frac{a\lambda}{N^{2/3}} - \frac{1}{12}\frac{(cs)^3}{N} + \frac{1}{4}\frac{a\lambda cs}{N} + \mathcal{O}(\frac{1}{N^{4/3}}) \\
N\phi(x,t) &= N\left(-2 + \log(-1)\right) - N\frac{1}{2}\frac{a\lambda}{N^{2/3}} - N\frac{1}{12}\frac{(cs)^3}{N} + N\frac{1}{4}\frac{a\lambda cs}{N} + \mathcal{O}(\frac{1}{N^{4/3}}) \\
&= N\left(-2 + \log(-1)\right) - \frac{N^{1/3}a\lambda}{2} - \frac{(cs)^3}{12} + \frac{a\lambda cs}{4} + \mathcal{O}(\frac{1}{N^{4/3}})
\end{aligned}$$

And hence we get for

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{c}{N^{1/3}} \qquad \text{and} \qquad x = 4 + \frac{a\lambda}{N^{2/3}}$$

The following expression for the derivative of the Laguerre polynomial

$$\begin{aligned}
\partial_z^l L_{N-k}^{(N\alpha)}(Nx) = \frac{-1}{2\pi\mathrm{i}}\frac{c}{N^{1/3}} \exp\left(-N\left(-2 + \log(-1)\right) + \frac{N^{1/3}a\lambda}{2}\right) \\
\cdot \int_C \exp\left(\frac{(cs)^3}{12} - \frac{a\lambda cs}{4} - \mathcal{O}(\frac{1}{N^{1/3}})\right) F_{k,l}\left(-1 + \frac{2^{(2/3)}s}{N^{(1/3)}}\right) \mathrm{d}s
\end{aligned}$$

Now chose $c := 2^{(2/3)}$ and $a := 2^{(4/3)}$ and so we have the following scaling

$$x = 4 + \frac{2^{(4/3)}\lambda}{N^{2/3}}, \qquad t = -1 + \frac{2^{(2/3)}s}{N^{1/3}} \tag{5.83}$$

drop $-\mathcal{O}\left(\frac{1}{N^{4/3}}\right)$ since we are interested in the limit. We now deform the contour C to get the Airy contour from definition 6. C is a circle around the origin with radius < 1 in the positive direction. We extend the contour so that it starts at $\infty e^{\frac{1}{3}\pi\mathrm{i}}$ encircles the origin in positive direction and ends in the conjugate sector $\infty e^{-\frac{1}{3}\pi\mathrm{i}}$. As a last step we change its direction such that C starts at $\infty e^{-\frac{1}{3}\pi\mathrm{i}}$ encircles the

origin in negative direction and ends $\infty e^{\frac{1}{3}\pi \mathrm{i}}$, which gives an additional -1. This the Airy contour.

$$\partial_z^l L_{N-k}^{(N\alpha)}(Nx) \approx \frac{2^{(2/3)}}{N^{1/3}}(-1)^{N+1}\exp\left(2N + N^{1/3}2^{(1/3)}\lambda\right) \cdot \frac{1}{2\pi \mathrm{i}}\int_C \exp\left(\frac{s^3}{3} - \lambda s\right) F_{k,l}\left(-1 + \frac{2^{2/3}s}{N^{1/3}}\right) \mathrm{d}s.$$

We see the Airy phase. Further notice that for this scaling

$$Nx = N\left(4 + \frac{2^{4/3}\lambda}{N^{2/3}}\right) = 4N + N^{1/3}2^{4/3}\lambda$$

and $\alpha = 0$ we have the follwing weights

$$\sqrt{\omega_\alpha(Nx)} = e^{-2N - N^{1/3}2^{1/3}\lambda}$$

and therefore

$$\partial_z^l L_{N-k}^{(N\alpha)}(Nx)\sqrt{\omega_\alpha(Nx)} \approx \frac{(-1)^{-N+1}2^{(2/3)}}{N^{1/3}}\frac{1}{2\pi \mathrm{i}}\int_C \exp\left(\frac{s^3}{3} - \lambda s\right) F_{k,l}\left(-1 + \frac{2^{2/3}s}{N^{1/3}}\right) \mathrm{d}s.$$

□

Recall from chapter 2 definition 6 the Airy function and it's derivative

$$\mathrm{Ai}(x) = \frac{1}{2\pi i}\int_C \exp\left(\frac{s^3}{3} - xs\right) \mathrm{d}s, \qquad \mathrm{Ai}'(x) = \frac{-1}{2\pi i}\int_C s\exp\left(\frac{s^3}{3} - xs\right) \mathrm{d}s$$

where C is the contour from lemma 14.

