ABSTRACT

LEE, HAN NA. Inference for Eigenvalues and Eigenvectors in Exponential Families of Random Symmetric Matrices. (Under the direction of Dr. Armin Schwartzman.)

Diffusion tensor imaging (DTI) data contains a $3 \times 3$ positive definite random matrix at every voxel (3D pixel). Motivated by the anatomical interpretation of the data, it is of interest to investigate the eigenvalues and eigenvectors of the mean parameter. This dissertation focuses on theory and methodology for analyzing matrix-variated data, based on the inference for the eigenstructure of the mean parameter.

In particular, a matrix-variate exponential family of distributions is presented for random positive definite matrices. Two particular examples of interest in this family that are developed in detail are the matrix-variate gaussian distribution and the matrix gamma distribution. An orthogonal equivariance property of the matrix-variate exponential family is investigated, and estimation and testing procedures for eigenvalues and eigenvectors of the mean matrix are derived. These estimation and testing procedures are carried out both in the one-sample and two-sample problems.

The dissertation is organized as follows. We define a matrix-variate exponential family and establish the properties of this family. Then, we estimate the eigenvalues and eigenvectors of the mean parameter $M$ under several restrictions, such as fixed eigenvalues and fixed eigenvectors. Using the likelihood of the exponential family defined in Chapter 2, maximum likelihood estimates of eigenvalues and eigenvectors are derived. According to the assumption we make for the eigenstructure, MLEs of eigenvalues and eigenvectors can be different. These assumptions for the eigenstructure are tested by likelihood ratio tests; therefore, test statistics for the hypotheses related to these assumptions are also derived.

Estimation and testing procedures for the eigenstructure of the mean parameters in the two-sample problem are also investigated. In two-sample cases, we are focusing on the comparison of two groups. Thus, the restriction for the eigenstructures are based on this comparison.

Although our research is focused on the inference for the mean parameter, assumptions on the nuisance parameters can affect estimation and testing. Thus, a section of this dissertation addresses the handling of nuisance parameters of the matrix-variate gaussian distribution and the matrix gamma distribution.

In the latter section of this dissertation, we investigate the asymptotic properties of the estimates and test statistics and their numerical performance via simulation studies. In addition, our estimates and test procedures are illustrated in a real DTI data analysis.
Inference for Eigenvalues and Eigenvectors in Exponential Families of Random Symmetric Matrices

by
Han Na Lee

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Statistics

Raleigh, North Carolina

2016

APPROVED BY:

Dr. David Dickey

Dr. Arnab Maity

Dr. Ana-Maria Staicu

Dr. Jack Silverstein

Dr. Armin Schwartzman

Chair of Advisory Committee
DEDICATION

To my parents
To my Love.
Han Na Lee was born in February 5, 1985, in An-Yang, Korea. She graduated from Seoul Foreign Language High School in 2003 and attended Yonsei University for her undergraduate study in Applied Statistics. She obtained her Bachelor’s degree in 2008, and attend the same university for her Master’s degree. She earned her M.S. in 2010, and she joined the North Carolina State University to pursue Ph.D. degree in statistics in the same year. Her doctoral dissertation research at NCSU is under the direction of Dr. Armin Schwartzman. She is expected to graduate with her Doctorate in Summer, 2016.
ACKNOWLEDGEMENTS

First, I would like to express my first and deepest gratitude to my advisor, Dr. Armin Schwartzman. It has been a pleasure working with you. I am grateful for your patience, respect, and warm concern. Without your guidance and financial and emotional support, I would not have been able to complete my study successfully.

I am especially grateful to Drs. David Dickey, Arnab Maity, Ana-Maria Staicu and Jack Silverstein for being my committee members. I greatly appreciate constructive feedback they have provided. I also want to thank to Dr. Huixia Judy Wang, my academic advisor. Her valuable advice and warm words were great help to me. Also, I would like to thank Drs. Kim Weems and Howard Bondell for their great help as a graduate program director.

I would like to thank to my friends in the department. It was a great time with you: To Eun Jeong, Janet, Mi, Meng, Kehui, Soyoung, Bongseok, Marcela, and many others; thank you!

I would like to express my deepest gratitude to my dearest family in Korea, Mom, Dad, and my brother. Also, I am really grateful to my great husband David Ha, my precious daughter Ellen, and parents-in-law. Your love and emotional support have been my biggest strength. It made me possible to complete my research work.

Finally, Our heavenly Father, I believe that everything is possible in you. You gave the most cherishable gift to my family and my life. Thank you for blessing me much more than I deserve.
# TABLE OF CONTENTS

List of Figures ................................................................. vii

Chapter 1 Introduction .................................................. 1

Chapter 2 Exponential Families of Symmetric Variate Random Matrices ................................................. 6
   2.1 Definition .............................................................. 6
   2.2 Expected value and Covariance Matrix of $T(Y)$ .............................................................. 7
   2.3 Orthogonal equivariance of the exponential family .......................................................... 11
   2.4 Eigenvalue Parameterization .................................................. 14

Chapter 3 Maximum Likelihood Estimators in the One-sample Problem .................................................... 17
   3.1 Likelihood for estimating $M$ ................................................. 17
   3.2 Inference for $M$: MLE and Asymptotic distribution .................................................. 17
   3.3 Inference for $U$ : MLE and Asymptotic distribution .................................................. 18
   3.4 Inference for $D$ : MLE and Asymptotic distribution .................................................. 22
   3.5 MLE of Eigenvectors and Eigenvalues with Known Multiplicities .................................. 25

Chapter 4 Testing in One-sample Problems ......................................................................................... 27
   4.1 Full matrix Test ............................................................ 27
   4.2 Test of eigenvectors with fixed eigenvalues in a non-increasing order .................................. 28
   4.3 Test of eigenvalues with unordered eigenvectors ............................................................. 28
   4.4 Test of eigenvectors and eigenvalues with known multiplicities ...................................... 29

Chapter 5 Estimation of eigenstructures in the Two-Sample Case ....................................................... 31
   5.1 Estimation under the same nuisance parameter assumption ........................................... 31
   5.2 Estimation under the different nuisance parameters ...................................................... 33

Chapter 6 Testing for eigenstructures in the Two-Sample Case ............................................................ 36
   6.1 Full Matrix test ............................................................ 36
   6.2 Testing for the eigenvectors ................................................ 37
   6.3 Testing for the eigenvalues ................................................ 38

Chapter 7 Estimating Nuisance Parameters ....................................................................................... 40
   7.1 Gaussian distribution with spherical covariance matrix ................................................ 40
      7.1.1 Test of Sphericity of Covariance matrices ........................................................ 40
      7.1.2 Estimating Covariance parameters in the one-sample case ................................... 41
      7.1.3 Estimating Covariance parameters in the two-sample case .................................. 42
   7.2 Matrix Gamma Distribution ......................................................................................... 43
      7.2.1 Estimating shape parameters in the one-sample case ........................................... 43
      7.2.2 Estimating shape parameters in the two-sample case ......................................... 44

Chapter 8 Numerical Study ........................................................................................................... 45
   8.1 Simulation Setting ........................................................ 45
8.1.1 Distribution of $\hat{U}$ and $\hat{D}$ in the one-sample case ........................................ 45
8.1.2 Null distribution of test statistics .......................................................... 45
8.1.3 Statistical Power of Tests ..................................................................... 46
8.1.4 Estimation of nuisance parameters ...................................................... 48
8.2 Simulation Results ..................................................................................... 48
  8.2.1 Distribution of MLEs ........................................................................ 48
  8.2.2 Distribution of P-values .................................................................... 50
  8.2.3 Statistical Power of Tests .................................................................. 52
  8.2.4 Multiplicity test ................................................................................ 60
  8.2.5 Estimation of other parameters ......................................................... 61

Chapter 9 Real Data Example .............................................................................. 65
  9.1 Data description ...................................................................................... 65
  9.2 Model checking ....................................................................................... 65
    9.2.1 Model fit ....................................................................................... 65
    9.2.2 Model checking for the Gaussian model ......................................... 69
  9.3 Analysis .................................................................................................... 71
    9.3.1 Gaussian distribution ................................................................. 71
    9.3.2 Log-normal distribution .............................................................. 72
    9.3.3 Matrix Gamma distribution .......................................................... 73
    9.3.4 P-value Maps ............................................................................... 75

Chapter 10 Summary and Discussion ................................................................. 79
  10.1 Summary ................................................................................................ 79
  10.2 Discussion .............................................................................................. 79

References ........................................................................................................... 81

Appendix ............................................................................................................. 83
  Appendix A Properties needed for the proof of Theorems ......................... 84
    A.1 Derivation of derivative rules in Example 2.2.2 ................................ 84
    A.2 Derivative of a scalar function with respect to a diagonal matrix ........ 85
    A.3 Other algebraic properties ................................................................. 85
**LIST OF FIGURES**

Figure 8.1  The distribution of $\hat{A}$: one-sample Gaussian case, with $n = 200$. ................................................. 49
Figure 8.2  The distribution of $\hat{A}$: one-sample matrix gamma case, with $n = 200$. ................................................. 49
Figure 8.3  The distribution of $\hat{D}$: one-sample Gaussian case, with $n = 200$. ....................................................... 49
Figure 8.4  The asymptotic distribution of $\hat{D}$: one-sample matrix gamma case, with $n = 200$. .......................... 50
Figure 8.5  The exact distribution of $\hat{D}$: one-sample matrix gamma case, with $n = 200$. Compared with gamma distribution. .......................................................... 50
Figure 8.6  Distribution of p-values for hypothesis testing scenarios: one-sample case, with $n_1 = n_2 = 100$. .......................................................... 51
Figure 8.7  Distribution of p-values for hypothesis testing scenarios: two-sample case, with $n_1 = n_2 = 100$. ................................. 51
Figure 8.8  ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Gaussian sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. ........................................ 52
Figure 8.9  ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Wishart sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. ........................................ 53
Figure 8.10 ROC curves for eigenvector changes: one-sample case, with $n = 50$, Gaussian sample. Four unit vectors for rotation matrices are chosen as (a) $a = (1, 0, 0, 0)$, (b) $\theta = \pi/12$ is fixed for all cases. ........................................ 54
Figure 8.11 ROC curves for eigenvector changes: one-sample case, with $n = 50$, Wishart sample. Four unit vectors for rotation matrices are chosen as (a) $a = (1, 0, 0, 0, 0)$, (b) $\theta = \pi/12$ is fixed for all cases. ........................................ 55
Figure 8.12 ROC curves for eigenvalue changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. ........................................ 56
Figure 8.13 ROC curves for eigenvalue changes: two-sample Wishart case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. ........................................ 57
Figure 8.14 ROC curves for eigenvector changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four unit vectors for rotation matrices are chosen as (a) $a = (1, 0, 0, 0, 0)$, (b) $\theta = \pi/12$ is fixed for all cases. ........................................ 58
Figure 8.15 ROC curves for eigenvector changes: two-sample Wishart case, with $n_1 = n_2 = 50$. Four unit vectors for rotation matrices are chosen as (a) $a = (1, 0, 0, 0, 0)$, (b) $\theta = \pi/12$ is fixed for all cases. ........................................ 59
Figure 8.16  ROC curves for multiplicity test: Gaussian and matrix gamma with \( n = 50 \). Four
eigenvalue-changing matrices are chosen as (a) \( \Delta D = \text{Diag}(0.8, 0.8, 0.4) \), (b) \( \Delta D =\)
\( \text{Diag}(0.8, -0.8, 0.4) \), (c) \( \Delta D = \text{Diag}(-0.4, -0.4, 0.4) \), (d) \( \Delta D = \text{Diag}(0.2, 0.2, 0.2) \). 60

Figure 8.17  ROC curves for multiplicity test: Gaussian and matrix gamma with \( n = 50 \). Four
eigenvector-changing matrices are chosen as (a) \( a = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) \), (b) \( a = (1, 0, 0) \), (c) \( a = (0, 1/\sqrt{2}, 1/\sqrt{2}) \), (d) \( a = (1/\sqrt{2}, 0, -1/\sqrt{2}) \). Figure (a) 61

Figure 8.18  The distribution of \( \hat{\sigma}^2 \): one-sample Gaussian case, with \( n = 100 \). 62

Figure 8.19  The distribution of \( \hat{\sigma}^2_{12} \) : two-sample Gaussian case, equal covariance assumption,
with \( n_1 = n_2 = 100 \). 62

Figure 8.20  The distribution of \( \hat{\sigma}_1^2 \) and \( \hat{\sigma}_2^2 \); two-sample Gaussian case, different covariance as-
sumption, with \( n_1 = n_2 = 100 \). 63

Figure 8.21  The distribution of \( \hat{\alpha} \); one-sample Wishart case, with \( n = 100 \). 63

Figure 8.22  The distribution of \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \); two-sample Wishart case, with \( n_1 = n_2 = 100 \). 64

Figure 9.1  Histograms of Mardia Test statistics to check normality 66

Figure 9.2  Histograms of Mardia Test statistics to check log-normality 66

Figure 9.3  Chi-square qq plot at voxel number=1020; Left : Boys, Right : Girls 67

Figure 9.4  Chi-square qq plot at voxel number=1000; Left : Boys, Right : Girls 67

Figure 9.5  Histograms of K-S P-values to check the gamma distribution assumption for boys 68

Figure 9.6  Histograms of K-S P-values to check the gamma distribution assumption for girls 68

Figure 9.7  Box’s M test for the equality of covariance matrices 69

Figure 9.8  The test of spherical covariance : Gaussian distribution 70

Figure 9.9  The test of multiplicity : \( k = 1 \) 71

Figure 9.10  The test of multiplicity : \( k = 2 \) 72

Figure 9.11  The p-value level plot in scale \(-\log_{10}(p)\) for multiplicity test at slice \( z = 49 \); \( k = 1 \),
matrix gamma distribution 72

Figure 9.12  The p-value level plot in scale \(-\log_{10}(p)\) for multiplicity test at slice \( z = 49 \); \( k = 2 \),
matrix gamma distribution 73

Figure 9.13  The distribution of p-values: two-sample Gaussian distribution with arbitrary covariance,
using weighted chi-square distribution 73

Figure 9.14  The distribution of p-values: two-sample log-normal distribution with arbitrary covariance,
using weighted chi-square distribution 74

Figure 9.15  Histogram of \( \log_{10}\frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\hat{\alpha}_1} \) 74

Figure 9.16  Scatterplot of \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) 75

Figure 9.17  The distribution of p-values: two-sample Wishart distribution 75

Figure 9.18  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 40 \); Gaussian distribution 76

Figure 9.19  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 49 \); Gaussian distribution 76

Figure 9.20  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 40 \); log-normal distribution 77

Figure 9.21  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 49 \); log-normal distribution 77

Figure 9.22  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 40 \); matrix gamma distribution 77

Figure 9.23  The p-value level plot in scale \(-\log_{10}(p)\) at slice \( z = 49 \); matrix gamma distribution 78
Chapter 1

Introduction

Consider a $p \times p$ random symmetric matrix $Y$ from some probability density function $f(Y; M)$, where $M$ is the mean parameter of this density. Here, $Y$ and $M$ are in the set of $p \times p$ symmetric matrices, denoted by $S_p$. Our goal is to estimate and test hypotheses about the mean parameter $M$, with some assumptions on the distribution of random matrix $Y$ and restrictions on the parameter space, such that $M$ is in certain subsets of $S_p$. In particular, our research is focused on estimating and testing the eigenvalues and eigenvectors of $M$, when we have restrictions such that the eigenvalues of $M$ are fixed, the set of eigenvectors of $M$ is fixed with consideration for the order, and the eigenvalues of $M$ have some specific multiplicities. For these inference problems, the observations are drawn from one population as a random sample.

In addition, we consider the comparison of two group means, by estimating and testing the eigenvalues and eigenvectors of each group mean parameter $M_1$ and $M_2$. For the two-sample case, we are interested in comparing the eigenstructure of two group means under restrictions for the eigenvalues and eigenvectors, such as the existence of particular multiplicities in eigenvalues, common eigenvalues, or common eigenvectors.

There are several situations where inferences for random symmetric matrix data are of interest. One of the related area is Diffusion Tensor Imaging (DTI), which is based on the measurement of the diffusion of water molecules in the brain tissue (Basser, Mattiello, and LeBihan, 1994). Water molecules diffuse differently along the tissues depending on the type, and the diffusion of water molecules at any given location refers to random displacements of molecules, which is also called Brownian motion. This random movement of water molecules in a given location, which is called “voxel”, is captured by a matrix called “diffusion tensor(DT)” (Basser and Pierpaoli, 1996; Le Bihan et al, 2001). A diffusion tensor can be thought of as the covariance matrix of Gaussian distribution, which models the physical Brownian motion of the water molecules in a voxel. Thus, the set of diffusion tensors from different subjects at a given voxel can be considered as a random sample from the population corresponding to that voxel. In addition, the diagonal elements of a diffusion tensor represent the diffusivity measured along each of the principal axes in 3D images, while the off-diagonal elements are the correlation of random motions between each pair of directions. For this reason, the form of a diffusion tensor is $3 \times 3$ symmetric positive definite matrix. From an anatomical view-
point, eigenvalues of the diffusion tensor represent information about the type of tissue and its condition, and
eigenvectors provide information about the spatial orientation of neural fibers (Basser, Mattiello, and LeBi-
han, 1994). Thus, statistical inference for the eigenstructure of random symmetric matrices can be useful in
the analysis of DTI data (Schwartzman, 2006; Zhu et al., 2007; Schwartzman et al., 2008; Schwartzman et
al., 2010).

Another motivating example can be found in social science, related to principal component analysis (PCA)
and factor analysis (FA) in classical multivariate analysis. PCA is a method for reorienting multivariate data
to explain as much of the information as possible with a small number of new variables called principal
components (PC) (Mardia, Kendt, and Bibby, 1979, page 213; Lattin, Carrol, and Green, 2003, page 121;
Johnson and Wichern, 2007, page 430). In PCA, the eigenvalues of the covariance matrix or correlation ma-
trix represent the amount of information explained by each principal component, while eigenvectors show
us the direction of variability. Factor analysis, which was originally developed by a psychologist (Spearman,
1904; Bartlett, 1937), is a method to describe the covariance relationships among many variables with a
small number of unobservable (latent) variables called factors (Mardia, Kendt, and Bibby, 1979, page 255;
Johnson and Wichern, 2007, page 481). To estimate the factor model, the principal components method
described above is commonly used.

