A Derivation of the Wishart and Singular Wishart Distributions

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Mathematics and Statistics
University of Saskatchewan
Saskatoon

By
Karly Stack

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Mathematics and Statistics
McLean Hall
106 Wiggins Rd
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5E6

Abstract

Multivariate statistical analysis is the area of statistics that is concerned with observations made on many variables. Determining how variables are related is a main objective in multivariate analysis. The covariance matrix is an essential part of understanding the dependence between variables. The distribution of the sample covariance matrix for a sample from a multivariate normal distribution, known as the Wishart distribution, is fundamental to multivariate statistical analysis. An important assumption of the well-known Wishart distribution is that the number of variables is smaller than the number of observations. In high-dimensions when the number of variables exceeds the number of observations, the Wishart matrix is singular and has a singular Wishart distribution. The purpose of this research is to rederive the Wishart and singular Wishart distributions and understand the mathematics behind each derivation.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Szmigielski for the continuous support of my Master's research. His motivation and immense knowledge aided me consistently throughout my studies.

I would also like to thank Prof. Soteros for the time and knowledgeable suggestions she offered for the writing of this thesis.

Contents

Chapter 1

Introduction

Multivariate statistical analysis is the area of statistics that is concerned with observations made on many variables [1]. An example of sample data that may be used for analysis is measurements of height and weight of individuals drawn randomly from a certain population. Determining how variables are related is a main objective in multivariate analysis. The covariance matrix is an essential part of understanding the dependence between variables. The distribution of the sample covariance matrix for a sample from a multivariate normal distribution, known as the Wishart distribution, is fundamental to multivariate statistical analysis [1].

In order to develop a better understanding of this chapter, we will introduce the underlying structure of the problem. A more detailed approach will be presented in chapter 2. Assume that a random vector $\mathbf{Y} = (X_1, X_2, \cdots, X_p)$ has a multivariate normal distribution, which we will denote $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$. Here $\boldsymbol{\mu} \in \mathbb{R}^{1\times p}$ is the population mean vector and $\Sigma \in \mathbb{R}^{p\times p}$ is the population covariance matrix. Consider a set of N independent observations on each of the p variates, which we will represent by a random matrix $\mathcal{X}$ having the independent random vectors $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ as its columns. We will demonstrate in chapter 2 that the sample covariance matrix $\mathcal{A}$ can be expressed as the difference between the matrix $\mathcal{X}\mathcal{X}^T$ and the matrix formed from the sample mean vector (see equation (2.1)). In order to simplify this expression, we first rotate $\mathcal{X}$ to create the rotated matrix $\mathcal{X}_r$. From the choice of the rotation matrix, we observe that $\mathcal{A}$ can be expressed as $N\mathcal{A} = \hat{\mathcal{X}}\hat{\mathcal{X}}^T$, where $\hat{\mathcal{X}}$ is the matrix obtained from removing the last column of $\mathcal{X}_r$. Thus, $\hat{\mathcal{X}}$ will have p rows and $M = N - 1$ independent random vectors $\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_M \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ as its columns. The density function of $\hat{\mathcal{X}}$ will be a starting point for the derivation of the distribution of the random Wishart matrix $\mathbf{\Omega} = \hat{\mathcal{X}}\hat{\mathcal{X}}^T$.

A fundamental assumption of the well-known Wishart distribution is that the dimension, p, is smaller than the number of observations, N. In high-dimensions when p is greater than N, the Wishart matrix is singular, where the reason for singularity is explained in chapter 2. For the case of a singular Wishart matrix, the random matrix has a singular Wishart distribution.

The research presented in this thesis is predominantly based on the derivation of the Wishart and singular Wishart distributions of J. Wishart [33] and R. A. Janik and M. A. Nowak [15]. Each of the distributions have been derived by several others, including the work of A. T. James [14] and I. Olkin and S. N. Roy [13] in their derivations of the Wishart distribution. The singular Wishart distribution has been derived by H. Uhlig [32], M. S. Srivastava [28], and Y. Yu and Y. Zhang [36], while a more recent derivation was provided by S.

Yu et al. [35] in 2014. Here we analyze the work of J. Wishart in his geometric derivation of the Wishart distribution and R. A. Janik and M. A. Nowak's algebraic approach to deriving both distributions. These publications were of interest because we wanted to find different ways of approaching the same problem. The fact that two completely different arguments can be used to derive identical distributions is an intriguing discovery.

J. Wishart's derivation of the Wishart distribution was of significant interest because of the alternative approach than that of R. A. Janik and M. A. Nowak. It is also the first derivation of the celebrated Wishart distribution and the work of J. Wishart has origins in random matrix theory [36]. The principal steps of the derivation are as follows. With the application of Fubini's theorem, we are able to consider the M observations on a single variate and conduct a geometrical change of variable. A substantial portion of chapter 3 is used to compute the Jacobian from this change of variable. After conducting this process for each variate and certain variables are integrated out, the result is a product of ratios of parallelogram volumes. J. Wishart provides an interesting result about such parallelogram volumes, which we present in Theorem 3.6.

The derivation of both the Wishart and singular Wishart distributions by R. A. Janik and M. A. Nowak uses an algebraic approach. The main point of the derivation is the fact that the density function of the random matrix $\mathbf{\Omega}$ satisfies a recurrence relation. We present extensive details about this recurrence relation and how to obtain a solution, more than that of the authors. The most interesting result of R. A. Janik and M. A. Nowak's paper was that the terms that appear in the recurrence relation can be expressed as a ratio of principal minors of the Wishart matrix Ω. We provide a different approach to proving this fact by using the Lewis Carroll identity. By incorporating a familiar identity, we are able to present a more precise proof of the theorem. We also compute the normalization coefficient of the singular Wishart distribution, which we compare to the derivation by S. Yu et al. [35] and this yields the same result.

R. A. Janik and M. A. Nowak [15] state that obtaining an expression for the singular Wishart distribution is a starting point for algorithms leading to fast reconstruction of the redundant information of the first M rows of the Wishart matrix. Y. Yu and Y. Zhang [36] explain that familiarity about the singular Wishart distribution is important in the study of bio-molecular interaction matrices, as knowledge of such a matrix is fundamental for quantitatively understanding how cells function. Future research interests include the application of the singular Wishart distribution in multivariate analysis, such as understanding the dependence between genes in DNA microarray experiments [7] and analyzing security returns in financial analysis [21].

Each of the distributions have statistical applications, two of which we will briefly introduce. The first will be the utilization of the Wishart distribution in the derivation of the generalized T^2 distribution and the second, an application of the singular Wishart distribution in Bayesian statistics.

Testing hypotheses about the mean vector $\boldsymbol{\mu}$ of the population and obtaining confidence intervals for the mean vector are important problems in multivariate statistics [20]. What is known as the T^2-statistic, proposed by Hotelling (1931), can be used for testing such hypotheses when the covariance matrix is unknown. In order to demonstrate the logistics of the statistic and how it is related to the Wishart distribution, we will introduce some assumptions. Suppose that $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ with sample mean vector $\bar{X} \in \mathbb{R}^{1\times p}$ and sample covariance matrix $A \in \mathbb{R}^{p\times p}$, then Hotelling's T^2-statistic is defined as $T^2 = N(\bar{X} - \boldsymbol{\mu})A^{-1}(\bar{X} - \boldsymbol{\mu})^T$ [20]. We will go into more details about the parameters, $\bar{X}$ and A, in the next chapter. Observe that if the population mean is zero, $\boldsymbol{\mu} = \mathbf{0}$, then one would expect the sample means, $\bar{X}$, to be close to zero and therefore the T^2-statistic would also be close to zero. If the population mean is unknown and we are testing the null hypothesis that $\boldsymbol{\mu} = \mathbf{0}$, we could reject the null when T^2 was sufficiently large. This threshold would be determined by the critical point on the T^2 distribution, in relation to the significance level. One of the fundamental assumptions for the T^2 distribution, which is proportional to the F-distribution, is that the matrix $\boldsymbol{\Omega} = N\mathcal{A}$ has a Wishart distribution [1].

In a paper by H. Uhlig [32], the basis of his motivation to obtain a distribution for a singular Wishart matrix is to update a Bayesian posterior when tracking a time-varying covariance matrix. A covariance matrix that is time-varying is common in financial analysis that uses time series data. H. Uhlig presents intricate calculations describing a posterior probability distribution, which is a probability distribution of a random variable that is conditional on evidence obtained from previous experiments. He explains, that the posterior probability for a covariance matrix A_t at time t is developed from one's knowledge of the distribution of the covariance matrix A_{t-1}, at time $t-1$. Thus, one can update a posterior probability for the covariance matrix as time is varying.

The following chapters will demonstrate the derivation of the Wishart and singular Wishart distributions. In chapter 2, we will go into greater detail about the assumptions of the model that were presented at the beginning of this chapter. That is, assuming our random variables have a multivariate normal distribution, we will make a transformation on our data matrix in order to derive the desired Wishart matrix. Also, we will address the fact that the Wishart matrix is singular when $p > N$. The derivation of the Wishart distribution, with the majority of the work following J. Wishart [33], will be demonstrated in chapter 3. This derivation will be based on a geometric argument, where the assumption that $p \leq N$ is crucial to the calculations. In section 3.2, we will consider a single variate and derive the Wishart distribution in this case. In the following section we will consider the bi-variate case and the last section of the chapter will be the derivation of the Wishart distribution for general p. For the final chapter, we will present the derivation of both the Wishart and Singular Wishart distributions. Computations throughout this chapter were motivated by the work of R. A. Janik and M A. Nowak [15].

Chapter 2

Samples of a Multivariate Normal Distribution

2.1 Introduction

The purpose of this chapter is to develop a starting point for the derivation of the distribution of a p-dimensional random Wishart matrix. By making a transformation of our matrix of sample values, we can express our sample covariance matrix as a matrix of inner products. This will allow us to obtain the density function as a prelude to our derivations, as well as make inferences about the rank of the sample covariance matrix.

2.2 The Density Function of $\hat{\mathcal{X}}$

Suppose that the elements of the random vector $\mathbf{Y} = (X_1, X_2, \ldots, X_p)$ are continuous random variables and define the set $\{\mathbf{y} = (x_1, x_2, \ldots, x_p) : x_k \in \mathbb{R}\}$, as the range of values the random vector can take. Assume the random variables $X_1, X_2, \ldots, X_p$ have a multivariate normal distribution with vector of means $\boldsymbol{\mu} \in \mathbb{R}^{1\times p}$ and covariance matrix $\Sigma \in \mathbb{R}^{p\times p}$. Then, their density function is given by, [23]

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{p}{2}}(\det \Sigma)^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})^T}, \qquad \mathbf{y} \in \mathbb{R}^p.$$

Here $\det \Sigma$ is the determinant of the matrix Σ and $\mathbf{y}$ denotes a row vector, while $\mathbf{y}^T$ is a column vector.

We will represent a set of N independent observations on each of the p variates, by the random matrix

$$\mathcal{X} = \begin{array}{c c} & \begin{array}{cccc} \mathbf{Y}_1^T & \mathbf{Y}_2^T & \cdots & \mathbf{Y}_N^T \end{array} \\ \begin{array}{c} \boldsymbol{\mathcal{X}}_1 \\ \boldsymbol{\mathcal{X}}_2 \\ \vdots \\ \boldsymbol{\mathcal{X}}_p \end{array} & \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1N} \\ X_{21} & X_{22} & \cdots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pN} \end{pmatrix} \end{array},$$

where $\mathbf{Y}_i^T$ is the vector of observations from the i^{th} sample and $\boldsymbol{\mathcal{X}}_j$ represents N independent observations of the j^{th} variable [23]. We will define $X = [x_{kn}] \in \mathbb{R}^{p\times M}$ as a matrix of possible values the random matrix $\mathcal{X}$ can take and we will denote $\mathbf{y}_i^T$ as the i^{th} column and $\mathbf{x}_j$ as the j^{th} row of X.

Then, the density function of the random matrix $\mathcal{X}$ is

$$\begin{aligned} f_{\mathcal{X}}(X) &= \prod_{n=1}^{N} f_{\mathbf{Y}_n}(\mathbf{y}_n) \\ &= \frac{1}{(2\pi)^{\frac{pN}{2}} (\det \Sigma)^{\frac{N}{2}}} \mathrm{e}^{-\frac{1}{2}\sum_{n=1}^{N}(\mathbf{y}_n-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{y}_n-\boldsymbol{\mu})^T}, \qquad \mathbf{y}_n \in \mathbb{R}^p \quad [23]. \end{aligned}$$

Throughout this chapter, we will be manipulating the matrices, vectors, and variables that are the possible values the corresponding random matrices, vectors, and variables can take. Any transformations done on these quantities, will correspond to the same transformation on the respective random quantity. For clarification, we will introduce the following notation:

- The random vector of sample means will be denoted by $\bar{\mathcal{X}}$ and the possible values this random vector can take will the denoted by $\bar{X}$
- The random sample covariance matrix will be denoted by $\mathcal{A}$ and the corresponding matrix of possible values will be denoted by A.

With this in mind, the entries of the vector of sample means $\bar{X} = [\bar{x}_k] \in \mathbb{R}^{1\times p}$ and the entries of the sample covariance matrix $A = [a_{kl}] \in \mathbb{R}^{p\times p}$ are defined by

$$\begin{aligned} \bar{x}_k &= \frac{1}{N}\sum_{n=1}^{N} x_{kn}, \\ a_{kl} &= \frac{1}{N}\sum_{n=1}^{N}(x_{kn}-\bar{x}_k)(x_{ln}-\bar{x}_l) = \frac{1}{N}\sum_{n=1}^{N} x_{kn}x_{ln} - \bar{x}_k\bar{x}_l \quad [1]. \end{aligned}$$

The second statement can be expressed in the following matrix form

$$N(A + \bar{X}^T\bar{X}) = XX^T.$$

Leading to an equation for the sample covariance matrix

$$NA = XX^T - N\bar{X}^T\bar{X}.$$

If we express this equation in terms of the corresponding random matrices and vectors, we obtain

$$N\mathcal{A} = \mathcal{X}\mathcal{X}^T - N\bar{\mathcal{X}}^T\bar{\mathcal{X}}. \tag{2.1}$$

The basis of our motivation from this point forward is to find a transformed matrix representation of A. In order to accomplish this, we will transform the data matrix X. If we choose a rotation matrix with the last column as $\left(1/\sqrt{N}, \quad 1/\sqrt{N}, \quad \ldots, \quad 1/\sqrt{N}\right)^T$, when this rotation matrix acts on X the last column of the resulting matrix will be proportional to the sample mean vector. This will allow us to simplify the equation given in (2.1) and the rank of our sample covariance matrix A will be clear.

When considering the N observations of a single variate, $x_{k1}, x_{k2}, \ldots, x_{kN}$, we have equations defining a hyperplane in $\mathbb{R}^N$

$$\bar{x}_k = \frac{1}{N}\sum_{n=1}^{N} x_{kn},$$

with the unit normal vector $\frac{1}{\sqrt{N}}(1, 1, \cdots, 1)^T$, as well as a sphere

$$a_{kk} = \frac{1}{N}\sum_{n=1}^{N} x_{kn}^2 - \bar{x}_k^2.$$

We will study the intersection of the hyperplane and the sphere by analyzing the line perpendicular to the hyperplane. To simplify the equation in (2.1), we will perform a rotation from the vector $(0, 0, \cdots, 1)^T$ to the vector $\frac{1}{\sqrt{N}}(1, 1, \cdots, 1)^T$, that is

$$R\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \frac{1}{\sqrt{N}}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix},$$

where $R \in \mathbb{R}^{N\times N}$ is a rotation matrix, i.e. $R^T = R^{-1}$ and $\det(R) = 1$. Now, the equations of the hyperplanes can be written as

$$N\bar{X} = (1, 1, \cdots, 1)\, X^T = \sqrt{N}\,(0, 0, \cdots, 1)\,(XR)^T,$$

resulting in

$$\sqrt{N}\bar{X} = (0, 0, \cdots, 1)\,(XR)^T. \tag{2.2}$$

Denote $XR = X_r = [\hat{x}_{kn}] \in \mathbb{R}^{p\times N}$, where each $\hat{x}_{kn}$ can be interpreted as the rotated observation from the original data matrix X. We will denote the corresponding random matrix as $\mathcal{X}_r$. Then, with the application of (2.2), we have

$$\sqrt{N}\bar{X} = (\hat{x}_{1N}, \hat{x}_{2N}, \cdots, \hat{x}_{pN}),$$

and

$$X_r X_r^T = (XR)(XR)^T = XRR^T X^T = XX^T. \tag{2.3}$$

From the transformation of the data matrix, we were able to obtain an explicit expression for the vector of sample means $\bar{X}$. That is, the sample mean vector is proportional to the last column of the rotated matrix X_r. This fact leads to the following lemma.

Lemma 2.1. *Let $\hat{X} = [\hat{x}_{km}] \in \mathbb{R}^{p\times M}$, with $M = N - 1$. Then,*

$$NA = \hat{X}\hat{X}^T,$$

where $\hat{X}$ is the matrix obtained from removing the last column of the rotated matrix X_r. The corresponding random matrix will be denoted as $\hat{\mathcal{X}} = [\hat{X}_{km}] \in \mathbb{R}^{p\times M}$, obtained from removing the last column of the random matrix $\mathcal{X}_r$. Then, we can express the random matrix $\mathcal{A}$ as

$$N\mathcal{A} = \hat{\mathcal{X}}\hat{\mathcal{X}}^T.$$

Proof. By (2.1) and (2.3),

$$\begin{aligned} NA &= X_r X_r^T - N\bar{X}^T\bar{X} \\ &= X_r X_r^T - \left(\hat{x}_{1N}, \hat{x}_{2N}, \cdots, \hat{x}_{pN}\right)^T \left(\hat{x}_{1N}, \hat{x}_{2N}, \cdots, \hat{x}_{pN}\right) \\ &= \hat{X}\hat{X}^T. \end{aligned}$$

□

Observe that from the previous lemma, the sample covariance matrix is expressed as a product of the transformed matrix $\hat{X} \in \mathbb{R}^{p\times M}$. Since the rank$(\hat{X}) \leq \min(M,p)$, then the rank$(A) \leq \min(M,p)$. In the case that the number of observations, N, is less than the number of variates, p, we have that the rank$(A) =$ rank$(\hat{X}) \leq M$. Notice that A is a p-dimensional matrix, thus when $p > N$, A will be singular.

Corollary 2.2. *NA is a Gram matrix, i.e.*

$$NA = \left[\hat{\mathbf{x}}_k \cdot \hat{\mathbf{x}}_l\right],$$

where $\hat{\mathbf{x}}_k$ is the k^{th} row of the matrix $\hat{X}$ and "$\cdot$" defines the inner product.

Proof.

$$\begin{aligned} NA &= \hat{X}\hat{X}^T \\ &= \begin{pmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_p \end{pmatrix} \begin{pmatrix} \hat{\mathbf{x}}_1 \\ \hat{\mathbf{x}}_2 \\ \vdots \\ \hat{\mathbf{x}}_p \end{pmatrix}^T \\ &= \begin{pmatrix} \hat{\mathbf{x}}_1\cdot\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_1\cdot\hat{\mathbf{x}}_2 & \dots & \hat{\mathbf{x}}_1\cdot\hat{\mathbf{x}}_p \\ \hat{\mathbf{x}}_2\cdot\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_2\cdot\hat{\mathbf{x}}_2 & \dots & \hat{\mathbf{x}}_2\cdot\hat{\mathbf{x}}_p \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{x}}_p\cdot\hat{\mathbf{x}}_1 & \hat{\mathbf{x}}_p\cdot\hat{\mathbf{x}}_2 & \dots & \hat{\mathbf{x}}_p\cdot\hat{\mathbf{x}}_p \end{pmatrix} \\ &= \left[\hat{\mathbf{x}}_k \cdot \hat{\mathbf{x}}_l\right]. \end{aligned}$$

□

As mentioned in the introduction of this chapter, we would like to derive the density function of the random Wishart matrix. We will define the random Wishart matrix by $\mathbf{\Omega} = N\mathcal{A}$, where a matrix of possible values $\mathbf{\Omega}$ can take will be denoted by Ω. Observe, that if the population mean $\boldsymbol{\mu}$ is known, we can shift each sample value by the population mean. Thus, without loss of generality we will assume $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_N \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$, giving the following density function

$$\begin{aligned} f_{\mathcal{X}}(X) &= \frac{1}{(2\pi)^{\frac{pN}{2}} (\det \Sigma)^{\frac{N}{2}}} \mathrm{e}^{-\frac{1}{2}\sum_{n=1}^{N} \mathbf{y}_n \Sigma^{-1} \mathbf{y}_n^T} \\ &= \frac{1}{(2\pi)^{\frac{pN}{2}} (\det \Sigma)^{\frac{N}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(X^T \Sigma^{-1} X)}. \end{aligned}$$

In the following theorem, we will show that the density function of the random matrix $\mathcal{X}$ is invariant under the rotation of $\mathcal{X}$ and the Jacobian of this transformation is equal to one.

We will introduce vertical bars to denote the product of the differentials, that is

$$|\mathrm{d}X| = \prod_{k=1}^{p} \prod_{m=1}^{M} \mathrm{d}x_{km}.$$

For symmetric matrices, we are only concerned with the $\frac{p(p+1)}{2}$ independent elements of the symmetric matrix, thus

$$|\mathrm{d}\Omega| = \prod_{\substack{k,l=1 \\ k \le l}}^{p} \mathrm{d}\omega_{kl}. \tag{2.4}$$

Theorem 2.3. *The density function of the random matrix $\mathcal{X}$ is invariant under the rotation of the random matrix, i.e.*

$$f_{\mathcal{X}_r}(X_r)|\mathrm{d}X_r| = f_{\mathcal{X}}(X)|\mathrm{d}X|.$$

Proof. For the exponent appearing in the density function, we have

$$\mathrm{Tr}((XR)^T \Sigma^{-1} (XR)) = \mathrm{Tr}(R^T X^T \Sigma^{-1} X R) = \mathrm{Tr}(\Sigma^{-1} X^T R^T R X) = \mathrm{Tr}(\Sigma^{-1} X^T X).$$

Define $\Psi : X \to X_r = XR$. Observe,

$$X_r = \begin{pmatrix} \mathbf{x}_1 R \\ \mathbf{x}_2 R \\ \vdots \\ \mathbf{x}_p R \end{pmatrix},$$

where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ are the rows of X. Then the Jacobian of this map will be

$$\det \begin{pmatrix} R & 0 & \ldots & 0 \\ 0 & R & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & R \end{pmatrix} = (\det R)^p.$$

Since R is an orthogonal matrix, $\det R = 1$. Thus, $f_{\mathcal{X}_r}(X_r)|\mathrm{d}X_r| = f_{\mathcal{X}}(X)|\mathrm{d}X|$. □

Theorem 2.4. *The random matrix $\hat{\mathcal{X}}$ has the following density function*

$$\hat{f}_{\hat{\mathcal{X}}}(\hat{X}) = \frac{1}{(2\pi)^{\frac{pM}{2}} (\det \Sigma)^{\frac{M}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)},$$

where $\Omega = NA = \hat{X}\hat{X}^T$.