Theorem 26 . *For the scaling*

$$x = 4 + \frac{2^{(4/3)}\lambda}{N^{2/3}}, \qquad t = -1 + \frac{2^{(2/3)}s}{N^{1/3}}$$

define the Laguerre functions

$$\phi_{N-k}(x) := \kappa_{N-k} L_{N-k}^{(N\alpha)}(Nx)\sqrt{\omega_\alpha(Nx)} \tag{5.84}$$

where κ_{N-k} is the normalization coefficient from (5.54) and explicitly

$$\kappa_{N-k} = N^{1/2} \tag{5.85}$$

and the weights are

$$\sqrt{\omega_\alpha(Nx)} = e^{-2N - N^{1/3}2^{1/3}\lambda}. \tag{5.86}$$

We have the following asymptotics

$$\phi_N(x) \approx N^{1/6}(-1)^{-N}2^{-1/3}\left\{\mathrm{Ai}(\lambda) - 3(2^{-1/3})\,\mathrm{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3})\right\} \tag{5.87}$$

$$\phi_{N-1}(x) \approx N^{1/6}(-1)^{-N+1}2^{-1/3}\left\{\mathrm{Ai}(\lambda) - 2^{-1/3}\,\mathrm{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3})\right\} \tag{5.88}$$

Proof. We start with lemma 14 and do a Taylor expansion for

$$F_{k,l}(t) = \frac{t^{l+k-1}}{(1-t)^{l+1}}$$

at -1 so

$$F_{k,l}\left(-1 + \frac{2^{2/3}s}{N^{1/3}}\right) = F_{k,l}(-1) + F'_{k,l}(-1)\left(\frac{2^{(2/3)}s}{N^{1/3}}\right) + \mathcal{O}(N^{-2/3})$$

case N**:** for $l = 0, k = 0$

$$F_{0,0}(t) = \frac{1}{t(1-t)}, \qquad F'_{0,0}(t) = \frac{2t-1}{t^2(1-t)^2}, \qquad F_{0,0}(-1) = -\frac{1}{2}, \qquad F'_{0,0}(-1) = -\frac{3}{4}$$

therefore

$$F_{0,0}\left(-1 + \frac{2^{2/3}s}{N^{1/3}}\right) = -\frac{1}{2} - \frac{3}{4}\left(\frac{2^{(2/3)}s}{N^{1/3}}\right) + \mathcal{O}(N^{-2/3})$$

Now we notice that the first integral of the sum is the Airy function and the second one $-\operatorname{Ai}'(s)$ therefore the integral becomes

$$-\frac{1}{2}\operatorname{Ai}(\lambda) + \frac{3}{2\cdot 2^{1/3}}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3})$$

and the N-th Laguerre asymptotic is

$$\begin{aligned} L_N^{(N\alpha)}(Nx)\sqrt{\omega_\alpha(Nx)} &\approx \frac{2^{(2/3)}}{N^{1/3}}(-1)^{-N+1} \\ &\quad \cdot \left\{ -\frac{1}{2}\operatorname{Ai}(\lambda) + \frac{3}{2\cdot 2^{1/3}}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3}) \right\} \\ &= \frac{1}{N^{1/3}}(-1)^{-N} \\ &\quad \cdot \left\{ 2^{-1/3}\operatorname{Ai}(\lambda) - \frac{3(2^{1/3})}{2}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3}) \right\} \\ &= \frac{1}{N^{1/3}}(-1)^{-N}2^{-1/3} \\ &\quad \cdot \left\{ \operatorname{Ai}(\lambda) - \frac{3}{2^{1/3}}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3}) \right\} \end{aligned}$$

case $N-1$**:** and $l = 0, k = 1$

$$F_{0,1}(t) = \frac{1}{(1-t)}, \qquad F'_{1,0}(t) = \frac{1}{(1-t)^2}, \qquad F'_{1,0}(-1) = \frac{1}{4}, \quad F_{1,0}(-1) = \frac{1}{2}$$