In practice, FA has contributed to advancing psychological research, which has been used extensively as a
data analytic technique for examining patterns of interrelationship, data reduction, hypothesis testing, and so
on (Rummel, 1970; Ford, Mccallum, and Tait, 1986; Molenaar et al., 1992; Fabrigar et al., 1999). Inferences
for eigenstructure of covariance matrices are relevant to FA. For instance, if we want to compare the pattern
of test scores of several different groups in terms of common factors, the effect of common factors can be
examined by testing equality of eigenvectors between the covariance matrices of the groups. Moreover, FA
is also commonly used in macroeconometrics. In dynamic factor models, for example, multivariate time-
series data is handled where the covariance matrix varies over time. As such, these time-varying multiple
covariance matrices can be compared in terms of factors by comparing their eigenvalues and eigenvectors
(Forni et al., 2000; Del Negro and Otrok, 2008).

We are considering techniques used to make inferences about the parametric distribution of the random sym-
metric matrix. Thus, it is helpful to first review some existing distributions for the matrix data. The problem
of inference for matrix-variate data has been treated before in the multivariate statistics literature, however
not as much for their eigenstructure. Information about the moments of specific parametric matrix-variate
distributions and some statistical properties of these distributions are provided in several books and papers;
matrix variate Gaussian distribution for rectangular random matrices (Gupta and Nagar, 1999, page 55),
Wishart distribution (Wishart, 1928; Mardia, Kent, and Bibby, 1979, page 66), Matrix Gamma distribution
(Gupta and Nagar, 1999, page 122), Spherical distributions and Elliptically contoured distributions (Gupta
and Nagar, 1999, page 315 and 322; Gupta and Varga, 1991, page 21). Our work focuses on making infer-
ences for random symmetric matrices by suggesting a distribution family that can embrace some commonly
used distributions for the symmetric matrix data, such as the spherical Gaussian, Wishart and matrix-Gamma
There are several decomposition methods applied to investigate the structure of covariance matrix. We can find several works in the literature that mention decomposition methods used in the estimation of a covariance matrix, such as Variance-correlation Decomposition (Barnard et al., 2000), Spectral Decomposition (Boik, 2002), and Cholesky Decomposition (Pourahmadi, 1999; Smith and Kohn, 2002; Chen and Dunson, 2003). Among these decompositions, Cholesky Decomposition is also commonly used to examine the structure of covariance matrices, especially in Time Series Analysis and Longitudinal Data Analysis. Cholesky Decomposition also provides a statistically meaningful reparameterization of a covariance matrix (Pourahmadi, 1999; Pourahmadi, 2007). In DTI, however, the eigenvalues and eigenvectors of a covariance matrix also have their own anatomical meanings. Since our work is motivated by DTI, we use a spectral decomposition.

There are a few works about inferences for the eigenstructure of \( p \times p \) random symmetric matrices when the sample data is assumed to be from a specific distribution. Mallows (1961) presented work about linear hypotheses involved in testing the eigenvectors of a single matrix in the case of a Gaussian distribution with orthogonally invariant covariance structure. This reference provides the concept of orthogonally invariant covariance structure, which means the distribution of a random matrix \( Y \) is the same as that of \( QYQ^T \) for any orthogonal matrix \( Q \). In addition, maximum likelihood estimator (MLE) and likelihood ratio test (LRT) for eigenvalues and eigenvectors of \( M \) under the assumption of symmetric matrix variate Gaussian distribution with orthogonally invariant covariance are derived in (Schwartzman, 2006; Schwartzman et al., 2008). However, these works do not model positive definite matrix data directly. Instead, a one-to-one matrix log transformation is applied to remove the positive-definite constraint before data analysis (Schwartzman, 2006; Fletcher and Joshi, 2007). After taking the matrix log transformation to the positive definite symmetric matrix data, the domain is expanded to the set of symmetric matrices. We attempt to include the positive definite constraint in the model so that we can include distributions such as the Wishart and matrix-Gamma.

In hypothesis testing problems, the LRT is one of the most commonly used methods. The distribution of the log likelihood ratio (LLR) depends on the geometry of the null and alternative parameter sets. It is well known that if the null set \( M_0 \) is nested in the alternative set \( M_a \), i.e. the null set \( M_0 \) is a subset of \( M_a \) and both sets are algebraic subsets of Euclidean space (such as affine subspaces or differentiable manifolds), then the LLR follows a \( \chi^2 \)-distribution asymptotically as the sample size \( n \to \infty \), with number of degrees of freedom \( d \) equal to the difference between the dimension of the alternative set and that of the null set (Mardia, Kent and Bibby, 1979, page 124; Lehmann and Romano, 2005, page 516). The number of degrees of freedom can also be written as the difference between the number of free parameters in the alternative set and the number of free parameters in the null set. Thus, by embedding the matrices of interest in Euclidean space we can also perform LRT for the parameter matrix. For example, the LLR for the eigenvalues and eigenvectors of \( M \) are asymptotically \( \chi^2 \) distributed when the null and alternative sets are closed embedded submanifolds corresponding to hypotheses about the eigenvalues and eigenvectors of \( M \) (Schwartzman, 2006; Schwartzman et al., 2008). We use this device as well, as we consider the same hypotheses tests but
different distributions.

We have assumed a parametric distribution for the matrix variate data to make inferences for the eigenstructure of symmetric random matrix using MLE and LRT, since this assumption has all the convenience of parametric models in terms of analytic expressions and can be efficiently applied to each of tens of thousands of voxels in DTI data. A symmetric matrix-variate Gaussian distribution with orthogonally invariant covariance can be considered for a symmetric random matrix $Y$. Otherwise, we can assume the distribution of a symmetric random matrix as Wishart distribution, since this is a natural model to consider for covariance matrices. Furthermore, we consider an exponential family, which includes many common distributions such as matrix-variate Gaussian, exponential, matrix gamma, and Wishart distribution (Fisher, 1934; Pitman, 1936; Koopman, 1936). Exponential families for univariate and multivariate data and inference methods for them are well-known (Mardia, Kent, and Bibby, 1979, page 45; Lehmann and Casella, 1998, page 23; Casella and Berger, 2001, page 217). Analogous to other multivariate exponential families, we define a matrix-variate version of exponential family. Analytical calculations are possible if the distribution is assumed to be equivariant under orthogonal transformations. With this exponential family form, we can estimate and test the eigenstructure of $M$ in the general exponential family using MLE and LRT. Then we can apply the result to each specific distribution.

In order to find analytical solutions, we define the concept of orthogonally equivariant (OE) family of distributions. OE family is a group family of probability distributions that is generated by the group action (rotation and reflection) of the orthogonal matrix group. As our problem is focused on the inferences for eigenvalues and eigenvectors, which are related to orthogonal matrices, by assuming that the probability distributions considered are in the OE family, we can make inferences for the eigenstructures of mean parameters more explicitly.

In our problem, the parameter space is, in general, a submanifold of Euclidean space. For instance, requiring the eigenvalue matrix to be diagonal creates a linear submanifold in the Euclidean space of square matrices of size $p$. On the other hand, the constraint that the eigenvector matrix should be orthogonal defines an algebraic curved submanifold. For this reason, we cannot use the usual method for calculating MLEs in a Euclidean space. Instead, we obtain the critical points by finding the score function in two ways: using the projection of the unrestricted gradient onto the manifold or tracing a curve directly on the manifold. Once we have the score function by either method, we obtain critical points by setting the score function equal to 0. These points maximize the likelihood because of convexity, so we get the MLEs. Using the MLEs derived from this procedure, we perform LRTs for the hypotheses related to the constraints on the eigenstructure of the parameter matrix $M$.

The organization of the dissertation is as follows. We first demonstrate the definition and properties of the matrix variate exponential family and give expressions for the moments of this family. Then we derive the MLE and LRT for the eigenvalues and eigenvectors of the mean parameter $M$ with some restrictions on the eigenstructures, based on the OE property of the exponential family. The inference problems are essentially the same as those considered by Schwartzman et al. (2008), but the solutions are given in greater generality,
being applicable in a more general distribution family. Numerical studies are also provided to check the validity of our derived estimators and testing procedures under non-asymptotic conditions. The methods are illustrated in a DTI data set also analyzed by Schwartzman et al. (2010). The goals of the analysis are to compare brain images and find regions of anatomic difference between two groups. Here we use our derived estimators and testing procedures instead of those derived under the Gaussian assumption. We show that changing the modeling assumptions has a substantial impact on the analysis results. All the simulations and data analysis in this article were performed in R.
Chapter 2

Exponential Families of Symmetric Variate Random Matrices

2.1 Definition

Suppose the probability density of a $p \times p$ symmetric random matrix $Y$ can be written in the form

$$f(Y; \Theta) = H(Y)e^{\{\text{tr}(\eta(\Theta)^T(Y)) - K(\eta(\Theta))\}}$$

(2.1)

where $\Theta$ is a parameter matrix. Assume that $K(\eta)$ is twice differentiable. Then, $Y$ is said to belong to an exponential family of distributions, and we call (2.1) the canonical form of the exponential family.

In our research, we focus on estimating and testing the mean parameter $M = E(T(Y))$, where $Y, M, T(Y)$, and $\eta(M)$ are symmetric. Any other parameters, such as variance, scale parameter, shape parameter and so on, are considered as nuisance parameters. Therefore, we rewrite (2.1) as a function of $M$, which has the form

$$f(Y; M) = H(Y)e^{\{\text{tr}(\eta(M)^T(Y)) - K(\eta(M))\}}.$$ 

(2.2)

Example 2.1.1. In the case of symmetric matrix variate Gaussian distribution with spherical covariance (Schwartzman et al, 2008), denoted as $Y \sim N_{pp}(M, \sigma^2 I_q)$, the probability density function is

$$f(Y; M) = \frac{1}{(2\pi)^{\frac{q}{2}} \sigma^q} e^{-\frac{1}{2\sigma^2} \text{tr}(Y-M)^2} = H(Y)e^{\frac{1}{2\sigma^2} \text{tr}(2YM-M^2)}$$

where $q = \frac{p(p+1)}{2}$. As the pdf of Gaussian distribution can be written in the canonical form, it is an exponential family. In this case, we have

$$H(Y) = \frac{1}{(2\pi)^{\frac{q}{2}} \sigma^q} e^{-\frac{1}{2\sigma^2} \text{tr}(Y^2)}, \quad \eta(M) = \frac{M}{\sigma^2}, \quad T(Y) = Y, \quad K(\eta(M)) = \frac{1}{2\sigma^2} \text{tr}(M^2).$$
Example 2.1.2. In the case of matrix gamma distribution, we have the pdf

$$f(Y; M) = \frac{|M|^{-\alpha}}{\alpha^{-\alpha p} \Gamma_p(\alpha)} |Y|^{\alpha - \frac{p+1}{2}} e^{\alpha \text{tr}(-M^{-1}Y)} = H(Y) e^{\alpha \{ -\text{tr}(M^{-1}Y) - \log |M| \}}.$$  

where $\Gamma_p(\alpha)$ is the multivariate gamma function defined as

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

As the pdf of matrix gamma distribution can be rewritten as the canonical form of exponential family, it is an exponential family. In this case, we have

$$H(Y) = \frac{|Y|^{\alpha - \frac{p+1}{2}}}{\alpha^{-\alpha p} \Gamma_p(\alpha)}, \quad \eta(M) = -\alpha M^{-1}, \quad T(Y) = Y, \quad K(\eta(M)) = \alpha \log |M|.$$  

2.2 Expected value and Covariance Matrix of $T(Y)$

To obtain the general form of the moments of $T(Y)$ in the exponential family, let us start from the definition of probability density function. Before we start, we will use a vectorization operator for any symmetric matrix $X$ defined in (Schwartzman et al, 2008), such that

$$\text{vecd}(X) = (\text{diag}(X)^T, \sqrt{2}\text{offdiag}(X)^T)^T$$

where the $\text{diag}(X)$ operator is defined as the vector whose elements are the diagonal elements of $X$, and $\text{offdiag}(X)$ is defined as the vector whose elements are the off-diagonal elements of $X$. Using this operator, the correlation between diagonal elements and off-diagonal elements of a matrix is seen more clearly. In addition, by multiplying $\sqrt{2}$ by the off-diagonal elements, the equation $\text{vecd}(A)^T \text{vecd}(B) = \text{Tr}(AB)$ is satisfied, so we can easily present the trace of a matrix product as an inner product of two vectors. We can also connect this definition to $\text{vec}$ operator, by choosing a duplication matrix $D$ such that $\text{vec}(Y) = D\text{vecd}(Y)$. For example, we have a duplication matrix $D$ for $p = 3$ such that

$$D = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 & 1/\sqrt{2} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2}
\end{bmatrix}^T.$$
Such a duplication matrix always exists for any $p$, although it is generally not unique.

**Theorem 2.1.** The expected value and covariance matrix of $T(Y)$ are as follows.

(i) $M = E[T(Y)] = \frac{\partial K(\eta(M))}{\partial \eta(M)}$.

(ii) $\text{Cov}[\text{vecd}(T(Y))] = \frac{\partial^2 K(\eta(M))}{\partial \text{vecd}(\eta(M)) \partial \text{vecd}(\eta(M))^T}$.

**Proof.** We will use two methods to prove this theorem.

**Method 1.** Using the moment generating function

(i) The moment generating function can be obtained as

$$M_T(S) = E\{e^{\text{tr}(ST(Y))}\}$$

$$= \int e^{\text{tr}(ST(Y))} H(Y) e^{\text{tr}(\eta T(Y)) - K(\eta)} dY$$

$$= e^{K(S+\eta) - K(\eta)} \int H(Y) e^{\text{tr}((S+\eta)T(Y)) - K(S+\eta)} dY$$

$$= e^{K(S+\eta) - K(\eta)}.$$

where the domain of $S$ is a set of symmetric matrices. Taking derivative of the moment generating function with respect to $S$, and substituting $S = 0$, we have

$$E(T(Y)) = \frac{\partial M_T(S)}{\partial S} \bigg|_{S=0} = \frac{\partial}{\partial S} e^{K(S+\eta) - K(\eta)} \bigg|_{S=0}$$

$$= e^{K(S+\eta) - K(\eta)} \frac{\partial K(S+\eta)}{\partial S} \bigg|_{S=0}$$

$$= \frac{\partial K(S+\eta)}{\partial S} \bigg|_{S=0}.$$

Setting $S + \eta = \nu$, $\frac{\partial \text{vecd}(\nu)}{\partial \text{vecd}(S)} = I_q$ and $\frac{\partial \text{vecd}(\eta)}{\partial \text{vecd}(\nu)} = I_q$. Thus, $\frac{\partial K(S+\eta)}{\partial \text{vecd}(S)} = \frac{\partial K(S+\eta)}{\partial \text{vecd}(\eta)} = \frac{\partial \text{vecd}(\nu)}{\partial \text{vecd}(\eta)}$. Using this fact, we have $\frac{\partial K(S+\eta)}{\partial S} = \frac{\partial K(S+\eta)}{\partial \eta}$. Thus,

(ii) To find the variance, we need the second moment. Let us compute the derivative of the moment generating function with respect to $\text{vecd}(S)$ instead of $S$. Then, we have

$$\frac{\partial M_T(S)}{\partial \text{vecd}(S)} = \frac{\partial}{\partial \text{vecd}(S)} e^{K(S+\eta) - K(\eta)}$$

$$= e^{K(S+\eta) - K(\eta)} \frac{\partial K(S+\eta)}{\partial \text{vecd}(S)}.$$

To obtain the second moment, take the derivative of $\frac{\partial M_T(S)}{\partial \text{vecd}(S)}$ with respect to $\text{vecd}(S)^T$. 

8
Another way of proving is to use the fact that

\[
\frac{\partial^2 M_T(S)}{\partial \vec{c}(S) \partial \vec{c}(S)^T} = \frac{\partial^2}{\partial \vec{c}(S) \partial \vec{c}(S)^T} [e^{K(S+\eta) - K(\eta)}] = e^{K(S+\eta) - K(\eta)} \left\{ \frac{\partial K(S + \eta)}{\partial \vec{c}(S)} \frac{\partial K(S + \eta)}{\partial \vec{c}(S)^T} \right\} + e^{K(S+\eta) - K(\eta)} \frac{\partial^2 K(S + \eta)}{\partial \vec{c}(S) \partial \vec{c}(S)^T}
\]

Using the relationship \( \frac{\partial K(S+\eta)}{\partial S} = \frac{\partial K(S+\eta)}{\partial \eta} \) shown in part (i), we have

\[
\frac{\partial^2 M_T(S)}{\partial \vec{c}(S) \partial \vec{c}(S)^T} = e^{K(S+\eta) - K(\eta)} \left\{ \frac{\partial K(S + \eta)}{\partial \vec{c}(\eta)} \frac{\partial K(S + \eta)}{\partial \vec{c}(\eta)^T} \right\} + e^{K(S+\eta) - K(\eta)} \frac{\partial^2 K(S + \eta)}{\partial \vec{c}(\eta) \partial \vec{c}(\eta)^T}
\]

From the above second derivative, we have

\[
E\{\vec{c}(T(Y))\vec{c}(T(Y))^T\} = \frac{\partial^2 M_T(S)}{\partial \vec{c}(S) \partial \vec{c}(S)^T} \bigg|_{S=0} = \frac{\partial K(\eta)}{\partial \vec{c}(\eta)} \frac{\partial K(\eta)}{\partial \vec{c}(\eta)^T} + \frac{\partial^2 K(\eta)}{\partial \vec{c}(\eta) \partial \vec{c}(\eta)^T}.
\]

Thus, the covariance matrix of \( T(Y) \) is

\[
\text{Cov}[\vec{c}(T(Y))] = E\{\vec{c}(T(Y))\vec{c}(T(Y))^T\} - E\{\vec{c}(T(Y))\}E\{\vec{c}(T(Y))\}^T = \left\{ \frac{\partial K(\eta)}{\partial \vec{c}(\eta)} \frac{\partial K(\eta)}{\partial \vec{c}(\eta)^T} \right\} + \frac{\partial^2 K(\eta)}{\partial \vec{c}(\eta) \partial \vec{c}(\eta)^T} - \left\{ \frac{\partial K(\eta)}{\partial \vec{c}(\eta)} \frac{\partial K(\eta)}{\partial \vec{c}(\eta)^T} \right\} = \frac{\partial^2 K(\eta)}{\partial \vec{c}(\eta) \partial \vec{c}(\eta)^T}.
\]

**Method 2.** Using the unit-mass property of pdf.

Another way of proving is to use the fact that \( \int f(Y; \Theta) dY = 1 \).