Proof. Recall from Theorem 2.3, we have that $f_{\mathcal{X}_r}(X_r)|\mathrm{d}X_r| = f_{\mathcal{X}}(X)|\mathrm{d}X|$. For simplicity, denote $B = \Sigma^{-1}$, then the density function of the random matrix $\mathcal{X}_r$ is given by,

$$f_{\mathcal{X}_r}(X_r) = \frac{1}{(2\pi)^{\frac{pN}{2}}(\det\Sigma)^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(X_r^T B X_r)}.$$

We will denote the columns of X_r as

$$X_r = \left(\hat{\mathbf{y}}_1^T, \hat{\mathbf{y}}_2^T, \dots, \hat{\mathbf{y}}_N^T\right),$$

where $\hat{\mathbf{y}}_n^T = [\hat{x}_{kn}] \in \mathbb{R}^{p\times 1}$. Similarily, we will express B in terms of its columns

$$B = \left(\mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_p^T\right),$$

with $\mathbf{b}_k^T \in \mathbb{R}^{p\times 1}$. Then, the matrix product appearing in the exponent of our density function can be simplified to

$$\begin{aligned}
X_r^T B X_r &= \begin{pmatrix} \hat{\mathbf{y}}_1\mathbf{b}_1^T & \hat{\mathbf{y}}_1\mathbf{b}_2^T & \dots & \hat{\mathbf{y}}_1\mathbf{b}_p^T \\ \hat{\mathbf{y}}_2\mathbf{b}_1^T & \hat{\mathbf{y}}_2\mathbf{b}_2^T & \dots & \hat{\mathbf{y}}_2\mathbf{b}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{y}}_N\mathbf{b}_1^T & \hat{\mathbf{y}}_N\mathbf{b}_2^T & \dots & \hat{\mathbf{y}}_N\mathbf{b}_p^T \end{pmatrix} \cdot \left(\hat{\mathbf{y}}_1^T, \hat{\mathbf{y}}_2^T, \dots, \hat{\mathbf{y}}_N^T\right) \\
&= \begin{pmatrix} \hat{\mathbf{y}}_1\mathbf{b}_1^T\hat{x}_{11} + \ldots + \hat{\mathbf{y}}_1\mathbf{b}_p^T\hat{x}_{p1} & & & \\ & \hat{\mathbf{y}}_2\mathbf{b}_1^T\hat{x}_{12} + \ldots + \hat{\mathbf{y}}_2\mathbf{b}_p^T\hat{x}_{p2} & & \\ & & \ddots & \\ & & & \hat{\mathbf{y}}_N\mathbf{b}_1^T\hat{x}_{1N} + \ldots + \hat{\mathbf{y}}_N\mathbf{b}_p^T\hat{x}_{pN} \end{pmatrix}.
\end{aligned}$$

Since we are taking the trace of this product, the diagonal elements are only of interest. Thus,

$$\begin{aligned}
\mathrm{Tr}(X_r^T B X_r) &= \hat{\mathbf{y}}_1\mathbf{b}_1^T\hat{x}_{11} + \ldots + \hat{\mathbf{y}}_1\mathbf{b}_1^T\hat{x}_{p1} + \ldots + \hat{\mathbf{y}}_N\mathbf{b}_1^T\hat{x}_{1N} + \ldots + \hat{\mathbf{y}}_N\mathbf{b}_p^T\hat{x}_{pN} \\
&= \mathrm{Tr}(\hat{X}^T B\hat{X}) + \hat{\mathbf{y}}_N B\hat{\mathbf{y}}_N^T.
\end{aligned}$$

Applying the above equation to our density function, we obtain

$$f_{\mathcal{X}_r}(X_r)|\mathrm{d}X_r| = \frac{1}{(2\pi)^{\frac{pN}{2}}(\det\Sigma)^{\frac{N}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\hat{X}^T B\hat{X})}\mathrm{e}^{\frac{1}{2}\hat{\mathbf{y}}_N^T B\hat{\mathbf{y}}_N}|\mathrm{d}\hat{X}||\mathrm{d}\mathbf{y}_N|.$$

Integrating our density function with respect to the p variables, $\hat{x}_{1N}, \hat{x}_{2N}, \dots, \hat{x}_{pN}$, will produce a factor of $(2\pi)^{\frac{p}{2}}(\det B)^{-\frac{1}{2}}$, as observed in Lemma C.1 in Appendix C. Hence, the density function of the random matrix $\hat{\mathcal{X}}$ is given by

$$\begin{aligned}
f_{\hat{\mathcal{X}}}(\hat{X}) &= \frac{1}{(2\pi)^{\frac{p(N-1)}{2}}(\det\Sigma)^{\frac{N-1}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\hat{X}^T B\hat{X})} \\
&= \frac{1}{(2\pi)^{\frac{pM}{2}}(\det\Sigma)^{\frac{M}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\hat{X}^T\Sigma^{-1}\hat{X})}.
\end{aligned}$$

Considering that the trace is invariant under cyclic permutations and $\Omega = NA = \hat{X}\hat{X}^T$,

$$f_{\hat{\mathcal{X}}}(\hat{X}) = \frac{1}{(2\pi)^{\frac{pM}{2}}(\det\Sigma)^{\frac{M}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)}. \qquad \square$$

**Notation

We will consider the above formula to be our starting point for future reductions. For simplicity, we will drop the hat from X in the formula. Thus, our density function reads

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}} (\det \Sigma)^{\frac{M}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)}, \tag{2.5}$$

with

$$X \in \mathbb{R}^{p\times M} \quad \text{and} \quad \Omega = XX^T.$$

2.3 Summary

The computations throughout this chapter allowed us to obtain a starting point for the derivation of the distribution of the random Wishart matrix $\mathbf{\Omega}$. That is, by rotating the random matrix $\mathcal{X}$, we generated a representation of our random sample covariance matrix $\mathcal{A}$ as a matrix of inner products. We then showed that the density function of $\mathcal{X}$ was invariant under the rotation of $\mathcal{X}$. Furthermore, we considered the density function of the rotated matrix $\mathcal{X}_r$ and integrated with respect to the last column of the rotated matrix. This enabled us to reduce the dimension and obtain the desired density function of the random matrix $\hat{\mathcal{X}} \in \mathbb{R}^{p\times M}$, for $M = N - 1$. Also, from the calculations we were able to deduce the rank of the random Wishart matrix, i.e. when $p > M$, rank $\mathbf{\Omega} = M$.

Chapter 3

A Geometric Derivation of the Wishart Distribution

3.1 Introduction

In this chapter we will obtain the Wishart distribution of a random matrix $\mathbf{\Omega}$ for the case of $p = 1$, $p = 2$, and for general p. With the application of Fubini's theorem, we will fix $p-1$ vectors and conduct a geometrical change of variable on the remaining vector. The majority of the work will involve the computation of the Jacobian from the transformation. The integration of part of the Jacobian, that involves the polar angles of the parameterization, gives the surface area of a sphere and the radius of the sphere is a ratio of parallelogram volumes. Once we have considered each vector and integrated out all of the angles we will see almost a complete cancellation of the parallelogram volumes and obtain the density function of the Wishart matrix. The calculations and geometrical argument presented throughout this chapter follow the derivation by J. Wishart [33]. We will now state the celebrated result of J. Wishart and the central outcome of this chapter, which appears in section 3.4, equation (3.23).

Consider the random matrix $\hat{\mathcal{X}}$, having the random vectors $\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_M \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ as its columns. The random matrix $\mathbf{\Omega} = \hat{\mathcal{X}}\hat{\mathcal{X}}^T$ is said to have a Wishart distribution if it has the following density function

$$F_{M>p}(\Omega) = \frac{\pi^{\frac{-p(p-1)}{4}}}{2^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{p} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Sigma)^{-\frac{M}{2}} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)}, \tag{3.1}$$

where $p < M$ [1]. Here, $\det \Sigma$ is the determinant of the covariance matrix and $\det \Omega$ the determinant of the Wishart matrix $\Omega = [\omega_{kl}] \in \mathbb{R}^{p \times p}$.

3.2 Univariate

In order to rederive the Wishart distribution for a univariate population, we will be sampling from our density function in (2.5),

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}} \sigma_{11}^{\frac{M}{2}}} \mathrm{e}^{-\frac{\omega_{11}}{2\sigma_{11}}},$$

with

$$X = \mathbf{x}_1 \in \mathbb{R}^{1 \times M}, \quad \omega_{11} = \mathbf{x}_1 \cdot \mathbf{x}_1, \tag{3.2}$$

and σ_{11} is the variance of the random vector $\boldsymbol{\mathcal{X}}_1$ as defined in Appendix A, definition A.7.

Observe this is a sphere in $\mathbb{R}^M$ of radius $\sqrt{\omega_{11}}$. Now, we will choose an arbitrary unit vector $\mathbf{g} \in \mathbb{R}^M$, then $\mathbf{x}_1$ can be parameterized as

$$\mathbf{x}_1 = \sqrt{\omega_{11}} R(\boldsymbol{\Theta}_1)\mathbf{g}, \tag{3.3}$$

where $R(\boldsymbol{\Theta}_1)$ is a rotation matrix. Define $\Psi_1 : \mathbb{R}^M \to \mathbb{R}^M$ as a map, such that $\Psi_1(\mathbf{u}_1) = \mathbf{x}_1$. Specifically,

$$\Psi_1 : \mathbf{u}_1 = \begin{pmatrix} \omega_{11} \\ \theta_{1,1} \\ \vdots \\ \theta_{1,M-1} \end{pmatrix}^T \longrightarrow \mathbf{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1M} \end{pmatrix}^T . \tag{3.4}$$

Consider the change of variable formula. Let g be an integrable function on $\mathbb{R}^M$, then

$$\begin{aligned} \int_{\mathbb{R}^M} g(\mathbf{x}_1)|\mathrm{d}\mathbf{x}_1| &= \int_{\mathbb{R}^M} g(\Psi_1(\mathbf{u}_1)\, |\det(\mathbf{D}\Psi_1(\mathbf{u}_1))|\, |\mathrm{d}\mathbf{u}_1| \\ &= \int_{\mathbb{R}^M} g(\Psi_1(\mathbf{u}_1)) \left[\det\left((\mathbf{D}\Psi_1(\mathbf{u}_1))^T (\mathbf{D}\Psi_1(\mathbf{u}_1))\right)\right]^{1/2} |\mathrm{d}\mathbf{u}_1|. \end{aligned} \tag{3.5}$$

The Jacobian appearing in (3.5) is therefore,

$$\left[\det\left((\mathbf{D}\Psi_1(\mathbf{u}_1))^T (\mathbf{D}\Psi_1(\mathbf{u}_1))\right)\right]^{1/2} = \left[\det\left(\begin{array}{c|c} \left\| \dfrac{\partial \mathbf{x}_1}{\partial \omega_{11}} \right\|^2 & \dfrac{\partial \mathbf{x}_1}{\partial \omega_{11}} \cdot \dfrac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} \\ \hline \dfrac{\partial \mathbf{x}_1}{\partial \theta_{1,i}} \cdot \dfrac{\partial \mathbf{x}_1}{\partial \omega_{11}} & \dfrac{\partial \mathbf{x}_1}{\partial \theta_{1,i}} \cdot \dfrac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} \end{array}\right)\right]^{1/2}, \tag{3.6}$$

where $1 \leq i, j \leq M-1$.

Lemma 3.1. *Consider the vector $\mathbf{x}_1$ as defined in* (3.3)*. For all $1 \leq j \leq M-1$*

$$\frac{\partial \mathbf{x}_1}{\partial \omega_{11}} \cdot \frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} = 0.$$

Proof. Observe that $\mathbf{x}_1 \cdot \mathbf{x}_1 = \omega_{11}$, then

$$\begin{aligned} \frac{\partial}{\partial \theta_{1,j}} (\mathbf{x}_1 \cdot \mathbf{x}_1) &= \frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} \cdot \mathbf{x}_1 + \mathbf{x}_1 \cdot \frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} \\ &= 2\left(\mathbf{x}_1 \cdot \frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}}\right) = 0. \end{aligned}$$

By the parameterization in 3.3, we see that $\dfrac{\partial \mathbf{x}_1}{\partial \omega_{11}} \propto \mathbf{x}_1$. Thus,

$$\frac{\partial \mathbf{x}_1}{\partial \omega_{11}} \cdot \frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} = 0. \qquad \square$$

Using Lemma 3.1, our Jacobian becomes

$$\left[\det\left(\left(\mathbf{D}\Psi_1(\mathbf{u}_1)\right)^T\left(\mathbf{D}\Psi_1(\mathbf{u}_1)\right)\right)\right]^{1/2} = \left[\det\begin{pmatrix} \left\|\frac{\partial \mathbf{x}_1}{\partial \omega_{11}}\right\|^2 & 0 \\ 0 & \frac{\partial \mathbf{x}_1}{\partial \theta_{1,i}}\cdot\frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}} \end{pmatrix}\right]^{1/2}$$
$$= \left\|\frac{\partial \mathbf{x}_1}{\partial \omega_{11}}\right\| \left[\det\left(\frac{\partial \mathbf{x}_1}{\partial \theta_{1,i}}\cdot\frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}}\right)\right]^{1/2}$$
$$= \frac{1}{2\sqrt{\omega_{11}}}\left[\det\left(\frac{\partial \mathbf{x}_1}{\partial \theta_{1,i}}\cdot\frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}}\right)\right]^{1/2}.$$

Observe the integration of the determinant appearing in the above result, with respect to each angle $\theta_{1,j}$, gives the surface area of a $(M-1)$-dimensional sphere having radius $\sqrt{\omega_{11}}$, thus

$$A_{M-1}\left(\sqrt{\omega_{11}}\right) = \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi}\left[\det\left(\frac{\partial \mathbf{x}_1}{\partial \theta_{1,i}}\cdot\frac{\partial \mathbf{x}_1}{\partial \theta_{1,j}}\right)\right]^{1/2} \mathrm{d}\theta_{1,M-1}\cdots \mathrm{d}\theta_{1,2}\mathrm{d}\theta_{1,1}$$
$$= \frac{2\pi^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}\right)}\left(\sqrt{\omega_{11}}\right)^{M-1}.$$

After this change of variables, we obtain the univariate Wishart distribution of a random variable ω

$$F_{M>1}(\omega_{11}) = \frac{1}{2^{\frac{M}{2}}\Gamma\left(\frac{M}{2}\right)}\sigma_{11}^{-\frac{M}{2}}\omega_{11}^{\frac{M-2}{2}}\mathrm{e}^{-\frac{\omega_{11}}{2\sigma_{11}}}.$$

3.3 Bivariate

For a bivariate population, we will be sampling from our density function in (2.5),

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}}(\det\Sigma)^{\frac{M}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)},$$

with

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \in \mathbb{R}^{2\times M},$$

and

$$\Omega = XX^T = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1\cdot\mathbf{x}_2 \\ \mathbf{x}_1\cdot\mathbf{x}_2 & \|\mathbf{x}_2\|^2 \end{pmatrix}$$
$$= \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}. \tag{3.7}$$

The 2-dimensional parallelogram in $\mathbb{R}^M$ spanned by vectors $\mathbf{x}_1, \mathbf{x}_2$, denoted $P(\mathbf{x}_1, \mathbf{x}_2)$, has a 2-dimensional volume of

$$v_2 = \mathrm{Vol}_2 P(\mathbf{x}_1, \mathbf{x}_2) = \left[\det(XX^T)\right]^{1/2} = (\det\Omega)^{1/2}.$$

Similarly, the volume of the 1-dimensional parallelogram spanned by $\mathbf{x}_1$ is given by

$$v_1 = \mathrm{Vol}_1 P(\mathbf{x}_1) = \sqrt{\mathbf{x}_1 \cdot \mathbf{x}_1} = \sqrt{\omega_{11}}.$$

Consider the expected value of the function $g(\boldsymbol{\Omega})$, with respect to density function $f_{\hat{\mathcal{X}}}(X)$, as defined in Appendix A Definition A.5

$$\begin{aligned} E(g) &= \int_{\mathbb{R}^{2M}} g(\Omega) f_{\hat{\mathcal{X}}}(X) |\mathrm{d}X| && (3.8)\\ &= \int_{\mathbb{R}^M} \left[\int_{\mathbb{R}^M} g(\Omega) f_{\hat{\mathcal{X}}}(X) |\mathrm{d}\mathbf{x}_2| \right] |\mathrm{d}\mathbf{x}_1|, && (3.9) \end{aligned}$$

where we used Fubini's Theorem, as presented in Appendix H. That is, we will fix the vector $\mathbf{x}_1$ and integrate with respect to $\mathbf{x}_2$. First, we will parameterize $\mathbf{x}_2$ using the following procedure. Consider a subspace $W \in \mathbb{R}^M$ such that $\mathbb{R}^M = W \oplus W^\perp$, where $W = \mathbb{R}\mathbf{x}_1$ and $W^\perp$ the orthogonal complement of W. Then, we can express $\mathbf{x}_2$ as

$$\mathbf{x}_2 = \alpha \mathbf{x}_1 + \mathbf{x}_2^\perp, \tag{3.10}$$

where $\mathbf{x}_2^\perp \in W^\perp$ and α a constant coefficient. Performing rotations of a unit vector $\mathbf{f} \in W^\perp$, that leave $W^\perp$ invariant, we obtain a unit $(M-2)$-dimensional sphere. Thus, an arbitrary point on a sphere of radius $||\mathbf{x}_2^\perp||$ can be parametrized as $||\mathbf{x}_2^\perp|| R(\Theta_2)\mathbf{f}$, where $R(\Theta_2)$ is a rotation matrix that leaves $W^\perp$ invariant. Therefore,

$$\mathbf{x}_2 = \alpha \mathbf{x}_1 + \left\| \mathbf{x}_2^\perp \right\| R(\boldsymbol{\Theta}_2)\mathbf{f}.$$

From (3.7) and the parameterization of $\mathbf{x}_2$ in (3.10), we can determine α and the length of $\mathbf{x}_2^\perp$ with the following computations

$$\mathbf{x}_2 \cdot \mathbf{x}_1 = (\alpha \mathbf{x}_1 + \mathbf{x}_2^\perp) \cdot \mathbf{x}_1 = \alpha(\mathbf{x}_1 \cdot \mathbf{x}_1).$$

Hence,

$$\alpha = \frac{\omega_{12}}{\omega_{11}}.$$

Then, we see that the length of $\mathbf{x}_2^\perp$ can be expressed as a ratio of parallelogram volumes, i.e.

$$\begin{aligned} \left\| \mathbf{x}_2^\perp \right\|^2 &= \mathbf{x}_2^\perp \cdot (\mathbf{x}_2 - \frac{\omega_{12}}{\omega_{11}} \mathbf{x}_1) \\ &= \mathbf{x}_2^\perp \cdot \mathbf{x}_2 - \frac{\omega_{12}}{\omega_{11}} \left(\mathbf{x}_2^\perp \cdot \mathbf{x}_1 \right) \\ &= \mathbf{x}_2^\perp \cdot \mathbf{x}_2 \\ &= (\mathbf{x}_2 - \frac{\omega_{12}}{\omega_{11}} \mathbf{x}_1) \cdot \mathbf{x}_2 \\ &= \omega_{22} - \frac{\omega_{12}^2}{\omega_{11}} \\ &= \frac{\det \Omega}{\omega_{11}} = \left(\frac{v_2}{v_1} \right)^2, && (3.11) \end{aligned}$$

resulting in a final parameterization for $\mathbf{x}_2$

$$\mathbf{x}_2 = \frac{\omega_{12}}{\omega_{11}}\mathbf{x}_1 + \frac{v_2}{v_1}R(\mathbf{\Theta}_2)\mathbf{f}. \tag{3.12}$$

Define $\Psi_2 : \mathbb{R}^M \to \mathbb{R}^M$ to be a map, such that $\Psi_2(\mathbf{u}_2) = \mathbf{x}_2$. Specifically,

$$\Psi_2 : \mathbf{u}_2 = \begin{pmatrix} \omega_{12} \\ \omega_{22} \\ \theta_{2,1} \\ \vdots \\ \theta_{2,M-2} \end{pmatrix}^T \longrightarrow \mathbf{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ x_{23} \\ \vdots \\ x_{2M} \end{pmatrix}^T.$$

Recall the change of variables formula from (3.5). For the above map, the computation of the Jacobian is as follows

$$\left[\det\left(\left(\mathbf{D}\Psi_2(\mathbf{u}_2)\right)^T \left(\mathbf{D}\Psi_2(\mathbf{u}_2)\right)\right)\right]^{1/2} = \left[\det\left(\begin{array}{cc|c} \left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\right\|^2 & \frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\cdot\frac{\partial \mathbf{x}_2}{\partial \omega_{22}} & \frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} \\ \frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\cdot\frac{\partial \mathbf{x}_2}{\partial \omega_{22}} & \left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\right\|^2 & \frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} \\ \hline \frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,i}} & \frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,i}} & \frac{\partial \mathbf{x}_2}{\partial \theta_{2,i}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} \end{array}\right)\right]^{1/2},$$

where $1 \le i, j \le M-2$.

Lemma 3.2. *Consider the vector $\mathbf{x}_2$ as defined in* (3.12)*. Then, for all $1 \le j \le M-2$,*

$$\frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} = 0, \quad \frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\cdot\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} = 0.$$

Proof. Observe from (3.11), we have that $\mathbf{x}_2^{\perp}\cdot\mathbf{x}_2^{\perp}$ does not depend on $\theta_{2,j}$. Then,

$$\frac{\partial}{\partial \theta_{2,j}}\left(\mathbf{x}_2^{\perp}\cdot\mathbf{x}_2^{\perp}\right) = 2\left(\frac{\partial \mathbf{x}_2^{\perp}}{\partial \theta_{2,j}}\cdot\mathbf{x}_2^{\perp}\right) = 0.$$

Similarly, $\mathbf{x}_2\cdot\mathbf{x}_1$ is independent of our angles

$$\begin{aligned}\frac{\partial}{\partial \theta_{2,j}}\left(\mathbf{x}_2\cdot\mathbf{x}_1\right) &= \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}}\cdot\mathbf{x}_1 + \mathbf{x}_2\cdot\frac{\partial \mathbf{x}_1}{\partial \theta_{2,j}} \\ &= \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}}\cdot\mathbf{x}_1 = 0.\end{aligned}$$

From (3.10), we obtain

$$\begin{aligned}\frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} &= \frac{\omega_{12}}{\omega_{11}}\frac{\partial \mathbf{x}_1}{\partial \theta_{2,j}} + \frac{\partial \mathbf{x}_2^{\perp}}{\partial \theta_{2,j}} \\ &= \frac{\partial \mathbf{x}_2^{\perp}}{\partial \theta_{2,j}},\end{aligned}$$

and

$$\frac{\partial \mathbf{x}_2}{\partial \omega_{12}} = \frac{1}{\omega_{11}}\mathbf{x}_1 + \frac{\partial \mathbf{x}_2^{\perp}}{\partial \omega_{12}}.$$

From the parameterization in (3.12), we have that $\dfrac{\partial \mathbf{x}_2^{\perp}}{\partial \omega_{12}} \propto \mathbf{x}_2^{\perp}$. Hence,

$$\frac{\partial \mathbf{x}_2}{\partial \omega_{12}} = \frac{1}{\omega_{11}}\mathbf{x}_1 + \beta \mathbf{x}_2^{\perp},$$

where β does not depend on $\theta_{2,j}$. Thus,

$$\begin{aligned}\frac{\partial \mathbf{x}_2}{\partial \omega_{12}} \cdot \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} &= \frac{1}{\omega_{11}}\left(\mathbf{x}_1 \cdot \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}}\right) + \beta\left(\mathbf{x}_2^{\perp} \cdot \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}}\right) \\ &= 0.\end{aligned}$$

By the same argument

$$\frac{\partial \mathbf{x}_2}{\partial \omega_{22}} \cdot \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} = 0. \qquad \square$$

Once applying Lemma 3.2, our Jacobian can be simplified to

$$\left[\det\left(\left(\mathbf{D}\Psi_2(\mathbf{u}_2)\right)^T\left(\mathbf{D}\Psi_2(\mathbf{u}_2)\right)\right)\right]^{1/2} = \left[\det\left(\begin{array}{cc|c} \left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\right\|^2 & \frac{\partial \mathbf{x}_2}{\partial \omega_{12}} \cdot \frac{\partial \mathbf{x}_2}{\partial \omega_{22}} & 0 \\ \frac{\partial \mathbf{x}_2}{\partial \omega_{12}} \cdot \frac{\partial \mathbf{x}_2}{\partial \omega_{22}} & \left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\right\|^2 & 0 \\ \hline 0 & 0 & \frac{\partial \mathbf{x}_2}{\partial \theta_{2,i}} \cdot \frac{\partial \mathbf{x}_2}{\partial \theta_{2,j}} \end{array}\right)\right]^{1/2}.$$

Calculating the partial derivatives of the above Jacobian, we obtain

$$\begin{aligned}\frac{\partial \mathbf{x}_2}{\partial \omega_{12}} &= \frac{\mathbf{x}_1}{v_1} - \omega_{12}\frac{1}{v_2 v_1}R(\mathbf{\Theta}_2)\mathbf{f}, \\ \frac{\partial \mathbf{x}_2}{\partial \omega_{22}} &= \frac{v_1}{2v_2}R(\mathbf{\Theta}_2)\mathbf{f}.\end{aligned}$$

Then

$$\begin{aligned}\left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{12}}\right\|^2 &= \frac{\omega_{22}}{v_2^2}, \\ \left\|\frac{\partial \mathbf{x}_2}{\partial \omega_{22}}\right\|^2 &= \frac{\omega_{11}}{4v_2^2}, \\ \frac{\partial \mathbf{x}_2}{\partial \omega_{12}} \cdot \frac{\partial \mathbf{x}_2}{\partial \omega_{22}} &= -\frac{\omega_{12}}{2v_2^2}.\end{aligned}$$

Resulting in the following

$$\left[\det\left((\mathbf{D}\Psi_2(\mathbf{u}_2))^T(\mathbf{D}\Psi_2(\mathbf{u}_2))\right)\right]^{1/2} = \left[\frac{1}{4}\left(\frac{1}{v_2^2}\right)^2 \det\begin{pmatrix} \omega_{22} & -\omega_{12} \\ -\omega_{12} & \omega_{11}\end{pmatrix} \det\left(\frac{\partial \mathbf{x}_2}{\partial\theta_{2,i}}\cdot\frac{\partial \mathbf{x}_2}{\partial\theta_{2,j}}\right)\right]^{1/2}$$
$$= \left[\frac{1}{4}\left(\frac{1}{v_2^2}\right)^2 v_2^2 \det\left(\frac{\partial \mathbf{x}_2}{\partial\theta_{2,i}}\cdot\frac{\partial \mathbf{x}_2}{\partial\theta_{2,j}}\right)\right]^{1/2}$$
$$= \frac{1}{2v_2}\left[\det\left(\frac{\partial \mathbf{x}_2}{\partial\theta_{2,i}}\cdot\frac{\partial \mathbf{x}_2}{\partial\theta_{2,j}}\right)\right]^{1/2}.$$

Observe that the integration of the determinant appearing in the above result, with respect to each angle $\theta_{2,j}$, gives the surface area of a $(M-2)$-dimensional sphere. The radius of this sphere is $\frac{v_2}{v_1}$, the length of $\|\mathbf{x}_2^{\perp}\|$ as observed in 3.12, and the area of its surface is therefore

$$A_{M-2}\left(\frac{v_2}{v_1}\right) = \frac{2\pi^{\frac{M-1}{2}}}{\Gamma\left(\frac{M-1}{2}\right)}\left(\frac{v_2}{v_1}\right)^{M-2}.$$

Thus, the expected value of the function g is now

$$E(g) = \int_{\mathbb{R}^{2M}} g(\Omega) f_{\hat{\mathcal{X}}}(X)\,|\mathrm{d}X| = C_1 \int_{\mathbb{R}^M}\left[\int_{\mathbb{R}^2} \frac{v_2^{M-3}}{v_1^{M-2}} g(\Omega) f_{\mathcal{X}}(X)\,\mathrm{d}\omega_{12}\mathrm{d}\omega_{22}\right]|\mathrm{d}\mathbf{x}_1|,$$

where $C_1 = \dfrac{\pi^{\frac{M-1}{2}}}{\Gamma\left(\frac{M-1}{2}\right)}$.