Evaluating the integral

$$\begin{aligned} &\frac{1}{2}\operatorname{Ai}(\lambda) - 2^{2/3}\,\frac{1}{4}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3}) \\ &\qquad = \frac{1}{2}\operatorname{Ai}(\lambda) - \frac{1}{2\cdot 2^{1/3}}\operatorname{Ai}'(\lambda)/N^{1/3} + O(N^{-2/3}) \end{aligned}$$

and we get the $N-1$ Laguerre asymptotic

$$\begin{aligned} L_{N-1}^{(N\alpha)}(Nx)e^{-2N-N^{1/3}2^{(1/3)}\lambda} \\ &\approx \frac{2^{(2/3)}}{N^{1/3}}(-1)^{-N+1}\cdot\left\{\frac{1}{2}\operatorname{Ai}(\lambda)-\frac{1}{2\cdot 2^{1/3}}\operatorname{Ai}'(\lambda)/N^{1/3}+O(N^{-2/3})\right\} \\ &= \frac{1}{N^{1/3}}(-1)^{-N+1}2^{-1/3}\left\{\operatorname{Ai}(\lambda)-2^{-1/3}\operatorname{Ai}'(\lambda)/N^{1/3}+O(N^{-2/3})\right\} \end{aligned}$$

And κ_{N-k} is as in (5.54) but for $\alpha=0$ we get for any $k\in\mathbb{Z}$

$$\kappa_{N-k} := \left(\frac{\Gamma(N(1+\alpha)+1-k)}{(N-k)!N^{N\alpha+1}}\right)^{-1/2} = \left(\frac{\Gamma(N+1-k)}{(N-k)!N}\right)^{-1/2} = N^{1/2}$$

and $\frac{N^{1/2}}{N^{1/3}} = N^{1/6}$ so we replace the factor. □

5.5.2 Airy kernel

Theorem 27 . *For the scaling*

$$x = 4+\frac{2^{4/3}\lambda_1}{N^{2/3}} \qquad \text{and} \qquad y = 4+\frac{2^{4/3}\lambda_2}{N^{2/3}}$$

we have

$$\begin{aligned} c_N\Big(\phi_N(x)\phi_{N-1}(y)-\phi_{N-1}(x)\phi_N(y)\Big) \\ &\approx (\operatorname{Ai}(x)\operatorname{Ai}'(y)-\operatorname{Ai}'(x)\operatorname{Ai}(y))+O(N^{-1/3}). \end{aligned}$$

where $c_N=\kappa_{N-1}/\kappa_N$.

Proof. By theorem 26 we have

$$\begin{aligned} \phi_N(x) &\approx N^{1/6}(-1)^{-N}2^{-1/3}\left\{\operatorname{Ai}(\lambda)-3(2^{-1/3})\operatorname{Ai}'(\lambda)/N^{1/3}+O(N^{-2/3})\right\} \\ \phi_{N-1}(x) &\approx N^{1/6}(-1)^{-N}2^{-1/3}\left\{-\operatorname{Ai}(\lambda)+2^{-1/3}\operatorname{Ai}'(\lambda)/N^{1/3}+O(N^{-2/3})\right\}. \end{aligned}$$

Let us introduce

$$K := N^{1/3}(-1)^{-2N}2^{-2/3}$$

and forget about the $O(N^{-2/3})$ term for the algebra

$$\begin{aligned} \phi_N(x)\phi_{N-1}(y) &\approx K\cdot\left\{\left(a_1\operatorname{Ai}(x)+a_2\operatorname{Ai}'(x)\frac{1}{N^{1/3}}\right)\left(b_1\operatorname{Ai}(y)+b_2\operatorname{Ai}'(y)\frac{1}{N^{1/3}}\right)\right\} \\ &= K\cdot\Big\{a_1b_1\operatorname{Ai}(x)\operatorname{Ai}(y)+a_1b_2\operatorname{Ai}(x)\operatorname{Ai}'(y)\frac{1}{N^{1/3}} \\ &\qquad +a_2b_1\operatorname{Ai}'(x)\operatorname{Ai}(y)\frac{1}{N^{1/3}}+a_2b_2\operatorname{Ai}'(x)\operatorname{Ai}'(y)\frac{1}{N^{2/3}}\Big\} \end{aligned}$$

and

$$\begin{aligned}\phi_N(y)\phi_{N-1}(x) &\approx K\cdot\left\{\left(a_1\operatorname{Ai}(y)+a_2\operatorname{Ai}'(y)\frac{1}{N^{1/3}}\right)\left(b_1\operatorname{Ai}(x)+b_2\operatorname{Ai}'(x)\frac{1}{N^{1/3}}\right)\right\}\\ &= K\Big\{a_1b_1\operatorname{Ai}(x)\operatorname{Ai}(y)+a_1b_2\operatorname{Ai}(y)\operatorname{Ai}'(x)\frac{1}{N^{1/3}})\\ &\qquad + a_2b_1\operatorname{Ai}'(y)\frac{1}{N^{1/3}}\operatorname{Ai}(x)+a_2b_2\operatorname{Ai}'(y)\operatorname{Ai}'(x)\frac{1}{N^{2/3}}\Big\}\end{aligned}$$