(i) We have

\[
\int h(Y) e^{\{\mu(\eta(M))T(Y)\} - K(\eta(M))} dY = 1.
\]

Drawing the function of parameters from the integral, we have

\[
e^{-K(\eta(M))} \int H(Y) e^{\{\mu(\eta(M))T(Y)\}} dY = 1.
\]

From the above equation, we can derive \( \int H(Y) e^{\{\mu(\eta(M))T(Y)\}} dY = e^{K(\eta(M))} \). Setting

\[F(\eta(M)) = \int H(Y) e^{\{\mu(\eta(M))T(Y)\}} dY,\]

we have \( K(\eta(M)) = \log F(\eta(M)) \).
Next, we take the derivative with respect to $\eta(M)$ on both sides. For the convenience, we denote $\eta = \eta(M)$.

$$\frac{\partial K}{\partial \eta} = \frac{\partial}{\partial \eta} [\log F(\eta)]$$

$$= \frac{1}{F(\eta)} \frac{\partial F(\eta)}{\partial \eta}$$

$$= \frac{1}{F(\eta)} \int T(Y) H(Y)e^{(\text{tr}(\eta^T(Y)))} dY$$

$$= \frac{\int T(Y) H(Y)e^{(\text{tr}(\eta^T(Y)))} dY}{F(\eta)} e^{-K(\eta)}$$

$$= \frac{\int T(Y) H(Y)e^{(\text{tr}(\eta^T(Y)))-K(\eta)}}{1} dY$$

$$= \mathbb{E}[T(Y)] = M$$

(ii) From (i), we have $\frac{\partial K}{\partial \text{vecd}(\eta)^T} = \text{vecd}(M)^T$. Therefore, let us set

$$\frac{\partial^2 K}{\partial \text{vecd}(\eta)^T \partial \text{vecd}(\eta)^T} = \frac{\partial}{\partial \text{vecd}(\eta)^T} \left[ \frac{\partial}{\partial \text{vecd}(\eta)^T} [\log F(\eta)] \right]$$

$$= \frac{\partial}{\partial \text{vecd}(\eta)^T} \left[ \frac{1}{F(\eta)} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} \right]$$

$$= -\frac{1}{F(\eta)^2} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} + \frac{1}{F(\eta)} \frac{\partial^2 F(\eta)}{\partial \text{vecd}(\eta)^T \partial \text{vecd}(\eta)^T}$$

The first term can be simplified as

$$\frac{1}{F(\eta)^2} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} = \frac{1}{F(\eta)} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T} \frac{\partial F(\eta)}{\partial \text{vecd}(\eta)^T}$$

$$= -\mathbb{E}\{\text{vecd}(T(Y))\} \mathbb{E}\{\text{vecd}(T(Y))\}^T$$

using the same logic in (i). Next, the second term can be rewritten as

$$\frac{1}{F(\eta)} \frac{\partial^2 F(\eta)}{\partial \text{vecd}(\eta)^T \partial \text{vecd}(\eta)^T} = \frac{1}{F(\eta)} \int \text{vecd}(T(Y)) \text{vecd}(T(Y))^T H(Y)e^{\text{vecd}(\eta)^T \text{vecd}(T(Y))} dY$$

$$= \frac{\int \text{vecd}(T(Y)) \text{vecd}(T(Y))^T H(Y)e^{\text{vecd}(\eta)^T \text{vecd}(T(Y))} dY}{F(\eta)} e^{-K(\eta)}$$

$$= \frac{\int \text{vecd}(T(Y)) \text{vecd}(T(Y))^T H(Y)e^{\text{vecd}(\eta)^T \text{vecd}(T(Y)) - K(\eta)}}{1} dY$$

$$= \mathbb{E}\{\text{vecd}(T(Y)) \text{vecd}(T(Y))^T \}$$

By combining these two terms, we have
In the case of symmetric matrix variate Gaussian distribution with spherical covariance shown in Example 2.1.1, we have \( \eta(M) = \frac{M}{\sigma}, T(Y) = Y \) and \( K(\eta(M)) = \text{tr}(\frac{M^2}{2\sigma}) = \text{tr}(\frac{\sigma^2}{2}) = \frac{\sigma^2}{2}\langle \text{vecd}(\eta)^T \text{vecd}(\eta) \rangle \). Therefore, \( E(T(Y)) = \frac{dK}{d\eta} = \sigma^2 \eta = \mathcal{M} \) and \( \text{Cov}(\text{vecd}(T(Y))) = (\frac{d^2K}{d\eta^2}) \langle \text{vecd}(\eta)^T \text{vecd}(\eta) \rangle = \sigma^2 I_q. \)

We use the derivative rules \( \frac{d\text{vecd}(\eta^{-1})}{d\text{vecd}(\eta)^T} = -D^T(\eta^{-1} \otimes \eta^{-1})D \), described in Appendix A. The covariance matrix has the form

\[
\text{Cov}(\text{vecd}(Y)) = \frac{\partial^2 K}{\partial \text{vecd}(\eta) \partial \text{vecd}(\eta)^T} = \frac{\partial}{\partial \text{vecd}(\eta)^T} \{-\alpha \text{vecd}(\eta^{-1})\} = -\alpha \{-D^T(\eta^{-1} \otimes \eta^{-1})D\} = \frac{1}{\alpha} D^T(M \otimes M)D.
\]

We have entry-wise variance and covariance formula for the matrix gamma case from other book (Gupta and Nagar, 1999), which is actually derived from the wishart distribution by modifying parameters, such that \( \text{Var}(Y_{ij}) = \frac{1}{\alpha} (\mu^2_{ij} + \mu_{i} \mu_{jj}) \) where \( \mu_{ij} \) is the \((i, j)\) th element of \( M \), and \( \text{Cov}(Y_{ij}, Y_{kl}) = \frac{1}{\alpha^2} (\mu_{ik} \mu_{jl} + \mu_{il} \mu_{jk}) \). The above matrix-variate formula provides the same answer.

### 2.3 Orthogonal equivariance of the exponential family

Orthogonal equivariance is an additional property imposed on the exponential family that facilitates obtaining analytical expressions for the MLEs and LLRs considered in this work. To make this explicit, some definitions and assumptions are given first.

**Definition 2.1.** A square matrix function \( a(X) \) of square matrix argument \( X \) is analytic if \( a(X) \) can be represented as the form of convergent power series such that \( a(X) = \sum_{k=-\infty}^{\infty} c_k X^k \) for some coefficients \( c_k \). In particular, an analytic function is infinitely differentiable.

Some examples of analytic functions are the inverse of \( X \), exponential and logarithm of \( X \), and polynomial functions of \( X \). In particular, an analytic function is infinitely differentiable.
Definition 2.2. For a square matrix $X$,
(i) A function $g(X)$ is orthogonally invariant if $g(X)$ satisfies $g(RXR^T) = g(X)$ for any orthogonal matrix $R$.
(ii) A squared matrix function $g(X)$ is orthogonally equivariant if $g(X)$ satisfies $g(RXR^T) = Rg(X)R^T$ for any orthogonal matrix $R$.

Note that an analytic function $a(X)$ is orthogonally equivariant, since an analytic function satisfies $a(PP^{-1}) = Pa(X)P^{-1}$ for any invertible matrix $P$ and an orthogonal matrix $R$ satisfies that $R^{-1} = R^T$.

Definition 2.3. If a pdf $f(Y; M)$ satisfies $f(Y; M) = f(RYR^T; RM R^T)$ for any orthogonal matrix $R$, the pdf is then said to be an orthogonal equivariant (OE) family.

Assumption 2.1. For a density function $f(Y; M)$ that has the form (2.2),
(i) $K(*)$ and $H(*)$ are orthogonally invariant functions.
(ii) $\eta(*)$ is an analytic function.
(iii) $T(*)$ is an OE family.

Proposition 2.1. Let $f(Y; M)$ have the form (2.2) and suppose that Assumption 2.1 holds. Then, the density of $Y$ is an OE family.

Proof. Let $\tilde{Y} = RY R^T$ for any orthogonal matrix $R$. We will obtain the density function of $\tilde{Y}$ by variable transformation. We have $Y = R^T\tilde{Y} R$ since $R$ is an orthogonal matrix, and we have $d\tilde{Y} = |R|^p |R^T|^pdY = |RR^T|^p dY = d\tilde{Y}$ using the property that if we have $X$ and $Y$ matrices such that $Y = AXB$ with nonsingular constant $p \times p$ matrices $A$ and $B$, $d\tilde{Y} = |A|^p |B|^pdX$ (Mathai, 1997, page 31). Thus, the Jacobian of this transformation is $|J| = |\frac{d\tilde{Y}}{d\tilde{Y}}| = 1$. Therefore, we can find the density of $\tilde{Y}$ as follows. As $T(*)$ is orthogonally equivariant and $H(*)$ is orthogonally invariant,

\[
f(\tilde{Y}) = H(R^T\tilde{Y} R) \exp[\text{tr}\{\eta(M)T(R^T\tilde{Y} R)\} - K\{\eta(M)\}] \ast |J|
\]

where $\tilde{M} = RM R^T$. Note that the density function of $\tilde{Y}$ has the same form as the original one, with the parameter matrix which is rotated with respect to $\tilde{R}$.

It can be seen that both the spherical Gaussian and matrix-Gamma distributions of examples 2.1.1 and 2.1.2 have OE density functions.

Example 2.3.1. From Example 2.1.1, the pdf of symmetric matrix variate Gaussian distribution with spher-
The pdf of matrix gamma distribution is an OE family.

**Example 2.3.2.** In the case of matrix gamma distribution, we have the pdf form

\[ f(Y; M) = H(Y)e^{\alpha(-\text{tr}(M^{-1}Y) - \log |M|)}, \]

with \( H(Y) = \frac{|Y|^{\alpha/2}}{\alpha\Gamma_p(\alpha)} \). Setting \( \tilde{Y} = RYR^T \) for some orthogonal matrix \( R \), we have

(i) \( H(\tilde{Y}) = \frac{|RYR^T|^{\alpha/2}}{\alpha\Gamma_p(\alpha)} = \frac{|Y|^{\alpha/2}}{\alpha\Gamma_p(\alpha)} = H(Y) \).

\( K(R\eta(M)R^T) = K(\eta(RMR^T)) = \alpha \log |RM| = K(\eta(M)) \).

Thus, \( H(Y) \) and \( K(\cdot) \) are orthogonally invariant.

(ii) \( \eta(M) = -\alpha M^{-1} \) is an analytic function.

(iii) \( T(Y) = Y \) is orthogonally equivariant.

Thus, all the three assumptions are satisfied, so the pdf of matrix gamma distribution is an OE family.

**Corollary 2.1.** Let \( W = RT(Y)R^T \) and \( \tilde{M} = RMR^T \) for any orthogonal matrix \( R \). Suppose the pdf of \( Y \) is an OE family. Then,

(i) \( E[W] = \frac{\partial K(\eta(\tilde{M}))}{\partial \eta(\tilde{M})} = \tilde{M} \)

(ii) \( \text{Cov} [\text{vec}(W)] = \frac{\partial^2 K(\eta(\tilde{M}))}{\partial \text{vec}(\eta(\tilde{M}))^T \partial \text{vec}(\eta(\tilde{M}))} \).

**Proof.** From the assumption, we have that the pdf \( f(Y) \) is orthogonally equivariant. Thus, by obtaining the distribution of \( \tilde{Y} = RYR^T \), we can find the mean and covariance of \( W = RT(Y)R^T \). Note that since \( T(Y) \) is a orthogonally equivariant function, \( W = T(\tilde{Y}) \). The pdf of \( \tilde{Y} \) is

\[ f(\tilde{Y}; \tilde{M}) = H(\tilde{Y})e^{\alpha \text{tr}(\eta(\tilde{M})T(\tilde{Y})) - K(\eta(\tilde{M}))}, \]

which also has the exponential family form. Thus, from this pdf, we can get

\[ E(T(\tilde{Y})) = \frac{\partial K(\eta(\tilde{M}))}{\partial \eta(\tilde{M})} = \tilde{M} = RMR^T \]
The following useful lemma establishes the moments of the random matrix $T(Y)$.

Suppose all the assumptions in Corollary 2.1 are satisfied and Lemma 2.1.

Example 2.3.3. In the case of symmetric matrix variate Gaussian distribution with spherical covariance, as shown in Example 2.1.1, we have $\eta(\bar{M}) = \frac{\bar{M}}{\sigma}$. Thus, $E(W) = \frac{\partial K(\eta(\bar{M}))}{\partial \eta(M)} = \sigma^2 \eta(\bar{M}) = \bar{M}$

$$\text{Cov}(\text{vecd}(W)) = \frac{\partial^2 K(\eta(\bar{M}))}{\text{vecd}(\eta(M)) \text{vecd}(\eta(M))} = \sigma^2 I_q.$$ 

Therefore, $E(W) = \frac{\partial K(\eta(\bar{M}))}{\partial \eta(M)} = -\alpha \bar{M}^{-1} = \bar{M}$

$$\text{Cov}(\text{vecd}(W)) = \frac{\partial^2 K(\eta(\bar{M}))}{\text{vecd}(\eta(M)) \text{vecd}(\eta(M))} = \frac{1}{\alpha} D \left( \bar{M} \otimes \bar{M} \right) D$$

Thus, each element of $W$ has the variance $\text{Var}(W_{ij}) = \frac{1}{2\alpha} (\bar{\mu}_{ij}^2 + \bar{\mu}_{ii} \bar{\mu}_{jj})$ where $\bar{\mu}_{ij}$ is the $(i, j)$th element of $\bar{M} = RAMR^T$ and $\text{Cov}(W_{ij}, W_{kl}) = \frac{1}{2\alpha} (\bar{\mu}_{ik} \bar{\mu}_{jl} + \bar{\mu}_{il} \bar{\mu}_{jk})$.

2.4 Eigenvalue Parameterization

Until the previous section, the structure of mean parameter $M$ was not specified. Now, consider the eigendstructure of $M$, denoted as $M = UDU^T$, where $D$ is a diagonal matrix whose elements are eigenvalues of $M$ and $U$ is an eigenvector matrix corresponding to $D$. Then, we can rewrite the pdf as a function of $U$ and $D$ as

$$f(Y; U, D) = H(Y) e \left\{ \text{tr}(U\eta(D)U^TT(Y)) - K(\eta(D)) \right\}. \quad (2.4)$$

The following useful lemma establishes the moments of the random matrix $T(Y)$ rotated by the eigenvectors.

Lemma 2.1. Suppose all the assumptions in Corollary 2.1 are satisfied and $W = RT(Y)R^T$. At $R = U^T$, (i) $E(W) = \frac{\partial K(\eta(\bar{M}))}{\partial \eta(M)} \bigg|_{R=U^T} = D$.

(ii) $\text{Cov}(\text{vecd}(W)) = \frac{\partial^2 K(\eta(\bar{M}))}{\text{vecd}(\eta(M)) \text{vecd}(\eta(M))} \bigg|_{R=U^T} = V$ is a diagonal matrix. i.e. all the elements in $W = U^TT(Y)U$ are mutually uncorrelated.
Proof. (i) We will prove this using Corollary 2.1. By proposition 2.1., the pdf of $Y$ is orthogonally equivariant. Therefore, we have

$$\frac{\partial K(\eta(\tilde{M}))}{\partial \eta(\tilde{M})} = \tilde{M}.$$ 

Assuming that $M = UDU^T$ by eigen decomposition, the expected value of $W$ can be obtained by evaluating the above equation at $R = U^T$. Thus, we have

$$\frac{\partial K(\eta(\tilde{M}))}{\partial \eta(\tilde{M})}\bigg|_{R=U^T} = \tilde{M}|_{R=U^T} = U^T MU = U^T UDU^T U = D.$$ 

(ii) From (i), we can derive that $$\frac{\partial K(\eta(\tilde{M}))}{\partial \text{vec}(\tilde{M})} = \text{vec}(\tilde{M}).$$ Therefore, this equation holds;

$$\frac{\partial^2 K(\eta(\tilde{M}))}{\partial \text{vec}(\eta(\tilde{M})) \partial \text{vec}(\eta(\tilde{M}))^T} = \frac{\partial \text{vec}(\tilde{M})}{\partial \text{vec}(\eta(\tilde{M}))^T}$$

Using the property of derivatives, $$\frac{\partial \text{vec}(\tilde{M})}{\partial \text{vec}(\eta(\tilde{M}))^T} = \left[ \frac{\partial \text{vec}(\eta(\tilde{M}))}{\partial \text{vec}(\eta(\tilde{M}))^T} \right]^{-1}.$$ Here, since the function $\eta$ is analytic, we can use the definition of analytic function to find the derivative. Recall

$$\eta(M) = \sum_{k=-\infty}^{\infty} c_k M^k.$$ 

For $k > 0$, we have

$$d(M^k) = dM \cdot M^{k-1} + M \cdot dM \cdot M^{k-2} + \ldots + M^{k-1} \cdot dM$$

$$\text{vec}(d(M^k)) = (M^{k-1} \otimes I)\text{vec}(dM) + (M^{k-2} \otimes M)\text{vec}(dM) + \ldots + (I \otimes M^{k-1})\text{vec}(dM)$$

$$\frac{\partial \text{vec}(M^k)}{\partial \text{vec}(M)^T} = (M^{k-1} \otimes I) + (M^{k-2} \otimes M) + \ldots + (I \otimes M^{k-1})$$

using matrix differential and property of vec operator.

Starting from the equation that $M^k \cdot M^{-k} = I$, we have

$$M^k \cdot M^{-k} = I$$

$$d(M^k)M^{-k} + M^k d(M^{-k}) = 0$$

$$d(M^{-k}) = -M^{-k}d(M^k)M^{-k}$$

$$\text{vec}(d(M^{-k})) = -(M^{-k} \otimes M^{-k})\text{vec}(dM^k)$$

$$\frac{\partial \text{vec}(M^{-k})}{\partial \text{vec}(M^k)^T} = -(M^{-k} \otimes M^{-k}).$$
Now we use the chain rule so that
\[
\frac{\partial \text{vec}(M^{-k})}{\partial \text{vec}(M)^T} = \frac{\partial \text{vec}(M^{-k})}{\partial \text{vec}(M)^T} \cdot \frac{\partial \text{vec}(M)}{\partial \text{vec}(M)^T}.
\]

Next, replacing \( M \) into \( \tilde{M} = UMU^T = D \), we can see that \( \frac{\partial \text{vec}(M^k)}{\partial \text{vec}(M)^T} \) is diagonal for all \( k \). Using a duplication matrix \( D \), we can obtain \( \frac{\partial \text{vec}(\tilde{M})}{\partial \text{vec}(M)^T} \) which is still diagonal for all \( k \). Thus, \( \frac{\partial \text{vec}(\tilde{M})}{\partial \text{vec}(M)^T} \) is diagonal since it can be represented as the product of diagonal matrices. By taking inverse of this, we can get the covariance matrix of \( \text{vec}(\tilde{M}) \), which is diagonal. ■

**Example 2.4.1.** In the case of symmetric matrix variate Gaussian distribution with spherical covariance, we have \( \eta(\tilde{M}) = \frac{\tilde{M}}{\sigma^2}, T(\tilde{Y}) = RYR^T \) and \( K(\eta(\tilde{M})) = \text{tr}(\frac{\tilde{M}^2}{2\sigma^2}) = \text{tr}(\frac{\alpha^2\tilde{\eta}^2}{2}) \), as shown in Example 2.3.3. Taking \( R = U^T \), we have \( \tilde{M} = MU = U^TUDU^T = D \). Therefore, \( E(W) = \sigma^2 \eta(D) = D \) and \( \text{Cov}(\text{vec}(W)) = \sigma^2 I_q \).