Consider the vector $\mathbf{x}_1$. Beginning with the parameterization in (3.3), the change of variables and integration of this vector will follow exactly as in the univariate case. Recalling that, $\sqrt{\omega_{11}} = v_1$, we obtain

$$E(g) = \int_{\mathbb{R}^{2M}} g(\Omega) f_{\hat{\mathcal{X}}}(X)\,|\mathrm{d}X|$$
$$= C_2 \int_{\mathbb{R}^3} \frac{v_2^{M-3}}{v_1^{M-2}} v_1^{M-2}\, g(\Omega) f_{\mathcal{X}}(X)\,\mathrm{d}\omega_{12}\mathrm{d}\omega_{22}\mathrm{d}\omega_{11}$$
$$= C_2 \int_{\mathbb{R}^3} v_2^{M-3}\, g(\Omega) f_{\hat{\mathcal{X}}}(X)\,\mathrm{d}\omega_{12}\mathrm{d}\omega_{22}\mathrm{d}\omega_{11},$$

where $C_2 = \dfrac{\pi^{\frac{2M-1}{2}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)}$. Then, we can simplify the presentation of the expected value of g, such that

$$E(g) = \int_{\mathbb{R}^3} g(\Omega) F_{M>2}(\Omega)|\mathrm{d}\Omega|,$$

where $F_{M>2}(\Omega)$ is the density function of the random matrix Ω, which can be expressed as

$$F_{M>2}(\Omega) = \frac{\pi^{-\frac{1}{2}}}{2^M\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)}(\det\Sigma)^{-\frac{M}{2}}(\det\Omega)^{\frac{M-3}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)}.$$

Thus, the random matrix $\mathbf{\Omega}$ has a Wishart distribution.

3.4 p - Variate

For the derivation of the Wishart distribution for a p-variate population, we will be sampling from our reduced probability measure (2.5)

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}}(\det\Sigma)^{\frac{M}{2}}}\mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)},$$

with

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{pmatrix} \in \mathbb{R}^{p\times M},$$

where we will assume the vectors $\mathbf{x}_k$ to be linearly independent. Again, we have that

$$\Omega = XX^T = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1\cdot\mathbf{x}_2 & \dots & \mathbf{x}_1\cdot\mathbf{x}_p \\ \mathbf{x}_2\cdot\mathbf{x}_1 & \|\mathbf{x}_2\|^2 & \dots & \mathbf{x}_2\cdot\mathbf{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_p\cdot\mathbf{x}_1 & \mathbf{x}_p\cdot\mathbf{x}_2 & \dots & \|\mathbf{x}_p\|^2 \end{pmatrix}$$
$$= \begin{pmatrix} \omega_{11} & \omega_{12} & \dots & \omega_{1p} \\ \omega_{21} & \omega_{22} & \dots & \omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \dots & \omega_{pp} \end{pmatrix}. \tag{3.13}$$

Consider the expected value, as defined in Appendix A Definition A.5, of the function $g(\mathbf{\Omega})$

$$E(g) = \int_{\mathbb{R}^{pM}} g(\Omega) f_{\hat{\mathcal{X}}}(X)|\mathrm{d}X|,$$

where $f_{\hat{\mathcal{X}}}(X)$ is the density function of the the random matrix $\hat{\mathcal{X}}$. Observe that for general p, the expected value involves pM integration variables. If we hold the vectors $\mathbf{x}_1, \dots, \mathbf{x}_{p-1}$ constant, we can integrate with respect to $\mathbf{x}_p$; then we integrate the resulting function with respect to $\mathbf{x}_{p-1}$, and so on. This application of Fubini's Theorem, with details given in Appendix H, allows us to compute multiple integrals by hand. Thus, the expectation of g can be viewed as

$$E(g) = \int_{\mathbb{R}^M} \cdots \left[\int_{\mathbb{R}^M} \cdots \left[\int_{\mathbb{R}^M} g(\Omega) f_{\hat{\mathcal{X}}}(X)|\mathrm{d}\mathbf{x}_p|\right] \cdots |\mathrm{d}\mathbf{x}_k|\right] \cdots |\mathrm{d}\mathbf{x}_1|. \tag{3.14}$$

We will now parameterize $\mathbf{x}_k$, for $1 \le k \le p$, using the following procedure. Consider a subspace $W \subset \mathbb{R}^M$, such that $\mathbb{R}^M = W \oplus W^\perp$, where $W = \mathrm{span}\{\mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}$ and $W^\perp$ is the orthogonal complement of W. Then, $\mathbf{x}_k$ can be expressed as

$$\mathbf{x}_k = \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l + \mathbf{x}_k^\perp, \tag{3.15}$$

where $\mathbf{x}_k^\perp \in W^\perp$ and each α_l are constant coefficients. Performing rotations about a unit vector $\mathbf{f} \in W^\perp$, that leave $W^\perp$ invariant, we obtain a unit $(M-k)$-dimensional sphere. Thus, any point on the sphere of radius $\|\mathbf{x}_k^\perp\|$ can be parameterized as $\|\mathbf{x}_k^\perp\| R(\Theta_k)\mathbf{f}$, where $R(\Theta_k)$ is a rotation matrix that leaves $W^\perp$ invariant.

Hence,

$$\mathbf{x}_k = \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l + \left\| \mathbf{x}_k^{\perp} \right\| R(\Theta_k)\mathbf{f}. \tag{3.16}$$

In the bi-variate case, we introduced formulas for volumes of 1 and 2-dimensional parallelograms. Now, we will define the volume of a k-parallelogram in $\mathbb{R}^M$.

Definition 3.3. (**Volume of a k-parallelogram in** $\mathbb{R}^M$). Let the k vectors $\mathbf{x}_1, \dots, \mathbf{x}_k$ be in $\mathbb{R}^M$, and let $B \in \mathbb{R}^{k \times M}$ be the matrix with these vectors as its rows: $B = (\mathbf{x}_1, \dots, \mathbf{x}_k)^T$. Then the *k-dimensional volume* of the parallelogram spanned by $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, denoted $P(\mathbf{x}_1, \dots, \mathbf{x}_k)$, is

$$v_k = \mathrm{Vol}_k P(\mathbf{x}_1, \dots, \mathbf{x}_k) = \left[\det\left(BB^T\right)\right]^{1/2}.$$

Theorem 3.4. *For a vector* $\mathbf{x}_k$*, as defined in* (3.15)

1. *The perpendicular distance from this vector to the subspace* $W = span(\mathbf{x}_1, \dots, \mathbf{x}_{k-1})$ *is*

$$\left\| \mathbf{x}_k^{\perp} \right\| = \frac{v_k}{v_{k-1}}$$

2. *Each coefficient,* α_l *for* $1 \le l \le k-1$*, must satisfy the following system*

$$\begin{pmatrix} \mathbf{x}_1 \cdot \mathbf{x}_1 & \mathbf{x}_2 \cdot \mathbf{x}_1 & \dots & \mathbf{x}_{k-1} \cdot \mathbf{x}_1 \\ \mathbf{x}_1 \cdot \mathbf{x}_2 & \mathbf{x}_2 \cdot \mathbf{x}_2 & \dots & \mathbf{x}_{k-1} \cdot \mathbf{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_1 \cdot \mathbf{x}_{k-1} & \mathbf{x}_2 \cdot \mathbf{x}_{k-1} & \dots & \mathbf{x}_{k-1} \cdot \mathbf{x}_{k-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k \cdot \mathbf{x}_1 \\ \mathbf{x}_k \cdot \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k \cdot \mathbf{x}_{k-1} \end{pmatrix}.$$

Proof. Consider the vector $\mathbf{x}_k$ as defined in (3.15), then

$$\mathbf{x}_k = \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l + \mathbf{x}_k^{\perp}.$$

Observe $W = \mathrm{span}(\mathbf{x}_1, \dots, \mathbf{x}_{k-1})$ and $\mathbf{x}_k^{\perp} \in W^{\perp}$. Then, for any $1 \le r \le k-1$,

$$0 = \mathbf{x}_k^{\perp} \cdot \mathbf{x}_r = (\mathbf{x}_k - \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l) \cdot \mathbf{x}_r, \tag{3.17}$$

and

$$\left\| \mathbf{x}_k^{\perp} \right\|^2 = \mathbf{x}_k^{\perp} \cdot (\mathbf{x}_k - \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l) = \mathbf{x}_k^{\perp} \cdot \mathbf{x}_k = (\mathbf{x}_k - \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l) \cdot \mathbf{x}_k. \tag{3.18}$$

From (3.17) and (3.18), we have a system of k linear equations

$$\sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l \cdot \mathbf{x}_r = \mathbf{x}_k \cdot \mathbf{x}_r, \tag{3.19}$$

$$\sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l \cdot \mathbf{x}_k + \left\| \mathbf{x}_k^{\perp} \right\|^2 = \mathbf{x}_k \cdot \mathbf{x}_k,$$

which can be written as

$$\begin{pmatrix} \mathbf{x}_1\cdot\mathbf{x}_1 & \mathbf{x}_2\cdot\mathbf{x}_1 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_1 & 0 \\ \mathbf{x}_1\cdot\mathbf{x}_2 & \mathbf{x}_2\cdot\mathbf{x}_2 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{x}_1\cdot\mathbf{x}_k & \mathbf{x}_2\cdot\mathbf{x}_k & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_k & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{k-1} \\ \left\|\mathbf{x}_k^{\perp}\right\|^2 \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k\cdot\mathbf{x}_1 \\ \mathbf{x}_k\cdot\mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k\cdot\mathbf{x}_k \end{pmatrix}. \tag{3.20}$$

We can determine $\left\|\mathbf{x}_k^{\perp}\right\|^2$ by the use of Cramer's rule, i.e.

$$\left\|\mathbf{x}_k^{\perp}\right\|^2 = \frac{\det\begin{pmatrix} \mathbf{x}_1\cdot\mathbf{x}_1 & \mathbf{x}_2\cdot\mathbf{x}_1 & \dots & \mathbf{x}_k\cdot\mathbf{x}_1 \\ \mathbf{x}_1\cdot\mathbf{x}_2 & \mathbf{x}_2\cdot\mathbf{x}_2 & \dots & \mathbf{x}_k\cdot\mathbf{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_1\cdot\mathbf{x}_k & \mathbf{x}_2\cdot\mathbf{x}_k & \dots & \mathbf{x}_k\cdot\mathbf{x}_k \end{pmatrix}}{\det\begin{pmatrix} \mathbf{x}_1\cdot\mathbf{x}_1 & \mathbf{x}_2\cdot\mathbf{x}_1 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_1 & 0 \\ \mathbf{x}_1\cdot\mathbf{x}_2 & \mathbf{x}_2\cdot\mathbf{x}_2 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{x}_1\cdot\mathbf{x}_k & \mathbf{x}_2\cdot\mathbf{x}_k & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_k & 1 \end{pmatrix}} = \left(\frac{\mathrm{Vol}_k P(\mathbf{x}_1,\dots,\mathbf{x}_k)}{\mathrm{Vol}_{k-1} P(\mathbf{x}_1,\dots,\mathbf{x}_{k-1})}\right)^2.$$

Thus, with the application of Definition (3.3), $\left\|\mathbf{x}_k^{\perp}\right\| = \dfrac{v_k}{v_{k-1}}$.

Recall the system of $k-1$ equations given in (3.19). Then, each coefficient, α_l, must satisfy the following

$$\begin{pmatrix} \mathbf{x}_1\cdot\mathbf{x}_1 & \mathbf{x}_2\cdot\mathbf{x}_1 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_1 \\ \mathbf{x}_1\cdot\mathbf{x}_2 & \mathbf{x}_2\cdot\mathbf{x}_2 & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_1\cdot\mathbf{x}_{k-1} & \mathbf{x}_2\cdot\mathbf{x}_{k-1} & \dots & \mathbf{x}_{k-1}\cdot\mathbf{x}_{k-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k\cdot\mathbf{x}_1 \\ \mathbf{x}_k\cdot\mathbf{x}_2 \\ \vdots \\ \mathbf{x}_k\cdot\mathbf{x}_{k-1} \end{pmatrix}.$$

□

From Theorem 3.4, we can express $\mathbf{x}_k$ as

$$\mathbf{x}_k = \sum_{l=1}^{k-1} \alpha_l \mathbf{x}_l + \frac{v_k}{v_{k-1}} R(\mathbf{\Theta}_k)\mathbf{f}. \tag{3.21}$$

Define $\Psi_k : \mathbb{R}^M \to \mathbb{R}^M$ as a map such that $\Psi_k(\mathbf{u}_k) = \mathbf{x}_k$. Specifically,

$$\Psi_k : \mathbf{u}_k = \begin{pmatrix} \omega_{1k} \\ \omega_{2k} \\ \vdots \\ \omega_{kk} \\ \theta_{k,1} \\ \vdots \\ \theta_{k,M-k} \end{pmatrix}^T \longrightarrow \mathbf{x}_k = \begin{pmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kM} \end{pmatrix}^T. \tag{3.22}$$

Consider the change of variables formula presented in (3.5). For the above map, the Jacobian will be

$$\left[\det\left(\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)^T\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)\right)\right]^{1/2} = \left[\det\left(\begin{array}{c|c} \dfrac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\omega_{rk}} & \dfrac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} \\ \hline \dfrac{\partial\mathbf{x}_k}{\partial\omega_{rk}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,i}} & \dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,i}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} \end{array}\right)\right]^{1/2},$$

for $1 \le l, r \le k$ and $1 \le i, j \le M-k$.

Theorem 3.5. *Consider the vector $\mathbf{x}_k$ as defined in* (3.21)*. For any $1 \le l \le k$ and $1 \le j \le M-k$, the following holds*

$$\frac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} = 0.$$

Proof. Consider the vector $\mathbf{x}_k$ as defined in (3.21),

$$\mathbf{x}_k = \sum_{r=1}^{k-1}\alpha_r\mathbf{x}_r + \frac{v_k}{v_{k-1}}R(\mathbf{\Theta}_k)\mathbf{f},$$

recall that $\dfrac{v_k}{v_{k-1}}R(\mathbf{\Theta}_k)\mathbf{f} = \mathbf{x}_k^{\perp}$. Observe,

$$\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} = \frac{\partial\mathbf{x}_k^{\perp}}{\partial\theta_{k,j}},$$

Since $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$ are independent of each ω_{lk} and $\dfrac{\partial\mathbf{x}_k^{\perp}}{\partial\omega_{lk}} \propto \mathbf{x}_k^{\perp}$

$$\frac{\partial\mathbf{x}_k}{\partial\omega_{lk}} = \sum_{r=1}^{k-1}\beta_r\mathbf{x}_r + \beta\mathbf{x}_k^{\perp},$$

where $\beta_r = \dfrac{\partial\alpha_r}{\partial\omega_{lk}}$. Considering the fact that the inner products, $\mathbf{x}_k^{\perp}\cdot\mathbf{x}_k^{\perp}$ and $\mathbf{x}_k\cdot\mathbf{x}_r$, are independent of each angle, $\theta_{k,j}$, we have

$$\frac{\partial\mathbf{x}_k^{\perp}}{\partial\theta_{k,j}}\cdot\mathbf{x}_k^{\perp} = 0, \quad \frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}}\cdot\mathbf{x}_r = 0,$$

for $1 \le r \le k-1$ and $1 \le j \le M-k$. Thus

$$\frac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} = \sum_{r=1}^{k-1}\beta_r\mathbf{x}_r\cdot\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} + \beta\mathbf{x}_k^{\perp}\cdot\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} = 0.$$

□

With the application of Theorem 3.5, we obtain the following simplified Jacobian

$$\left[\det\left(\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)^T\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)\right)\right]^{1/2} = \left[\det\left(\begin{array}{c|c} \dfrac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\omega_{rk}} & 0 \\ \hline 0 & \dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,i}}\cdot\dfrac{\partial\mathbf{x}_k}{\partial\theta_{k,j}} \end{array}\right)\right]^{1/2}.$$

Theorem 3.6. *Consider the vector $\mathbf{x}_k$ as previously defined. Then for any $1 \le l, r \le k$, we have*

$$\left[\det\left(\frac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\frac{\partial\mathbf{x}_k}{\partial\omega_{rk}}\right)\right]^{1/2} = \frac{1}{2v_k}.$$

Proof. Consider the vector $\mathbf{x}_k$ associated with the map from (3.22), $\Psi_k(\mathbf{u}_k) = \mathbf{x}_k$. As observed, the Jacobian of this map will be

$$\left[\det\left((\mathbf{D}\Psi_k(\mathbf{u}_k))^T (\mathbf{D}\Psi_k(\mathbf{u}_k))\right)\right]^{1/2},$$

where

$$(\mathbf{D}\Psi_k(\mathbf{u}_k)) = \begin{pmatrix} \dfrac{\partial \mathbf{x}_k}{\partial \omega_{1k}} & \cdots & \dfrac{\partial \mathbf{x}_k}{\partial \omega_{kk}} & \dfrac{\partial \mathbf{x}_k}{\partial \theta_{k,1}} & \cdots & \dfrac{\partial \mathbf{x}_k}{\partial \theta_{k,M-k}} \end{pmatrix},$$

and from Theorem 3.5,

$$(\mathbf{D}\Psi_k(\mathbf{u}_k))^T (\mathbf{D}\Psi_k(\mathbf{u}_k)) = \left(\begin{array}{c|c} \dfrac{\partial \mathbf{x}_k}{\partial \omega_{lk}} \cdot \dfrac{\partial \mathbf{x}_k}{\partial \omega_{rk}} & 0 \\ \hline 0 & \dfrac{\partial \mathbf{x}_k}{\partial \theta_{k,i}} \cdot \dfrac{\partial \mathbf{x}_k}{\partial \theta_{k,j}} \end{array}\right),$$

for $1 \leq l, r \leq k$ and $1 \leq i, j \leq M - k$. Consider

$$\left((\mathbf{D}\Psi_k(\mathbf{u}_k))^T (\mathbf{D}\Psi_k(\mathbf{u}_k)\right)^{-1} = (\mathbf{D}\Psi_k(\mathbf{u}_k))^{-1} \left((\mathbf{D}\Psi_k(\mathbf{u}_k))^{-1}\right)^T.$$

For a differentiable function $f(x_1, x_2, \ldots, x_n)$, we will define ∇f as the gradient of f, that is

$$\nabla f = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \dfrac{\partial f}{\partial x_2} & \cdots & \dfrac{\partial f}{\partial x_n} \end{pmatrix}.$$

Thus,

$$(\mathbf{D}\Psi_k(\mathbf{u}_k))^{-1} = \begin{pmatrix} \dfrac{\partial \mathbf{u}_k}{\partial x_{k1}} & \cdots & \dfrac{\partial \mathbf{u}_k}{\partial x_{kM}} \end{pmatrix} = \begin{pmatrix} \nabla \omega_{1k} \\ \vdots \\ \nabla \omega_{kk} \\ \nabla \theta_{k,1} \\ \vdots \\ \nabla \theta_{k,M-k} \end{pmatrix}.$$

Then our inverse Jacobian becomes,

$$\left((\mathbf{D}\Psi_k(\mathbf{u}_k))^T (\mathbf{D}\Psi_k(\mathbf{u}_k))\right)^{-1} = \left(\begin{array}{c|c} \nabla \omega_{lk} \cdot \nabla \omega_{rk} & 0 \\ \hline 0 & \nabla \theta_{k,i} \cdot \nabla \theta_{k,j} \end{array}\right).$$

Observe we have the following relationships

$$\omega_{lk} = \mathbf{x}_l \cdot \mathbf{x}_k.$$

Therefore,

$$\nabla \omega_{lk} = \begin{cases} \mathbf{x}_l & l \neq k \\ 2\mathbf{x}_k & l = k \end{cases},$$

which results in the following

$$\nabla\omega_{lk}\cdot\nabla\omega_{rk} = \begin{pmatrix} \mathbf{x}_1\cdot\mathbf{x}_1 & \mathbf{x}_1\cdot\mathbf{x}_2 & \dots & 2\mathbf{x}_1\cdot\mathbf{x}_k \\ \mathbf{x}_2\cdot\mathbf{x}_1 & \mathbf{x}_2\cdot\mathbf{x}_2 & \dots & 2\mathbf{x}_2\cdot\mathbf{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ 2\mathbf{x}_k\cdot\mathbf{x}_1 & 2\mathbf{x}_k\cdot\mathbf{x}_2 & \dots & 4\mathbf{x}_k\cdot\mathbf{x}_k \end{pmatrix}.$$

Observe $\det\left(\nabla\omega_{lk}\cdot\nabla\omega_{rk}\right) = \left(2\cdot\mathrm{Vol}_k(\mathbf{x}_1,\dots,\mathbf{x}_k)\right)^2 = (2v_k)^2$. Hence,

$$\left[\det\left(\frac{\partial\mathbf{x}_k}{\partial\omega_{lk}}\cdot\frac{\partial\mathbf{x}_k}{\partial\omega_{rk}}\right)\right]^{1/2} = \left[\det\left(\nabla\omega_{lk}\cdot\nabla\omega_{rk}\right)\right]^{-1/2} = \frac{1}{2v_k}. \qquad \square$$

Applying Theorem 3.6 we have

$$\left[\det\left(\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)^T\left(\mathbf{D}\Psi_k(\mathbf{u}_k)\right)\right)\right]^{1/2} = \frac{1}{2v_k}\left[\det\left(\frac{\partial\mathbf{x}_k}{\partial\theta_{k,i}}\cdot\frac{\partial\mathbf{x}_k}{\partial\theta_{k,j}}\right)\right]^{1/2}.$$

As observed in the bi-variate case, the integration of the of the determinant appearing in above result gives the surface area of a $(M-k)$-dimensional sphere. The radius of the sphere is $\frac{v_k}{v_{k-1}}$, the length of $\mathbf{x}_k^{\perp}$, and the area of its surface is therefore

$$A_{M-k}\left(\frac{v_k}{v_{k-1}}\right) = \frac{2\pi^{\frac{M-k+1}{2}}}{\Gamma\left(\frac{M-k+1}{2}\right)}\left(\frac{v_k}{v_{k-1}}\right)^{M-k}.$$

Hence the change of variables and integration of the $M-k$ angles from the k^{th}-variate produces

$$\frac{\pi^{\frac{M-k+1}{2}}}{\Gamma\left(\frac{M-k+1}{2}\right)}\frac{v_k^{M-k-1}}{v_{k-1}^{M-k}}|\mathrm{d}\omega_k|,$$

where we will define $|\mathrm{d}\omega_k| = \mathrm{d}\omega_{1k}\mathrm{d}\omega_{2k}\cdots\mathrm{d}\omega_{kk}$.

For the computation of our entire density function, we will begin by letting $k=p$ and fixing each vector $\mathbf{x}_1,\dots\mathbf{x}_{p-1}$. If we conduct the same change of variables and integration of the $M-p$ angles as we presented for general k, we will obtain

$$\frac{\pi^{\frac{M-p+1}{2}}}{\Gamma\left(\frac{M-p+1}{2}\right)}\frac{v_p^{M-p-1}}{v_{p-1}^{M-p}}|\mathrm{d}\omega_p|.$$

Thus, the expected value of g, as presented in (3.14), is

$$\begin{aligned} E(g) &= \int_{\mathbb{R}^{pM}} g(\Omega)f_{\hat{\mathcal{X}}}(X)|\mathrm{d}X| \\ &= C_1\int_{\mathbb{R}^M}\cdots\left[\int_{\mathbb{R}^M}\cdots\left[\int_{\mathbb{R}^p}\frac{v_p^{M-p-1}}{v_{p-1}^{M-p}}g(\Omega)f_{\hat{\mathcal{X}}}(X)|\mathrm{d}\omega_p|\right]|\mathrm{d}\mathbf{x}_{p-1}|\right]\cdots|\mathrm{d}\mathbf{x}_1|, \end{aligned}$$

where $C_1 = \dfrac{\pi^{\frac{M-p+1}{2}}}{\Gamma\left(\frac{M-p+1}{2}\right)}$.

Now consider the vector $\mathbf{x}_{p-1}$ and fix each vector $\mathbf{x}_1,\cdots,\mathbf{x}_{p-2}$. The change of variables and integration of the $M-p+1$ angles will produce

$$\frac{\pi^{\frac{M-(p-1)+1}{2}}}{\Gamma\left(\frac{M-(p-1)+1}{2}\right)}\frac{v_{p-1}^{M-(p-1)-1}}{v_{(p-1)-1}^{M-(p-1)}}|\mathrm{d}\omega_{p-1}|.$$

Observe, the expected value of g will be

$$
\begin{aligned}
E(g) &= \int_{\mathbb{R}^{pM}} g(\Omega) f_{\hat{\mathcal{X}}}(X)|\mathrm{d}X| \\
&= C_2 \int_{\mathbb{R}^M} \cdots \left[\int_{\mathbb{R}^M} \left[\int_{\mathbb{R}^{2p-1}} \frac{v_p^{M-p-1}}{v_{p-1}^{M-p}} \frac{v_{p-1}^{M-p}}{v_{p-2}^{M-p+1}} g(\Omega) f_{\hat{\mathcal{X}}}(X) |\mathrm{d}\omega_p||\mathrm{d}\omega_{p-1}| \right] |\mathrm{d}\mathbf{x}_{p-2}| \right] \cdots |\mathrm{d}\mathbf{x}_1|,
\end{aligned}
$$

where $C_2 = \dfrac{\pi^{\frac{M-p+2}{2}}}{\Gamma\left(\frac{M-p+2}{2}\right)} \dfrac{\pi^{\frac{M-p+1}{2}}}{\Gamma\left(\frac{M-p+1}{2}\right)}$.