together

$$\begin{aligned}&\phi_N(x)\phi_{N-1}(y)-\phi_N(y)\phi_{N-1}(x)\\ &\quad\approx K\cdot\left\{(a_1b_2-a_2b_1)\operatorname{Ai}(x)\operatorname{Ai}'(y)\frac{1}{N^{1/3}}+(a_2b_1-a_1b_2)\operatorname{Ai}'(x)\operatorname{Ai}(y)\frac{1}{N^{1/3}}\right\}\\ &\quad= K(a_1b_2-a_2b_1)\left\{\operatorname{Ai}(x)\operatorname{Ai}'(y)\frac{1}{N^{1/3}}-\operatorname{Ai}'(x)\operatorname{Ai}(y)\frac{1}{N^{1/3}}\right\}\end{aligned}$$

We conclude

$$a_1=1,\quad a_2=-3(2^{-1/3}),\qquad b_1=-1,\qquad b_2=2^{-1/3}$$

and

$$(a_1b_2-a_2b_1)=2^{-1/3}-3(2^{-1/3})=-2^{(2/3)}. \tag{5.89}$$

We get an additional sign change from the first coefficient of the Laguerre polynomials and we have $c_N=\kappa_{N-1}/\kappa_N=1$. Using theorem 26 we get the approximation

$$\begin{aligned}&c_N\phi_N(x)\phi_{N-1}(y)-\phi_{N-1}(x)\phi_N(y)\\ &\quad\approx N^{1/3}(-1)^{-2N}2^{-2/3}\left\{\frac{2^{(2/3)}}{N^{1/3}}(\operatorname{Ai}(x)\operatorname{Ai}'(y)-\operatorname{Ai}'(x)\operatorname{Ai}(y))+O(N^{-2/3})\right\}\\ &\quad=(\operatorname{Ai}(x)\operatorname{Ai}'(y)-\operatorname{Ai}'(x)\operatorname{Ai}(y))+O(N^{-1/3}).\end{aligned}$$

□

The scaling of the kernel

For the scaling

$$x=4+\frac{2^{4/3}\lambda_1}{N^{2/3}}\qquad\text{and}\qquad y=4+\frac{2^{4/3}\lambda_2}{N^{2/3}}$$

then

$$N(x-y)=2(2N)^{1/3}(x-y)$$

and we have by the Christoffel-Darboux formula and theorem 27

$$\begin{aligned}K_N(Nx,Ny)&=c_N\frac{\phi_N(x)\phi_{N-1}(y)-\phi_{N-1}(x)\phi_N(y)}{N(x-y)}\\ &=\frac{(\operatorname{Ai}(x)\operatorname{Ai}'(y)-\operatorname{Ai}'(x)\operatorname{Ai}(y))+O(N^{-1/3})}{2(2N)^{1/3}(x-y)}\end{aligned}$$

since we get the convergence to the Airy kernel as $N \to \infty$

$$\begin{aligned} &2(2N)^{1/3} K_N(4N + 2(2N)^{1/3}x, 4N + 2(2N)^{1/3}y) \\ &\qquad = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y} + O(N^{-1/3}). \end{aligned} \tag{5.90}$$

We derived the following theorem.

Theorem 28 . *(Airy kernel) We assume $\alpha = n - N = 0$. After the scaling of x*

$$\mu + \sigma x, \qquad \text{where} \qquad \mu = 4N,\ \sigma = 2(2N)^{1/3} \tag{5.91}$$

the rescaled kernel converges to the Airy kernel

$$\lim_{N\to\infty} \sigma K_N\left(\mu + \sigma x, \mu + \sigma y\right) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}. \tag{5.92}$$

Let us check for $n = N$ with the formula from theorem 22

$$\begin{aligned} \mu &= \left(\sqrt{n} + \sqrt{N}\right)^2 = \left(2\sqrt{N}\right)^2 = 4N \\ \sigma &= (\sqrt{n} + \sqrt{N})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{N}}\right)^{1/3} = (2\sqrt{N})\left(\frac{2}{\sqrt{N}}\right)^{1/3} = 2(2N)^{1/3}. \end{aligned}$$