**Example 2.4.2.** In the case of matrix gamma distribution, we have \( \eta(\tilde{M}) = -\alpha \tilde{M}^{-1}, T(\tilde{Y}) = RYR^T \), and \( K(\eta(\tilde{M})) = \alpha \log |\tilde{M}| = \alpha \{ p \log \alpha - \log |-\tilde{\eta}| \} \), as shown in Example 2.3.4. Taking \( R = U^T \), we have \( \tilde{M} = MU = U^TUDU^T = D \). Therefore, \( E(W) = -\alpha \eta(D)^{-1} = D \) and the covariance matrix has the form
\[
\text{Cov}(\text{vec}(W)) = \frac{1}{\alpha} D^T (D \otimes D) D.
\]
Thus, we can see that the covariance of \( W \) is a diagonal matrix. In addition, each element of \( W \) has the variance \( \text{Var}(w_{ij}) = \frac{1}{2\alpha}(d_{ij}^2 + d_{ii}d_{jj}) \) where \( d_{ij} \) is the \((i,j)\)th element of \( D \). By replacing \( M \) to \( D \), this entry-wise formula is obtained from Example 2.2.2. In addition, the above formula provides the same answer.
Chapter 3

Maximum Likelihood Estimators in the One-sample Problem

3.1 Likelihood for estimating $M$

Let $Y_1, \ldots, Y_n$ be a random sample of size $n$ from the density (2.2). In terms of this sample with size $n$, the likelihood function with respect to $M$ is

$$l(M; Y_1, \ldots, Y_n) \propto \text{tr} \left\{ \eta(M) \sum_{i=1}^{n} T(Y_i) \right\} - nK(\eta(M))$$

$$= n \text{tr} \left\{ \eta(M) \bar{T}(Y) \right\} - nK(\eta(M))$$

where $\bar{T}(Y) = \frac{1}{n} \sum_{i=1}^{n} T(Y_i)$. Here, we are interested in the eigenstructure of the mean parameter. Replacing $M$ as $M = UDU^T$, we have the likelihood with respect to $U$ and $D$,

$$l(M; Y_1, \ldots, Y_n) \propto n \text{tr} \left\{ \eta(M) \bar{T}(Y) \right\} - nK(\eta(M))$$

$$= n \text{tr} \left\{ U\eta(D)U^T \bar{T}(Y) \right\} - nK(\eta(D))$$

where the simplified expression in the second row follows from Assumption 2.1. Note that the last term does not depend on $U$. Some matrix properties that are necessary to derive the estimates of $M$, $U$, and $D$ are given in Appendix A.3.

3.2 Inference for $M$: MLE and Asymptotic distribution

To fix ideas, we first present the inference for the mean parameter $M$, without any restriction on the eigenstructure. Using the central limit theorem, we can obtain the following result.

**Theorem 3.1.** Let $Y_1, \ldots, Y_n$ be i.i.d. samples from the pdf $f(Y; M)$ in (2.2). If there is no restriction for
estimating $M$,
(i) The MLE of $M$ is $\hat{M} = \bar{T}(Y)$.
(ii) The asymptotic distribution of $\hat{M}$ is
\[
\sqrt{n}(\text{vecd}(\hat{M}) - \text{vecd}(M)) \rightarrow^d N(0, \text{Cov}(\text{vecd}(T(Y))))
\]
where $\text{Cov}(\text{vecd}(T(Y)))$ is the covariance matrix of $\text{vecd}(T(Y))$ in the exponential family given in Theorem 2.1.

Proof. (i) Starting from the likelihood with respect to $M$, we have
\[
l(M; Y_1, \ldots, Y_n) \propto \text{tr}(\eta(M)\bar{T}(Y)) - nK(\eta(M)).
\]
Taking a derivative with respect to $\eta(M)$, we have
\[
\frac{dl(M; Y_1, \ldots, Y_n)}{d\eta} = \frac{d}{d\eta}n\text{tr}(\eta(M)\bar{T}(Y)) - nK(\eta(M))
\]
\[
= n\bar{T}(Y) - n\frac{d}{d\eta}K(\eta(M))
\]
Setting this gradient function equal to 0, we have $\frac{d}{d\eta}K(\eta(M)) = \bar{T}(Y)$. As $\frac{dK(\eta(M))}{d\eta} = E[T(Y)] = M$ from Theorem 2.1, the MLE of $M$ is $\hat{M} = \bar{T}(Y)$.

(ii) Since the estimator for $M$ is a sample mean of the data, we can apply the central limit theorem to find the asymptotic distribution of $\hat{M}$. Then, the asymptotic distribution is
\[
\sqrt{n}(\text{vecd}(\hat{M}) - \text{vecd}(M)) \rightarrow^d N(0, \text{Cov}(\text{vecd}(T(Y))))
\]
where $\text{Cov}(\text{vecd}(T(Y)))$ is a covariance matrix of $\text{vecd}(T(Y))$ in the exponential family, obtained in Theorem 2.1. ■

3.3 Inference for $U$: MLE and Asymptotic distribution

Now, we focus on the inference for the eigenstructure. According to the restrictions for the structure of the mean parameter $M = UDU^T$, we have different estimates. The following result gives the MLE of the eigenvectors and its asymptotic distribution.

Theorem 3.2. Let $Y_1, \ldots, Y_n$ be i.i.d. samples from the pdf $f(Y; M)$ in (2.2) and assume the pdf of $Y$ is an OE family. Suppose that $D$ is fixed as $D_0$, with diagonal entries in a non-increasing order, with $k$ distinct
eigenvalues and corresponding multiplicities \( m_j, j = 1, ..., k \).

(i) Let the decomposition \( \bar{T}(Y) = V \Lambda V^T \) be chosen so that the \( \Lambda \) and \( D_0 \) have their diagonal entries in the same rank order. Then, the MLE of \( U \) is \( \hat{U} = V Q \), where \( Q \in \mathbb{O}_p \) satisfies the condition that \( Q D_0 Q^T = D_0 \), i.e. \( Q \) is a block diagonal matrix with the orthogonal blocks of size \( m_j \). If the diagonal entries of \( D_0 \) are distinct, then \( Q \) has diagonal entries \( \pm 1 \).

(ii) Suppose all the elements of \( D_0 \) are distinct so \( M \) has unique eigenvectors \( U \) up to sign. Let \( \hat{U} \) be chosen to minimize the norm \( ||U \hat{U}^T - I_p|| \) and let \( \hat{A} = \log(U^T \hat{U}) \in \mathbb{A}_p \) be a \( p \times p \) antisymmetric matrix. Then, as the sample size \( n \) gets larger, the off-diagonal entries of \( \hat{A} \) denoted by \( \{\hat{a}_{ij}\}_{i \neq j} \) are asymptotically independent, and the asymptotic distribution of \( \{\hat{a}_{ij}\}_{i \neq j} \) is given by

\[
\frac{\sqrt{n\hat{a}_{ij}}}{2\sqrt{[\{\eta(D_0)\}_i - \eta(D_0)\}_j]^2\text{Var}(w_{ij})^{-1}}} \rightarrow_d N(0, 1)
\]

where \( w_{ij} \) is the \((i, j)\)th element of \( W = U^T \bar{T}(Y)U \).

Proof. (i) Starting from the log likelihood with respect to \( U \) and \( D \), we have

\[
l(U, D; Y_1, ..., Y_n) \propto n \text{tr}(U \eta(D)U^T \bar{T}(Y)) - n K(\eta(D)).
\]

As \( K(\eta(D)) \) is a function of \( D \) only, the second term is not affected by the change of \( U \). Therefore, from the likelihood, we can see that \( \hat{U} \) maximizes the function

\[
g(U) = n \text{tr}\{U \eta(D_0)U^T \bar{T}(Y)\}
\]

with the constraint that \( U \) is an orthogonal matrix. Note that we assumed \( D = D_0 \), which is fixed. Then, the critical points of \( g(U) \) with this restriction are the points where the gradient of \( g(U) \) is orthogonal to the tangent space, which has the form of \( \hat{U} = UA \) for \( A \in \mathbb{A}_p \), the set of antisymmetric matrices.

To find these points, we first obtain the score function in two ways: using the projection of the gradient into the tangent space or setting the curve that is in our manifold. Then, setting the score function equal to zero, we can find the critical points.

**Method 1.** Using the projection of gradient function to the tangent space.

The gradient of \( g(U) \) with respect to \( U \) is

\[
\nabla g(U) = \frac{\partial g(U)}{\partial U} = n \frac{\partial}{\partial U} \text{tr}\{U \eta(D_0)U^T \bar{T}(Y)\} = 2n(\bar{T}(Y)U \eta(D_0) + \bar{T}(Y)U \eta(D_0)) = 2n(\bar{T}(Y)U \eta(D_0)).
\]

Now, we will project our gradient of \( g(U) \) into the tangent space. This means that we need to find \( S \) such
that \( S = \text{argmin}_{A \in \mathcal{A}_p} \| \nabla g(U) - UA \|^2 \). Simplifying this norm, we have

\[
\| \nabla g(U) - UA \|^2 = \| 2n(\bar{T}(Y)U \eta(D_0)) - UA \|^2 \\
= \| U(2nU^T \bar{T}(Y)U \eta(D_0) - A) \|^2 \\
= \text{tr}(U(2nU^T \bar{T}(Y)U \eta(D_0) - A)(2nU^T \bar{T}(Y)U \eta(D_0) - A)^TU^T) \\
= \| 2nU^T \bar{T}(Y)U \eta(D_0) - A \|^2
\]

Set \( \bar{W} = U^T \bar{T}(Y)U \). This is the projection of the gradient onto the tangent space, which is the score function.

**Method 2.** Using the curve on the manifold of parameters.

The other approach is to find the score, as the unique tangent vector whose inner product with any direction \( A \) is the directional derivative in that direction. To get the gradient of \( g(U) \) with the constraint, set a curve \( Q(t) = Ue^{\hat{A}t} \) such that \( \hat{A} \in \mathcal{A}_p \), which passes through \( U \) in the direction of \( \hat{A} \). Then, the function \( g(U) \) restricted to the curve is represented as

\[
g(Q(t)) = n[\text{tr}(Ue^{\hat{A}t} \eta(D_0)e^{\hat{A}t}U^T \bar{T}(Y))]
\]

Taking the derivative of \( g(Q(t)) \) with respect to \( t \) and evaluating at \( t = 0 \), we get the directional derivative in the direction of \( A \) as follows.

\[
\frac{dg(Q(t))}{dt} |_{t=0} = n \frac{d}{dt}[\text{tr}(Ue^{\hat{A}t} \eta(D_0)e^{\hat{A}t}U^T \bar{T}(Y))]|_{t=0} \\
= n \text{tr} [\dot{A}e^{\hat{A}t} \eta(D_0)e^{\hat{A}t}U^T \bar{T}(Y)U + e^{\hat{A}t} \eta(D_0)e^{\hat{A}t} \dot{A}^T U^T \bar{T}(Y)U]_{t=0} \\
= n \text{tr} [\dot{A}e^{\hat{A}t} \eta(D_0)e^{\hat{A}t} \bar{W} + e^{\hat{A}t} \eta(D_0)e^{\hat{A}t} \dot{A} \bar{W}]_{t=0} \quad \text{(Set } \bar{W} = U^T \bar{T}(Y)U) \\
= n \text{tr} [\dot{A} \eta(D_0) \bar{W} + \eta(D_0) \dot{A} \bar{W}] \\
= n \text{tr} [\dot{A} \eta(D_0) \bar{W} - \dot{A} \bar{W} \eta(D_0)] \quad \text{As } \dot{A}^T = -\dot{A} \\
= n \text{tr} [\dot{A} (\eta(D_0) \bar{W} - \bar{W} \eta(D_0))]
\]

This is the inner product of \( \dot{A} \) and \( S = n(\eta(D_0) \bar{W} - \bar{W} \eta(D_0)) \). Thus, \( S \) is the score function for \( A \), which is the same with the result from the first approach.

**Finding the MLE.** Setting the score function equal to 0, we have the condition that
\[ \hat{U}^T \hat{T}(Y) \hat{U} \eta(D_0) = \eta(D_0) \hat{U}^T \hat{T}(Y) \hat{U}. \]

As \( \hat{T}(Y) \) is a symmetric matrix, it can be decomposed as \( \hat{T}(Y) = VA \hat{V}^T \) using the eigen decomposition. Then, the equation above can be written as
\[ U^T V A \hat{V}^T U \eta(D_0) = \eta(D_0) U^T V A \hat{V}^T U. \]

Set \( Q = V^T U \). \( Q \) is an orthogonal matrix because \( QQ^T = V^T UU^T V = I \). Then, the above equation is \( Q \Lambda Q^T \eta(D_0) = \eta(D_0) \Lambda Q^T \). By premultiplying \( Q^T \) and postmultiplying \( Q \) both sides, we get \( \Lambda Q^T \eta(D)Q = Q^T \eta(D_0)Q \Lambda \). From this equation, we can find that \( \Lambda Q^T \eta(D_0)Q \) is symmetric and \( \Lambda \) and \( Q^T \eta(D_0)Q \) commute. In addition, \( \Lambda \) is a diagonal matrix with the elements in non-increasing order. Therefore, using proposition 5 in Appendix A, \( Q^T \eta(D_0)Q \) must be a diagonal matrix. This condition is simplified because \( \eta \) is an analytic function. \( Q \eta(D_0)Q^T = \eta(QD_0Q^T) \) and so \( \eta(QD_0Q^T) \) is also diagonal. To make it diagonal, \( QD_0Q^T \) should be a diagonal matrix. Thus, we can say that every critical point of the likelihood function is of the form \( \hat{U} = VQ \), where \( Q \in \mathbb{O}_p \) satisfies the condition that \( QD_0Q^T \) is a diagonal matrix.

Among several choices of \( Q \), we need to find those critical points that maximize the likelihood. From the likelihood function, we can find that \( \text{tr}(Q^T \Lambda Q \eta(D_0)) \) should be maximized. That means, \( QD_0Q^T \) should have the same order as \( \Lambda \). We assumed that \( D_0 \) and \( \Lambda \) have the same order, so \( Q \) must be a block diagonal matrix with orthogonal blocks, where the blocks correspond to the multiplicity blocks of \( D_0 \). In particular, if the eigenvalues in \( D_0 \) are distinct, then the orthogonal blocks are of size 1, which means that \( Q \) is diagonal with diagonal entries \( \pm 1 \).

(ii) From the gradient of \( g(U) \) in the direction of \( \hat{A} \), we have the score function for \( A \) as \( S = n(\eta(D_0) \hat{W} - \hat{W} \eta(D_0)) \). Thus, the covariance of \( \text{vecd}(S) \) is the Fisher information for \( \text{vecd}(A) \), which can be used to find the asymptotic distribution of \( \hat{A} \). Note that \( S \) is an antisymmetric matrix, so all the diagonal entries are 0. On the other hand, each off-diagonal entry \( s_{ij} \) in \( S \) can be represented as \( s_{ij} = n[\{\eta(D_0)_i - \eta(D_0)_j\} \hat{W}_{ij}] \) for \( i \neq j \). Since the expected value of score function is zero, \( E(s_{ij}) = 0 \) for all \( i \neq j \). To find the covariance of \( S \), we need to find the covariance of \( \hat{W} = U^T \hat{T}(Y)U \) first. From Lemma 2.1, we have the covariance of \( W \) by setting \( R = U^T \), and this covariance matrix of \( W \), denoted by \( \mathcal{V} \), is a diagonal matrix. As the covariance matrix of \( \hat{W} = U^T \hat{T}(Y)U = \frac{1}{n} \sum W_i \) is just \( \frac{2}{n} \), all the elements of \( \hat{W} \) are mutually independent. Therefore, all the entries of \( S \) are independent, and the variance of each elements of \( S \) can be represented as \( \text{Var}(s_{ij}) = n^2[\{\eta(D_0)_i - \eta(D_0)_j\}^2 \text{Var}(\hat{W}_{ij})] = n[\{\eta(D_0)_i - \eta(D_0)_j\}^2 \text{Var}(w_{ij})] \).

Note that the Fisher information matrix is \( I[\text{vecd}(A)] = \text{cov}[\text{vecd}(S)] \). Then, the information for the off-diagonal elements can be represented as \( I(\sqrt{2}a_{ij}) = \text{Var}(\sqrt{2}s_{ij}) = 2 \text{Var}(s_{ij}) \). Since the variance of \( \sqrt{2}a_{ij} \) is the inverse of this information, \( \text{Var}(\sqrt{2}a_{ij}) = \frac{1}{2 \text{Var}(s_{ij})} \). So, we can find the asymptotic variance of \( a_{ij} \) as \( \text{Var}(a_{ij}) = \frac{1}{4 \text{Var}(s_{ij})} \). Thus, asymptotically, the entries \( a_{ij} \) are independently \( N(0, [4 \text{Var}(s_{ij})]^{-1}) \).

Note that the variance of each off-diagonal entry \((i, j)\) of \( A \) increases without bound as \( \eta(D_0)_i \) and \( \eta(D_0)_j \) get closer. In the limit, when \( \eta(D_0)_i = \eta(D_0)_j \), the variance goes to infinity, and \( \eta(D_0)_i \) and \( \eta(D_0)_j \) can no
longer be estimated separately because the two parameters become unidentifiable. In fact, in that situation, the dimension of the parameter space goes down by 1.