We will continue to implement the same process, by considering the vector $\mathbf{x}_{p-2}$ and so on until the last vector $\mathbf{x}_1$. As observed in previous sections, the change of variables and integration of the angles from the parameterization of $\mathbf{x}_1$ will produce

$$
\frac{\pi^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}\right)} v_1^{M-2}.
$$

Thus,

$$
E(g) = \int_{\mathbb{R}^{pM}} g(\Omega) f_{\hat{\mathcal{X}}}(X)|\mathrm{d}X| = C_p \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \frac{v_p^{M-p-1}}{v_{p-1}^{M-p}} \frac{v_{p-1}^{M-p}}{v_{p-2}^{M-p+1}} \cdots \frac{v_2^{M-3}}{v_1^{M-2}} v_1^{M-2} g(\Omega) f_{\mathcal{X}}(X) |\mathrm{d}\Omega|
$$

where $C_p = \dfrac{\pi^{\frac{p(2M-p+1)}{4}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)\cdots\Gamma\left(\frac{M-p+1}{2}\right)}$. Observe that we will see almost a complete cancellation of parallelogram volumes, with only the largest volume, v_p^{M-p-1}, remaining.

Then, the expected value of the function g can be simplified to

$$
E(g) = \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} g(\Omega) F_{M>p}(\Omega) |\mathrm{d}\Omega|,
$$

where $F_{M>p}(\Omega)$ is the density function of the random matrix $\mathbf{\Omega}$, which can be expressed as

$$
F_{M>p}(\Omega) = \frac{\pi^{\frac{-p(p-1)}{4}}}{2^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{p} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Sigma)^{-\frac{M}{2}} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Omega)}. \tag{3.23}
$$

From the form of the density function in (3.23), the random matrix $\mathbf{\Omega}$ has a Wishart distribution.

3.5 Summary

As presented in this chapter, the distribution of the Wishart matrix can be obtained by a geometrical argument. After conducting a change of variable and computing the Jacobian of each transformation, we observed that integration with respect to each angle resulted in the surface area of a sphere. The radius of the sphere was a ratio of parallelogram volumes, which we proved by the use of Cramer's rule. Upon integration of the angles, there was an abundance of parallelogram volumes. After cancellations, the resulting volume was in fact the determinant of our Wishart matrix and we obtained the desired Wishart distribution.

Chapter 4

An Algebraic Derivation of the Wishart and Singular Wishart Distributions

4.1 Introduction

This chapter will present R. A. Janik and M. A. Nowak's [15] approach to deriving both the Wishart and singular Wishart distributions. We will introduce the Dirac delta in order to express the density function of the random matrix $\mathbf{\Omega}$. After simplifications, we demonstrate that the density function of $\mathbf{\Omega}$ satisfies a recurrence relation. The terms in the recurrence relation can be expressed as ratios of minors from the original matrix Ω. We took a different approach than the authors, by using the Lewis Carroll identity to prove the theorem. Dependent on the number of variates compared to the number of observations, the solution of the recurrence relation will be proportional to either the Wishart or singular Wishart distribution. The main outcome of this chapter will appear in Theorem 4.8, which we will now introduce. The notation appearing in the second result will be explained in more detail later in the chapter.

Consider the random matrix $\hat{\mathcal{X}}$, having the random vectors $\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_M \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ as its columns. We can simplify the presentation of the density functions by setting $\Sigma = I$, where an explanation is given at the beginning of section 4.2. Then, the density function of the random matrix $\mathbf{\Omega} = \hat{\mathcal{X}}\hat{\mathcal{X}}^T$ can take the following forms

1. When $M > p$, the random matrix $\mathbf{\Omega}$ has a Wishart distribution given by the density function

$$F_{M>p}(\Omega) = \frac{\pi^{\frac{-p(p-1)}{4}}}{2^{\frac{pM}{2}} \prod_{k=1}^{p} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}(\Omega)}.$$

2. When $M < p$, the random matrix $\mathbf{\Omega}$ has a singular Wishart distribution given by the density function

$$F_{M<p}(\Omega) = \frac{\pi^{\frac{M(M+1)}{4}}}{(2\pi)^{\frac{pM}{2}} \prod_{k=1}^{M} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega_{[M]})^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}\Omega} \prod_{\substack{l,r=M+1 \\ l\leq r}}^{p} \delta\left(\frac{\det \Omega_{[M]\cup\{l,r\}}}{\det \Omega_{[M]}}\right).$$

4.2 Derivation

The density function of the random matrix $\hat{\mathcal{X}}$, having the random vectors $\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_M \sim \mathcal{N}_p(\mathbf{0}, \Sigma)$ as its columns, is given by

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}}(\det\Sigma)^{\frac{M}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(X^T\Sigma^{-1}X)}.$$

As mentioned in the introduction of this chapter, we will introduce the Dirac delta in order to establish the density function of the Wishart matrix $\mathbf{\Omega}$. We will use the integral representation of the Dirac delta in our density function. After this, we will need to simplify our function in order to integrate with respect to our sample matrix X. For the purpose of simplifying these calculations, we will rescale the matrix X, that is we will make the transformation that $\tilde{X} = \Sigma^{-\frac{1}{2}}X$. In order to determine the Jacobian of this transformation, observe

$$\tilde{X} = \begin{pmatrix} \Sigma^{-\frac{1}{2}}\mathbf{y}_1^T & \Sigma^{-\frac{1}{2}}\mathbf{y}_2^T & \ldots & \Sigma^{-\frac{1}{2}}\mathbf{y}_M^T \end{pmatrix},$$

where $\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_M^T$ are the columns of X. Hence, the Jacobian of this transformation will be $(\det\Sigma)^{\frac{M}{2}}$. We will drop the tilde from the formula for simplicity, thus the density function of the random matrix $\hat{\mathcal{X}}$ becomes

$$f_{\hat{\mathcal{X}}}(X) = \frac{1}{(2\pi)^{\frac{pM}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}(XX^T)},$$

with

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_p \end{pmatrix} \in \mathbb{R}^{p\times M},$$

and

$$\Omega = XX^T = \begin{pmatrix} \|\mathbf{x}_1\|^2 & \mathbf{x}_1\cdot\mathbf{x}_2 & \ldots & \mathbf{x}_1\cdot\mathbf{x}_p \\ \mathbf{x}_2\cdot\mathbf{x}_1 & \|\mathbf{x}_2\|^2 & \ldots & \mathbf{x}_2\cdot\mathbf{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_p\cdot\mathbf{x}_1 & \mathbf{x}_p\cdot\mathbf{x}_2 & \ldots & \|\mathbf{x}_p\|^2 \end{pmatrix}$$

$$= \begin{pmatrix} \omega_{11} & \omega_{12} & \ldots & \omega_{1p} \\ \omega_{21} & \omega_{22} & \ldots & \omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{p1} & \omega_{p2} & \ldots & \omega_{pp} \end{pmatrix}.$$

Since we would like to obtain the density function for the random matrix $\mathbf{\Omega}$, we may incorporate the Dirac delta. Thus, the density function of a p-dimensional random matrix $\mathbf{\Omega}$ is given by

$$F(\Omega) = \int_{\mathbb{R}^{pM}} f_{\hat{\mathcal{X}}}(X)\delta(\Omega - XX^T)|\mathrm{d}X|,$$

where $\delta(\Omega - XX^T)$ is a multidimensional Dirac delta [6]. We will now introduce an integral representation, with detail given in Appendix D, for a multidimensional Dirac delta

$$\delta(\Omega - XX^T) = \frac{2^{\frac{p(p-1)}{2}}}{(2\pi)^{\frac{p(p+1)}{2}}} \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\mathrm{Tr}[T(\Omega - XX^T)]} |\mathrm{d}T|,$$

where T is a real symmetric $p \times p$ matrix. Inserting the integral representation of the Dirac delta into our density function we obtain

$$\begin{aligned} F(\Omega) &= C \int_{\mathbb{R}^{pM}} \left[\int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{Tr}(-\frac{1}{2}XX^T + \mathrm{i}T\Omega - \mathrm{i}TXX^T)} |\mathrm{d}T| \right] |\mathrm{d}X| \\ &= C \int_{\mathbb{R}^{pM}} \left[\int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}[(1+\mathrm{i}2T)XX^T]} \mathrm{e}^{\mathrm{i}\mathrm{Tr}(T\Omega)} |\mathrm{d}T| \right] |\mathrm{d}X|, \end{aligned}$$

where $C = \dfrac{2^{\frac{p(p-1)}{2}}}{(2\pi)^{\frac{p(p+1+M)}{2}}}$. Observe that with the application of Fubini's theorem (before we take each $N_j \to \infty$, as explained in Appendix D), we can fix T and integrate with respect to X, that is

$$F(\Omega) = C \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\mathrm{Tr}(T\Omega)} \left[\int_{\mathbb{R}^{pM}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}[(1+\mathrm{i}2T)XX^T]} |\mathrm{d}X| \right] |\mathrm{d}T|.$$

To eliminate the factor of 2, we will do a simple change of variable, $T \to 2T$

$$F(\Omega) = C' \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T\Omega)} \left[\int_{\mathbb{R}^{pM}} \mathrm{e}^{-\frac{1}{2}\mathrm{Tr}[(1+\mathrm{i}T)XX^T]} |\mathrm{d}X| \right] |\mathrm{d}T|,$$

where $C' = \dfrac{2^{-p}}{(2\pi)^{\frac{p(p+1+M)}{2}}}$. Observe that the integration with respect to X is a real Gaussian integral, which produces $(2\pi)^{pM/2} \det(1 + \mathrm{i}T)^{-\frac{M}{2}}$, as shown in Appendix C. Thus,

$$F(\Omega) = C'' \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \det(1 + \mathrm{i}T)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T\Omega)} |\mathrm{d}T|, \tag{4.1}$$

where $C'' = \dfrac{2^{-p}}{(2\pi)^{\frac{p(p+1)}{2}}}$. From this point forward we will conduct multiple integrations of the above $\dfrac{p(p+1)}{2}$ dimensional integral. We will denote this integral as

$$G_{p,M}(\Omega) = \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \det(1 + \mathrm{i}T)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T\Omega)} |\mathrm{d}T|. \tag{4.2}$$

Consider the following decomposition of the matrices T and Ω

$$T = \left(\begin{array}{c|c} t_{11} & \mathbf{t} \\ \hline \mathbf{t}^T & T_{p-1} \end{array} \right) \qquad \Omega = \left(\begin{array}{c|c} \omega_{11} & \boldsymbol{\omega} \\ \hline \boldsymbol{\omega}^T & \Omega_{p-1} \end{array} \right), \tag{4.3}$$

where t_{11} and ω_{11} are the entries from the first row and first column of each respective matrix, $\mathbf{t}$ and $\boldsymbol{\omega}$ are $(p-1)$-dimensional row vectors and the bottom right $(p-1) \times (p-1)$ submatrices are symmetric. Observe that the determinant of some matrix D, that is decomposed in the above way and has an invertible submatrix, can be expressed as follows

$$\det(D) = \det\left(\begin{array}{c|c} d_{11} & \mathbf{d} \\ \hline \mathbf{d}^T & D_{p-1} \end{array}\right)$$
$$= \det\left(\left(\begin{array}{c|c} 1 & \mathbf{d} \\ \hline \mathbf{0} & D_{p-1} \end{array}\right)\left(\begin{array}{c|c} d_{11} - \mathbf{d}D_{p-1}^{-1}\mathbf{d}^T & \mathbf{0} \\ \hline D_{p-1}^{-1}\mathbf{d}^T & I_{p-1} \end{array}\right)\right),$$

where I_{p-1} is a $(p-1)$-dimensional identity matrix. Thus

$$\det(D) = \det(D_{p-1}) \cdot \det(d_{11} - \mathbf{d}D_{p-1}^{-1}\mathbf{d}^T).$$

Observe the second term is the determinant of a 1-dimensional matrix

$$\det(D) = \det(D_{p-1}) \cdot (d_{11} - \mathbf{d}D_{p-1}^{-1}\mathbf{d}^T). \tag{4.4}$$

In our case, we can use this property to factor the matrix $1 + \mathrm{i}T$. Observe that the submatrix $1 + \mathrm{i}T_{p-1}$, is invertible. This can be shown by first diagonalizing T_{p-1}, that is

$$T_{p-1} = UDU^{-1},$$

for an orthogonal matrix U and diagonal matrix D. Then,

$$1 + \mathrm{i}T_{p-1} = 1 + \mathrm{i}UDU^{-1} = UU^{-1} + \mathrm{i}UDU^{-1} = U(1 + \mathrm{i}D)U^{-1},$$

observe that an inverse exists since $1 + \mathrm{i}D$ is a diagonal matrix with all nonzero entries. Thus, from the property in (4.4) we can rewrite $\det(1 + \mathrm{i}T)$, which gives

$$G_{p,M}(\Omega) = \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \left((1 + \mathrm{i}t_{11} + \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T) \det(1 + \mathrm{i}T_{p-1})\right)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T\Omega)} \mathrm{d}t_{11}|\mathrm{d}\mathbf{t}||\mathrm{d}T_{p-1}|.$$

Note that the trace of the product $T\Omega$ can be written as

$$\mathrm{Tr}(T\Omega) = \mathrm{Tr}\left(\begin{array}{c|c} t_{11}\omega_{11} + \mathbf{t}\boldsymbol{\omega}^T & t_{11}\boldsymbol{\omega} + \mathbf{t}\Omega_{p-1} \\ \hline \omega_{11}\mathbf{t}^T + T_{p-1}\boldsymbol{\omega}^T & \mathbf{t}^T\boldsymbol{\omega} + T_{p-1}\Omega_{p-1} \end{array}\right)$$
$$= t_{11}\omega_{11} + \mathbf{t}\boldsymbol{\omega}^T + \mathrm{Tr}(\mathbf{t}^T\boldsymbol{\omega} + T_{p-1}\Omega_{p-1})$$
$$= t_{11}\omega_{11} + \mathbf{t}\boldsymbol{\omega}^T + \boldsymbol{\omega}\mathbf{t}^T + \mathrm{Tr}(T_{p-1}\Omega_{p-1}). \tag{4.5}$$

Then, by rewriting the exponent involving $\mathrm{Tr}(T\Omega)$ and applying Fubini's Theorem, we have

$$G_{p,M}(\Omega) = \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \int_{\mathbb{R}^{p-1}} [\det(1 + \mathrm{i}T_{p-1})]^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T + \boldsymbol{\omega}\mathbf{t}^T)} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}\Omega_{p-1})} \times$$
$$\left[\int_{\mathbb{R}} (1 + \mathrm{i}t_{11} + \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}t_{11}\omega_{11}} \mathrm{d}t_{11}\right] |\mathrm{d}\mathbf{t}||\mathrm{d}T_{p-1}|.$$

Now, we would like to integrate with respect to t_{11}.

Lemma 4.1. *The following integral*

$$I = \int_{\mathbb{R}} (1 + \mathrm{i}t_{11} + \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}t_{11}\omega_{11}} \mathrm{d}t_{11},$$

is defined and evaluates to

$$I = \frac{2\pi}{\Gamma\left(\frac{M}{2}\right)} \left(\frac{\omega_{11}}{2}\right)^{\frac{M}{2}-1} \mathrm{e}^{-\frac{1}{2}\omega_{11}(1+\mathbf{t}(1+\mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)}.$$

Proof. In order to evaluate this integral we will use the following integral representation of the gamma function, with details given in Appendix E,

$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tw}(k + \mathrm{i}t)^{-v} \mathrm{d}t = \frac{2\pi}{\Gamma(v)} w^{v-1} \mathrm{e}^{-wk}, \tag{4.6}$$

where $Re\, k > 0$; $w > 0$; $Re\, v > 0$;. Consider the integral we want to evaluate

$$I = \int_{\mathbb{R}} (1 + \mathrm{i}t_{11} + \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)^{-\frac{M}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}t_{11}\omega_{11}} \mathrm{d}t_{11}.$$

With use of the property in (4.6), we have that

$$k = 1 + \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T, \quad w = \frac{\omega_{11}}{2}, \quad v = \frac{M}{2}.$$

Next, we will show the conditions are satisfied. Observe that $\omega_{11} = \|\mathbf{x}_1\|^2$ where $\mathbf{x} \neq \mathbf{0}$, thus $\omega_{11} > 0$ and so $w > 0$. Since $M \in \mathbb{Z}^+$, $Re\, v > 0$. For $Re\, k > 0$, consider

$$\begin{aligned}
\mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T &= \mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}(1 - \mathrm{i}T_{p-1})^{-1}(1 - \mathrm{i}T_{p-1})\mathbf{t}^T \\
&= \mathbf{t}((1 + \mathrm{i}T_{p-1})(1 - \mathrm{i}T_{p-1}))^{-1}(1 - \mathrm{i}T_{p-1})\mathbf{t}^T \\
&= \mathbf{t}(1 + T_{p-1}^2)^{-1}(1 - \mathrm{i}T_{p-1})\mathbf{t}^T \\
&= \mathbf{t}(1 + T_{p-1}^2)^{-1}\mathbf{t}^T - \mathrm{i}\mathbf{t}(1 + T_{p-1}^2)^{-1}T_{p-1}\mathbf{t}^T.
\end{aligned}$$

Now, we will diagonalize T_{p-1}, i.e.

$$T_{p-1} = UDU^{-1} \implies T_{p-1}^2 = UDU^{-1}UDU^{-1} = UD^2U^{-1}$$

where U is an orthogonal matrix and D is a diagonal matrix of the eigenvalues of T_{p-1}. Then,

$$(1 + T_{p-1}^2)^{-1} = (UU^{-1} + UD^2U^{-1})^{-1} = (U(1 + D^2)U^{-1})^{-1} = U(1 + D^2)^{-1}U^{-1}.$$

Since the above matrix has strictly positive eigenvalues, the matrix is positive definite. Thus

$$Re\,(\mathbf{t}(1 + \mathrm{i}T_{p-1})^{-1}\mathbf{t}^T) > 0.$$

Hence all of our conditions are satisfied, so our integral evaluates to

$$I = \frac{2\pi}{\Gamma\left(\frac{M}{2}\right)} \left(\frac{\omega_{11}}{2}\right)^{\frac{M}{2}-1} \mathrm{e}^{-\frac{1}{2}\omega_{11}(1+\mathbf{t}(1+\mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)}. \qquad \square$$

Observe with the application of the above Lemma, our integral $G_{p,M}(\Omega)$ is

$$G_{p,M}(\Omega) = \frac{2\pi}{\Gamma\left(\frac{M}{2}\right)}\left(\frac{\omega_{11}}{2}\right)^{\frac{M}{2}-1} \mathrm{e}^{-\frac{1}{2}\omega_{11}} \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \det(1+\mathrm{i}T_{p-1})^{-M/2}\mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}\Omega_{p-1})}\times$$
$$\left[\int_{\mathbb{R}^{p-1}} \mathrm{e}^{-\frac{1}{2}\omega_{11}(\mathbf{t}(1+\mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)+\frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T+\boldsymbol{\omega}\mathbf{t}^T)}|\mathrm{d}\mathbf{t}|\right]|\mathrm{d}T_{p-1}|.$$

From here, we would like to integrate with respect to the vector $\mathbf{t}$. We will show that this will be a Gaussian integral.

Lemma 4.2. *The integral*

$$I = \int_{\mathbb{R}^{p-1}} \mathrm{e}^{-\frac{1}{2}\omega_{11}(\mathbf{t}(1+\mathrm{i}T_{p-1})^{-1}\mathbf{t}^T)+\frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T+\boldsymbol{\omega}\mathbf{t}^T)}|\mathrm{d}\mathbf{t}|,$$

is Gaussian and is equal to

$$I = \pi^{\frac{p-1}{2}}\left(\frac{2}{\omega_{11}}\right)^{\frac{p-1}{2}} \det(1+iT_{p-1})^{1/2}e^{-\frac{1}{2\omega_{11}}\boldsymbol{\omega}(1+\mathrm{i}T_{p-1})\boldsymbol{\omega}^T}.$$

Proof. For simplicity, we will denote $B=(1+\mathrm{i}T_{p-1})^{-1}$, notice that $Re\,B$ is a symmetric positive definite matrix. Then, we have

$$I = \int_{\mathbb{R}^{p-1}} \mathrm{e}^{-\frac{1}{2}\mathrm{i}\omega_{11}(\mathbf{t}B\mathbf{t}^T)+\frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T+\boldsymbol{\omega}\mathbf{t}^T)}|\mathrm{d}\mathbf{t}|.$$

Observe that the exponent can be written as a shift, i.e.

$$I = C\int_{\mathbb{R}^{p-1}} \mathrm{e}^{-\frac{1}{2}\omega_{11}(\mathbf{t}+\mathbf{a})B(\mathbf{t}^T+\mathbf{a}^T)}|\mathrm{d}\mathbf{t}|, \tag{4.7}$$

where C is a normalization constant independent of $\mathbf{t}$ and the vector $\mathbf{a}$ is $p-1$ dimensional. So we need,

$$Ce^{-\frac{1}{2}\omega_{11}(\mathbf{t}B\mathbf{t}^T+\mathbf{t}B\mathbf{a}^T+\mathbf{a}B\mathbf{t}^T+\mathbf{a}B\mathbf{a}^T)} = \mathrm{e}^{-\frac{1}{2}\omega_{11}(\mathbf{t}B\mathbf{t}^T)+\frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T+\boldsymbol{\omega}\mathbf{t}^T)}.$$

Thus

$$C\cdot\mathrm{e}^{-\frac{1}{2}\omega_{11}\mathbf{a}B\mathbf{a}^T} = 1 \implies C = \mathrm{e}^{\frac{1}{2}\omega_{11}\mathbf{a}B\mathbf{a}^T},$$

and

$$-\frac{1}{2}\omega_{11}(\mathbf{t}B\mathbf{a}^T+\mathbf{a}B\mathbf{t}^T) = \frac{1}{2}\mathrm{i}(\mathbf{t}\boldsymbol{\omega}^T+\boldsymbol{\omega}\mathbf{t}^T),$$

which gives

$$-\omega_{11}\mathbf{a}B = \mathrm{i}\boldsymbol{\omega} \implies \mathbf{a} = -\frac{\mathrm{i}}{\omega_{11}}\boldsymbol{\omega}B^{-1},$$

and

$$C = \mathrm{e}^{-\frac{1}{2\omega_{11}}\boldsymbol{\omega}B^{-1}\boldsymbol{\omega}^T}.$$

Next, make a change of variable such that $\mathbf{x}=\sqrt{\frac{\omega_{11}}{2}}(\mathbf{t}+\mathbf{a})$. Then, after a shift to the real axis, our integral from (4.7) becomes

$$I = \mathrm{e}^{-\frac{1}{2\omega_{11}}\boldsymbol{\omega}B^{-1}\boldsymbol{\omega}^T}\int_{\mathbb{R}^{p-1}} \mathrm{e}^{-\mathbf{x}B\mathbf{x}^T}\left(\frac{2}{\omega_{11}}\right)^{\frac{p-1}{2}}|\mathrm{d}\mathbf{x}|.$$

Observe the resulting integral is Gaussian, as shown in Appendix C, which gives

$$
\begin{aligned}
I &= \mathrm{e}^{-\frac{1}{2\omega_{11}}\boldsymbol{\omega} B^{-1}\boldsymbol{\omega}^T} \left(\frac{2}{\omega_{11}}\right)^{\frac{p-1}{2}} \pi^{\frac{p-1}{2}} (\det B)^{-1/2} \\
&= \left(\frac{2}{\omega_{11}}\right)^{\frac{p-1}{2}} \pi^{\frac{p-1}{2}} \det(1+iT_{p-1})^{1/2} e^{-\frac{1}{2\omega_{11}}\boldsymbol{\omega}(1+\mathrm{i}T_{p-1})\boldsymbol{\omega}^T}. \qquad \square
\end{aligned}
$$

Making use of the Lemma 4.2, we obtain the following expression for $G_{p,M}(\Omega)$,

$$
G_{p,M}(\Omega) = C_{p,M} \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \omega_{11}^{\frac{M-p-1}{2}} \det(1+\mathrm{i}T_{p-1})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}\Omega_{p-1})} \mathrm{e}^{-\frac{1}{2}(\omega_{11}+\frac{\boldsymbol{\omega}\boldsymbol{\omega}^T}{\omega_{11}}+\frac{\mathrm{i}\boldsymbol{\omega} T_{p-1}\boldsymbol{\omega}^T}{\omega_{11}})} |\mathrm{d}T_{p-1}|.
$$

Observe that $\boldsymbol{\omega} T_{p-1}\boldsymbol{\omega}^T = \mathrm{Tr}(T_{p-1}\boldsymbol{\omega}^T\boldsymbol{\omega})$, then

$$
G_{p,M}(\Omega) = C_{p,M}\, \omega_{11}^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}(\omega_{11}+\frac{\boldsymbol{\omega}\boldsymbol{\omega}^T}{\omega_{11}})} \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \det(1+\mathrm{i}T_{p-1})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}(\Omega_{p-1}-\frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}))} |\mathrm{d}T_{p-1}|,
$$

where $C_{p,M} = \dfrac{2\pi}{\Gamma\left(\frac{M}{2}\right)} \pi^{\frac{p-1}{2}} 2^{\frac{p-M+1}{2}}$.