The density of the Airy point process

The density of the *Airy point process* is

$$\begin{aligned} \lim_{y\to x} \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y} &= \mathrm{Ai}(x)\,\mathrm{Ai}''(x) - \mathrm{Ai}'(x)\,\mathrm{Ai}'(x) \\ &= -x(\mathrm{Ai}(x)^2 - (\mathrm{Ai}'(x))^2. \end{aligned} \tag{5.93}$$

5.5.3 Tracy-Widom limit

Theorem 29 (Tracy-Widom law)**.** *As $N, n \to \infty$ and $N/n = 1 \in (0, \infty)$ we have*

$$\begin{aligned} &\lim_{N,n\to\infty} \mathbb{P}\left(\frac{\lambda_1 - \mu_{n,N}}{\sigma_{n,N}} \le t\right) \\ &\qquad = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_t^\infty \cdots \int_t^\infty \det\left(\mathrm{K}_{\mathrm{Ai}}(x_i, x_j)\right)_{i,j=1}^k \prod_{j=1}^k \mathrm{d}x_j \end{aligned} \tag{5.94}$$

where the scaling is

$$\mu_{n,N} = \left(\sqrt{n} + \sqrt{N}\right)^2 \tag{5.95}$$

$$\sigma_{n,N} = (\sqrt{n} + \sqrt{N})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{N}}\right)^{1/3}. \tag{5.96}$$

Proof. By theorem 28

$$\lim_{N\to\infty} \sigma K_N\left(\mu+\sigma x, \mu+\sigma y\right) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}$$

Consider now $\lambda_1 \geq \lambda_2 \geq \dots$ then the event

$$\prod_{j=1}^{\infty} \chi_{(-\infty,t]}(\lambda_i) = \chi_{(-\infty,t]}(\lambda_1) = \{\text{no particles in } (t,\infty)\}.$$

Therefore we have by the formula for $g : \mathbb{R} \to [0,1]$

$$\mathbb{E}\left[\prod_{j=1}^{\infty} g(\lambda_i)\right] = \det(I - gK)_{L^2(\mathbb{R})}$$

that

$$\mathbb{E}\left[\prod_{j=1}^{\infty} \chi_{(-\infty,t]}\right] = \det(I - gK)_{L^2(\mathbb{R})}$$

and so

$$\begin{aligned}\mathbb{P}\left(\text{no particles in } (t,\infty)\right) &= \det\left(I - \chi_{(-\infty,t]} K_{\mathrm{Ai}}\right) \\ &= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_t^{\infty} \cdots \int_t^{\infty} \det\left(\mathrm{K}_{\mathrm{Ai}}(x_i,x_j)\right)_{i,j=1}^{k} \prod_{j=1}^{k} \mathrm{d}x_j\end{aligned} \tag{5.97}$$

which is the Tracy-Widom law. □

Appendix A

Appendix

A.1 List of symbols

$x \ll y$	x much smaller than y
β	Dyson index.
$\widehat{\mathbb{P}}(\mathrm{d}x)$	Probability measure for ordered eigenvalues.
$\mathbb{P}(\mathrm{d}x)$	Probability measure for unordered eigenvalues.
$\mathbb{R}^p_{\geq,+}$	Space for p ordered eigenvalues on $\mathbb{R}_+$.
$\mathcal{S}_n$	Symmetric group of size n
$\mathcal{P}_k$	The space of all partitions of k.
$\triangle(\lambda)$	Vandermonde determinant
$\mathbb{H}^+_p$	Space of $p \times p$ positive Hermitian matrices.
$\mathbb{S}^+_p$	Space of $p \times p$ positive symmetric matrices.
$\mathbb{U}_p$	Unitary group with dimension p.
$\mathbb{O}_p$	Orthogonal group with dimension p.
$W_p(n, \Sigma_p)$	Wishart distribution of with n degress of freedom.
$\chi_S, \mathbf{1}_S$	Indicator function of the set S.
$\mathcal{S}_p$	Schatten-von Neumann class.
$\Gamma_p^{(\beta)}$	Generalized multivariate gamma function.
${}_tF_q^{(\beta)}[S]$	hypergeometric function of matrix argument.
$C_k^{(\beta)}$	Jack polynomial respectively zonal polynomial.