Example 3.3.1. In the case of symmetric matrix variate Gaussian distribution with spherical covariance, we have \( \bar{T}(Y) = \bar{Y}, \eta(D) = \frac{D}{\sigma^2} \) and \( \bar{Y} \sim N_{pp}(M, \frac{\sigma^4}{n} I_q) \). By using the property that if \( X \sim N_{pp}(M, \frac{\sigma^2}{m} I_q) \), then \( U^T X U \sim N_{pp}(U^T MU, \frac{\sigma^2}{m} I_q) \) for all \( U \in \mathbb{O}_p \), we have \( \bar{W} = U^T \bar{Y} U \sim N_{pp}(D, \frac{\sigma^2}{n} I_q) \). This means that the covariance matrix of \( \bar{W} \) is a diagonal matrix and so the off-diagonal elements are mutually independent. Since the off-diagonal elements \( \bar{W}_{ij} \)’s follows \( N(0, \frac{\sigma^2}{2n}) \) independently, the variance of each entry in score function \( s_{ij} \) is \( \text{Var}(s_{ij}) = \frac{n^2(d_i - d_j)^2}{\sigma^4} \text{Var}(\bar{W}_{ij}) = \frac{n^2(d_i - d_j)^2}{\sigma^4} \frac{\sigma^2}{2n} = \frac{n(d_i - d_j)^2}{2\sigma^2}. \) Thus, \( \text{Var}(a_{ij}) = \{4\text{Var}(s_{ij})\}^{-1} = \frac{\sigma^2}{2n(d_i - d_j)^2} \) and so \( \sqrt{n}a_{ij} \sim N(0, \frac{\sigma^2}{2(d_i - d_j)^2}) \).

Example 3.3.2. In the matrix gamma case, we have \( \bar{T}(Y) = \bar{Y}, \eta(D) = -\alpha D^{-1} \) and \( \bar{Y} \sim MG(n\alpha, \frac{M}{n\alpha}) \). Thus, \( \bar{W} = U^T \bar{Y} U \sim MG(n\alpha, \frac{U^T MU}{n\alpha}) \), which is the same distribution with \( MG(n\alpha, \frac{D}{n\alpha}) \). Using the result of Example 2.2.2, we can calculate that the variance of each element in \( W \) is \( \text{Var}(\bar{W}_{ij}) = \frac{1}{2n\alpha}(D_{ij}^2 + D_{ii}D_{jj}) = \frac{d_{ij}}{2n\alpha}. \) Therefore, the variance of each entry in score function is \( \text{Var}(s_{ij}) = n^2 \{\eta(D)_{i} - \eta(D)_{j}\}^2 \text{Var}(\bar{W}_{ij}) \). Thus, \( \text{Var}(a_{ij}) = \{4\text{Var}(s_{ij})\}^{-1} = \frac{d_{ij}}{2\alpha(d_i - d_j)^2} \), and so the asymptotic distribution of \( a_{ij} \) is \( \sqrt{n}a_{ij} \rightarrow_d N(0, \frac{d_{ij}}{2\alpha(d_i - d_j)^2}) \).

3.4 Inference for \( D : \text{MLE and Asymptotic distribution} \)

Now we will move to the inference for the eigenvalues of mean parameter. Before we start, we need to properly define the derivative of a scalar function with respect to a diagonal matrix. The derivative of a scalar function with respect to a general matrix is defined elementwise, but since \( D \) is a diagonal matrix and we cannot take the derivative with respect to 0 in the off-diagonal part, we cannot use the general derivative. We adopt the following definition for the derivative with respect to the diagonal matrix.

Definition 3.1. \( \text{Diag}(A) \) is the diagonal matrix whose diagonal elements are the diagonal elements of \( A \) if \( A \) is a squared matrix, and \( A \) if \( A \) is a vector.

Definition 3.2. If \( X \) is a diagonal matrix and \( g(*) \) is a scalar function, the partial derivative of \( g(*) \) with respect to \( X \) is defined as

\[
\frac{\partial g(X)}{\partial X} = \text{Diag}(\frac{\partial g(X)}{\partial \text{diag}(X)}).
\]

Theorem 3.3. Let \( Y_1, ..., Y_n \) be i.i.d. samples from (2.2) and suppose that Assumption 2.1 is satisfied. Then,

(i) Let \( \bar{W} = U^T \bar{T}(Y) U \). If we assume that \( U = U_0 \) is fixed, then the MLE of \( D \) is

\[
\hat{D} = \text{Diag}(U_0^T \bar{T}(Y) U_0) = \text{Diag}(\bar{W}).
\]

(ii) If we assume that \( U = U_0 \) is fixed, the asymptotic distribution of diagonal entries of \( \hat{D} = \text{Diag}(\bar{W}) \)
denoted by \( \{ \hat{d}_i \} \) is
\[
\frac{\sqrt{n}(\hat{d}_i - d_i)}{\sqrt{\text{Var}(w_{ii})}} \rightarrow_d N(0, 1)
\]
where \( w_{ii} \) is the \( i \)th diagonal element of \( W = U^T T(Y) U \).

**Proof.** We will use the same two methods to find the score function with respect to our parameter, like as in the case of \( \hat{U} \).

**Method 1.** Using the projection of gradient function to the tangent space
Starting from the likelihood, we have
\[
l(U, D; Y_1, ..., Y_n) \propto n \text{tr}(U \eta(D) U^T \tilde{T}(Y)) - nK(\eta(D)).
\]
Thus, the gradient function with respect to \( D \) is
\[
\nabla l_D = n \frac{\partial}{\partial D} \text{tr}(U \eta(D) U^T \tilde{T}(Y)) - n \frac{\partial K}{\partial D}.
\]

Even though the gradient function can be represented as a function of \( D \), we cannot find the closed form of MLE of \( D \) from this gradient. Instead, we set \( B = \eta(D) \) and find a gradient function of \( l(B) = n \text{tr}(UBU^T \tilde{T}(Y)) - nK(B) \) with respect to \( B \). As \( D \) is a diagonal matrix and \( \eta(\ast) \) is an analytic function, \( B \) is also a diagonal matrix. The gradient of \( l(B) \) can be written as
\[
\frac{\partial l(B)}{\partial B} = n \frac{\partial}{\partial B} \text{tr}(BU^T \tilde{T}(Y)U) - n \frac{\partial K(B)}{\partial B}.
\]
The first term is equal to \( nU^T \tilde{T}(Y)U \) using the derivative of trace, so
\[
\nabla l_B = \frac{\partial l(B)}{\partial B} = n \left[ U^T \tilde{T}(Y)U - \frac{\partial K(B)}{\partial B} \right].
\]
As we know that \( B \) is a diagonal matrix, the tangent space should be the space of diagonal matrices denoted by \( \mathbb{D}_p \). Therefore, \( \hat{B} \) should be the value of the matrix \( B \) that makes the gradient function orthogonal to the tangent space \( \mathbb{D}_p \).

As in the case of \( U \), we can find the projection of the gradient function with respect to \( B \) onto the tangent space \( \mathbb{D}_p \). This means that we need to find \( C = \arg \min_{C \in \mathbb{D}_p} ||\nabla l_B - C||^2 \). Setting \( W = U^T \tilde{T}(Y)U \) and simplifyfying this norm, we have
\[
||\nabla l_B - C||^2 = \left| \left| n \left[ U^T \tilde{T}(Y)U - \frac{\partial K(B)}{\partial B} - C \right] \right| \right|^2 = n^2 \left| \left| W - \frac{\partial K(B)}{\partial B} - C \right| \right|^2.
\]
Thus, by the proposition 7 in the appendix A, $C = \text{Diag}(W - \frac{\partial K(B)}{\partial B})$. By Lemma 2.1, $C = \text{Diag}(W) - D$, which is the score function for $B = \eta(D)$.

**Method 2.** Drawing a curve on the manifold.

The other approach is to set a curve on the manifold of parameters, as in the case of $\hat{U}$. To get the gradient of $l(B)$ with the constraint, set a curve $C(t) = B + \hat{B}t$ such that $\hat{B} \in \mathbb{D}_p$, which passes through $B$ in the direction of $\hat{B}$. Then, the function is represented as

$$l(C(t)) = n[\text{tr}(U(B + \hat{B}t)U^T\bar{T}(Y)) - K(B + \hat{B}t)].$$

Taking the derivative of $l(C(t))$ with respect to $t$ and evaluating at $t = 0$, we get the gradient of $l(B)$ as follows;

$$\frac{dl(C(t))}{dt} |_{t=0} = n \frac{d}{dt}[\text{tr}(U(B + \hat{B}t)U^T\bar{T}(Y)) - K(B + \hat{B}t)] |_{t=0}$$

$$= n \left[ \text{tr}(UBU^T\bar{T}(Y)) - \text{tr} \left( \frac{dK(B + \hat{B}t)dB}{dt} \hat{B} \right) \right] |_{t=0}$$

$$= n \left[ \text{tr}(\bar{W}\hat{B}) - \text{tr} \left( \frac{dK(B)}{dB} \hat{B} \right) \right]$$

$$= n \text{tr} \left\{ \left( \bar{W} - \frac{dK(B)}{dB} \hat{B} \right) \hat{B} \right\}$$

$$= n \text{tr} \left[ \text{Diag} \left\{ \left( \bar{W} - \frac{dK(B)}{dB} \right) \hat{B} \right\} \right]$$

$$= n \text{tr} \left[ \text{Diag}(\bar{W}) - \frac{dK(B)}{dB} \hat{B} \right]$$

(Since $B$ and $\hat{B}$ are diagonal).

This is the inner product of $\hat{B}$ and the score function of $B = \eta(D)$, $S_B = n(\text{Diag}(\bar{W}) - \frac{dK(B)}{dB})$. By Lemma 2.1, we have $\frac{dK(B)}{dB} = D$ and so $S_B = n(\text{Diag}(\bar{W}) - D)$.

**Finding the MLE.** Setting the score function obtained above equal to zero,

$$\frac{\partial K(B)}{\partial B} = D = \text{Diag}(U^T\bar{T}(Y)U).$$

Thus, if we use $U = U_0$, then $\hat{D} = \text{Diag}(U_0^T\bar{T}(Y)U_0) = \text{Diag}(\bar{W})$.

(ii) From the gradient function, we have $\hat{D} = \text{Diag}(\bar{W})$, according to the value of $U$. Thus, the vari-
ance of each diagonal element in $\hat{D}$ is the same as the variance of each diagonal element of $\bar{W}$. Since we have the form of the covariance matrix of $\bar{W} = U^T \mathbf{T}(Y) U$ in Theorem 3.2, we can find the asymptotic distribution of the diagonal elements of $\hat{D}$ as

\[
\frac{\sqrt{n}(\hat{d}_i - d_i)}{\sqrt{\text{Var}(w_{ii})}} \xrightarrow{d} N(0, 1) \quad \blacksquare.
\]

**Example 3.4.1.** In the case of symmetric matrix variate Gaussian distribution with spherical covariance, we have $\bar{T}(Y) = \bar{Y}$. Thus, we have $\hat{D} = \text{Diag}(U^T \bar{Y} U)$. Using the property that if $X \sim N_{pp}(M, \sigma_n^2 I_q)$ then $U^T X U \sim N_{pp}(U^T M U, \sigma_n^2 I_q)$ for all $U \in \mathbb{O}_p$, we have $U^T \bar{Y} U \sim N_{pp}(D, \sigma_n^2 I_q)$. Thus, since we have $\text{Var}(\hat{D}) = \text{Var}(\text{Diag}(\bar{W}))$, we can find the asymptotic distribution of each diagonal element as

\[
\hat{d}_i \sim N \left( d_i, \frac{\sigma_n^2}{n} \right).
\]

**Example 3.4.2.** In the matrix gamma case, we have $\bar{T}(Y) = \bar{Y}$. Thus, we have $\hat{D} = \text{Diag}(U^T \bar{Y} U)$. Using the property of matrix gamma distribution, $W = U^T \bar{Y} U \sim MG(n\alpha, \frac{D}{n\alpha})$. Because the diagonal elements of matrix gamma distribution are independently gamma distributed, we have the exact distribution of each diagonal entries as

\[
\hat{d}_i \sim \text{Gamma} \left( n\alpha, \frac{d_i}{n\alpha} \right).
\]

In addition, since we have $\text{Var}(\hat{D}) = \text{Var}(\text{Diag}(\bar{W}))$, we can find that $\text{Var}(w_{ii}) = \frac{1}{n\alpha} \{d_i\}^2$, as well as the asymptotic distribution of each diagonal element as

\[
\hat{d}_i \sim N \left( d_i, \frac{\{d_i\}^2}{n\alpha} \right).
\]

### 3.5 MLE of Eigenvectors and Eigenvalues with Known Multiplicities

In this section, we assume that we do not know the eigenvalues or the eigenvectors, but the order of eigenvalues is known. To state the results, we first need to define how to perform block averages.

**Definition 3.3.** Define the **block average** of a diagonal matrix $B$ according to the multiplicities $m_1, \ldots, m_k$, as the block diagonal matrix formed by partitioning $B$ into $k$ diagonal blocks of sizes $m_1, \ldots, m_k$ and replacing the diagonal entries in each block by the mean of the diagonal entries in each block. Denote this term as $\text{blk}_{m_1, \ldots, m_k}(B)$. If we denote the original diagonal blocks as $B_{m_1}, \ldots, B_{m_k}$ and cumulative multiplicities $e_0 = 0, e_j = \sum_{i=1}^{j} m_i$, then the $j$th diagonal block of $\text{blk}_{m_1, \ldots, m_k}(B)$ has the form $(1/m_j) \sum_{i=e_{j-1}+1}^{e_j} b_{i} I_{m_j}$.

**Definition 3.4.** For any square matrix $A$, define $\text{Blk}_{m_1, \ldots, m_k}(A) = \text{blk}_{m_1, \ldots, m_k}(\text{Diag}(A))$. Denote the diagonal block of $A$ with respect to the multiplicities $m_1, \ldots, m_k$ as $A_{jj}, j = 1, \ldots, k$. Then, each diagonal
block of $\text{Blk}_{m_1,\ldots,m_k}(A)$ has the form $(1/m_j)tr(A_{jj})I_{m_j}$. If $A$ is a diagonal matrix, then $\text{Blk}_{m_1,\ldots,m_k}(A) = \text{blk}_{m_1,\ldots,m_k}(A)$.

**Theorem 3.4.** Let $M_{m_1,\ldots,m_k}$ be $M_{m_1,\ldots,m_k} = \{ M = UDU^T : U ∈ O_p, D ∈ D_p, d_1 ≥ \ldots ≥ d_k, \text{mult. } m_1,\ldots,m_k \}$ and $\check{T}(Y) = VΛV^T$ be an eigen- decomposition where the diagonal elements of $Λ$ are in non-increasing order. Suppose that $D$ is unknown but the order and multiplicities of $D$ are known. Then, The MLE of $M$ is $\hat{M} = \hat{U}\hat{D}\hat{U}^T$, where

(i) $\hat{U}$ is any matrix that has the form $\hat{U} = VQ$ and $Q ∈ O_p$ satisfies the condition that $Q\hat{D}Q^T = \hat{D}$. $Q$ must be a block diagonal matrix with orthogonal blocks of the size $m_1,\ldots,m_k$ in that order.

(ii) $\hat{D} = \text{Blk}_{m_1,\ldots,m_k}(Λ)$.

Note that $\hat{D}$ is unique regardless of the choice of $Q$.

**Proof.** This scenario is similar to those in Theorem 3.2 and 3.3, except for one condition: we only know the multiplicities of $D$. Given $D$, the MLE of $U$ is the same as in Theorem 3.2, since the form of $U$ only depend on the multiplicities of $D$. Next, to find the MLE of $D$, the tangent space that should be orthogonal to the gradient function is $T = \{ E : E = \text{Blk}_{m_1,\ldots,m_k}(A), A ∈ S_p \}$ instead of $D_p$. Thus, the equation for $D$ should be changed, by using the $\text{Blk}$ operator. Note that $T$ is a subspace of $D_p$, so any matrices in $T$ are also diagonal matrices. Thus, by changing all the equations in Theorem 3.3 using $\text{Blk}$ operator and choosing $Q = I$, i.e. $\hat{U} = V$, we have

$$\text{Blk}(D) = \text{Blk}(\hat{U}^T\check{T}(Y)\hat{U})$$

As the multiplicities of $D$ are fixed, $\text{Blk}(D) = D$. Thus,

$$\hat{D} = \text{Blk}(\hat{U}^T\check{T}(Y)\hat{U}).$$

Now, using the estimated $\hat{U}$ and $\check{T}(Y) = VΛV^T$, the above equation is simplified as

$$\hat{D} = \text{Blk}(Λ).$$

**Example 3.5.1.** In the case of symmetric matrix variate Gaussian distribution with spherical covariance, we have $\check{T}(Y) = \check{Y}$, $η(D) = \frac{D}{σ^2}$ and $\check{Y} ∼ N_{pp}(M, \frac{σ^2}{m}I_q)$. Thus, for the estimated values, we have $\hat{U} = VQ$ and $\hat{D} = \text{Blk}(\hat{U}^T\check{Y}\hat{U})$.

**Example 3.5.2.** In the matrix gamma case, we have $\check{T}(Y) = \check{Y}$, $η(D) = -αD^{-1}$. Thus, for the estimated values, we have $\hat{U} = VQ$ and $\hat{D} = \text{Blk}(\hat{U}^T\check{Y}\hat{U})$. 

26
Chapter 4

Testing in One-sample Problems

4.1 Full matrix Test

First, we start with the hypothesis test of $H_0 : M = M_0$ versus $H_a : M \neq M_0$. The LRT statistic for this hypothesis is

$$2(l_a - l_0) = 2n[\text{tr}((\eta(\bar{T}(Y)) - \eta(M_0))\bar{T}(Y)) + K(\eta(M_0)) - K(\eta(\bar{T}(Y)))],$$

and it follows a $\chi^2_q$ distribution asymptotically as $n$ gets large. This is obtained by noticing that under $H_0$, the MLE for $M$ is just $\hat{M} = M_0$, while the MLE under the alternative is $\hat{M} = \bar{T}(Y)$ obtained in section 3.1. This test shows whether the mean parameter has a specific value.