Theorem 4.3. *The function $G_{p,M}(\Omega)$ satisfies the following recurrence relation*

$$
G_{p,M}(\Omega) = C_{p,M}\, \omega_{11}^{\frac{M-p-1}{2}} \mathrm{e}^{-\omega_{11}-\frac{\boldsymbol{\omega}\boldsymbol{\omega}^T}{\omega_{11}}} G_{p-1,M-1}(\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}),
$$

where $C_{p,M} = \dfrac{2\pi}{\Gamma\left(\frac{M}{2}\right)} \pi^{\frac{p-1}{2}} 2^{\frac{p-M+1}{2}}$.

Proof. Observe that with the application of Lemma 4.1 and 4.2, we have

$$
G_{p,M}(\Omega) = C_{p,M}\, \omega_{11}^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}(\omega_{11}+\frac{\boldsymbol{\omega}\boldsymbol{\omega}^T}{\omega_{11}})} \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \det(1+\mathrm{i}T_{p-1})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}(\Omega_{p-1}-\frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}))} |\mathrm{d}T_{p-1}|,
$$

where $C_{p,M} = \dfrac{2\pi}{\Gamma\left(\frac{M}{2}\right)} \cdot \pi^{\frac{p-1}{2}} 2^{\frac{p-M+1}{2}}$. Notice that the integral appearing in the above expression is a function of a $p-1$ dimensional symmetric matrix $\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}$, having $p-1$ variables sampled $M-1$ times, with the first entry of the resulting matrix being strictly greater than zero. Thus, from the relationship given in (4.2), we have

$$
\int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \det(1+\mathrm{i}T_{p-1})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}(\Omega_{p-1}-\frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}))} |\mathrm{d}T_{p-1}| = G_{p-1,M-1}(\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}).
$$

Hence, we can express $G_{p,M}(\Omega)$ as a recurrence relation,

$$
G_{p,M}(\Omega) = C_{p,M}\, \omega_{11}^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}(\omega_{11}+\frac{\boldsymbol{\omega}\boldsymbol{\omega}^T}{\omega_{11}})} G_{p-1,M-1}(\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}). \tag{4.8}
$$

$\square$

We will now explicitly outline the steps for the solution to the recurrence relation. First we will introduce notation for the reduced matrix $\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}$, i.e.

$$
\Omega_{p-1} - \frac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}} = \Omega^{(p-1)}. \tag{4.9}
$$

Then the distribution of any $k \times k$ matrix of the $(p-k)^{th}$-step of the recursion, denoted by $\Omega^{(k)}$ and decomposed as such

$$\Omega^{(k)} = \left(\begin{array}{c|c} \omega_{11}^{(k)} & \boldsymbol{\omega}^{(k)} \\ \hline (\boldsymbol{\omega}^{(k)})^T & \Omega_{k-1}^{(k)} \end{array}\right),$$

is used to generate the next matrix in the recursion by the formula:

$$\Omega^{(k-1)} = \Omega_{k-1}^{(k)} - \frac{(\boldsymbol{\omega}^{(k)})^T \boldsymbol{\omega}^{(k)}}{\omega_{11}^{(k)}}.$$

Now we will outline a procedure to develop the solution to the recurrence relation.

1. We will begin at formula (4.2), then

$$G_{p-1,M-1}(\Omega^{(p-1)}) = \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \det(1+iT_{p-1})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-1}\Omega^{(p-1)})} |\mathrm{d}T_{p-1}|.$$

2. Now, we will decompose the matrices T_{p-1} and $\Omega^{(p-1)}$, as done in (4.3), that is

$$T_{p-1} = \left(\begin{array}{c|c} t_{11} & \mathbf{t} \\ \hline \mathbf{t}^T & T_{p-2} \end{array}\right) \qquad \Omega^{(p-1)} = \left(\begin{array}{c|c} \omega_{11}^{(p-1)} & \boldsymbol{\omega}^{(p-1)} \\ \hline (\boldsymbol{\omega}^{(p-1)})^T & \Omega_{p-2}^{(p-1)} \end{array}\right),$$

   where the row vectors $\mathbf{t}$ and $\boldsymbol{\omega}^{(p-1)}$ are $(p-2)$-dimensional and the superscript, $(p-1)$, denotes the 1^{st} step of the recursion. Since $\omega_{11}^{(p-1)}$ is a ratio of leading principal minors of our original matrix Ω, of which we will show later in Theorem 4.7, we have that $\omega_{11}^{(p-1)} > 0$.

3. From the property given in (4.4), we can rewrite the determinant appearing in $G_{p-1,M-1}(\Omega^{(p-1)})$ and the trace as observed in (4.5)

$$G_{p-1,M-1}(\Omega^{(p-1)}) = \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \left((1+\mathrm{i}t_{11}+\mathbf{t}(1+\mathrm{i}T_{p-2})^{-1}\mathbf{t}^T)\det(1+\mathrm{i}T_{p-2})\right)^{-\frac{M-1}{2}} \times$$
$$\mathrm{e}^{\frac{1}{2}\mathrm{i}(t_{11}\omega_{11}^{(p-1)}+\mathbf{t}(\boldsymbol{\omega}^{(p-1)})^T+\boldsymbol{\omega}^{(p-1)}\mathbf{t}^T+\mathrm{Tr}(T_{p-2}\Omega_{p-2}^{(p-1)}))}\mathrm{d}t_{11}|\mathrm{d}\mathbf{t}||\mathrm{d}T_{p-2}|.$$

4. Next, we would like to integrate with respect to t_{11}, thus we can apply Lemma 4.1

$$G_{p-1,M-1}(\Omega^{(p-1)}) = \frac{2\pi}{\Gamma\left(\frac{M-1}{2}\right)}\left(\frac{\omega_{11}^{(p-1)}}{2}\right)^{\frac{M-1}{2}-1} \mathrm{e}^{-\frac{1}{2}\omega_{11}^{(p-1)}} \times$$
$$\int_{\mathbb{R}^{\frac{(p-1)(p-2)}{2}}} \det(1+\mathrm{i}T_{p-2})^{-\frac{M-1}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-2}\Omega_{p-2}^{(p-1)})} \times$$
$$\left[\int_{\mathbb{R}^{p-2}} \mathrm{e}^{-\frac{1}{2}\omega_{11}^{(p-1)}(\mathbf{t}(1+\mathrm{i}T_{p-2})^{-1}\mathbf{t}^T)+\frac{1}{2}\mathrm{i}(\mathbf{t}(\boldsymbol{\omega}^{(p-1)})^T+\boldsymbol{\omega}^{(p-1)}\mathbf{t}^T)}|\mathrm{d}\mathbf{t}|\right]|\mathrm{d}T_{p-2}|.$$

5. Observe the integration with respect to $\mathbf{t}$ is Gaussian, so we will use Lemma 4.2

$$G_{p-1,M-1}(\Omega^{(p-1)}) = C_{p-1,M-1}\,(\omega_{11}^{(p-1)})^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\omega_{11}^{(p-1)}+\frac{\boldsymbol{\omega}^{(p-1)}(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}}\right)} \times$$
$$\int_{\mathbb{R}^{\frac{(p-1)(p-2)}{2}}} \det(1+\mathrm{i}T_{p-2})^{-\frac{M-2}{2}} \mathrm{e}^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_{p-2}\Omega^{(p-2)})}|\mathrm{d}T_{p-2}|,$$

   where $C_{p-1,M-1} = \dfrac{2\pi}{\Gamma\left(\frac{M-1}{2}\right)}\pi^{\frac{p-2}{2}}2^{\frac{p-M+1}{2}}$ and $\Omega^{(p-2)} = \Omega_{p-2}^{(p-1)} - \dfrac{(\boldsymbol{\omega}^{(p-1)})^T\boldsymbol{\omega}^{(p-1)}}{\omega_{11}^{(p-1)}}$.

6. Notice that the integral appearing in the above function describes $G_{p-1,M-1}(\Omega^{(p-1)})$ as a function of a reduced $(p-2)\times(p-2)$ matrix $\Omega^{(p-2)}$. Hence,

$$G_{p-1,M-1}(\Omega^{(p-1)}) = C_{p-1,M-1}\left(\omega_{11}^{(p-1)}\right)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\left(\omega_{11}^{(p-1)}+\frac{\boldsymbol{\omega}^{(p-1)}(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}}\right)} G_{p-2,M-2}(\Omega^{(p-2)}).$$

Then, we can write a more explicit expression for $G_{p,M}(\Omega)$,

$$G_{p,M}(\Omega) = C_2\left[\prod_{k=0}^{1}\omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{1}\left(\omega_{11}^{(p-k)}+\frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} G_{p-2,M-2}(\Omega^{(p-2)}),$$

where $C_2 = C_{p,M}C_{p-1,M-1}$.

7. Depending on the number of variables compared to the sample size, there will be two cases. The first, when $M > p$, will be the Wishart distribution as derived in the previous chapter. While the second, when $M < p$, will be the singular Wishart distribution.

Case 1 (Wishart): For $M > p$, we will consider $p-1$ steps of the recursion. Thus, we will repeat steps 1-6, $p-1$ times, giving

$$G_{p,M}(\Omega) = C_{p-1}\left[\prod_{k=0}^{p-2}\omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{p-2}\left(\omega_{11}^{(p-k)}+\frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} G_{1,M-(p-1)}(\Omega^{(1)}),$$

where $C_{p-1} = \prod_{k}^{p-2} C_{p-k,M-k}$. Now, from 4.2, we can express $G_{1,M-(p-1)}(\Omega^{(1)})$ as

$$G_{1,M-(p-1)}(\Omega^{(1)}) = \int_{\mathbb{R}} \det(1+\mathrm{i}T_1)^{-\frac{M-(p-1)}{2}} e^{\frac{1}{2}\mathrm{i}\mathrm{Tr}(T_1\Omega^{(1)})}\mathrm{d}T_1,$$

observe each matrix is 1-dimensional, i.e. $T_1 = t_{11}$ and $\Omega^{(1)} = \omega_{11}^{(1)}$, then

$$G_{1,M-(p-1)}(\Omega^{(1)}) = \int_{\mathbb{R}} (1+\mathrm{i}t_{11})^{-\frac{M-(p-1)}{2}} e^{\frac{1}{2}\mathrm{i}t_{11}\omega_{11}^{(1)}}\mathrm{d}t_{11}.$$

Now, if we consider Lemma 4.1, we can evaluate this integral, giving

$$G_{1,M-(p-1)}(\Omega^{(1)}) = \frac{2\pi}{\Gamma\left(\frac{M-(p-1)}{2}\right)}\left(\frac{\omega_{11}^{(1)}}{2}\right)^{\frac{M-(p-1)}{2}-1} \mathrm{e}^{-\frac{1}{2}\omega_{11}^{(1)}}.$$

Lemma 4.4. *When $M > p$, the function $G_{p,M}(\Omega)$ can be expressed as*

$$G_{p,M}(\Omega) = C_p\left[\prod_{k=0}^{p-1}\omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{p-2}\left(\omega_{11}^{(p-k)}+\frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} \mathrm{e}^{-\frac{1}{2}\omega_{11}^{(1)}},$$

where $C_p = \dfrac{(2\pi)^p\, 2^{\frac{p(p-M+1)}{2}}\, \pi^{\frac{p(p-1)}{4}}}{\Gamma\left(\frac{M}{2}\right)\cdots\Gamma\left(\frac{M-p+1}{2}\right)}$.

Case 2 (Singular Wishart): For $M < p$, we will repeat steps 1-6, M times, then

$$G_{p,M}(\Omega) = C_M \left[\prod_{k=0}^{M-1} \omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega_{11}^{(p-k)} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} G_{p-M,0}(\Omega^{(p-M)}),$$

where $C_M = \prod_{k=0}^{M-1} C_{p-k,M-k}$. Observe from (4.2), we have

$$G_{p-M,0}(\Omega^{(p-M)}) = \int_{\mathbb{R}^{\frac{(p-M)(p-M+1)}{2}}} \mathrm{e}^{\frac{1}{2}\mathrm{iTr}(T_{p-M}\Omega^{(p-M)})} |\mathrm{d}T_{p-M}|, \tag{4.10}$$

where T_{p-M} is all real $(p-M)\times(p-M)$ symmetric matrices and the determinant has disappeared, thus the integral is not convergent in any ordinary sense. We will show that the above function is proportional to the Dirac delta.

Recall from Appendix D, the multidimensional Dirac delta can be expressed as

$$\delta(\Omega^{(p-M)}) = \frac{2^{\frac{(p-M)(p-M-1)}{2}}}{(2\pi)^{\frac{(p-M)(p-M+1)}{2}}} \int_{\mathbb{R}^{\frac{(p-M)(p-M+1)}{2}}} \mathrm{e}^{\mathrm{iTr}(T\Omega^{(p-M)})} |\mathrm{d}T|,$$

where T is all real $(p-M)$-dimensional symmetric matrices. Then,

$$\frac{(2\pi)^{\frac{(p-M)(p-M+1)}{2}}}{2^{\frac{(p-M)(p-M-1)}{2}}}\delta(\Omega^{(p-M)}) = \int_{\mathbb{R}^{\frac{(p-M)(p-M+1)}{2}}} \mathrm{e}^{\mathrm{iTr}(T\Omega^{(p-M)})} |\mathrm{d}T|.$$

We will rescale T such that $T = \frac{1}{2}\hat{T}$,

$$\begin{aligned}\frac{(2\pi)^{\frac{(p-M)(p-M+1)}{2}}}{2^{\frac{(p-M)(p-M-1)}{2}}}\delta(\Omega^{(p-M)}) &= 2^{-\frac{(p-M)(p-M+1)}{2}} \int_{\mathbb{R}^{\frac{(p-M)(p-M+1)}{2}}} \mathrm{e}^{\frac{1}{2}\mathrm{iTr}(\hat{T}\Omega^{(p-M)})} |\mathrm{d}\hat{T}| \\ &= 2^{-\frac{(p-M)(p-M+1)}{2}} G_{p-M,0}(\Omega^{(p-M)}).\end{aligned}$$

Thus,

$$G_{p-M,0}(\Omega^{(p-M)}) = \frac{(2\pi)^{\frac{(p-M)(p-M+1)}{2}}}{2^{\frac{(p-M)(p-M-1)}{2}}} 2^{\frac{(p-M)(p-M+1)}{2}} \delta(\Omega^{(p-M)}).$$

From this representation, we have that

$$G_{p,M} = C'_M \left[\prod_{k=0}^{M-1} \omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega_{11}^{(p-k)} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} \delta\left(\Omega^{(p-M)}\right),$$

where $C'_M = \dfrac{\pi^{M+\frac{M(2p-M-1)}{4}+\frac{(p-M)(p-M+1)}{2}} 2^{p+\frac{p(p-M+1)}{2}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)\cdots\Gamma\left(\frac{1}{2}\right)}$.

Lemma 4.5. *When $M < p$, the function $G_{p,M}(\Omega)$ can be expressed as*

$$G_{p,M}(\Omega) = C'_M \left[\prod_{k=0}^{M-1} \omega_{11}^{(p-k)}\right]^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega_{11}^{(p-k)} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega_{11}^{(p-k)}}\right)} \delta\left(\Omega^{(p-M)}\right),$$

where $C'_M = \dfrac{\pi^{M+\frac{M(2p-M-1)}{4}+\frac{(p-M)(p-M+1)}{2}} 2^{p+\frac{p(p-M+1)}{2}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)\cdots\Gamma\left(\frac{1}{2}\right)}$.

Now, we would like to write each distribution in terms of the original matrix Ω. Interestingly enough, we can express each element of a reduced matrix, $\Omega^{(k)}$, as ratios of minors from the original matrix Ω. The proof of this can be done by induction, which we will show. For now, let us dive into more notation.

Definition 4.6. We will define $\Omega_{[k]}$, for $1 \leq k \leq M$ when $p > M$ and $1 \leq k \leq p$ when $p \leq M$, as the upper left hand $k \times k$ submatrix of Ω. For each $l, r > k$ consider the $(k+1) \times (k+1)$ matrix, denoted $\Omega_{[k]\cup\{l,r\}}$, obtained by adjoining the $k+1$ entries of the l^{th} row and r^{th} column of Ω to the submatrix $\Omega_{[k]}$, i.e.

$$\Omega_{[k]\cup\{l,r\}} = \left(\begin{array}{ccc|c} & & & \omega_{1r} \\ & \Omega_{[k]} & & \vdots \\ & & & \omega_{kr} \\ \hline \omega_{l1} & \dots & \omega_{lk} & \omega_{lr} \end{array}\right).$$

When $k = 0$, we will define $\det \Omega_{[0]} = 1$ and $\det \Omega_{[0]\cup\{l,r\}} = \omega_{lr}$.

We will begin by introducing some assumptions about the matrix Ω. Observe when $M > p$ we have a Gram matrix, thus Ω is positive definite. When $p > M$, our matrix is only semi-positive definite, but since the $\operatorname{rank}(\Omega) = M$, we will have an M-by-M block that is positive definite. We will assume that this block exists in the upper left hand corner of our matrix, i.e. $\Omega_{[M]}$ is positive definite. Then, all leading principal minors of $\Omega_{[M]}$ are positive definite.

Theorem 4.7. *Any element of the matrix $\Omega^{(p-k)}$, from the k^{th} step of the recurrence relation, can be written as a ratio of minors from the original matrix Ω, i.e.*

$$\omega_{lr}^{(p-k)} = \frac{\det \Omega_{[k]\cup\{l+k,r+k\}}}{\det \Omega_{[k]}},$$

where $0 \leq k \leq M$ when $p > M$ and $0 \leq k \leq p$ when $p \leq M$. Also, $l, r > k$.

Proof. For $k = 0$ the statement in true, since we have

$$\omega_{lr}^{(p)} = \frac{\det \Omega_{[0]\cup\{l,r\}}}{\det \Omega_{[0]}} = \omega_{lr},$$

where by definition $\det \Omega_{[0]} = 1$. Assume the relationship stands for $k-1$ and we will verify for k. Recall the relationship from (4.9)

$$\Omega^{(k-1)} = \Omega_{k-1}^{(k)} - \frac{(\boldsymbol{\omega}^{(k)})^T \boldsymbol{\omega}^{(k)}}{\omega_{11}^{(k)}}.$$

Hence,

$$\Omega^{(p-k)} = \Omega_{p-k}^{(p-k+1)} - \frac{(\boldsymbol{\omega}^{(p-k+1)})^T \boldsymbol{\omega}^{(p-k+1)}}{\omega_{11}^{(p-k+1)}}.$$

Then, each element of the matrix can be viewed as

$$\omega_{lr}^{(p-k)} = \left(\Omega_{p-k}^{(p-k+1)} - \frac{(\boldsymbol{\omega}^{(p-k+1)})^T \boldsymbol{\omega}^{(p-k+1)}}{\omega_{11}^{(p-k+1)}}\right)_{lr}. \tag{4.11}$$

Leading to an expression for each element

$$\omega_{lr}^{(p-k)} = \omega_{l+1,r+1}^{(p-k+1)} - \frac{\omega_{l+1,1}^{(p-k+1)}\omega_{1,r+1}^{(p-k+1)}}{\omega_{11}^{(p-k+1)}}, \tag{4.12}$$

where the shift in l and r comes from the fact the element in the l^{th} row and r^{th} column of the new matrix is made up from the $l+1$ row and $r+1$ column of the previous matrix, i.e. $\omega_{11}^{(p-k)} = \omega_{22}^{(p-k+1)} - \frac{\omega_{21}^{(p-k+1)}\omega_{12}^{(p-k+1)}}{\omega_{11}^{(p-k+1)}}$. Also, we will introduce a comma in some situations in order to communicate the appropriate row and column of Ω that we are interested in.

Recall, we assumed the relationship to be true for $k-1$, so the right hand side of (4.12) is

$$\begin{aligned}
\omega_{l+1,r+1}^{(p-k+1)} - \frac{\omega_{l+1,1}^{(p-k+1)}\omega_{1,r+1}^{(p-k+1)}}{\omega_{11}^{(p-k+1)}} &= \frac{\det \Omega_{[k-1]\cup\{l+k,r+k\}}}{\det \Omega_{[k-1]}} - \frac{\det \Omega_{[k-1]\cup\{l+k,k\}}}{\det \Omega_{[k-1]}} \times \\
&\quad \frac{\det \Omega_{[k-1]\cup\{k,r+k\}}}{\det \Omega_{[k-1]}} \cdot \frac{\det \Omega_{[k-1]}}{\det \Omega_{[k-1]\cup\{k,k\}}} \\
&= \frac{\det \Omega_{[k-1]\cup\{l+k,r+k\}}}{\det \Omega_{[k-1]}} - \frac{\det \Omega_{[k-1]\cup\{l+k,k\}} \det \Omega_{[k-1]\cup\{k,r+k\}}}{\det \Omega_{[k-1]} \det \Omega_{[k]}} \\
&= C\left(\det \Omega_{[k]} \det \Omega_{[k-1]\cup\{l+k,r+k\}} - \det \Omega_{[k-1]\cup\{l+k,k\}} \det \Omega_{[k-1]\cup\{k,r+k\}}\right),
\end{aligned}$$

where $C = \dfrac{1}{\det \Omega_{[k-1]} \det \Omega_{[k]}}$. Thus we want to prove the following identity

$$\det \Omega_{[k-1]} \det \Omega_{[k]\cup\{l+k,r+k\}} = \det \Omega_{[k]} \det \Omega_{[k-1]\cup\{l+k,r+k\}} - \det \Omega_{[k-1]\cup\{l+k,k\}} \det \Omega_{[k-1]\cup\{k,r+k\}}.$$

We will show that with some shifting of rows and columns we can apply the Lewis Carroll identity, as presented in Appendix G. Observe

$$\det \Omega_{[k]\cup\{l+k,r+k\}} = \det \left(\begin{array}{ccc|cc}
 & & & \omega_{1k} & \omega_{1,r+k} \\
 & \Omega_{[k-1]} & & \vdots & \vdots \\
 & & & \omega_{k-1,k} & \omega_{k-1,r+k} \\
\hline
\omega_{k1} & \cdots & \omega_{k,k-1} & \omega_{kk} & \omega_{k,r+k} \\
\omega_{l+k,1} & \cdots & \omega_{l+k,k-1} & \omega_{l+k,k} & \omega_{l+k,r+k}
\end{array}\right).$$

Now we would like to move the last row to the first row and the last column to the first column, thus we are doing k row interchanges and k column interchanges

$$\begin{aligned}
\det \Omega_{[k]\cup\{l+k,r+k\}} &= (-1)^{2k} \det \left(\begin{array}{c|ccc|c}
\omega_{l+k,r+k} & \omega_{l+k,1} & \cdots & \omega_{l+k,k-1} & \omega_{l+k,k} \\
\hline
\omega_{1,r+k} & & & & \omega_{1k} \\
\vdots & & \Omega_{[k-1]} & & \vdots \\
\omega_{k-1,r+k} & & & & \omega_{k-1,k} \\
\hline
\omega_{k,r+k} & \omega_{k,1} & \cdots & \omega_{k,k-1} & \omega_{kk}
\end{array}\right) \qquad (4.13) \\
&= \det D.
\end{aligned}$$

In order to apply the Lewis Carroll identity, we will develop some more notation. We will denote the submatrix from which the $i_1, i_2, \ldots, i_n$ rows and $j_1, j_2, \ldots, j_n$ columns of $D \in \mathbb{R}^{k+1 \times k+1}$ are removed by $D^{i_1,i_2,\ldots,i_n}_{j_1,j_2,\ldots,j_n}$. Then the identity states

$$\det D \cdot \det D^{1,k+1}_{1,k+1} = \det D^1_1 \cdot \det D^{k+1}_{k+1} - \det D^{k+1}_1 \cdot \det D^1_{k+1}.$$

Consider $\Omega_{[k-1]}$, from (4.13) we see that this is the matrix D but with the first and last row removed as well as the first and last column, so $\det \Omega_{[k-1]} = \det D^{1,k+1}_{1,k+1}$. Also, $\det \Omega_{[k]} = \det \Omega_{[k-1]\cup\{k,k\}} = \det D^1_1$. Observe, if we move the last row and column of $\Omega_{[k-1]\cup\{l+k,r+k\}}$ to the first row and column and take the determinant of the resulting matrix, we will obtain $(-1)^{2(k-1)} \det D^{k+1}_{k+1}$. Also, by moving the last row of $\Omega_{[k-1]\cup\{l+k,k\}}$ to the first row and taking the determinant, this is equivalent to $(-1)^{k-1} \det D^{k+1}_1$. Similarly, by moving the last column of $\Omega_{[k-1]\cup\{k,r+k\}}$ to the first column, we have $\det \Omega_{[k-1]\cup\{k,r+k\}} = (-1)^{k-1} \det D^1_{k+1}$. Thus,

$$\begin{aligned}
\det \Omega_{[k-1]} \det \Omega_{[k]\cup\{l+k,r+k\}} &= \det D^{1,k+1}_{1,k+1} \cdot \det D \\
&= \det D^1_1 \cdot \det D^{k+1}_{k+1} - \det D^{k+1}_1 \cdot \det D_{k+1} \\
&= \det \Omega_{[k]} \cdot \det \Omega_{[k-1]\cup\{l+k,r+k\}} - \\
&\quad (-1)^{-(k-1)} \det \Omega_{[k-1]\cup\{l+k,k\}} \cdot (-1)^{-(k-1)} \det \Omega_{[k-1]\cup\{k,r+k\}} \\
&= \det \Omega_{[k]} \det \Omega_{[k-1]\cup\{l+k,r+k\}} - \det \Omega_{[k-1]\cup\{l+k,k\}} \det \Omega_{[k-1]\cup\{k,r+k\}}.
\end{aligned}$$