A.2 Combinatorics

Definition 30 [Mac95, pp. 19-21]**.** *We define the elementary symmetric polynomials $e_k(x_1, \ldots, x_n)$ for $k \leq n$:*

$$e_k(X_1, \ldots, X_n) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} X_{j_1} \cdots X_{j_k}$$

and the complete homogeneous symmetric polynomial $m_k(x_1, \ldots, x_n)$ for $k \leq n$:

$$m_k(X_1, \ldots, X_n) = \sum_{1 \leq j_1 \leq j_2 \leq \cdots \leq j_k \leq n} X_{j_1} \cdots X_{j_k}$$

In the case $k > n$

$$e_k(X_1, \ldots, X_n) = m_k(X_1, \ldots, X_n) = 0$$

and in the case $k = 0$

$$e_0(X_1, \ldots, X_n) = m_0(X_1, \ldots, X_n) = 1.$$

Example 8 : For $n = 2$

$$\begin{aligned} e_1(X_1, X_2) &= X_1 + X_2 \\ e_2(X_1, X_2) &= X_1 X_2 \end{aligned}$$

For $n = 3$

$$\begin{aligned} e_1(X_1, X_2, X_3) &= X_1 + X_2 + X_3 \\ e_2(X_1, X_2, X_3) &= X_1 X_2 + X_1 X_3 + X_1 X_3 \\ e_3(X_1, X_2, X_3) &= X_1 X_2 X_3. \end{aligned}$$

In general we have for $e_k(X_1, \ldots, X_n)$ exactly $\binom{n}{k}$ terms.

For $n = 2$

$$\begin{aligned} m_1(X_1, X_2) &= X_1 + X_2 \\ m_2(X_1, X_2) &= X_1^2 + X_1 X_2 + X_2^2 \end{aligned}$$

For $n = 3$

$$\begin{aligned} m_1(X_1, X_2, X_3) &= X_1 + X_2 + X_3 \\ m_2(X_1, X_2, X_3) &= X_1^2 + X_2^2 + X_3^2 + X_1 X_2 + X_1 X_3 + X_2 X_3 \\ m_3(X_1, X_2, X_3) &= X_1^3 + X_2^3 + X_3^3 + X_1^2 X_2 + X_1^2 X_3 \\ &\quad + X_2^2 X_1 + X_2^2 X_3 + X_3^2 X_1 + X_3^2 X_2 + X_1 X_2 X_3. \end{aligned}$$

Remark 39: *We usually omit the endpoints in the sum and just write*

$$\sum_{j_1 \leq j_2 \leq \cdots \leq j_k}$$

Lemma 15 (Combinatorics for Fredholm determinants)**.**

$$\begin{aligned}(1-tX_1)\cdots(1-tX_n) &= \sum_{k=0}^{n}(-t)^n e_k(X_1,X_2,\ldots,X_n)\\ &= \sum_{k=0}^{n}(-t)^n \sum_{j_1<j_2<\cdots<j_k}^{n} X_{j_1}X_{j_2}\cdots X_{j_k}\end{aligned}$$

Proof. For $n=2$

$$(1-tX_1)(1-tX_2) = \sum_{k=0}^{2}(-t)^k e_k(X_1,X_2,X_3).$$

For $n=3$

$$\begin{aligned}&(1-tX_1)(1-tX_2)(1-tX_3)\\ &\qquad = (1-tX_1-tX_2+t^2X_1X_2)(1-tX_3)\\ &\qquad = (1-tX_1-tX_2-tX_3+t^2X_1X_2+t^2X_1X_3+t^2X_2X_3\\ &\qquad\quad -t^3X_1X_2X_3)\\ &\qquad = \sum_{k=0}^{3}(-t)^k e_k(X_1,X_2,X_3).\end{aligned}$$

By induction

$$\begin{aligned}(1-tX_1)\cdots(1-tX_n)(1-tX_{n+1}) &= (1-tX_{n+1})\sum_{k=0}^{n}(-t)^n e_k(X_1,X_2,\ldots,X_n)\\ &= \sum_{k=0}^{n}(-t)^n e_k(X_1,X_2,\ldots,X_n)\\ &\qquad + x_{n+1}\sum_{k=0}^{n}(-t)^{n+1} e_k(X_1,X_2,\ldots,X_n)\\ &= \sum_{k=0}^{n+1}(-t)^{n+1} e_k(X_1,X_2,\ldots,X_n,X_{n+1})\\ &= \sum_{k=0}^{n+1}(-t)^{n+1} \sum_{j_1<j_2<\cdots<j_k}^{n+1} X_{j_1}X_{j_2}\cdots X_{j_{k-1}}X_{j_k}\end{aligned}$$

where we used the fact that the sums of $x_{n+1}e_k(X_1,X_2,\ldots,X_n)$ and the sum of $e_k(X_1,X_2,\ldots,X_n)$ share no term. □