Example 4.1.1. From the Gaussian case, we have $\eta(M) = \frac{M}{\sigma^2}, \bar{T}(Y) = \bar{Y}$, and $K(\eta(M)) = \text{tr}(\frac{M^2}{2\sigma^2})$. The test statistic for this test is

$$2(l_a - l_0) = 2n[\text{tr}((\eta(\bar{Y}) - \eta(M_0))\bar{Y}) + \text{tr}(\frac{M_0^2}{2\sigma^2}) - \text{tr}(\frac{\bar{Y}^2}{2\sigma^2})] = \frac{n}{\sigma^2}\text{tr}((\bar{Y} - M_0)^2).$$

Example 4.1.2. In the case of matrix gamma distribution, we have $\eta(M) = -\alpha M^{-1}, \bar{T}(Y) = \bar{Y}$, and $K(\eta(M)) = \alpha \log |M|$. The test statistic for this test is given by

$$2(l_a - l_0) = 2n[\text{tr}((\eta(\bar{Y}) - \eta(M_0))\bar{Y}) + K(\eta(M_0)) - K(\eta(\bar{Y}))] = 2n[\text{tr}(-\alpha(\bar{Y}^{-1} - M_0^{-1})\bar{Y}) + \alpha \log |M_0| - \alpha \log |\bar{Y}|] = 2n\alpha \left[\text{tr}(M_0^{-1}\bar{Y}) - p + \log \frac{|M_0|}{|\bar{Y}|}\right].$$
4.2 Test of eigenvectors with fixed eigenvalues in a non-increasing order

Here, we assume that eigenvalues are known and distinct, so they can be ordered. Thus, the problem of testing whether a set of eigenvectors is equal to \( U_0 \) is formulated as a test of whether the columns of \( U_0 \) are the eigenvectors corresponding to the ordered eigenvalues in \( D_0 \). Now, we will consider that the eigenvalues are fixed, i.e. with restriction \( D = D_0 \).

Let \( M_{D_0} \) be \( \mathbb{M}_{D_0} = \{ M = UD_0U^T : U \in \mathbb{O}_p \} \) and \( \bar{\Phi}(Y) = V\Lambda V^T \) be an eigendecomposition where the diagonal elements of \( \Lambda \) are in non-increasing order. Then, by Theorem 3.2, the MLE of \( M \) is \( \hat{M} = \hat{U}D_0\hat{U}^T \), where \( \hat{U} \) is any matrix which has the form \( \hat{U} = VQ \) and \( Q \in \mathbb{O}_p \) satisfies that \( QD_0Q^T = D_0 \).

Using this MLE, the LRT statistic for testing \( H_0 : M = U_0D_0U_0^T \) vs \( H_a : M \in \mathbb{M}_{D_0} \) is

\[
2(l_a - l_0) = 2n[\text{tr}(\hat{U}\eta(D_0)\hat{U}^T\bar{\Phi}(Y)) - \text{tr}(U_0\eta(D_0)U_0^T\bar{\Phi}(Y)) - K(\eta(D_0)) + K(\eta(D_0))] \\
= 2n[\text{tr}(Q\eta(D_0)Q^T\Lambda - U_0\eta(D_0)U_0^T\bar{\Phi}(Y))] \\
= 2n[\text{tr}(\eta(D_0)(\Lambda - U_0^T\bar{\Phi}(Y)U_0))]
\]

and it follows \( \chi^2_{q - \sum l_i \frac{m_i(m_i+1)}{2}} \) asymptotically, where \( m_i \) is the multiplicity corresponding to \( i \)th distinct eigenvalue of \( M \) (Schwartzman, 2008).

Example 4.2.1. From the Gaussian case, we have \( \hat{U} = VQ \), where \( \bar{\Phi}(Y) = \bar{Y} \), and so \( \bar{Y} = V\Lambda V^T \). The test statistic for this hypothesis is

\[
2(l_a - l_0) = 2n[\text{tr}((\hat{U}\eta(D_0)\hat{U}^T - U_0\eta(D_0)U_0^T)\bar{\Phi}(Y))] \\
= \frac{2n}{\sigma^2} [\text{tr}((\hat{U}D_0\hat{U}^T - U_0D_0U_0^T)\bar{Y})] \\
= \frac{2n}{\sigma^2} [\text{tr}(D_0(\Lambda - U_0^T\bar{Y}U_0))].
\]

Example 4.2.2. In the case of matrix gamma distribution, we have the same solution as \( \hat{U} = VQ \), \( \bar{\Phi}(Y) = \bar{Y} \), and so \( \bar{Y} = V\Lambda V^T \). We have \( \eta(D) = -\alpha D^{-1} \) and \( K(\eta(D)) = \alpha \log |D| \), so the test statistic for this hypothesis is

\[
2(l_a - l_0) = 2n[\text{tr}((\hat{U}\eta(D_0)\hat{U}^T - U_0\eta(D_0)U_0^T)\bar{\Phi}(Y))] \\
= 2n[-\alpha \text{tr}((\hat{U}D_0^{-1}\hat{U}^T\bar{Y} - U_0D_0^{-1}U_0^T\bar{Y})] \\
= 2n\alpha [\text{tr}(D_0^{-1}(U_0^T\bar{Y}U_0 - \Lambda)).
\]

4.3 Test of eigenvalues with unordered eigenvectors

Consider the case that the eigenvectors are fixed as in Section 3.4. An associated test is whether the eigenvalues of the mean parameter \( M \) are different from some fixed values while the eigenvectors are fixed.
Let $\mathcal{M}_{U_0} = \{M = U_0DU_0^T : D \in \mathbb{D}_p\}$. Then, under the hypothesis that $H_0 : M \in \mathcal{M}_{U_0}$, the MLE of $D$ is $\hat{D} = \text{Diag}(U_0^T T(Y) U_0)$, using the result of Theorem 3.3 with restriction $U = U_0$.

The LRT statistic for testing $H_0 : M = U_0D_0U_0^T$ vs $H_a : M \in \mathcal{M}_{U_0}$ is as follows. As the MLE for the alternative hypothesis is $\hat{M} = U_0\hat{D}U_0^T$, we have

$$2(l_a - l_0) = 2n[\text{tr}(U_0\eta(\hat{D})U_0^T \bar{T}(Y)) - \text{tr}(U_0\eta(D_0)U_0^T \bar{T}(Y) - K(\eta(\hat{D})) + K(\eta(D_0))]$$

and it asymptotically follows $\chi^2_p$, since the number of parameters tested is $p$.

**Example 4.3.1.** From the Gaussian case, we have $\eta(D) = \frac{D}{\sigma^2}$, $\bar{T}(Y) = \bar{Y}$, and $K(\eta(D)) = \text{tr}(\frac{D^2}{2\sigma^2})$. In addition, $\hat{D} = \text{Diag}(U_0^T \bar{Y} U_0)$. The LRT statistic for this test is

$$2(l_a - l_0) = \frac{2n}{\sigma^2} \left[ \text{tr}\{U_0\hat{D}U_0^T \bar{Y} - U_0D_0U_0^T \bar{Y}\} - \text{tr}\left\{\frac{\hat{D}^2 - D_0^2}{2}\right\} \right]$$

$$= \frac{2n}{\sigma^2} \left[ \text{tr}\{U_0^T \bar{Y} U_0 \hat{D} - U_0^T \bar{Y} U_0 D_0\} - \text{tr}\left\{\frac{\hat{D}^2 - D_0^2}{2}\right\} \right]$$

$$= \frac{2n}{\sigma^2} \left[ \text{tr}\left\{\hat{D}^2 - \hat{D}D_0 - \frac{\hat{D}^2}{2} + \frac{D_0^2}{2}\right\} \right]$$

$$= \frac{n}{\sigma^2} \|\hat{D} - D_0\|^2.$$ 

As $\text{tr}(A) = \text{tr} \{\text{Diag}(A)\}$, $\text{tr}\{U_0^T \bar{Y} U_0 \hat{D}\} = \text{tr}\{\text{Diag}(U_0^T \bar{Y} U_0) \hat{D}\}$. We know that $\hat{D}$ is a diagonal matrix, so from Proposition 1 in Appendix A.3,

$$\text{tr}\{U_0^T \bar{Y} U_0 \hat{D}\} = \text{tr}\{\text{Diag}(U_0^T \bar{Y} U_0) \hat{D}\} = \text{tr}\{\text{Diag}(U_0^T \bar{Y} U_0) \hat{D}\}.$$

We know that $\hat{D} = \text{Diag}(U_0^T \bar{Y} U_0)$, thus $\text{tr}\{U_0^T \bar{Y} U_0 \hat{D}\} = \text{tr}(\hat{D})^2$. Similarly, $\text{tr}\{U_0^T \bar{Y} U_0 D_0\} = \text{tr}\{\hat{D}D_0\}$. Thus, the formula can be simplified as the final line.

**Example 4.3.2.** In the case of the matrix gamma distribution, we have $\eta(D) = -\alpha D^{-1}$, $\bar{T}(Y) = \bar{Y}$, and $K(\eta(D)) = \alpha \log |D|$. Thus, $\hat{D} = \text{Diag}(U_0^T \bar{Y} U_0)$. The test statistic for this test is

$$2(l_a - l_0) = 2n\alpha \left[ \text{tr}\{U_0^T \bar{Y} U_0 (D_0^{-1} - \hat{D})^{-1}\} + \log \frac{|D_0|}{|\hat{D}|} \right].$$

### 4.4 Test of eigenvectors and eigenvalues with known multiplicities

In Section 3.5, we estimated $U$ and $D$ with the assumption that we only know the multiplicities. Using these MLE values, we can test whether the eigenvalues of the matrix $M$ have particular multiplicities. Using the
In the case of matrix gamma distribution, we have Example 4.4.2. The test statistic is

\[ 2(l_a - l_0) = 2n[\text{tr}(\eta(M)\hat{T}(Y)) - \text{tr}(\hat{U}\eta(D)\hat{U}^T\hat{T}(Y)) - K(\eta(M)) + K(\eta(D))] \]

\[ = 2n[\text{tr}\{(\eta(\hat{T}(Y)) - \hat{U}\eta(D)\hat{U}^T\hat{T}(Y)\} - K(\eta(\hat{T}(Y))) + K(\eta(D))] \]

\[ = 2n[\text{tr}\{(\eta(\Lambda) - \eta(D))\Lambda\} - K(\eta(\Lambda)) + K(\eta(D))] \quad (5) \]

where \( \hat{D} \) is the estimated value from Theorem 3.4, and \( \hat{M} = \bar{T}(Y) = V\Lambda V^T \) is the solution that maximizes the likelihood without any restriction. It follows \( \chi^2_{\sum_{i=1}^{k} \frac{m_i(m_i+1)}{2} - k} \) asymptotically. The number of degrees of freedom is the difference between the dimension of alternative parameter space \( q \) and the dimension of \( M_{m_1,...,m_k} \) which is equal to \( k + q - \sum_{i=1}^{k} \frac{m_i(m_i+1)}{2} \).

**Example 4.4.1.** From the Gaussian case, we have \( \eta(D) = \frac{D}{\sigma^2} \), \( \bar{T}(Y) = \bar{Y} \), \( \bar{Y} = V\Lambda V^T \) and \( K(\eta(D)) = \text{tr}\left(\frac{D^2}{2\sigma^2}\right) \). Therefore, \( \hat{D} = \text{Blk}(Q^T\Lambda Q) \) and \( \hat{U} = VQ \). Using the technique presented in Example 4.1.1, the test statistic for this hypothesis is

\[ 2(l_a - l_0) = \frac{2n}{\sigma^2} \left[ \text{tr}(\bar{Y}\bar{Y}) - \text{tr}(\hat{U}\hat{D}\hat{U}^T\bar{Y}) - \text{tr}\left(\frac{\bar{Y}^2}{2}\right) + \text{tr}\left(\frac{\hat{D}^2}{2}\right) \right] \]

\[ = \frac{n}{\sigma^2} \left[ \text{tr}\{2(\bar{Y}^2) - 2Q\hat{D}Q^T\Lambda - (\bar{Y}^2) + \hat{D}^2\} \right] 

\[ = \frac{n}{\sigma^2} \left[ \text{tr}\{\bar{Y}^2 - 2\hat{D}Q^T\Lambda Q + \hat{D}^2\} \right] 

\[ = \frac{n}{\sigma^2} \left[ \text{tr}\{\Lambda^2 - 2\hat{D}\Lambda + \hat{D}^2\} \right] 

\[ = \frac{n}{\sigma^2} ||\Lambda - \hat{D}||^2. \]

**Example 4.4.2.** In the case of matrix gamma distribution, we have \( \eta(D) = -\alpha D^{-1} \), \( \bar{T}(Y) = \bar{Y} \), \( \bar{Y} = V\Lambda V^T \) and \( K(\eta(D)) = \alpha \log|D| \). Thus, \( \hat{D} = \text{Blk}(Q^T\Lambda Q) \) and \( \hat{U} = VQ \). The test statistic for this hypothesis is

\[ 2(l_a - l_0) = 2n[\text{tr}\{(\eta(\bar{T}(Y)) - \hat{U}\eta(D)\hat{U}^T\bar{T}(Y)\} - K(\eta(\bar{T}(Y))) + K(\eta(D))] \]

\[ = 2\alpha [\text{tr}(\hat{U}\hat{D}^{-1}\hat{U}^T\bar{Y} - I_p) - \log|\Lambda| + \log|\hat{D}|] \]

\[ = 2\alpha \left[ \text{tr}(\hat{Q}\hat{D}^{-1}Q^T\Lambda) - p + \log\frac{|\hat{D}|}{|\Lambda|} \right]. \]
Chapter 5

Estimation of eigenstructures in the Two-Sample Case

Until the previous section, we only considered some problems for one-sample cases. Now, we will expand to the two-sample case problems. Let \( Y_1, ..., Y_{n_1} \) and \( Y_{n_1+1}, ..., Y_{n} \), \( n = n_1 + n_2 \) be two independent i.i.d. samples from two exponential family distributions \( f(Y; M_1) \) and \( f(Y; M_2) \) and \( M_1, M_2 \) be the mean for each sample. Here, we are interested in estimating and testing the pair of means \( M = (M_1, M_2) \). Note that, in practice, the inference depends on nuisance parameters, such as variance parameters and shape parameters. Thus, we need to consider assumptions for the nuisance parameter.

5.1 Estimation under the same nuisance parameter assumption

In order to obtain analytical expressions, here we make an assumption that the nuisance parameters are the same. We start from the joint likelihood function of the two samples. The form of the joint likelihood function is

\[
L = \prod_{i=1}^{n_1} H(Y_i) e^{\text{tr}(\eta(M_1)\bar{T}_1(Y)) - K(\eta(M_1))} \prod_{j=n_1+1}^{n} H(Y_j) e^{\text{tr}(\eta(M_2)\bar{T}_2(Y)) - K(\eta(M_2))}.
\]

With the assumption that \( \eta_1(D) = \eta_2(D) = \eta(D) \), the log likelihood is

\[
\log L(M; Y_1, ..., Y_n) \propto n_1(\text{tr}(\eta(M_1)\bar{T}_1(Y)) - K(\eta(M_1))) + n_2(\text{tr}(\eta(M_2)\bar{T}_2(Y)) - K(\eta(M_2)))
\]

\[
\propto n_1\text{tr}(U_1\eta(D)U_1^T\bar{T}_1(Y)) + n_2\text{tr}(U_2\eta(D)U_2^T\bar{T}_2(Y)) - (n_1 + n_2)K(\eta(D)).
\]

**Theorem 5.1.** Define the set \( \mathcal{M}_{2,D} = \{(M_1, M_2) : M_1 = U_1DU_1^T, M_2 = U_2DU_2^T, U_1, U_2 \in \mathbb{O}_p, D \in \mathbb{D}_p, \text{mult. } m_1, ..., m_k, \text{ diag. entries of } D \text{ are in non-increasing order} \} \) and let \( \bar{Y}_1 = V_1\Lambda_1V_1^T, \bar{Y}_2 = V_2\Lambda_2V_2^T \) be eigen decompositions of two sample means, where the diagonal elements of \( \Lambda_1, \Lambda_2 \) are in non-increasing order. Then, the maximum likelihood estimator of \( M = (M_1, M_2) \) is given by:

\[
\hat{M} = \left( \frac{1}{n_1} \sum_{i=1}^{n_1} \bar{Y}_i, \frac{1}{n_2} \sum_{j=n_1+1}^{n} \bar{Y}_j \right).
\]

With this estimator, the log likelihood is maximized.

31
order. Suppose that \( D \) is unknown but the order and multiplicities of \( D \) are known and \( \eta(\cdot) \) are the same for two groups. Then, the MLE of \( M_1 \) and \( M_2 \) are \( \hat{M}_1 = \hat{U}_1 \hat{D} \hat{U}_1^T \) and \( \hat{M}_2 = \hat{U}_2 \hat{D} \hat{U}_2^T \), where \( \hat{U}_1, \hat{U}_2 \) and \( \hat{D} \) satisfy these conditions:

(i) \( \hat{U}_1 \) and \( \hat{U}_2 \) are any matrices that have the form \( \hat{U}_1 = V_1 Q_1 \) and \( \hat{U}_2 = V_2 Q_2 \), where \( Q_1, Q_2 \in \mathcal{O}_p \)

(ii) \( \hat{D} = \text{Blk}_{m_1, \ldots, m_k}(\frac{n_1 \Lambda_1 + n_2 \Lambda_2}{n_1 + n_2}) \).

**Proof.** Taking the derivative with respect to \( U_1, U_2 \) and \( D \), we have

\[
\nabla l_{U_1} = \frac{\partial l}{\partial U_1} = n_1 \frac{\partial}{\partial U_1} \text{tr}(U_1 \eta(D) U_1^T \hat{T}_1(Y)) = 2n_1 \text{tr}(\hat{T}_1(Y) U_1^T \eta(D))
\]

\[
\nabla l_{U_2} = \frac{\partial l}{\partial U_2} = n_2 \frac{\partial}{\partial U_2} \text{tr}(U_2 \eta(D) U_2^T \hat{T}_2(Y)) = 2n_2 \text{tr}(\hat{T}_2(Y) U_2^T \eta(D))
\]

\[
\nabla l_D = \frac{\partial l}{\partial D} = \frac{\partial}{\partial D} \{n_1 \text{tr}(U_1 \eta(D) U_1^T \hat{T}_1(Y))\} + \frac{\partial}{\partial D} \{n_2 \text{tr}(U_2 \eta(D) U_2^T \hat{T}_2(Y))\} - n \frac{\partial K(\eta(D))}{\partial D}
\]

We can see that the gradients of \( U_1 \) and \( U_2 \) have the same form as the one-sample case, and each gradient function does not depend on other variables. Therefore, using the logic of Theorem 3.2, we can estimate \( U_1 \) and \( U_2 \) separately as \( \hat{U}_1 = V_1 Q_1 \) and \( \hat{U}_2 = V_2 Q_2 \), with the same logic as in the one-sample case.