Hence,

$$\omega^{(p-k)}_{lr} = \frac{\det \Omega_{[k]\cup\{l+k,r+k\}}}{\det \Omega_{[k]}}. \qquad \square$$

From the above Theorem, we can express the terms in both of our distributions as ratios of determinants. Consider,

$$\omega^{(p-k)}_{11} = \frac{\det \Omega_{[k]\cup\{1+k,1+k\}}}{\det \Omega_{[k]}} = \frac{\det \Omega_{[k+1]}}{\det \Omega_{[k]}}.$$

Then, for the product from $G_{p,M}(\Omega)$, given in Lemma 4.4,

$$\begin{aligned}
\prod_{k=0}^{p-1} \omega^{p-k}_{11} &= \prod_{k=0}^{p-1} \frac{\det \Omega_{[k+1]}}{\det \Omega_{[k]}} \\
&= \frac{\det \Omega_{[1]}}{\det \Omega_{[0]}} \cdot \frac{\det \Omega_{[2]}}{\det \Omega_{[1]}} \cdots \frac{\det \Omega_{[p-1]}}{\det \Omega_{[p-2]}} \cdot \frac{\det \Omega_{[p]}}{\det \Omega_{[p-1]}} = \frac{\det \Omega_{[p]}}{\det \Omega_{[0]}} \\
&= \det \Omega_{[p]},
\end{aligned}$$

and, for the product from $G_{p,M}(\Omega)$, given in Lemma 4.5,

$$\prod_{k=0}^{M-1} \omega^{(p-k)}_{11} = \prod_{k=0}^{M-1} \frac{\det \Omega_{[k+1]}}{\det \Omega_{[k]}} = \frac{\det \Omega_{[M]}}{\det \Omega_{[0]}} = \det \Omega_{[M]}.$$

Next, we will show that the expression in the exponent that appears in both distributions is the trace of the original matrix Ω. Recall we have the relationship in (4.9), from which we will take the trace of both sides

$$\begin{aligned}\operatorname{Tr}\Omega^{(k-1)} &= \operatorname{Tr}\left(\Omega^{(k)}_{k-1} - \frac{(\boldsymbol{\omega}^{(k)})^T\boldsymbol{\omega}^{(k)}}{\omega^{(k)}_{11}}\right)\\ &= \operatorname{Tr}\Omega^{(k)}_{k-1} - \frac{\boldsymbol{\omega}^{(k)}(\boldsymbol{\omega}^{(k)})^T}{\omega^{(k)}_{11}},\end{aligned}$$

since $\operatorname{Tr}((\boldsymbol{\omega}^{(k)})^T\boldsymbol{\omega}^{(k)}) = \boldsymbol{\omega}^{(k)}(\boldsymbol{\omega}^{(k)})^T$. Observe from the decomposition of $\Omega^{(k)}$, $\operatorname{Tr}\Omega^{(k)} = \omega^{(k)}_{11} + \operatorname{Tr}\Omega^{(k)}_{k-1}$

$$\operatorname{Tr}\Omega^{(k-1)} = \operatorname{Tr}\Omega^{(k)} - \omega^{(k)}_{11} - \frac{\boldsymbol{\omega}^{(k)}(\boldsymbol{\omega}^{(k)})^T}{\omega^{(k)}_{11}}.$$

Thus

$$\omega^{(k)}_{11} + \frac{\boldsymbol{\omega}^{(k)}(\boldsymbol{\omega}^{(k)})^T}{\omega^{(k)}_{11}} = \operatorname{Tr}\Omega^{(k)} - \operatorname{Tr}\Omega^{(k-1)}.$$

For the Wishart case, we have

$$-\frac{1}{2}\sum_{k=0}^{p-2}\left(\omega^{(p-k)}_{11} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega^{(p-k)}_{11}}\right) - \frac{1}{2}\omega^{(1)}_{11} = -\frac{1}{2}\sum_{k=0}^{p-2}\left(\operatorname{Tr}\Omega^{(p-k)} - \operatorname{Tr}\Omega^{(p-k-1)}\right) - \frac{1}{2}\omega^{(1)}_{11}.$$

Observe the right hand side is an alternating sum, thus we will only be left with the first and last term, that is

$$\begin{aligned}-\frac{1}{2}\sum_{k=0}^{p-2}\left(\omega^{(p-k)}_{11} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega^{(p-k)}_{11}}\right) - \frac{1}{2}\omega^{(1)}_{11} &= -\frac{1}{2}(\operatorname{Tr}\Omega - \operatorname{Tr}\Omega^{(1)}) - \frac{1}{2}\omega^{(1)}_{11}\\ &= -\frac{1}{2}\operatorname{Tr}\Omega,\end{aligned}$$

since $\Omega^{(1)} = \omega^{(1)}_{11}$.

Then, for the singular Wishart case,

$$-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega^{(p-k)}_{11} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega^{(p-k)}_{11}}\right) = -\frac{1}{2}\sum_{k=0}^{M-1}\left(\operatorname{Tr}\Omega^{(p-k)} - \operatorname{Tr}\Omega^{(p-k-1)}\right).$$

which again is an alternating sum, thus

$$-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega^{(p-k)}_{11} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega^{(p-k)}_{11}}\right) = -\frac{1}{2}\left(\operatorname{Tr}\Omega - \operatorname{Tr}\Omega^{(p-M)}\right).$$

From the Dirac delta, the second trace vanishes

$$-\frac{1}{2}\sum_{k=0}^{M-1}\left(\omega^{(p-k)}_{11} + \frac{\boldsymbol{\omega}^{(p-k)}(\boldsymbol{\omega}^{(p-k)})^T}{\omega^{(p-k)}_{11}}\right) = -\frac{1}{2}\operatorname{Tr}\Omega.$$

We will now obtain an expression for the mulitdimensional Dirac delta, $\delta(\Omega^{(p-M)})$, that appears in Lemma 4.5. Recall that $\Omega^{(p-M)}$ is a $(p-M)$ dimensional symmetric matrix with $\frac{(p-M)(p-M+1)}{2}$ independent elements. We will express each independent entry of this matrix as a ratio of determinants. Thus for any l, r

$$\omega^{(p-M)}_{lr} = \frac{\det\Omega_{[M]\cup\{l+M,r+M\}}}{\det\Omega_{[M]}},$$

resulting in

$$\delta(\Omega^{(p-M)}) = \prod_{\substack{l,r=M+1 \\ l\le r}}^{p} \delta\left(\frac{\det \Omega_{[M]\cup\{l,r\}}}{\det \Omega_{[M]}}\right).$$

In the following theorem we will present both the Wishart and singular Wishart distribution, taking into account the previous findings.

Theorem 4.8. *The density function of the random matrix* $\mathbf{\Omega}$ *can take the following forms*

1. *When* $M > p$*, the random matrix* $\mathbf{\Omega}$ *has a Wishart distribution given by the density function*

$$F_{M>p}(\Omega) = \frac{\pi^{\frac{-p(p-1)}{4}}}{2^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{p} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}(\Omega)}.$$

2. *When* $M < p$*, the random matrix* $\mathbf{\Omega}$ *has a singular Wishart distribution given by the density function*

$$F_{M<p}(\Omega) = \frac{\pi^{\frac{M(M+1)}{4}}}{(2\pi)^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{M} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega_{[M]})^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}\Omega} \prod_{\substack{l,r=M+1 \\ l\le r}}^{p} \delta\left(\frac{\det \Omega_{[M]\cup\{l,r\}}}{\det \Omega_{[M]}}\right).$$

Proof. Recall from (4.1) we have that

$$F(\Omega) = \frac{2^{-p}}{(2\pi)^{\frac{p(p+1)}{2}}} G_{p,M}(\Omega).$$

1. As observed in Lemma 4.4, when $p \le M$, $G_{p,M}(\Omega)$ can be expressed as

$$G_{p,M}(\Omega) = \frac{(2\pi)^p \, 2^{\frac{p(p-M+1)}{2}} \pi^{\frac{p(p-1)}{4}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)\cdots\Gamma\left(\frac{M-p+1}{2}\right)} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}\Omega}.$$

   Thus,

$$F_{M>p}(\Omega) = \frac{\pi^{\frac{-p(p-1)}{4}}}{2^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{p} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega)^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}(\Omega)}.$$

2. As observed in Lemma 4.5, when $p > M$, we have that

$$G_{p,M}(\Omega) = \frac{\pi^{M+\frac{M(2p-M-1)}{4}+\frac{(p-M)(p-M+1)}{2}} 2^{p+\frac{p(p-M+1)}{2}}}{\Gamma\left(\frac{M}{2}\right)\Gamma\left(\frac{M-1}{2}\right)\cdots\Gamma\left(\frac{1}{2}\right)} (\det \Omega_{[M]})^{\frac{M-p-1}{2}} e^{-\frac{1}{2}\operatorname{Tr}\Omega} \prod_{\substack{l,r=M+1 \\ l\le r}}^{p} \delta\left(\frac{\det \Omega_{[M]\cup\{l,r\}}}{\det \Omega_{[M]}}\right).$$

   Thus,

$$F_{M<p}(\Omega) = \frac{\pi^{\frac{M(M+1)}{4}}}{(2\pi)^{\frac{pM}{2}} \displaystyle\prod_{k=1}^{M} \Gamma\left(\frac{M-k+1}{2}\right)} (\det \Omega_{[M]})^{\frac{M-p-1}{2}} \mathrm{e}^{-\frac{1}{2}\operatorname{Tr}\Omega} \prod_{\substack{l,r=M+1 \\ l\le r}}^{p} \delta\left(\frac{\det \Omega_{[M]\cup\{l,r\}}}{\det \Omega_{[M]}}\right). \qquad \square$$

4.3 Summary

R. A. Janik and M. A. Nowak's approach to the derivation of the distributions required intricate calculations. With the application of the integral representation of the Dirac delta and integration of the data matrix X, we observed that the density function of the random matrix $\mathbf{\Omega}$ satisfied a recurrence relation. The steps of the recurrence relation were outlined explicitly, which produced a complicated expression of terms. In fact, the elements of a reduced matrix at any step of the recurrence could be expressed by the ratio of minors from the original matrix Ω. This result allowed a simplification of the density function and we were able to prove that the Wishart matrix had one of two density functions. For the case of $p < M$ the random matrix had a Wishart distribution and for $p > M$ the random matrix had a singular Wishart distribution.

Chapter 5

Concluding Remarks and Future Work

The derivations provided throughout this thesis were approached in entirely different manners. The work of J. Wishart uses a geometric argument, while R. A. Janik and M. A. Nowak use an algebraic approach. These publications were of interest to us because we wanted to find different ways of approaching the same problem. It is interesting to discover that two completely different arguments can be used to derive identical distributions.

One remaining topic of interest is the applications of the distributions in multivariate analysis. We introduced two applications of the distributions in statistics, such as the need for the Wishart distribution in the derivation of the generalized T^2 distribution and the application of the singular Wishart distribution in Bayesian statistics. But, the open question is how can the singular Wishart distribution be applied to high-dimensional data analysis? For statistical methods that require assumptions about the covariance matrix, specifically discriminant analysis, is there a need for knowing such a distribution?

In the introduction of this thesis, we offered two examples of high-dimensional problems, such as analyzing security returns in financial analysis [21] and classifying tumors using DNA microarray experiments [7]. The use of microarrays is becoming more prevelant in the area of biomedical research. Commonly in microarray experiments, there may be an abundance of variables, leading to an enormous covariance matrix [9]. When considering DNA microarray data, there is availability of gene expressions on thousands of genes, but very few individuals. Even though the genes are correlated, most of the analysis carried out on the data ignore these correlations [29]. Learning more about analysis in this field and how the singular Wishart distribution could be applied is a compelling topic for future research.

References

[1] T. W. Anderson. *An introduction to multivariate statistical analysis*, chapter 2,3,5,7. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2003.

[2] Jushan Bai and Shuzhong Shi. Estimating high dimensional covariance matrices and its applications. *Columbia University Academic Commons*, 2011.

[3] Richard Beals and Roderick Wong. *Special functions*, volume 126 of *Cambridge Studies in Advanced Mathematics*, chapter 2. Cambridge University Press, Cambridge, 2010. A graduate text.

[4] Tim Bollerslev, Robert F Engle, and Daniel B Nelson. *Handbook of Econometrics, Volume IV*, chapter 49, pages 2961–2967. North Holland, 1994.

[5] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*, chapter 1. Springer Science & Business Media, 2006.

[6] Santanu Chakraborty. Some applications of Diracs delta function in statistics for more than one random variable. *Applications and Applied Mathematics*, 3(1):2008, 2008.

[7] Sandrine Dudoit, Jane Fridlyand, and Terence P Speed. Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97(457):77–87, 2002.

[8] Warren J. Ewens and Gregory R. Grant. *Statistical methods in bioinformatics: an introduction*, chapter 2. Statistics for Biology and Health. Springer-Verlag, New York, 2001.

[9] Jianqing Fan and Runze Li. Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *arXiv preprint math/0602133*, 2006.

[10] Israel M Gelfand and GE Shilov. *Generalized functions volume I: Properties and operations*, chapter 2, pages 34–38. Academic Press, New York, 1964.

[11] Roger A Horn and Charles R Johnson. *Matrix analysis*, chapter 3, pages 216–217. Cambridge University Press, 2012.

[12] John H Hubbard and Barbara B Hubbard. *Vector calculus, linear algebra, and differential forms: A unified approach*. Matrix Editions, 2009.

[13] S. N. Roy I. Olkin. On multivariate distribution theory. *The Annals of Mathematical Statistics*, 25(2):329–339, 1954.

[14] A. T. James. Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics*, 25(1):40–75, 1954.

[15] Romuald A Janik and Maciej A Nowak. Wishart and anti-Wishart random matrices. *Journal of Physics A: Mathematical and General*, 36(12):3629, 2003.

[16] Christian Krattenthaler. Advanced determinant calculus. In *The Andrews Festschrift*, pages 349–426. Springer, 2001.

[17] Tatsuya Kubokawa and Muni S. Srivastava. Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *Journal of Multivariate Analysis*, 99(9):1906–1928, 2008.

[18] Mei-Ling Ting Lee. *Analysis of microarray gene expression data*, chapter 2,3. Springer Science & Business Media, 2007.

[19] Edgard M Maboudou-Tchao and Vincent Agboto. Monitoring the covariance matrix with fewer observations than variables. *Computational Statistics & Data Analysis*, 64:99–112, 2013.

[20] Robb J. Muirhead. *Aspects of multivariate statistical theory.* John Wiley & Sons, Inc., New York, 1982. Wiley Series in Probability and Mathematical Statistics.

[21] CR Narayanaswamy and D Raghavarao. Principal component analysis of large dispersion matrices. *Applied Statistics*, pages 309–316, 1991.

[22] Pavel Okunev and Charles R Johnson. Necessary and sufficient conditions for existence of the LU factorization of an arbitrary matrix. *arXiv preprint math/0506382*, 2005.

[23] C. Radhakrishna Rao. *Linear statistical inference and its applications*, chapter 8. John Wiley & Sons, Inc., New York-London-Sydney, 1965.

[24] Sheldon M. Ross. *Introduction to probability models*, chapter 1,2. Elsevier/Academic Press, Amsterdam, ninth edition, 2014.

[25] Clement Joseph Savant and Ezra C Levy. *Fundamentals of the Laplace transformation*, chapter 2,5. McGraw-Hill, 1962.

[26] Thomas Schmelzer and Lloyd N Trefethen. Computing the Gamma function using contour tntegrals and rational approximations. *SIAM Journal on Numerical Analysis*, 45(2):558–571, 2007.

[27] Michael Gambier Smith. *Laplace transform theory.* D. Van Norstrand Company LTD, 1966.

[28] Muni S Srivastava. Singular Wishart and multivariate beta distributions. *The Annals of Statististics*, 31(5):1537–1560, 2003.

[29] Muni S Srivastava. Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, 35(2):251–272, 2005.

[30] Muni S Srivastava. Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, 37(1):53–86, 2007.

[31] Muni S Srivastava and Yasunori Fujikoshi. Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, 2006.

[32] Harald Uhlig. On singular Wishart and singular multivariate Beta distributions. *The Annals of Statistics*, pages 395–405, 1994.

[33] John Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, pages 32–52, 1928.

[34] Charlie C Xiang and Yidong Chen. cDNA microarray technology and its applications. *Biotechnology Advances*, 18(1):35–46, 2000.

[35] Soonyu Yu, Jaena Ryu, and Kyoohong Park. A derivation of anti-wishart distribution. *Journal of Multivariate Analysis*, 131:121 – 125, 2014.

[36] Yi-Kuo Yu and Yi-Cheng Zhang. On the anti-Wishart distribution. *Physica A: Statistical Mechanics and its Applications*, 312(1):1–22, 2002.

Appendix A

Basics of Probability

The following definitions on basic concepts in probability were obtained from S. M. Ross [24] and J. H. Hubbard et. al [12].

Definition A.1. The *sample space* S of an experiment contains all possible outcomes of the experiment. A subset H of S ($H \subset S$) is called an *event*.

Definition A.2. Let S be the sample space of outcomes of an experiment. A *random variable* is a function $X : S \to \mathbb{R}$.

Definition A.3. A random variable X is said to be a *continuous random variable* if there exists a nonnegative function $f(x)$, defined for all real $x \in \mathbb{R}$, having the property that for any set of B of real numbers

$$\text{Prob}(X \in B) = \int_B f(x)\mathrm{d}x.$$

Definition A.4. The function $f(x)$ appearing in Definition A.3 is called the *probability density function* of the random variable X.

Definition A.5. For a continuous random variable X having a probability density function $f(x)$, the *expected value* of X is defined by

$$E(X) = \int_{\mathbb{R}} x f(x)\mathrm{d}x.$$

Definition A.6. The *variance* of a random variable X, denoted $Var(X)$, is given by the formula

$$Var(X) = E\big((X - E(X))^2\big) = E(X^2) - (E(X))^2.$$

Definition A.7. The *standard deviation* of a random variable X, denoted $SD(X)$, is given by

$$SD(X) = \sqrt{Var(X)}.$$

Definition A.8. If X and Y are random variables, their *covariance*, denoted $Cov(X,Y)$, is

$$Cov(X,Y) = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big).$$

Definition A.9. Let X and Y be random variables. Their *correlation coefficient*, denoted $corr(X,Y)$, is given by

$$corr(X,Y) = \frac{Cov(X,Y)}{SD(X)SD(Y)}$$

It follows that $|corr(X,Y)| \leq 1$.

Definition A.10. A *random vector* $\mathbf{X} = \big(X_1, X_2, \ldots, X_p\big)^T$ is a vector of jointly distributed random variables. The expectation of a random vector, denoted $E(\mathbf{X})$, is given by

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix}.$$

Definition A.11. The *covariance matrix*, denoted Σ, of a random vector $\mathbf{X}$ is

$$\begin{aligned}\Sigma &= E\Big(\big(\mathbf{X}-E(\mathbf{X})\big)\big(\mathbf{X}-E(\mathbf{X})\big)^T\Big)\\ &= \begin{pmatrix} Var(X_1) & Cov(X_1,X_2) & \cdots & Cov(X_1,X_p)\\ Cov(X_2,X_1) & Var(X_2) & \cdots & Cov(X_2,X_p)\\ \vdots & \vdots & \ddots & \vdots\\ Cov(X_p,X_1) & Cov(X_p,X_2) & \dots & Var(X_p)\end{pmatrix}.\end{aligned}$$

Definition A.12. Let B be a $p\times p$ square matrix. The matrix B is called *symmetric* if the (i,j) entry is equal to the (j,i) entry

$$b_{ij}=b_{ji}\quad 1\le i,j\le p.$$

Definition A.13. Let B be a symmetric $p\times p$ real matrix. Then the matrix B is said to be *positive semi-definite* if $\mathbf{z}^TB\mathbf{z}\ge 0$ for every non-zero column vector $\mathbf{z}\in\mathbb{R}^p$.

Lemma A.14. *If Σ is the covariance matrix of a random vector $\mathbf{X}$, then Σ is positive semi-definite.*

Proof. Observe Σ is symmetric since

$$Cov(X_i,X_j)=Cov(X_j,X_i)\quad 1\le i,j\le p.$$

Consider a non-zero column vector $\mathbf{z}\in\mathbb{R}^p$

$$\begin{aligned}\mathbf{z}^T\Sigma\mathbf{z} &= \mathbf{z}^TE\Big(\big(\mathbf{X}-E(\mathbf{X})\big)\big(\mathbf{X}-E(\mathbf{X})\big)\Big)\mathbf{z}^T\\ &= E\Big(\mathbf{z}^T\big(\mathbf{X}-E(\mathbf{X})\big)\big(\mathbf{X}-E(\mathbf{X})\big)^T\mathbf{z}\Big)\\ &= E\big(YY^T\big)\ge 0,\end{aligned}$$

where $Y=\mathbf{z}^T\big(\mathbf{X}-E(\mathbf{X})\big)$. $\square$

Definition A.15. Let X be a $p\times N$ matrix of p random variables each sampled N times

$$X=[x_{kn}]\quad 1\le k\le p,\quad 1\le n\le N.$$

The *sample mean* of the k^{th}-variate, denoted $\bar{x}_k$ is given by

$$\bar{x}_k=\frac{1}{N}\sum_{n=1}^{N}x_{kn}.$$

Definition A.16. Let X be a $p\times N$ matrix of p random variables each sampled N times. The *sample covariance matrix* is a $p\times p$ matrix A with entries

$$a_{kl}=\frac{1}{N}\sum_{n=1}^{N}(x_{kn}-\bar{x}_k)(x_{ln}-\bar{x}_l)\quad 1\le k,l\le p.$$

Appendix B

Surface Area of Spheres

Observe,

$$\int_{\mathbb{R}} e^{-x^2}\mathrm{d}x = \sqrt{\pi}.$$

Then,

$$\left(\int_{\mathbb{R}} e^{-x^2}\mathrm{d}x\right)^2 = \int_{\mathbb{R}} e^{-x^2}\mathrm{d}x \int_{\mathbb{R}} e^{-y^2}\mathrm{d}y,$$

using Fubini's Theorem, with details given in Appendix H, we have

$$= \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\mathrm{d}(x,y),$$

let $r^2 = x^2 + y^2$

$$\begin{aligned} &= \int_0^\infty e^{-r^2} A_1(1)\, r\, \mathrm{d}r \\ &= A_1(1) \int_0^\infty e^{-r^2} r\, \mathrm{d}r, \end{aligned}$$

where $A_1(1)$ is the surface area of a unit 1-dimensional sphere or, similarly, the circumference of the unit circle. Thus,

$$A_1(1) = \frac{\left(\int_{\mathbb{R}} e^{-x^2} dx\right)^2}{\int_0^\infty e^{-r^2} r\, \mathrm{d}r}.$$

To compute the denominator, we will institute a change of variable. That is,

$$u = -r^2 \qquad\qquad \mathrm{d}u = -2\, \mathrm{d}r.$$

Then,

$$\int_0^\infty e^{-r^2} r\, \mathrm{d}r = \frac{-1}{2} \int_0^\infty e^u\, \mathrm{d}u = \frac{-1}{2}(e^\infty - e^0) = \frac{1}{2}.$$

So,

$$A_1(1) = \frac{\pi}{1/2} = 2\pi.$$

As expected, this is exactly the circumference of a unit circle. Consider

$$\left(\int_{\mathbb{R}} e^{-x^2}\mathrm{d}x\right)^3 = \int_{\mathbb{R}^3} e^{-(x^2+y^2+z^2)}\mathrm{d}(x,y,z)$$

let $r^2 = x^2 + y^2 + z^2$

$$\begin{aligned} &= \int_0^\infty e^{-r^2} A_2(1)\, r^2\, \mathrm{d}r \\ &= A_2(1) \int_0^\infty e^{-r^2}\, r^2\, \mathrm{d}r, \end{aligned}$$

where $A_2(1)$ is the surface area of a 2-dimensional unit sphere. Observe

$$A_2(1) = \frac{\left(\int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x\right)^3}{\int_0^\infty e^{-r^2} r^2\,\mathrm{d}r}.$$

For the computation of the denominator, apply integration by parts. That is,

$$w = r \qquad \mathrm{d}v = re^{-r^2}\mathrm{d}r$$
$$\mathrm{d}w = \mathrm{d}r \qquad v = \frac{-1}{2}e^{-r^2}.$$