A.3 Code examples in Matlab

A.3.1 Marchenko-Pastur example

```
rng(15012022)

testV = [[5000, 2000];[500, 100]];

for j=1:length(testV)

    n = testV(j:j,1:1);
    p = testV(j:j,2:2);

    gamma=p/n;
    rep = 10; % number of replicas
    sEig = zeros(p*rep,1); % total list of eigenvalues initialized

    % edge points of the MP distribution
    a=(1-sqrt(gamma))^2;
    b=(1+sqrt(gamma))^2;

    for r=1:rep
        X = randn(n,p);
        s = std(X(:));
        S=X'*X/n;
        pos = (r-1)*p+1; % position of the eigenvalues of the ...
            current replica
        sEig(pos:pos+p-1,1)=eig(S);
    end
    figure;
    % create histogram
    binSize = gamma/rep;
    edges = min(sEig):binSize:max(sEig); % bin edge positions
    h=histogram(sEig,edges,'Normalization','pdf');
    set(h,'FaceAlpha',0.3,'EdgeAlpha',0.5);

    % create MP density

    x = h.BinEdges;
    index = find(x<a|x>b); % find points outside support
    x(index)=[];
    x =[a x b];

    mpf =@(x,a,b,gamma) (1./(2*pi*x*gamma).*sqrt((b-x).*(x-a)));
    mp = mpf(x,a,b,gamma);

    % draw MP and real eigenvalues
    hold on
    plot(x,mp,'b','LineWidth',1.5);
    line([1 1], ...
        ylim,'Color','red','LineWidth',0.5,'LineStyle','--');
    title(strcat(sprintf('n = %i', n),sprintf(' p = %i', p)))
    hold off

```

```
end
```

A.3.2 Screeplot example

```
rng(15012022)

testV = [[50 1000];[100, 100]];

for j=1:length(testV)

    p = testV(j:j,1:1);
    n = testV(j:j,2:2);

    eigBig = 5:-0.5:2; % create spiked eigenvalues
    eigOnes = ones(1,p-length(eigBig));
    eigT = sort([eigBig eigOnes]); % eigenvalues

    Sigma = diag([eigBig eigOnes]);
    mu = zeros(1,p);

    X = mvnrnd(mu,Sigma,n); % draw n times p multivarite normal
    X = X-mean(X); % subtract the mean
    S = cov(X);

    figure
    plot(1:p,eig(S),'LineWidth',1.5);
    xlabel('Position')
    ylabel('Eigenvalue')
    title(strcat(sprintf('n = %i', n),sprintf(' p = %i', p)))

    hold on;
    plot(1:p,eigT,'LineStyle','--','LineWidth',1, 'Color',[1,0,0]);
    hold off;

end
```

Bibliography

[Ahl79] Lars V. Ahlfors. *Complex Analysis*. McGraw-Hill, 1979.

[BDT16] Jinho Baik, Percy Deift, and Suidan Toufic. *Combinatorics and Random Matrix Theory*. American Mathematical Society, 2016.

[Bor09] Folkmar Bornemann. On the numerical evaluation of fredholm determinants. *Mathematics of Computation*, 79(270):871–915, Sep 2009.

[Bre11] Haim Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer-Verlag New York, 2011.

[BS06] Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[BV12] Alex Bloemendal and Bálint Virág. Limits of spiked random matrices i. *Probability Theory and Related Fields*, 156(3–4):795–825, Sep 2012.

[Chi78] Theodore Seio Chihara. *An Introduction to Orthogonal Polynomials*. Gordon and Breach, 1978.

[CHP17] Tony Cai, Xiao Han, and Guangming Pan. Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices. 2017.

[DE02] Ioana Dumitriu and Alan Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, Nov 2002.

[Dei00] Percy Deift. *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach (Courant Lecture Notes)*. American Mathematical Society, 2000.

[DEKV13] Alexander Dubbs, Alan Edelman, Plamen Koev, and Praveen Venkataramana. The beta-wishart ensemble. *Journal of Mathematical Physics*, 54(8):083507, Aug 2013.

[DES04] Ioana Dumitriu, Alan Edelman, and Gene Shuman. Mops: Multivariate orthogonal polynomials (symbolically). *Journal of Symbolic Computation*, 42:587–620, 06 2004.

[EK18] Noureddine El Karoui. *The Oxford Handbook of Random Matrix Theory: Multivariate Statistics*. Oxford University Press, 2018.