Now, we will choose \( Q_1 = Q_2 = I \) and plug \( \hat{U}_1 = V_1 \) and \( \hat{U}_2 = V_2 \) into the gradient function with respect to \( D \). Similar to the one-sample case described in Theorem 3.3, the gradient function with respect to \( B = \eta(D) \) can be simplified as

\[
\nabla l_B = n_1 \frac{\partial}{\partial B} \{\text{tr}(U_1 B U_1^T \hat{T}_1(Y))\} + n_2 \frac{\partial}{\partial B} \{\text{tr}(U_2 B U_2^T \hat{T}_2(Y))\} - n \frac{\partial K(B)}{\partial B}
\]

\[
= n_1 \hat{U}_1^T \hat{T}_1(Y) \hat{U}_1 + n_2 \hat{U}_2^T \hat{T}_2(Y) \hat{U}_2 - n \frac{\partial K(B)}{\partial B}.
\]

This gradient function should be orthogonal to the surface defined by the constraint that \( B \) is a diagonal matrix and has the given multiplicities. Thus, the tangent space is \( \mathcal{T} = \{B; B = \text{Blk}_{m_1, \ldots, m_k}(A), A \in \mathcal{S}_p\} \) and so we can use the same logic of Theorem 3.5. From the previous chapter, we know that \( \frac{\partial K(B)}{\partial B} = D \).

Thus, we have

\[
n\text{Blk}_{m_1, \ldots, m_k}(\frac{\partial K(B)}{\partial B}) = \text{Blk}_{m_1, \ldots, m_k}(n_1 \hat{U}_1^T \hat{T}_1(Y) \hat{U}_1 + n_2 \hat{U}_2^T \hat{T}_2(Y) \hat{U}_2)
\]

\[
n\text{Blk}_{m_1, \ldots, m_k}(\frac{\partial K(B)}{\partial B}) = \text{Blk}_{m_1, \ldots, m_k}(n_1 \Lambda_1 + n_2 \Lambda_2)
\]

\[
\hat{D} = \text{Blk}_{m_1, \ldots, m_k}(\frac{n_1 \Lambda_1 + n_2 \Lambda_2}{n}). \quad \blacksquare
\]
5.2 Estimation under the different nuisance parameters

Now, we assume that $\eta_1(D) \neq \eta_2(D)$. Then, the log likelihood function is

$$l(M; Y_1,...,Y_n) \propto n_1 \text{tr}(U_1\eta_1(D)U_1^T\bar{T}(Y)) + n_2 \text{tr}(U_2\eta_2(D)U_2^T\bar{T}(Y)) - n_1 K(\eta_1(D)) - n_2 K(\eta_2(D)).$$

**Theorem 5.2.** Let $\bar{Y}_1 = V_1\Lambda_1 V_1^T$, $\bar{Y}_2 = V_2\Lambda_2 V_2^T$ be eigen decompositions of two sample means, where the diagonal elements of $\Lambda_1, \Lambda_2$ are in non-increasing order. Suppose that $D$ has the multiplicities $m_1,...,m_k$ and is the same for two groups but $\eta(\ast)$ are different. Then, the MLE of $M_1$ and $M_2$ are $\hat{M}_1 = \hat{U}_1\hat{D}\hat{U}_1^T$ and $\hat{M}_2 = \hat{U}_2\hat{D}\hat{U}_2^T$, where $\hat{U}_1, \hat{U}_2,$ and $\hat{D}$ satisfy these conditions:

(i) $\hat{U}_1 = V_1Q_1$ and $\hat{U}_2 = V_2Q_2$, where $Q_1, Q_2 \in \mathbb{R}^p$ satisfy the condition that $Q_1\hat{D}Q_1^T = \hat{D}$.

(ii) $\hat{D}_{ii} = BL_{m_1,...,m_k} \left[\frac{n_1(\Lambda_{1})_{ii}\frac{\partial\eta_1(D)}{\partial D}_{ii} + n_2(\Lambda_{2})_{ii}\frac{\partial\eta_2(D)}{\partial D}_{ii}}{n_1\frac{\partial\eta_1(D)}{\partial D}_{ii} + n_2\frac{\partial\eta_2(D)}{\partial D}_{ii}}\right].$

**Proof.** Taking the derivative with respect to $U_1, U_2$ and $D$, we have

$$\nabla l_{U_1} = \frac{\partial l}{\partial U_1} = n_1 \frac{\partial}{\partial U_1} \text{tr}(U_1\eta_1(D)U_1^T\bar{T}_1(Y))$$

$$= 2n_1 (\bar{T}_1(Y)U_1^T\eta_1(D))$$

$$\nabla l_{U_2} = \frac{\partial l}{\partial U_2} = n_2 \frac{\partial}{\partial U_2} \text{tr}(U_2\eta_2(D)U_2^T\bar{T}_2(Y))$$

$$= 2n_2 (\bar{T}_2(Y)U_2^T\eta_2(D))$$

$$\nabla l_D = \frac{\partial l}{\partial D} = \frac{\partial n_1 \{\text{tr}(U_1\eta_1(D)U_1^T\bar{T}_1(Y))\}}{\partial D} + \frac{\partial n_2 \{\text{tr}(U_2\eta_2(D)U_2^T\bar{T}_2(Y))\}}{\partial D} - n_1 \frac{\partial K(\eta_1(D))}{\partial D} - n_2 \frac{\partial K(\eta_2(D))}{\partial D}.$$
have \( \frac{\partial K(B_1)}{\partial B_1} = \frac{\partial K(B_2)}{\partial B_2} = D \). Using this fact, we have

\[
\frac{\partial l}{\partial D_{ii}} = n_1 \text{tr} \left[ \frac{\partial}{\partial B_1} \left\{ \text{tr}(U_1 B_1 U_1^T T(Y)) \right\} \right] - n_1 \text{tr} \left[ \frac{\partial K(B_1)}{\partial B_1} \frac{\partial B_1}{\partial D_{ii}} \right] + n_2 \text{tr} \left[ \frac{\partial}{\partial B_2} \left\{ \text{tr}(U_2 B_2 U_2^T T(Y)) \right\} \right] - n_2 \text{tr} \left[ \frac{\partial K(B_2)}{\partial B_2} \frac{\partial B_2}{\partial D_{ii}} \right]
\]

This process can be applied to all \( i \), so all the diagonal elements of gradient function should be equal to 0 to make the above gradient orthogonal to the tangent space. Thus,

\[
0 = \frac{\partial l}{\partial D_{ii}} = n_1 \left[ (U_1^T \bar{T}_1(Y) U_1)_{ii} \frac{\partial B_{1i}}{\partial D_{ii}} - D_{ii} \frac{\partial B_{1i}}{\partial D_{ii}} \right] + n_2 \left[ (U_2^T \bar{T}_2(Y) U_2)_{ii} \frac{\partial B_{2i}}{\partial D_{ii}} - D_{ii} \frac{\partial B_{2i}}{\partial D_{ii}} \right]
\]

\[
D_{ii} = \frac{n_1 (U_1^T \bar{T}_1(Y) U_1)_{ii} \frac{\partial B_{1i}}{\partial D_{ii}} + n_2 (U_2^T \bar{T}_2(Y) U_2)_{ii} \frac{\partial B_{2i}}{\partial D_{ii}}}{n_1 \frac{\partial B_{1i}}{\partial D_{ii}} + n_2 \frac{\partial B_{2i}}{\partial D_{ii}}}
\]

for all \( i \). Now, by substituting \( \hat{U}_1 = V_1 \) and \( \hat{U}_2 = V_2 \), as estimated in the previous steps into \( U_1 \) and \( U_2 \), we get the final result as

\[
\hat{D}_{ii} = \frac{n_1 (A_1)_{ii} \frac{\partial B_{1i}}{\partial D_{ii}} + n_2 (A_2)_{ii} \frac{\partial B_{2i}}{\partial D_{ii}}}{n_1 \frac{\partial B_{1i}}{\partial D_{ii}} + n_2 \frac{\partial B_{2i}}{\partial D_{ii}}}
\]

**Example 5.2.1.** From the Gaussian case, we have \( \eta(D) = \frac{D}{\sigma^2}, \bar{T}(Y) = \bar{Y} \), and \( K(\eta(D)) = \text{tr}(\frac{D^2}{2\sigma^2}) \). \( U_1 \) and \( U_2 \) can be obtained by the same process as the one-sample case. Next, we have the likelihood for two different groups in gaussian case as

\[
l(U_1, U_2, D, \sigma_1^2, \sigma_2^2) \propto \frac{n_1}{2\sigma_1^2} \text{tr} \left[ 2U_1 DU_1^T \bar{Y}_1 - (U_1 DU_1^T)^2 \right] + \frac{n_2}{2\sigma_2^2} \text{tr} \left[ 2U_2 DU_2^T \bar{Y}_2 - (U_2 DU_2^T)^2 \right].
\]

As such, the gradient function for \( D \) is

\[
\nabla l_D = \frac{n_1}{2\sigma_1^2} \left[ -2U_1^T \bar{Y}_1 U_1 + 2D \right] + \frac{n_2}{2\sigma_2^2} \left[ -2U_2^T \bar{Y}_2 U_2 + 2D \right].
\]
To make this function orthogonal to the space of diagonal matrices, this gradient should be projected to the diagonal space. Thus, we have the score function for \( D \), which is \( S(D) = \text{Diag}(\nabla l_D) \), and the MLE can be obtained by setting the score function equal to 0.

\[
\text{Blk}\left[ \left( \frac{n_1}{\sigma_1^2} + \frac{n_2}{\sigma_2^2} \right) D \right] = \text{Blk}\left\{ \frac{n_1}{\sigma_1^2} (U_1^T \bar{Y}_1 U_1) \right\} + \text{Blk}\left\{ \frac{n_2}{\sigma_2^2} (U_2^T \bar{Y}_2 U_2) \right\}
\]

\[
\hat{D} = \frac{n_1}{\sigma_1^2} \text{Blk}(U_1^T \bar{Y}_1 U_1) + \frac{n_2}{\sigma_2^2} \text{Blk}(U_2^T \bar{Y}_2 U_2)
\]

By substituting \( \hat{U}_1 = V_1 \) and \( \hat{U}_2 = V_2 \) into \( U_1 \) and \( U_2 \), we get the final estimate for \( D \) as

\[
\hat{D} = \frac{n_1}{\sigma_1^2} \text{Blk}(\Lambda_1) + \frac{n_2}{\sigma_2^2} \text{Blk}(\Lambda_2)
\]

\[
\text{Example 5.2.2.} \quad \text{In the case of matrix gamma distribution, we have } \eta(D) = -\alpha D^{-1}, \bar{T}(Y) = \bar{Y}, \text{ and } K(\eta(D)) = \alpha \log |D|. \text{ For } U_1 \text{ and } U_2, \text{ it has the same process as the one-sample case. Next,}
\]

\[
l(U_1, U_2, D, \alpha_1^2, \alpha_2^2) \propto n_1\alpha_1 \{- \log |D| - \text{tr}(U_1 D^{-1} U_1^T \bar{Y}_1)\} + n_2\alpha_2 \{- \log |D| - \text{tr}(U_2 D^{-1} U_2^T \bar{Y}_2)\}.
\]

Therefore, the gradient function for \( D \) is

\[
\nabla l_D = n_1\alpha_1 \left[ -D^{-1} + D^{-1} U_1^T \bar{Y}_1 U_1 D^{-1} \right] + n_2\alpha_2 \left[ -D^{-1} + D^{-1} U_2^T \bar{Y}_2 U_2 D^{-1} \right]
\]

Setting the score function for \( D \) equal to 0, we have

\[
\text{Blk}[n_1\alpha_1 D^{-1} U_1^T \bar{Y}_1 U_1 D^{-1} + n_2\alpha_2 D^{-1} U_2^T \bar{Y}_2 U_2 D^{-1}] = \text{Blk}[n_1\alpha_1 D^{-1} U_1^T \bar{Y}_1 U_1 D^{-1} + n_2\alpha_2 D^{-1} U_2^T \bar{Y}_2 U_2 D^{-1}]
\]

\[
\hat{D} = \frac{n_1\alpha_1 \text{Blk}(U_1^T \bar{Y}_1 U_1) + n_2\alpha_2 \text{Blk}(U_2^T \bar{Y}_2 U_2)}{n_1\alpha_1 + n_2\alpha_2}
\]

By substituting \( \hat{U}_1 = V_1 \) and \( \hat{U}_2 = V_2 \) into \( U_1 \) and \( U_2 \), we get the final estimation for \( D \) as

\[
\hat{D} = \frac{n_1\alpha_1 \text{Blk}(\Lambda_1) + n_2\alpha_2 \text{Blk}(\Lambda_2)}{n_1\alpha_1 + n_2\alpha_2}
\]

35
Chapter 6

Testing for eigenstructures in the Two-Sample Case

6.1 Full Matrix test

First we start with the test of $H_0 : M_1 = M_2$ versus $H_a : M_1 \neq M_2$. This is a test of whether the means of two groups are equal. Under $H_0$, the MLE for $M_1, M_2$ is just the MLE of one-sample case with sample size $n = n_1 + n_2$. By setting $\bar{T}(Y) = \frac{n_1 \bar{T}_1(Y) + n_2 \bar{T}_2(Y)}{n}$, we have $\hat{M}_1 = \hat{M}_2 = \bar{T}(Y)$. Then, the LRT statistic for this hypothesis is

$$2(l_a - l_0) = 2\left[ n_1 \text{tr}(\eta(\bar{T}_1(Y))\bar{T}_1(Y)) + n_2 \text{tr}(\eta(\bar{T}_2(Y))\bar{T}_2(Y)) - n_1 K(\eta(\bar{T}_1(Y))) - n_2 K(\eta(\bar{T}_2(Y))) - n \text{tr}(\eta(\bar{T}(Y))\bar{T}(Y)) + n K(\eta(\bar{T}(Y))) \right],$$

and it follows a $\chi^2_q$ distribution asymptotically, since the number of parameters tested is $q$.

Example 6.1.1. From the Gaussian case, we have $\eta(M) = \frac{M}{\sigma^2}, \bar{T}(Y_1) = \bar{Y}_1, \bar{T}(Y_2) = \bar{Y}_2$, and $K(\eta(M)) = \text{tr}(M^2 / 2\sigma^2)$. The test statistic for this test is

$$2(l_a - l_0) = \frac{2}{\sigma^2} \left[ n_1 \text{tr}(\bar{Y}_1^2) + n_2 \text{tr}(\bar{Y}_2^2) - \frac{n_1}{2} \text{tr}(\bar{Y}_1^2) - \frac{n_2}{2} \text{tr}(\bar{Y}_2^2) - n \text{tr}(\bar{Y}^2) + \frac{n}{2} \text{tr}(\bar{Y}^2) \right]$$

$$= \frac{1}{\sigma^2} \left[ n_1 \text{tr}(\bar{Y}_1^2) + n_2 \text{tr}(\bar{Y}_2^2) - n \text{tr}(\bar{Y}^2) \right]$$

$$= \frac{1}{n\sigma^2} \left[ n_1 \text{tr}(\bar{Y}_1^2) + n_2 \text{tr}(\bar{Y}_2^2) - n \text{tr}(\bar{Y}^2) \right]$$

$$= \frac{n_1 n_2}{n \sigma^2} \text{tr}((\bar{Y}_1 - \bar{Y}_2)^2).$$

Example 6.1.2. In the case of matrix gamma distribution, we have $\eta(M) = -\alpha M^{-1}, \bar{T}(Y_1) = \bar{Y}_1, \bar{T}(Y_2) =$
\( \tilde{Y}_2 \), and \( K(\eta(M)) = \alpha \log |M| \). The test statistic for this test is given by

\[
2(l_a - l_0) = 2\alpha \left[ -n_1 \log |\tilde{Y}_1| - n_2 \log |\tilde{Y}_2| - \text{tr} \left( \sum_{i=1}^{n_1} \tilde{Y}_1^{-1} \tilde{Y}_i \right) - \text{tr} \left( \sum_{j=n_1+1}^{n} \tilde{Y}_2^{-1} \tilde{Y}_j \right) + n \log |\tilde{Y}| + \text{tr} \left( \sum_{i=1}^{n_1} \tilde{Y}_1^{-1} \tilde{Y}_i \right) \right]
\]

6.2 Testing for the eigenvectors

Now, we move to the testing for the eigenstructure of \( M_1 \) and \( M_2 \). A test of eigenvectors can be performed by testing the hypothesis that \( H_0 : M_1 = M_2 \) versus \( H_a : (M_1, M_2) \in \mathbb{M}_{2,D} \). This is the test of equal sets of eigenvectors when the two mean parameters have the same eigenvalues. Under \( H_0 \), the MLE for \( M_1, M_2 \) is just the MLE in the one-sample case with sample size \( n = n_1 + n_2 \). Setting \( \hat{T}(Y) = \frac{n_1 T_1(Y) + n_2 T_2(Y)}{n} = V\Lambda V^T \), we have \( M_1 = M_2 = \hat{\Upsilon} \hat{\Lambda}^* \hat{\Upsilon}^T \), where \( \hat{\Upsilon} = VQ \) and \( \hat{\Lambda}^* = \text{Blk}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot) \) are calculated by the one-sample case formula in Theorem 3.5, with the sample size \( n = n_1 + n_2 \). Then, the test statistic is

\[
2(l_a - l_0) = 2[n_1 \text{tr}(\hat{U}_1\eta(\hat{D})\hat{U}_1^T\hat{T}_1(Y)) + n_2 \text{tr}(\hat{U}_2\eta(\hat{D})\hat{U}_2^T\hat{T}_2(Y)) - nK(\eta(\hat{D})) - n\text{tr}(\eta(\hat{D})\hat{A}_1) + n_2 \text{tr}(\eta(\hat{D})\hat{A}_2) - n\text{tr}(\eta(\hat{D}^*)\Lambda) + n\{K(\eta(\hat{D}^*)) - K(\eta(\hat{D}))\}],
\]

and it follows \( \chi^2_{q - \sum_{i=1}^{k_i} m_i(m_i + 1)/2} \) asymptotically. The number of degrees of freedom is the difference between the dimension of the alternative parameter set \( (k + 2q - \sum_{i=1}^{k_i} m_i(m_i + 1)) \) and that of the null parameter set \( (q + k - \sum_{i=1}^{k_i} m_i(m_i + 1)/2) \).