Thus,

$$\begin{aligned}\int_0^\infty e^{-r^2} r^2\,\mathrm{d}r &= -\frac{r}{2}e^{-r^2}\Big|_0^\infty + \frac{1}{2}\int_0^\infty e^{-r^2}\mathrm{d}r\\ &= \frac{1}{2}\int_0^\infty e^{-r^2}\mathrm{d}r,\end{aligned}$$

let $r = \sqrt{t} \implies dr = \frac{1}{2}t^{\frac{1}{2}-1}\,\mathrm{d}t$

$$= \frac{1}{4}\int_0^\infty t^{\frac{1}{2}-1}e^{-t}\,\mathrm{d}t,$$

which is the integral representation of the Gamma function

$$= \frac{1}{4}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{4}.$$

The surface area of the 2-dimensional unit sphere is therefore,

$$A_2(1) = \frac{(\sqrt{\pi})^3}{\sqrt{\pi}/4} = 4\pi.$$

Then, for any natural number N, we have

$$\begin{aligned}A_{N-1}(1) &= \frac{\left(\int_{\mathbb{R}} e^{-x^2}dx\right)^N}{\int_0^\infty e^{-r^2} r^{N-1}\,dr}\\ &= \frac{(\sqrt{\pi})^N}{\Gamma\left(\frac{N}{2}\right)/2}\\ &= \frac{2(\sqrt{\pi})^N}{\Gamma\left(\frac{N}{2}\right)}.\end{aligned}$$

Appendix C

Gaussian Integrals

Recall from Appendix B that we have

$$\int_{\mathbb{R}} \mathrm{e}^{-x^2} dx = \sqrt{\pi}.$$

Then for any variable $a > 0$, independent of x, consider

$$\int_{\mathbb{R}} \mathrm{e}^{-ax^2} dx,$$

we will make a change of variable such that $\tilde{x} = \sqrt{a}x$, then

$$\begin{aligned}
\int_{\mathbb{R}} \mathrm{e}^{-ax^2} dx &= \int_{\mathbb{R}} \mathrm{e}^{-\tilde{x}^2} \frac{1}{\sqrt{a}} \mathrm{d}\tilde{x} \\
&= \sqrt{\frac{\pi}{a}}.
\end{aligned}$$

Now we will consider the 2-dimensional case, that is

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-(x^2+y^2)} \mathrm{d}x\mathrm{d}y.$$

With the application of Fubini's theorem, where details are given in Appendix H, we have

$$\begin{aligned}
\int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-(x^2+y^2)} \mathrm{d}x\mathrm{d}y &= \int_{\mathbb{R}} \mathrm{e}^{-y^2} \left[\int_{\mathbb{R}} \mathrm{e}^{-x^2} \mathrm{d}x \right] \mathrm{d}y \\
&= \sqrt{\pi} \int_{\mathbb{R}} \mathrm{e}^{-y^2} \mathrm{d}y \\
&= (\sqrt{\pi})^2.
\end{aligned}$$

Thus, for general k we have

$$\int_{\mathbb{R}^k} \mathrm{e}^{-(x_1^2+x_2^2+\cdots+x_k^2)} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_k = (\sqrt{\pi})^k.$$

Now, consider

$$I = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-(a_{11}x^2+a_{12}xy+a_{22}y^2)} \mathrm{d}x\mathrm{d}y.$$

In order to integrate with respect to x, we will fix y and complete the square in x

$$I = \int_{\mathbb{R}} \mathrm{e}^{-(a_{22}y^2)} \left[\int_{\mathbb{R}} \mathrm{e}^{-a_{11}(x+\frac{a_{12}}{a_{11}}y)^2+\frac{a_{12}^2}{a_{11}}y^2} \mathrm{d}x \right] \mathrm{d}y,$$

now, make a change of variable such that $\tilde{x} = \sqrt{a_{11}}(x + \frac{a_{12}}{a_{11}}y)$, then

$$\begin{aligned}
I &= \int_{\mathbb{R}} \mathrm{e}^{-a_{22}y^2} \mathrm{e}^{\frac{a_{12}^2}{a_{11}}y^2} \left[\int_{\mathbb{R}} \mathrm{e}^{-\tilde{x}^2} \frac{1}{\sqrt{a_{11}}} \mathrm{d}\tilde{x} \right] \mathrm{d}y \\
&= \sqrt{\frac{\pi}{a_{11}}} \int_{\mathbb{R}} \mathrm{e}^{-(a_{22}-\frac{a_{12}^2}{a_{11}})y^2}.
\end{aligned}$$

Then make the final change of variable, $\tilde{y} = \sqrt{a_{22} - \frac{a_{12}^2}{a_{11}}} y$

$$\begin{aligned} I &= \sqrt{\frac{\pi}{a_{11}}} \int_{\mathbb{R}} \mathrm{e}^{-\tilde{y}^2} \frac{1}{\sqrt{a_{22} - \frac{a_{12}^2}{a_{11}}}} \mathrm{d}\tilde{y} \\ &= \frac{(\sqrt{\pi})^2}{\sqrt{a_{11}}} \sqrt{\frac{a_{11}}{a_{22}a_{11} - a_{12}^2}} \\ &= \sqrt{\frac{\pi^2}{a_{22}a_{11} - a_{12}^2}}. \end{aligned}$$

Observe that if we write $\mathbf{x} = (x, y)$ and $A = [a_{lr}] \in \mathbb{R}^{2\times 2}$ such that A is a symmetric positive definite matrix, we have

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 = \mathbf{x}A\mathbf{x}^T.$$

Then

$$\int_{\mathbb{R}^2} \mathrm{e}^{-\mathbf{x}A\mathbf{x}^T} |\mathrm{d}x| = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathrm{e}^{-(a_{11}x^2 + a_{12}xy + a_{22}y^2)} \mathrm{d}x\mathrm{d}y = \sqrt{\frac{\pi^2}{\det A}}.$$

Lemma C.1. *For any vector $\mathbf{x} \in \mathbb{R}^{1\times k}$ and symmetric positive definite matrix $A = [a_{lr}] \in \mathbb{R}^{k\times k}$, the following holds*

$$\int_{\mathbb{R}^k} \mathrm{e}^{-\mathbf{x}A\mathbf{x}^T} |\mathrm{d}\mathbf{x}| = \sqrt{\frac{\pi^k}{\det A}}.$$

Proof. Since A is a symmetric matrix, we will diagonalize A such that

$$A = UDU^T,$$

where U is an orthogonal matrix and D a diagonal matrix. Then,

$$\begin{aligned} \int_{\mathbb{R}^k} \mathrm{e}^{-\mathbf{x}A\mathbf{x}^T} |\mathrm{d}\mathbf{x}| &= \int_{\mathbb{R}^k} \mathrm{e}^{-\mathbf{x}UDU^T\mathbf{x}^T} |\mathrm{d}x| \\ &= \int_{\mathbb{R}^k} \mathrm{e}^{-(\mathbf{x}U)D(\mathbf{x}U)^T} |\mathrm{d}\mathbf{x}|. \end{aligned}$$

We will make a change of variables such that $\tilde{\mathbf{x}} = \mathbf{x}U$, thus

$$\begin{aligned} \int_{\mathbb{R}^k} \mathrm{e}^{-\mathbf{x}A\mathbf{x}^T} |\mathrm{d}\mathbf{x}| &= \int_{\mathbb{R}^k} \mathrm{e}^{-\tilde{\mathbf{x}}D\tilde{\mathbf{x}}^T} |\det U||\mathrm{d}\tilde{\mathbf{x}}| \\ &= \int_{\mathbb{R}^k} \mathrm{e}^{-(x_1^2 d_1 + x_2^2 d_2 + \cdots + x_k^2 d_2)} \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_k \\ &= \sqrt{\frac{\pi^k}{d_1 d_2 \cdots d_k}}. \end{aligned}$$

Observe that

$$\det A = \det(ODO^T) = \det(O)\det(D)\det(O^T) = \det(D) = d_1 d_2 \cdots d_k.$$

Hence,

$$\int_{\mathbb{R}^k} \mathrm{e}^{-\mathbf{x}A\mathbf{x}^T} |\mathrm{d}\mathbf{x}| = \sqrt{\frac{\pi^k}{\det A}}.$$

□

Theorem C.2. *For a matrix $X \in \mathbb{R}^{k\times M}$ and symmetric postive definite matrix $A \in \mathbb{R}^{k\times k}$, the integral*

$$I = \int_{\mathbb{R}^{kM}} e^{-\mathrm{Tr}(X^T AX)}|\mathrm{d}X|$$

is Gaussian and evaluates to

$$I = (\pi)^{\frac{kM}{2}}(\det A)^{-\frac{M}{2}}.$$

Proof. We will denote the columns of X as

$$X = \begin{pmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \dots & \mathbf{y}_M^T \end{pmatrix},$$

and the columns of A

$$A = \begin{pmatrix} \mathbf{a}_1^T & \mathbf{a}_2^T & \dots & \mathbf{a}_k^T \end{pmatrix}.$$

Consider the product appearing in the exponent

$$\begin{aligned} X^T AX &= \begin{pmatrix} \mathbf{y}_1\mathbf{a}_1^T & \mathbf{y}_1\mathbf{a}_2^T & \dots & \mathbf{y}_1\mathbf{a}_k^T \\ \mathbf{y}_2\mathbf{a}_1^T & \mathbf{y}_2\mathbf{a}_2^T & \dots & \mathbf{y}_2\mathbf{a}_k^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_M\mathbf{a}_1^T & \mathbf{y}_M\mathbf{a}_2^T & \dots & \mathbf{y}_M\mathbf{a}_k^T \end{pmatrix} \cdot \begin{pmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \dots & \mathbf{y}_M^T \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{y}_1 A\mathbf{y}_1^T & & & \\ & \mathbf{y}_2 A\mathbf{y}_2^T & & \\ & & \ddots & \\ & & & \mathbf{y}_M A\mathbf{y}_M^T \end{pmatrix}. \end{aligned}$$

Since the exponent in our integral involves the trace of the above product, the diagonal elements are only of interest. Thus,

$$\mathrm{Tr}(X^T AX^T) = \mathbf{y}_1 A\mathbf{y}_1^T + \mathbf{y}_2 A\mathbf{y}_2^T + \ldots + \mathbf{y}_M A\mathbf{y}_M^T.$$

Hence we can write

$$\begin{aligned} \int_{\mathbb{R}^{kM}} \mathrm{e}^{-\mathrm{Tr}(X^T AX)}|\mathrm{d}X| &= \int_{\mathbb{R}^k} \cdots \left[\int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^k} \mathrm{e}^{-(\mathbf{y}_1 A\mathbf{y}_1^T + \mathbf{y}_2 A\mathbf{y}_2^T + \ldots + \mathbf{y}_M A\mathbf{y}_M^T)}|\mathrm{d}\mathbf{y}_1|\right]|\mathrm{d}\mathbf{y}_2|\right]\cdots|\mathrm{d}\mathbf{y}_M| \\ &= \sqrt{\frac{\pi^k}{\det A}}\int_{\mathbb{R}^k} \cdots \left[\int_{\mathbb{R}^k} \mathrm{e}^{-(\mathbf{y}_2 A\mathbf{y}_2^T + \ldots + \mathbf{y}_M A\mathbf{y}_M^T)}|\mathrm{d}\mathbf{y}_2|\right]\cdots|\mathrm{d}\mathbf{y}_M| \\ &= \left(\frac{\pi^k}{\det A}\right)^{\frac{M}{2}}. \end{aligned}$$

□

Appendix D

Integral Representation of the Dirac Delta

Definition D.1. [10] A function $f_N(x)$ is a *delta convergent sequence* if:

a) For all $M > 0$ and for $|a| \leq M$ and $|b| \leq M$, the quantities

$$\left| \int_a^b f_N(t)\mathrm{d}t \right|$$

must be bounded by a constant independent of a, b, or N.

b) For any fixed, nonzero a and b, we must have

$$\lim_{N\to\infty} \int_a^b f_N(t)\mathrm{d}t = \begin{cases} 0 & a < b < 0, \quad 0 < a < b \\ 1 & a < 0 < b \end{cases} .$$

Consider the following sequence

$$\begin{aligned} f_N(x) &= \frac{1}{2\pi} \int_{-N}^{N} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t \\ &= \frac{1}{2\pi} \left(\frac{\mathrm{e}^{\mathrm{i}tx}}{ix} \Big|_{-N}^{N} \right) \\ &= \frac{1}{2\pi \mathrm{i} x} \left(\cos(Nx) + \mathrm{i}\sin(Nx) - \cos(Nx) + \mathrm{i}\sin(Nx)\right) \\ &= \frac{1}{\pi} \frac{\sin(Nx)}{x}. \end{aligned}$$

Consider the following

$$\int_a^b f_N(x)\mathrm{d}x = \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x}\mathrm{d}x.$$

Next, make a change of variable such that $y = Nx$, then

$$\begin{aligned} \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x}\mathrm{d}x &= \frac{1}{\pi} \int_{Na}^{Nb} \frac{\sin\, y}{y} N \frac{\mathrm{d}y}{N} \\ &= \frac{1}{\pi} \int_{Na}^{Nb} \frac{\sin\, y}{y}\mathrm{d}y. \end{aligned}$$

In order to show the second condition of Definition D.1, we will have three cases. For $a < b < 0$

$$\lim_{N\to\infty} \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x}\mathrm{d}x = \lim_{N\to\infty} \frac{1}{\pi} \int_{Na}^{Nb} \frac{\sin\, y}{y}\mathrm{d}y = \frac{1}{\pi} \int_{-\infty}^{-\infty} \frac{\sin\, y}{y}\mathrm{d}y = 0.$$

Similarly, when $0 < a < b$

$$\lim_{N\to\infty} \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x}\mathrm{d}x = \frac{1}{\pi} \int_{\infty}^{\infty} \frac{\sin\, y}{y}\mathrm{d}y = 0.$$

Lastly, for $a < 0 < b$, we have

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x}\mathrm{d}x &= \lim_{N\to\infty} \frac{1}{\pi} \int_{Na}^{Nb} \frac{sin\, y}{y}\mathrm{d}y \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin\, y}{y}\mathrm{d}y. \end{aligned}$$

Observe that this is the integral of the sinc function and, for $x \neq 0$, is equal to π [10]. Thus,

$$\lim_{N\to\infty} \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x} \mathrm{d}x = 1.$$

For the first condition, observe

$$\left| \frac{1}{\pi} \int_a^b \frac{\sin(Nx)}{x} \mathrm{d}x \right| = \left| \frac{1}{\pi} \int_{aN}^{bN} \frac{\sin(y)}{y} \mathrm{d}y \right|,$$

is bounded uniformly in a and b for all N. Therefore, $f_N(x)$ satisfies the necessary conditions to be a delta convergent sequence.

Lemma D.2. [10] *For a delta convergent sequence $f_N(x)$, we have*

$$\lim_{N\to\infty} f_N(x) = \delta(x).$$

Proof. Let $f_N(x)$ be a delta convergent sequence. Consider the following sequence of functions

$$F_N(x) = \int_{-1}^{x} f_N(\tau)\mathrm{d}\tau.$$

Observe from the properties of a delta convergent sequences, as N increases we have

$$\lim_{N\to\infty} F_N(x) = \lim_{N\to\infty} \int_{-1}^{x} f_N(\tau)\mathrm{d}\tau = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases}$$

Futhermore, these functions $F_N(x)$ are bounded uniformly. In the sense of generalized functions, the sequence of functions converges to the step function, i.e.

$$\lim_{N\to\infty} F_N(x) = \theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases}$$

Observe, from the Fundamental Theorem of Calculus, $f_N(x) = F_N'(x)$. Thus,

$$\lim_{N\to\infty} f_N(x) = \theta'(x).$$

Since $\theta'(x) = \delta(x)$, we have that $f_N(x)$ converges to $\delta(x)$. □

Definition D.3. [10] The *space of test functions* denoted by K is the set of all real *test functions* $\phi(x)$ which are smooth and have compact support. Then, for every test function $\phi \in K$, we have

$$\langle \delta(x), \phi(x) \rangle = \phi(0).$$

Observe with the application Lemma D.2, the Dirac delta is given by

$$\delta(x) = \lim_{N\to\infty} \frac{\sin(Nx)}{\pi x} = \lim_{N\to\infty} \left(\frac{1}{2\pi} \int_{-N}^{N} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}t \right).$$

The above limit is not defined point-wise, but rather in the weak sense. That is, for a test function $\phi \in K$, we have

$$\lim_{N\to\infty} \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{-N}^{N} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}t \right] \phi(x)\mathrm{d}x = \phi(0).$$

Since we observed that the one dimensional Dirac delta can be expressed using an integral, consider the vector case, specifically a p-dimensional vector. That is,

$$\begin{aligned}\int_{\mathbb{R}^p} \mathrm{e}^{\mathrm{i}\mathbf{t}\cdot\mathbf{x}}|\mathrm{d}\mathbf{t}| &= \int_{\mathbb{R}^p} \mathrm{e}^{\mathrm{i}(t_1x_1+t_2x_2+\ldots+t_px_p)}\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_p \\ &= \int_{\mathbb{R}} e^{\mathrm{i}t_1x_1}\mathrm{d}t_1 \int_{\mathbb{R}} e^{\mathrm{i}t_2x_2}\mathrm{d}t_2 \cdots \int_{\mathbb{R}} e^{\mathrm{i}t_px_p}\mathrm{d}t_p \\ &= 2\pi\delta(x_1)2\pi\delta(x_2)\cdots 2\pi\delta(x_p) \\ &= (2\pi)^p\delta(\mathbf{x}).\end{aligned}$$

Thus, for a p-dimensional vector we have

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} \mathrm{e}^{\mathrm{i}\mathbf{t}\cdot\mathbf{x}}|\mathrm{d}\mathbf{t}|.$$

We are interested in the representation of the Dirac delta for symmetric matrices of size $p\times p$. First, consider the following integral

$$\int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\mathrm{Tr}\, TX}|\mathrm{d}T|,$$

where T is a symmetric matrix. Thus we will have $\dfrac{p(p+1)}{2}$ integrations. Observe the exponent can be simplified as

$$\begin{aligned}\mathrm{Tr}\, TX &= \sum_{i,j=1}^{p} t_{ij}x_{ij} \\ &= \sum_{i\leq j}^{p} t_{ij}x_{ij} + \sum_{i>j}^{p} t_{ij}x_{ij}.\end{aligned}$$

By switching indices on the second sum, we can write it as $i<j$

$$\mathrm{Tr}\, TX = \sum_{i\leq j}^{p} t_{ij}x_{ij} + \sum_{i<j}^{p} t_{ji}x_{ji}.$$

Since both matrices are symmetric, $t_{ij}x_{ij} = t_{ji}x_{ji}$

$$\begin{aligned}\mathrm{Tr}\, TX &= \sum_{i\leq j}^{p} t_{ij}x_{ij} + \sum_{i<j}^{p} t_{ij}x_{ij} \\ &= 2\sum_{i<j}^{p} t_{ij}x_{ij} + \sum_{i=1}^{p} t_{ii}x_{ii}.\end{aligned}$$

Hence, our integral can be expressed as

$$\int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\mathrm{Tr}\, TX}|\mathrm{d}T| = \int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \mathrm{e}^{\mathrm{i}2\sum_{i<j}^{p} t_{ij}x_{ij}}\mathrm{d}t_{12}\mathrm{d}t_{13}\cdots\mathrm{d}t_{p-1\,p}\int_{\mathbb{R}^p} \mathrm{e}^{\mathrm{i}\sum_{i=1}^{p} t_{ii}x_{ii}}\mathrm{d}t_{11}\mathrm{d}t_{22}\cdots\mathrm{d}t_{pp}.$$

Now let $2t_{ij} = \hat{t_{ij}}$ for $i<j$, then $\mathrm{d}t_{ij} = 2\,\mathrm{d}\hat{t_{ij}}$

$$\begin{aligned}\int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\mathrm{Tr}\, TX}|\mathrm{d}T| &= \frac{1}{2^{\frac{p(p-1)}{2}}}\int_{\mathbb{R}^{\frac{p(p-1)}{2}}} \mathrm{e}^{\mathrm{i}\sum_{i<j}^{p} \hat{t}_{ij}x_{ij}}\mathrm{d}\hat{t}_{12}\mathrm{d}\hat{t}_{13}\cdots\mathrm{d}\hat{t}_{p-1\,p}\int_{\mathbb{R}^p} \mathrm{e}^{\mathrm{i}\sum_{i=1}^{p} t_{ii}x_{ii}}\mathrm{d}t_{11}\mathrm{d}t_{22}\cdots\mathrm{d}t_{pp} \\ &= \left(\frac{2\pi}{2}\right)^{\frac{p(p-1)}{2}}\delta(x_{12})\delta(x_{13})\cdots\delta(x_{p-1\,p})(2\pi)^p\delta(x_{11})\delta(x_{22})\cdots\delta(x_{pp}) \\ &= \frac{(2\pi)^{\frac{p(p+1)}{2}}}{2^{\frac{p(p-1)}{2}}}\delta(X).\end{aligned}$$

Therefore, we have a integral representation for the mutlidimensional Dirac delta

$$\delta(X) = \frac{2^{\frac{p(p-1)}{2}}}{(2\pi)^{\frac{p(p+1)}{2}}} \int_{\mathbb{R}^{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\operatorname{Tr} TX} |\mathrm{d}T|,$$

and for a test function $\phi(X) \in K$, we have

$$\lim_{N_j \to \infty} \frac{2^{\frac{p(p-1)}{2}}}{(2\pi)^{\frac{p(p+1)}{2}}} \int_{\mathbb{R}^{pM}} \left[\int_{-N_1}^{N_1} \cdots \int_{-N_j}^{N_j} \cdots \int_{-N_{\frac{p(p+1)}{2}}}^{N_{\frac{p(p+1)}{2}}} \mathrm{e}^{\mathrm{i}\operatorname{Tr} TX} |\mathrm{d}T| \right] \phi(X) |\mathrm{d}X| = \phi(0).$$

Appendix E

The Laplace Transform

Definition E.1. [27] The *Laplace transform* of a function $f(t)$ is denoted by $\hat{f}(s)$ and is defined by the integral

$$\hat{f}(s) = \int_0^\infty \mathrm{e}^{-st} f(t)\mathrm{d}t,$$

where $s \in \mathbb{C}$, $t \geq 0$ and f is locally integrable and such that the integral exists for a given s.

Consider the Laplace transform for the function $f(t) = t^\alpha$. In order for $f(t) \in L^1(\text{loc})$, we need Re $\alpha > -1$. Then,

$$\hat{f}(s) = \int_0^\infty \mathrm{e}^{-st} t^\alpha \mathrm{d}t, \qquad \text{Re } s > 0.$$

We will make the substitution such that $u = st$, then $\mathrm{d}u = s\,\mathrm{d}t$ and

$$\begin{aligned} \hat{f}(s) &= \int_0^\infty \mathrm{e}^{-u} \left(\frac{u}{s}\right)^\alpha \frac{\mathrm{d}u}{s} \\ &= s^{-\alpha-1} \int_0^\infty \mathrm{e}^{-u} u^\alpha \mathrm{d}u. \end{aligned}$$

In order to evaluate this integral we will use the integral representation of the Gamma function.

Lemma E.2. [3] *For Re $z > 0$*

$$\Gamma(z) = \int_0^\infty \mathrm{e}^{-t} t^{z-1} \mathrm{d}t.$$

Thus, for Re $\alpha > -1$ and Re $s > 0$

$$\hat{f}(s) = s^{-\alpha-1}\Gamma(\alpha+1).$$

Lemma E.3. [25] *The integral formula for the inverse of the Laplace transformation, at the points where f is continuous, is given by the line integral*

$$f(t) = \frac{1}{2\pi \mathrm{i}} \lim_{R\to\infty} \int_{\sigma - iR}^{\sigma + iR} e^{st} \hat{f}(s)\mathrm{d}s,$$

where the contour for the inversion integral is shown in Figure E.1.

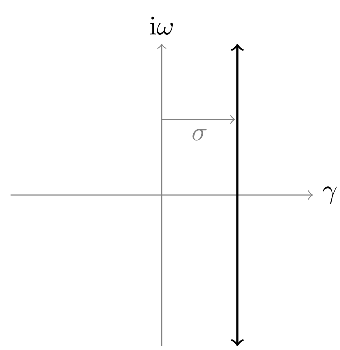

Figure E.1: Contour for inversion intgeral

Recall for $f(t) = t^\alpha$, we found $\hat{f}(s) = s^{-\alpha-1}\Gamma(\alpha+1)$, then

$$t^\alpha = \frac{\Gamma(\alpha+1)}{2\pi\mathrm{i}} \lim_{R\to\infty} \int_{\sigma-iR}^{\sigma+iR} \mathrm{e}^{st} s^{-\alpha-1} \mathrm{d}s.$$

We will let $s = k + \mathrm{i}x$, where $\operatorname{Re} k > 0$. Then, $\mathrm{d}s = \mathrm{i}\,\mathrm{d}x$ and

$$\begin{aligned} t^\alpha &= \frac{\Gamma(\alpha+1)}{2\pi} \int_{\mathbb{R}} (k+\mathrm{i}x)^{-\alpha-1} \mathrm{e}^{(k+\mathrm{i}x)t} \mathrm{d}x \\ &= \frac{\Gamma(\alpha+1)}{2\pi} \mathrm{e}^{kt} \int_{\mathbb{R}} (k+\mathrm{i}x)^{-\alpha-1} \mathrm{e}^{\mathrm{i}xt} \mathrm{d}x. \end{aligned}$$

Thus,

$$\int_{\mathbb{R}} (k+\mathrm{i}x)^{-\alpha-1} \mathrm{e}^{\mathrm{i}xt} \mathrm{d}x = \frac{2\pi}{\Gamma(\alpha+1)} t^\alpha \mathrm{e}^{-kt},$$

where $\operatorname{Re} \alpha > -1$, $\operatorname{Re} k > 0$ and $t > 0$.