[Enf73] Per Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Mathematica*, 130:309 – 317, 1973.

[For10] Peter J. Forrester. *Log-Gases and Random Matrices*. London Mathematical Society Monographs, 2010.

[Gau04] Walter Gautschi. *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press, 2004.

[Ger14] Fischer Gerd. *Lineare Algebra*. Springer Spektrum, 2014.

[GGK00] Israel Gohberg, Seymour Goldberg, and Nahum Krupnik. *Traces and Determinants of Linear Operators*. Birkhäuser Verlag, 2000.

[GN00] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000.

[Goo63] N. R. Goodman. Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction). *The Annals of Mathematical Statistics*, 34(1):152 – 177, 1963.

[HB14] Leonhard Held and Daniel Sabanés Bové. *Applied Statistical Inference: Likelihood and Bayes*. Springer, 2014.

[Jam64] Alan T. James. Distributions of Matrix Variates and Latent Roots Derived from Normal Samples. *The Annals of Mathematical Statistics*, 35(2):475 – 501, 1964.

[Joh01a] Kurt Johansson. Universality of the local spacing distribution in certain ensembles of hermitian wigner matrices. *Communications in Mathematical Physics*, 215(3):683–705, Jan 2001.

[Joh01b] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295 – 327, 2001.

[Joh05] Kurt Johansson. Random matrices and determinantal processes. *Mathematical Statistical Physics*, 11 2005.

[Koo13] Tom H. Koornwinder. Orthogonal polynomials. *Computer Algebra in Quantum Field Theory*, page 145–170, 2013.

[Lax02] Peter D. Lax. *Functional Analysis*. Wiley-Interscience, 2002.

[LNV18] Giacomo Livan, Marcel Novaes, and Pierpaolo Vivo. Introduction to random matrices. *SpringerBriefs in Mathematical Physics*, 2018.

[LX09] Fei Li and Yifeng Xue. Zonal polynomials and hypergeometric functions of quaternion matrix argument. *Communication in Statistics- Theory and Methods*, 38, 02 2009.

[Ma08] Zongming Ma. Accuracy of the tracy-widom limit for the largest eigenvalue in white wishart matrices. 2008.

[Mac95] Ian G. Macdonal. *Symmetric Functions and Hall Polynomials*. Oxford University Press, 2 edition, 1995.

[McS21] Colin McSwiggen. The harish-chandra integral: An introduction with examples. *L'Enseignement Mathématique*, 67(3):229–299, 2021.

[MF89] Bru Marie-France. Diffusions of perturbed principal component analysis. *Journal of Multivariate Analysis*, 29(1):127–136, 1989.

[Olv97] Frank W. J. Olver. *Asymptotics and Special Functions*. A K Peters, Wellesley, MA, 1997.

[Par98] Beresford N. Parlett. *The Symmetric Eigenvalue Problem*. Society for Industrial and Applied Mathematics, 1998.

[Pas20] Leonid Pastur. On random matrices arising in deep neural networks. gaussian case. 2020.

[PSG18] Jeffrey Pennington, Samuel S. Schoenholz, and Surya Ganguli. The emergence of spectral universality in deep networks. 2018.

[RVA04] T. Ratnarajah, R Vaillancourt, and M Alvo. Complex random matrices and applications. 02 2004.

[Sim77] Barry Simon. Notes on infinite determinants of hilbert space operators. *Advances in Mathematics*, 24(3):244–273, 1977.

[Sim05] Barry Simon. *Trace ideals and their applications*. American Mathematical Society, 2 edition, 2005.

[Sos01] Alexander Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. 2001.

[Spi99] Michael Spivak. *A Comprehensive Introduction to Differential Geometry*, volume 1. Publish or Perish, 1999.

[Sze75] Gabor Szegö. *Orthogonal Polynomials*. American Mathematical Society, 4 edition, 1975.

[Tak84] Akimichi Takemura. *Zonal Polynomials (Lecture Notes - Monograph Series)*, volume 4. Michigan State University, 1984.

[Tao13] Terence Tao. Terence tao's blog: The harish-chandra-itzykson-zuber integral formula. `https://terrytao.wordpress.com/2013/02/08/the-harish-chandra-itzykson-zuber-integral-formula/`, 2013. Accessed: 2021-12-12.

[Wai19] Marin J. Wainwright. *High-Dimensional Statistics*. Cambridge University Press, 2019.

[Won01] R. Wong. *Asymptotic Approximations of Integrals*. Society for Industrial and Applied Mathematics, 2001.