Appendix F

LU Factorization

Definition F.1. [11] Let $A \in \mathbb{R}^{n\times n}$. The presentation $A = LU$, in which $L \in \mathbb{R}^{n\times n}$ is lower triangular and $U \in \mathbb{R}^{n\times n}$ is upper triangular, is called an *LU factorization of A*.

Theorem F.2. [11] *Suppose that $A \in \mathbb{R}^{n\times n}$ and rank $A = k$. If $A_{[j]}$, for all $1 \leq j \leq k$, is nonsingular, then A has an LU factorization. Furthermore, if $k = n$ then A has an LU factorization if and only if A and all of its leading principal submatrices are nonsingular.*

Example F.3. Consider the following matrix

$$A = \begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix}.$$

First we will check if an *LU* factorization exists, thus we will compute the leading principal minors

$$\det(a_{11}) = 2 \neq 0 \quad \det\ A = -10 \neq 0.$$

Since they are nonzero, an LU decomposition exists. So,

$$\begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}.$$

Now, we will multiply the right hand side to obtain the following system of equations

$$\begin{aligned} l_{11}u_{11} &= 2 \\ l_{11}u_{12} &= 3 \\ l_{21}u_{11} &= 6 \\ l_{21}u_{12} + l_{22}u_{22} &= 4. \end{aligned}$$

Observe there are more unknowns than the number of equations, so we will arbitrarily assign the diagonal elements of L to be 1, that is $l_{11} = 1$ and $l_{22} = 1$. Then,

$$\begin{aligned} u_{11} &= 2 \\ u_{12} &= 3 \\ l_{21} &= 3 \\ u_{22} &= -5, \end{aligned}$$

which gives the following *LU* factorization

$$\begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & -5 \end{pmatrix}.$$

If we were to take the original matrix A and conduct Gaussian elimination until the matrix is in upper triangular form, without switching any rows or columns, we would obtain U. We will see that the elementary matrices obtained from conducting the row operations can be used to find L. Then,

$$\begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix} R_2 \to R_2 - 3R_1 \begin{pmatrix} 2 & 3 \\ 0 & -5 \end{pmatrix} = U.$$

Observe that from the single row operation we have the following elementary matrix,

$$E_{21} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}.$$

Then we have that $E_{21}A = U$, which implies $A = E_{21}^{-1}U$. Therefore, we see that $E_{21}^{-1} = L$, i.e.

$$E_{21}^{-1} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} = L.$$

Hence, we get the same LU factorization as before,

$$A = \begin{pmatrix} 2 & 3 \\ 6 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & -5 \end{pmatrix} = LU$$

Example F.4. Now, we will consider

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix},$$

where A has the following principal minors

$$\begin{aligned} \det(a_{11}) &= 1 \neq 0 \\ \det(a_{11}a_{22} - a_{12}a_{21}) &= 2 \neq 0 \\ \det(A) &= 1 \cdot \det \begin{pmatrix} 8 & 14 \\ 6 & 13 \end{pmatrix} - 2 \cdot \det \begin{pmatrix} 3 & 14 \\ 2 & 13 \end{pmatrix} + 4 \cdot \det \begin{pmatrix} 3 & 8 \\ 2 & 6 \end{pmatrix} \\ &= 6 \neq 0. \end{aligned}$$

Thus, an LU factorization exists, that is

$$\begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{23} & 0 \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix}.$$

We will perform Gaussian elimination on the matrix A in order to make it upper triangular

$$\begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} \xrightarrow[R_3 \to R_3 - 2R_1]{R_2 \to R_2 - 3R_1} \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 2 & 5 \end{pmatrix} R_3 \to R_3 - R_2 \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} = U.$$

From each row operation we have the following elementary matrices,

$$E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad E_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \quad E_{32} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Thus we have that

$$E_{32}E_{31}E_{21}A = U \implies A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U.$$

In order to find the inverse of each elementary matrix, we will switch the sign of any off diagonal element. So,

$$\begin{aligned} E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} = L. \end{aligned}$$

Thus we have an LU factorization for A,

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 3 & 8 & 14 \\ 2 & 6 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} = LU.$$

Lemma F.5. [11] *Let $A \in \mathbb{R}^{n\times n}$ and supposed that $A = LU$ is an LU factorization. For any block 2-by-2 partition*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}, \quad U = \begin{pmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{pmatrix},$$

with A_{11}, L_{11}, $U_{11} \in \mathbb{R}^{k\times k}$ and $k \leq n$, we have $A_{11} = L_{11}U_{11}$. Consequently, each leading principal submatrix of A has an LU factorization in which the factors are the corresponding leading principal submatrices of L and U.

If we consider the matrix Ω as presented in Chapter 4, we will see that during the process of our recurrence relation we use an LU factorization. Observe that from the decomposition in (4.3) we have

$$\Omega = \left(\begin{array}{c|c} \omega_{11} & \boldsymbol{\omega} \\ \hline \boldsymbol{\omega}^T & \Omega_{p-1} \end{array}\right).$$

We will write this as an LU factorization

$$\Omega = LU = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{\boldsymbol{\omega}^T}{\omega_{11}} & L_1 \end{array}\right) \cdot \left(\begin{array}{c|c} \omega_{11} & \boldsymbol{\omega} \\ \hline 0 & U_1 \end{array}\right).$$

So, we need $L_1U_1 = \Omega_{p-1} - \dfrac{\boldsymbol{\omega}^T\boldsymbol{\omega}}{\omega_{11}}$, which is exactly what we defined as the next matrix in our recurrence relation, as we observed in (4.9), i.e. $L_1U_1 = \Omega^{(p-1)}$. Thus, we can factorize this resulting matrix

$$\Omega^{(p-1)} = L_1U_1 = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}} & L_2 \end{array}\right) \cdot \left(\begin{array}{c|c} \omega_{11}^{(p-1)} & \boldsymbol{\omega}^{(p-1)} \\ \hline 0 & U_2 \end{array}\right),$$

where $L_2U_2 = \Omega_{p-2}^{(p-1)} - \dfrac{(\boldsymbol{\omega}^{(p-1)})^T\boldsymbol{\omega}^{(p-1)}}{\omega_{11}^{(p-1)}}$. Again, this is defined as $\Omega^{(p-2)}$. Hence, we have

$$\Omega = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{\boldsymbol{\omega}^T}{\omega_{11}} & \begin{array}{c|c} 1 & 0 \\ \hline \frac{(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}} & \begin{array}{c|c} 1 & 0 \\ \hline \frac{(\boldsymbol{\omega}^{(p-2)})^T}{\omega_{11}^{(p-2)}} & L_3 \end{array} \end{array} \end{array}\right) \cdot \left(\begin{array}{c|c} \omega_{11} & \boldsymbol{\omega} \\ \hline 0 & \begin{array}{c|c} \omega_{11}^{(p-1)} & \boldsymbol{\omega}^{(p-1)} \\ \hline 0 & \begin{array}{c|c} \omega_{11}^{(p-2)} & \boldsymbol{\omega}^{(p-2)} \\ \hline 0 & U_3 \end{array} \end{array} \end{array}\right).$$

We will continue this process for each step of our recurrence relation. When $p \leq M$ we will see that Ω will have a complete LU factorization where L is unipotent, having ones on the diagonal. When $p > M$, our recurrence stops after M steps, thus the factorization will produce a lower triangular matrix L that is only partially unipotent.

For the first case, $p \leq M$, we will conduct the factorization process for $p-1$ steps, giving

$$\Omega = \left(\begin{array}{c|c|c|c|c} 1 & \multicolumn{4}{c}{0} \\ \hline \multirow{4}{*}{$\frac{\boldsymbol{\omega}^T}{\omega_{11}}$} & 1 & \multicolumn{3}{c}{0} \\ \cline{2-5} & \multirow{3}{*}{$\frac{(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}}$} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & 1 & 0 \\ \cline{4-5} & & & \frac{(\boldsymbol{\omega}^{(2)})^T}{\omega_{11}^{(2)}} & L_{p-1} \end{array}\right) \cdot \left(\begin{array}{c|c|c|c|c} \omega_{11} & \multicolumn{4}{c}{\boldsymbol{\omega}} \\ \hline \multirow{4}{*}{0} & \omega_{11}^{(p-1)} & \multicolumn{3}{c}{\boldsymbol{\omega}^{(p-1)}} \\ \cline{2-5} & \multirow{3}{*}{0} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & \omega_{11}^{(2)} & \boldsymbol{\omega}^{(2)} \\ \cline{4-5} & & & 0 & U_{p-1} \end{array}\right),$$

where

$$\Omega^{(2)} = \left(\begin{array}{c|c} 1 & 0 \\ \hline \frac{(\boldsymbol{\omega}^{(2)})^T}{\omega_{11}^{(2)}} & L_{p-1} \end{array}\right) \cdot \left(\begin{array}{c|c} \omega_{11}^{(2)} & \boldsymbol{\omega}^{(2)} \\ \hline 0 & U_{p-1} \end{array}\right).$$

So,

$$L_{p-1}U_{p-1} = \Omega_1^{(2)} - \frac{(\boldsymbol{\omega}_{\mathbf{21}}^{(2)})^T \boldsymbol{\omega}_{\mathbf{12}}^{(2)}}{\omega_{11}^{(2)}} = \Omega^{(1)} = \omega_{11}^{(1)}.$$

If we assign $L_{p-1} = 1$ and $U_{p-1} = \omega_{11}^{(1)}$, we have

$$\Omega = \left(\begin{array}{c|c|c|c|c} 1 & \multicolumn{4}{c}{0} \\ \hline \multirow{4}{*}{$\frac{\boldsymbol{\omega}^T}{\omega_{11}}$} & 1 & \multicolumn{3}{c}{0} \\ \cline{2-5} & \multirow{3}{*}{$\frac{(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}}$} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & 1 & 0 \\ \cline{4-5} & & & \frac{(\boldsymbol{\omega}^{(2)})^T}{\omega_{11}^{(2)}} & 1 \end{array}\right) \cdot \left(\begin{array}{c|c|c|c|c} \omega_{11} & \multicolumn{4}{c}{\boldsymbol{\omega}} \\ \hline \multirow{4}{*}{0} & \omega_{11}^{(p-1)} & \multicolumn{3}{c}{\boldsymbol{\omega}^{(p-1)}} \\ \cline{2-5} & \multirow{3}{*}{0} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & \omega_{11}^{(2)} & \boldsymbol{\omega}^{(2)} \\ \cline{4-5} & & & 0 & \omega_{11}^{(1)} \end{array}\right).$$

Thus, Ω accepts an LU factorization having a unipotent lower triangular matrix.

We will conduct the same proces when $p > M$, except we will stop after M steps. Hence, we will have

$$\Omega = \left(\begin{array}{c|c|c|c|c} 1 & \multicolumn{4}{c}{0} \\ \hline \multirow{4}{*}{$\frac{\boldsymbol{\omega}^T}{\omega_{11}}$} & 1 & \multicolumn{3}{c}{0} \\ \cline{2-5} & \multirow{3}{*}{$\frac{(\boldsymbol{\omega}^{(p-1)})^T}{\omega_{11}^{(p-1)}}$} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & 1 & 0 \\ \cline{4-5} & & & \frac{(\boldsymbol{\omega}^{(p-(M-1))})^T}{\omega_{11}^{(p-(M-1))}} & L_M \end{array}\right) \cdot \left(\begin{array}{c|c|c|c|c} \omega_{11} & \multicolumn{4}{c}{\boldsymbol{\omega}} \\ \hline \multirow{4}{*}{0} & \omega_{11}^{(p-1)} & \multicolumn{3}{c}{\boldsymbol{\omega}^{(p-1)}} \\ \cline{2-5} & \multirow{3}{*}{0} & \ddots & \multicolumn{2}{c}{\ddots} \\ \cline{3-5} & & \multirow{2}{*}{$\ddots$} & \omega_{11}^{(p-(M-1))} & \boldsymbol{\omega}^{(p-(M-1))} \\ \cline{4-5} & & & 0 & U_M \end{array}\right).$$

Observe that again L and U will have the same form as before, that is

$$L_{M+1}U_{M+1} = \Omega_{p-M}^{(p-M+1)} - \frac{(\boldsymbol{\omega}^{(p-M+1)})^T \boldsymbol{\omega}^{(p-M+1)}}{\omega_{11}^{(p-M+1)}} = \Omega^{(p-M)},$$

but from our Dirac delta appearing in our recurrence relation, that is $\delta\left(\Omega^{(p-M)}\right)$, we see this resulting matrix vanishes. Thus,

$$L_M = [0], \quad U_M = [0].$$

Hence, our LU factorization is of the form

$$\Omega = \left(\begin{array}{c|c|c|c|c}
1 & \multicolumn{4}{c}{0} \\
\hline
 & 1 & \multicolumn{3}{c}{0} \\
\cline{2-5}
\dfrac{\boldsymbol{\omega}^T}{\omega_{11}} & & \ddots & \multicolumn{2}{c}{\ddots} \\
\cline{3-5}
 & \dfrac{\left(\boldsymbol{\omega}^{(p-1)}\right)^T}{\omega_{11}^{(p-1)}} & & 1 & 0 \\
\cline{4-5}
 & & \ddots & \dfrac{\left(\boldsymbol{\omega}^{(p-(M-1))}\right)^T}{\omega_{11}^{(p-(M-1))}} & 0
\end{array}\right) \cdot \left(\begin{array}{c|c|c|c|c}
\omega_{11} & \multicolumn{4}{c}{\boldsymbol{\omega}} \\
\hline
 & \omega_{11}^{(p-1)} & \multicolumn{3}{c}{\boldsymbol{\omega}^{(p-1)}} \\
\cline{2-5}
0 & & \ddots & \multicolumn{2}{c}{\ddots} \\
\cline{3-5}
 & 0 & \ddots & \omega_{11}^{(p-(M-1))} & \boldsymbol{\omega}^{(p-(M-1))} \\
\cline{4-5}
 & & & 0 & 0
\end{array}\right).$$

Appendix G

Lewis Carroll Identity

Theorem G.1. [16] *Let $A \in \mathbb{R}^{n\times n}$. Denote the submatrix of A in which the rows $l_1, l_2, \ldots, l_k$ and columns $r_1, r_2, \ldots, r_k$ have been removed by $A^{l_1,l_2,\ldots,l_k}_{r_1,r_2,\ldots,r_k}$. Then the following holds*

$$\det A \cdot \det A^{1,n}_{1,n} = \det A^1_1 \cdot \det A^n_n - \det A^n_1 \cdot \det A^1_n.$$

Proof. Consider the following matrix $A \in \mathbb{R}^{\Bbbk\times\Bbbk}$

$$A = \left(\begin{array}{c|ccc|c} a_{11} & a_{12} & \ldots & a_{1,n-1} & a_{1n} \\ \hline a_{21} & & & & a_{2n} \\ \vdots & & B & & \vdots \\ a_{n-1,1} & & & & a_{2,n-1} \\ \hline a_{n1} & a_{n2} & \ldots & a_{n,n-1} & a_{nn} \end{array}\right)$$

$$= \left(\begin{array}{c|c|c} a_{11} & \mathbf{l} & a_{1n} \\ \hline \mathbf{r}^T & B & \mathbf{j}^T \\ \hline a_{n1} & \mathbf{k} & a_{nn} \end{array}\right).$$

We will compute the righthand side of the identity in order to find an explicit expression for each determinant. For the above matrix, observe by removing the last row and last column we obtain the submatrix A^n_n. Then the determinant of this submatrix will be

$$\det A^n_n = \det \left(\begin{array}{c|c} a_{11} & \mathbf{l} \\ \hline \mathbf{r}^T & B \end{array}\right) = \det B \cdot (a_{11} - \mathbf{l}B^{-1}\mathbf{r}^T).$$

Similarly, by removing the first row and column from A, we obtain A^1_1 and the following determinant

$$\det A^1_1 = \det \left(\begin{array}{c|c} B & \mathbf{j}^T \\ \hline \mathbf{k} & a_{nn} \end{array}\right) = \det B \cdot (a_{nn} - \mathbf{k}B^{-1}\mathbf{j}^T),$$

and

$$\det A^n_1 = \det \left(\begin{array}{c|c} \mathbf{l} & a_{1n} \\ \hline B & \mathbf{j}^T \end{array}\right).$$

Now, we would like to make $n-2$ interchanges by moving the first row to the last row, giving

$$\det A^n_1 = (-1)^{n-2} \det \left(\begin{array}{c|c} B & \mathbf{j}^T \\ \hline \mathbf{l} & a_{1n} \end{array}\right) = (-1)^{n-2} \det B(a_{1n} - \mathbf{l}B^{-1}\mathbf{j}^T).$$

By the same logic, we will consider

$$\det A^1_n = \left(\begin{array}{c|c} \mathbf{r}^T & B \\ \hline a_{n1} & \mathbf{k} \end{array}\right),$$

and move the first column to last column

$$\det A_n^1 = (-1)^{n-2} \left(\begin{array}{c|c} B & \mathbf{r}^T \\ \hline \mathbf{k} & a_{n1} \end{array} \right) = (-1)^{n-2} \det B (a_{n1} - \mathbf{k} B^{-1} \mathbf{r}^T).$$

Next, we want to find an expression for $\det A$. Denote C as the upper left $(n-1) \times (n-1)$ submatrix of A, i.e.

$$C = \left(\begin{array}{c|c} a_{11} & \mathbf{l} \\ \hline \mathbf{r}^T & B \end{array} \right).$$

Then,

$$A = \left(\begin{array}{cc|c} \multicolumn{2}{c|}{C} & \begin{array}{c} a_{1n} \\ \mathbf{j}^T \end{array} \\ \hline a_{n1} & \mathbf{k} & a_{nn} \end{array} \right) = \left(\begin{array}{c|c} C & \hat{\mathbf{j}}^T \\ \hline \hat{\mathbf{k}} & a_{nn} \end{array} \right),$$

and

$$\det A = \det C \cdot (a_{nn} - \hat{\mathbf{k}}^T C^{-1} \hat{\mathbf{j}}^T). \tag{G.1}$$

In order to determine C^{-1}, we must find a matrix such that

$$\left(\begin{array}{c|c} a_{11} & \mathbf{l} \\ \hline \mathbf{r}^T & B \end{array} \right) \cdot \left(\begin{array}{c|c} e & \mathbf{f} \\ \hline \mathbf{g}^T & Y \end{array} \right) = I,$$

thus we have

$$\begin{aligned} e &= \frac{1}{a_{11} - \mathbf{l} B^{-1} \mathbf{r}^T} \\ \mathbf{f} &= -e\, \mathbf{l} B^{-1} \\ \mathbf{g}^T &= -e\, B^{-1} \mathbf{r}^T \\ Y &= B^{-1} + e\, B^{-1} \mathbf{r}^T \mathbf{l} B^{-1}. \end{aligned}$$

Now, consider the product appearing in G.1

$$\begin{aligned} \hat{\mathbf{k}}^T C^{-1} \hat{\mathbf{j}}^T &= \left(\begin{array}{c|c} a_{n1} & \mathbf{k} \end{array} \right) \cdot \left(\begin{array}{c|c} e & \mathbf{f} \\ \hline \mathbf{g}^T & Y \end{array} \right) \cdot \left(\begin{array}{c|c} a_{1n} & \mathbf{j} \end{array} \right)^T \\ &= e\, a_{n1} a_{1n} - e\, a_{1n} \mathbf{k} B^{-1} \mathbf{r}^T - e\, a_{n1} \mathbf{l} B^{-1} \mathbf{j}^T + \mathbf{k} B^{-1} \mathbf{j}^T + e\, \mathbf{k} B^{-1} \mathbf{r}^T \mathbf{l} B^{-1} \mathbf{j}^T \\ &= e\, a_{1n} (a_{n1} - \mathbf{k} B^{-1} \mathbf{r}^T) - e\, (a_{n1} - \mathbf{k} B^{-1} \mathbf{r}^T) \mathbf{l} B^{-1} \mathbf{j}^T + \mathbf{k} B^{-1} \mathbf{j}^T \\ &= e\, (a_{n1} - \mathbf{k} B^{-1} \mathbf{r}^T)(a_{1n} - \mathbf{l} B^{-1} \mathbf{j}^T) + \mathbf{k} B^{-1} \mathbf{j}^T. \end{aligned}$$

So, we can express G.1 as

$$\begin{aligned} \det A &= \det B \cdot (a_{11} - \mathbf{l} B \mathbf{r}^T)(a_{nn} - \mathbf{k} B^{-1} \mathbf{j}^T - e(a_{n1} - \mathbf{k} B^{-1} \mathbf{r}^T)(a_{1n} - \mathbf{l} B^{-1} \mathbf{j}^T)) \\ &= \frac{\det B}{e} \cdot \left(\frac{\det A_1^1}{\det B} - e \frac{\det A_n^1 \cdot \det A_1^n}{(\det B)^2} \right) \\ &= \frac{\det A_1^1}{e} - \frac{\det A_n^1 \cdot \det A_1^n}{\det B} \end{aligned}$$

From our definition of e, we see that $e = \dfrac{\det B}{\det A_n^n}$. Thus,

$$\det A = \frac{\det A_n^n \cdot \det A_1^1}{\det B} - \frac{\det A_n^1 \cdot \det A_1^n}{\det B}.$$

Observe that $\det B = \det A_{1,n}^{1,n}$, hence

$$\det A \cdot \det A_{1,n}^{1,n} = \det A_n^n \cdot \det A_1^1 - \det A_n^1 \cdot \det A_1^n. \qquad \square$$

We will demonstrate the above theorem with an example

Example G.2. Consider the following matrix

$$A = \begin{pmatrix} 1 & 4 & 20 & 3 \\ 2 & 7 & 9 & 10 \\ 13 & 15 & 5 & 8 \\ 4 & 11 & 6 & 7 \end{pmatrix}.$$

For the lefthand side of the identity we have that

$$\begin{aligned} \det A \cdot \det A_{1,4}^{1,4} &= \det \begin{pmatrix} 1 & 4 & 20 & 3 \\ 2 & 7 & 9 & 10 \\ 13 & 15 & 5 & 8 \\ 4 & 11 & 6 & 7 \end{pmatrix} \cdot \det \begin{pmatrix} 7 & 9 \\ 15 & 5 \end{pmatrix} \\ &= 7801 \cdot -100 \\ &= -780,100. \end{aligned}$$

Then, we see the computation of the right hand of the identity side gives the same result, i.e.

$$\begin{aligned} \det A_1^1 \cdot \det A_4^4 - \det A_1^4 \cdot \det A_4^1 = \det &\begin{pmatrix} 7 & 9 & 10 \\ 15 & 5 & 8 \\ 11 & 6 & 7 \end{pmatrix} \cdot \det \begin{pmatrix} 1 & 4 & 20 \\ 2 & 7 & 9 \\ 13 & 15 & 5 \end{pmatrix} - \\ \det &\begin{pmatrix} 4 & 20 & 3 \\ 7 & 9 & 10 \\ 15 & 5 & 8 \end{pmatrix} \cdot \det \begin{pmatrix} 2 & 7 & 9 \\ 13 & 15 & 5 \\ 4 & 11 & 6 \end{pmatrix} \\ = &\, 106 \cdot -892 - 1668 \cdot 411 \\ = &\, -780,100. \end{aligned}$$

Appendix H

Fubini's Theorem

Theorem H.1. [12] *Let f be an integrable function on $\mathbb{R}^n \times \mathbb{R}^m$ and suppose that for each $\mathbf{x} \in \mathbb{R}^n$, the function $\mathbf{y} \mapsto f(\mathbf{x}, \mathbf{y})$ is integrable. Then the function*

$$\mathbf{x} \mapsto \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) |\mathrm{d}\mathbf{y}|,$$

is integrable and

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{x}, \mathbf{y}) |\mathrm{d}\mathbf{x}||\mathrm{d}\mathbf{y}| = \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^M} f(\mathbf{x}, \mathbf{y}) |\mathrm{d}\mathbf{y}| \right] |\mathrm{d}\mathbf{x}|.$$

Example H.2. Consider the function $f(y) = \mathrm{e}^{-y^2}$. Let us integrate this function over the triangle

$$T = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \mid 0 \leq x \leq y \leq 1 \right\}.$$

Fubini's theorem allows us to write this integral as an iterated one-dimensional integral

$$\int_0^1 \left[\int_x^1 \mathrm{e}^{-y^2} \mathrm{d}y \right] \mathrm{d}x \quad \text{and} \quad \int_0^1 \left[\int_0^y \mathrm{e}^{-y^2} \mathrm{d}x \right] \mathrm{d}y.$$

The first integral cannot be computed in an elementary sense, as the function does not have an elementary antiderivative. Consider the computation of the second integral

$$\begin{aligned} \int_0^1 \left[\int_0^y \mathrm{e}^{-y^2} \mathrm{d}x \right] \mathrm{d}y &= \int_0^1 y \mathrm{e}^{-y^2} \mathrm{d}y \\ &= -\frac{1}{2} \left. \mathrm{e}^{-y^2} \right|_0^1 = \frac{1}{2} \left(1 - \frac{1}{e} \right). \end{aligned}$$

Thus,

$$\iint_T e^{-y^2} \mathrm{d}y \mathrm{d}x = \frac{1}{2} \left(1 - \frac{1}{e} \right).$$