Example 6.2.1. In the spherical Gaussian case, we have \( \eta(D) = \frac{D}{\sigma^2} \), \( \hat{T}(Y_1) = \tilde{Y}_1 \), \( \hat{T}(Y_2) = \tilde{Y}_2 \), \( \tilde{Y}_i = V_i\Lambda_i V_i^T \), \( \tilde{Y}(Y) = \tilde{Y} = \frac{n_1 \tilde{Y}_1 + n_2 \tilde{Y}_2}{n} = V\Lambda V^T \) and \( K(\eta(D)) = \text{tr}(\frac{D^2}{2\sigma^2}) \). Thus, we have \( \hat{D}^* = \text{Blk}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot) \), \( \hat{\Upsilon} = \text{Blk}(\cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot, \cdot) \), \( \hat{\Upsilon}_1 = V_1Q_1 \), and \( \hat{\Upsilon}_2 = V_2Q_2 \). The test statistic for this hypothesis is

\[
2(l_a - l_0) = \frac{2}{\sigma^2} \left[ n_1 \text{tr}(\hat{D}\hat{A}_1) + n_2 \text{tr}(\hat{D}\hat{A}_2) - n\text{tr}(\hat{D}^*\Lambda) + \frac{n}{2} \{\text{tr}((\hat{D}^*\hat{A})^2 - \hat{D}^2)\} \right]
\]

\[
= \frac{2}{\sigma^2} \left[ \text{tr}\{(n_1\Lambda_1 + n_2\Lambda_2)\hat{D} - n\Lambda\hat{D}^*\} + \frac{n}{2} \{\text{tr}((\hat{D}^*\hat{A})^2 - \hat{D}^2)\} \right]
\]

\[
= \frac{n}{\sigma^2} \text{tr}(\hat{D}^2 - \hat{D}^2^*)
\]
Example 6.2.2. In the case of matrix gamma distribution, we have the assumption that \( \alpha \) is the same for two groups. We also have \( B = \eta(D) = -\alpha D^{-1} \), \( \bar{T}(Y_1) = \bar{Y}_1 \), \( \bar{T}(Y_2) = \bar{Y}_2 \), \( \bar{Y}_i = V_i \Lambda_i V_i^T \), \( \bar{T}(Y) = \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = V \Lambda V^T \) and \( K(\eta(D)) = \alpha \log |D| \). Thus, \( \hat{D}^* = \text{Blk}_{m_1, \ldots, m_k}(\Lambda), \hat{D} = \text{Blk}_{m_1, \ldots, m_k}(\frac{n_1 \Lambda_1 + n_2 \Lambda_2}{n}) \), \( \hat{U}_1 = V_1 Q_1 \), and \( \hat{U}_2 = V_2 Q_2 \). The test statistic for this hypothesis is

\[
2(l_a - l_0) = 2\alpha \left[ \text{tr}(-n_1 \hat{D}^{-1} \Lambda_1 - n_2 \hat{D}^{-1} \Lambda_2 + n \hat{D}^* - \Lambda) + n \{ \log |\hat{D}|^2 - \log |\hat{D}|^2 \} \right] \\
= 2\alpha \left[ \text{tr}(n\Lambda \hat{D}^* - (n_1 \Lambda_1 + n_2 \Lambda_2) \hat{D}^{-1} + n \log \frac{|\hat{D}|^2}{|\hat{D}|} \right]
\]

6.3 Testing for the eigenvalues

The next test is a test of whether the means of the two groups have the same eigenvalues, while differing on their eigenvectors. The test is \( H_0 : (M_1, M_2) \in M_{d,2} \) versus \( H_a : (M_1, M_2) \notin M_{d,2} \). The MLE for \( M_1 \) and \( M_2 \) under the alternative hypothesis are just \( \hat{M}_1 = \bar{T}_1(Y) \), \( \hat{M}_2 = \bar{T}_2(Y) \). Therefore, the test statistic is

\[
2(l_a - l_0) = 2[n_1 \text{tr}(\eta(\bar{T}_1(Y))) + n_2 \text{tr}(\eta(\bar{T}_2(Y))) - n_1 K(\eta(\bar{T}_1(Y))) - n_2 K(\eta(\bar{T}_2(Y))) + n \sum_{i=1}^{k} m_i(m_i + 1)] \\
= 2\left[ n_1 \text{tr}(\eta(\Lambda_1) \Lambda_1) + n_2 \text{tr}(\eta(\Lambda_2) \Lambda_2) - n_1 K(\eta(\Lambda_1)) - n_2 K(\eta(\Lambda_2)) \right] \\
= 2(n_1 \text{tr}(\bar{T}(\Lambda_1) - \Lambda_1) + n_2 \text{tr}(\hat{D}(\Lambda_2) - \Lambda_2) + K(\eta(D))]
\]

and it follows \( \chi^2_{\sum_{i=1}^{k} m_i(m_i + 1)} \) asymptotically. The degrees of freedom is the difference between the dimension of the alternative parameter set \( 2q \) and that of the null parameter set \( (k + 2q - \sum_{i=1}^{k} m_i(m_i + 1)) \).

Example 6.3.1. From the Gaussian case, we have \( \eta(D) = \frac{D}{\sigma^2}, \bar{T}(Y_1) = \bar{Y}_1, \bar{T}(Y_2) = \bar{Y}_2, \bar{Y}_i = V_i \Lambda_i V_i^T \) and \( K(\eta(D)) = \text{tr}(\frac{D^2}{2\sigma^2}) \). So \( \hat{D} = \text{Blk}_{m_1, \ldots, m_k}(\frac{n_1 \Lambda_1 + n_2 \Lambda_2}{n}), \hat{U}_1 = V_1 Q_1 \), and \( \hat{U}_2 = V_2 Q_2 \). The test statistic for this test is

\[
2(l_a - l_0) = \frac{2}{\sigma^2} \left[ n_1 \text{tr}(\Lambda_1) + n_2 \text{tr}(\Lambda_2) - \frac{1}{2} n_1 \text{tr}(\Lambda_1^2) - \frac{1}{2} n_2 \text{tr}(\Lambda_2^2) - n_1 \text{tr}(\hat{D} \Lambda_1) - n_2 \text{tr}(\hat{D} \Lambda_2) + \frac{n}{2} \text{tr}(\hat{D}) \right] \\
= \frac{2}{\sigma^2} \left[ n_1 \text{tr}(\Lambda_1^2) + n_2 \text{tr}(\Lambda_2^2) - \frac{n}{2} \text{tr}(\hat{D}) \right] \\
= \frac{1}{\sigma^2} \left[ n_1 \text{tr}^2(\Lambda_1) + n_2 \text{tr}^2(\Lambda_2) - n \text{tr}(\hat{D}) \right]
\]

Using the same technique in Example 4.1.1.

Example 6.3.2. In the case of matrix gamma distribution, we have \( B = \eta(D) = -\alpha D^{-1} \), \( \bar{T}(Y_1) = \bar{Y}_1 \), \( \bar{T}(Y_2) = \bar{Y}_2 \), \( \bar{M}_i = \bar{Y}_i = V_i \Lambda_i V_i^T \) and \( K(\eta(D)) = \alpha \log |D| \). Thus, \( \frac{\partial K(B)}{\partial B} = -\alpha B^{-1} = D \) and so
\[ \hat{D} = \text{Blk}_{m_1,\ldots,m_k}(n_1\Lambda_1^{-1} + n_2\Lambda_2^{-1}) \], \( \hat{U}_1 = V_1Q_1 \), and \( \hat{U}_2 = V_2Q_2 \). The test statistic for this test is given by

\[
2(l_a - l_0) = 2[n_1\text{tr}(-\alpha\Lambda_1^{-1}\Lambda_1) + n_2\text{tr}(-\alpha\Lambda_2^{-1}\Lambda_2) - n_1\alpha \log |\Lambda_1| - n_2\alpha \log |\Lambda_2| \\
- n_1\text{tr}(-\alpha\hat{D}^{-1}\Lambda_1) - n_2\text{tr}(-\alpha\hat{D}^{-1}\Lambda_2) + n\alpha \log |\hat{D}|]
\]

\[
= 2\alpha[n_1\text{tr}(\hat{D}^{-1}\Lambda_1) + n_2\text{tr}(\hat{D}^{-1}\Lambda_2) + n \log |\hat{D}| \\
- n_1p - n_2p - n_1 \log |\Lambda_1| - n_2 \log |\Lambda_2|]
\]

\[
= 2\alpha \left[ n_1\text{tr}(\hat{D}^{-1}\Lambda_1) + n_2\text{tr}(\hat{D}^{-1}\Lambda_2) + \log \frac{|\hat{D}|^n}{|\Lambda_1|^{n_1}|\Lambda_2|^{n_2}} - np \right].
\]
Chapter 7

Estimating Nuisance Parameters

In addition to estimating $M$ and calculating related test statistics, in practice, we need to estimate other parameters such as variance components, scale parameters, and shape parameters. Thus, we need to estimate these parameters under the restrictions we have made for the mean parameter.

7.1 Gaussian distribution with spherical covariance matrix

7.1.1 Test of Sphericity of Covariance matrices

In the Gaussian distribution example, we assume that the covariance matrix is spherical. Thus, estimation and tests for $M$ and $\sigma^2$ are derived based on this assumption. In data analysis, however, we need to test whether this assumption holds. Here, we use a special case of the test of orthogonal invariance described in Schwartzman et al. (2008). The test statistic considered there is

$$ T_o = nq \log(\hat{\sigma}^2) - n \log(1 - p\hat{\tau}) - n \log |S|, $$

where $S$ is the sample covariance matrix and $\tau$ is an additional parameter required by the orthogonally invariant Gaussian distribution. This test statistic follows a $\chi^2$ distribution with $df = \frac{q(q+1)}{2} - 2$, since the dimension of the parameter space under $H_a$ is $\frac{q(q+1)}{2}$ and the dimension of null parameter space is 2.

To test sphericity, we set $\tau = 0$ in the above test. Then, the null hypothesis for this test is that the covariance matrix is spherical. As $\tau = 0$, the test statistic for the sphericity of a covariance matrix is simplified as

$$ T_s = nq \log(\hat{\sigma}^2) - n \log |S|, $$

which follows $\chi^2$ distribution with $df = \frac{q(q+1)}{2} - 1$. The number of degrees of freedom changes because the dimension of the null parameter space is reduced to 1.
7.1.2 Estimating Covariance parameters in the one-sample case

If the null hypothesis of the sphericity test is not rejected, the next step is to find an appropriate way to estimate the variance parameter. In the case of Gaussian distribution with a spherical covariance matrix that has the likelihood

\[ l(M, \sigma^2; Y) = -\frac{nq}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \text{tr}\{ (Y_i - M)^2 \}, \]

we have the variance parameter \( \sigma^2 \). To find the distribution of \( M \), we also need to estimate \( \sigma^2 \). Here, we use the MLE for \( \sigma^2 \) described in Schwartzman et al.(2008):

\[ \hat{\sigma}^2 = \frac{1}{qn} \sum_{i=1}^{n} \text{tr}(Y_i - \bar{Y})^2 + \frac{1}{q} \text{tr}(\bar{Y} - \hat{M})^2 \] (7.1)

where \( \hat{M} \) is the MLE under any of the eigenvalue or eigenvector restrictions. This form is obtained by taking the derivative with respect to \( \sigma^2 \) and setting it equal to 0.

**Proposition 7.1.** Let \( \hat{\sigma}^2 \) be given by (7.1). Then,

\[ qn\hat{\sigma}^2/\sigma^2 \sim \chi^2_{q(n-1)+dfr} \]

where \( dfr \) is the difference between the dimension of \( \bar{Y} \) and the dimension of \( \hat{M} \).

**Proof.** From Schwartzman et al. (2008), if the spherical Gaussian assumption holds, then \( s^2 = \frac{1}{qn} \sum_{i=1}^{n} \text{tr}(Y_i - \bar{Y})^2 \) has the distribution \( qns^2/\sigma^2 \sim \chi^2(q(n-1)) \). Next, we use the likelihood ratio test statistic for testing \( H_0 : M \in M_0 \) with a restricted parameter space \( M_0 \) versus \( H_a : \) No restriction, to obtain the distribution of the remaining term. That is because the asymptotic distribution of the LRT statistic is known as a \( \chi^2 \) distribution. We have

\[ 2(l_a - l_0) = \frac{n}{\sigma^2} [\text{tr}(2\bar{Y}^2 - \bar{Y}^2) - \text{tr}(2\bar{Y} \hat{M} - \hat{M}^2)] \]

\[ = \frac{n}{\sigma^2} [\text{tr}(\bar{Y}^2 - 2\bar{Y} \hat{M} + \hat{M}^2)] \]

\[ = \frac{n}{\sigma^2} [\text{tr}\{ (\bar{Y} - \hat{M})^2 \}] \]

Therefore, the LRT statistic is \( 2(l_a - l_0) = n\text{tr}(\bar{Y} - \hat{M})^2/\sigma^2 \). Since the LRT statistic follows \( \chi^2 \) distribution asymptotically, we can conclude that \( n\text{tr}\{ (\bar{Y} - \hat{M})^2 \}/\sigma^2 \) follows a \( \chi^2 \) distribution. Since the degrees of freedom is the difference between the dimension of alternative parameter space and that of null parameter space, the number of degrees of freedom is equal to the difference between the dimension of \( \bar{Y} \) and dimension of \( \hat{M} \) (denoted by \( df_r \)).

Note now that \( \hat{\sigma}^2 \) is equal to the sum of \( s^2 \) and \( 2(l_a - l_0) \) and the two are independent. Since the sum of two
independent $\chi^2$-distributed statistics also follows a $\chi^2$ distribution, $\hat{\sigma}^2$ follows a $\chi^2$ distribution with degrees of freedom $q(n - 1) + df_r$.

7.1.3 Estimating Covariance parameters in the two-sample case

For the two-sample Gaussian case, we have two options: equal or unequal covariance matrix assumption. In general, equal covariance matrix assumption is preferred because this assumption makes it possible to use the Hotelling’s $T^2$ test when we compare two group means, which is simpler and has an exact distribution. If the covariance matrices are not equal, we should use other alternative tests such as Behrens-Fisher tests, which is more complicated and its distribution is only approximate.

If we assume that the covariance matrices of the two groups are equal, then we need an estimate of the pooled covariance matrix $\hat{\Sigma}_{12} = \hat{\sigma}^2_{12} I_p$, which has the form

$$\hat{\sigma}^2_{12} = s^2_{12} + \frac{1}{q(n_1 + n_2)}[n_1 \text{tr}(\bar{Y}_1 - \hat{M}_1)^2 + n_2 \text{tr}(\bar{Y}_2 - \hat{M}_2)^2],$$

where

$$s^2_{12} = \frac{1}{q(n_1 + n_2)} \left[ \sum_{i=1}^{n_1} \text{tr}(Y_i - \bar{Y}_1)^2 + \sum_{i=1}^{n_2} \text{tr}(Y_i - \bar{Y}_2)^2 \right]$$

and $\hat{M}_1, \hat{M}_2$ are the MLEs under the relevant eigenvalue and eigenvector restrictions (Schwartzman et al., 2008). Like the one-sample case, $\{n_1 \text{tr}(\bar{Y}_1 - \hat{M}_1)^2 + n_2 \text{tr}(\bar{Y}_2 - \hat{M}_2)^2\}/\sigma^2$ also follows the $\chi^2$ distribution, with degrees of freedom that depend on the differences between the dimension of $(\bar{Y}_1, \bar{Y}_2)$ and that of $(\hat{M}_1, \hat{M}_2)$ (denoted by $df_{r_2}$).

If we assume that two variance parameters are different, we can estimate two covariance parameters separately. For $i = 1, 2$, we have

$$\hat{\sigma}^2_i = \frac{1}{qn_i} \sum_{j=1}^{n_i} \text{tr}(Y_{ij} - \bar{Y}_i)^2 + \frac{1}{q} \text{tr}(\bar{Y}_i - \hat{M}_i)^2,$$

where $\hat{M}_i$s are the MLE of $M_i$ under the restriction. The procedure is the same as the one-sample case described in Proposition 7.1, since each group can be thought and treated as one sample.

In addition, using the property that $s^2_i = \frac{1}{qn_i} \sum_{i=1}^{n_i} \text{tr}(Y_i - \bar{Y}_i)^2$ has the distribution $q n_i s^2_i / \sigma^2 \sim \chi^2(n_i - 1)$ for $i = 1, 2$, as mentioned in the proof of Proposition 7.1, $n_i \text{tr}(\bar{Y}_i - \hat{M}_i)^2 / \sigma^2$ also follows the $\chi^2$ distribution, with degrees of freedom which is the difference between the dimension of $\bar{Y}_i$ and the dimension of $\hat{M}_i$ (denoted by $df_{r_i}$). As described in Proposition 7.1, note that $\hat{\sigma}^2_i$ is equal to the sum of $s^2_i$ and $2(l_a - l_0)$ and the two are independent. Since the sum of two independent $\chi^2$-distributed statistics also follows a $\chi^2$ distribution, $qn_i \hat{\sigma}^2_i / \sigma^2$ follows a $\chi^2$ distribution with degrees of freedom $q(n_i - 1) + df_{r_i}$. 

42
7.2 Matrix Gamma Distribution

7.2.1 Estimating shape parameters in the one-sample case

Recall that the pdf of matrix gamma distribution described in Example 2.1.2 is

\[ f(Y; \alpha, M) = \frac{|M|^{-\alpha}}{\alpha^{-\alpha p}\Gamma_p(\alpha)}|Y|^{\alpha-p+1/2}e^{\alpha \text{tr}(-M^{-1}Y)} \]

where \( \alpha \) is a shape parameter. In one-sample cases, \( \alpha \) can be estimated from the first derivative of log likelihood. Starting from the log likelihood with respect to \( \alpha \) and \( M \), we have

\[ l(M, \alpha) \propto n \left[ -\log \Gamma_p(\alpha) + \alpha \left\{ p \log \alpha - \log |M| + \frac{1}{n} \sum_{i=1}^{n} \log |Y_i| - \text{tr}(M^{-1}\bar{Y}) \right\} \right] \]

Taking the derivative with respect to \( \alpha \), we have

\[ \frac{\partial l}{\partial \alpha} = -n \log |M| - n \frac{\partial \log \Gamma_p(\alpha)}{\partial \alpha} + np \log \alpha + np + \frac{1}{n} \sum_{i=1}^{n} \log |Y_i| - n \text{tr}(M^{-1}\bar{Y}) \]

Thus, the MLE of \( \alpha \) should satisfy the condition that the first derivative of likelihood evaluated by \( M = \hat{M} \), the MLE under each restriction, is equal to 0. Therefore, the MLE of \( \alpha \) satisfies the equation

\[ p \log \alpha - \frac{\partial \log \Gamma_p(\alpha)}{\partial \alpha} = \log |\hat{M}| + \text{tr}(\hat{M}^{-1}\bar{Y}) - \frac{1}{n} \sum_{i=1}^{n} \log |Y_i| - p. \]

Since there is no analytical expression for multivariate gamma function, there is no closed-form solution for \( \alpha \). However, the equation for \( \alpha \) presented above is numerically well behaved. Therefore, if a numerical solution is desired, it can be found using an approximation method. Thus, we estimated \( \hat{\alpha} \) using this formula and numerical computation such as Newton’s method.

As \( \hat{\alpha} \) is the MLE, we can find the asymptotical distribution of the estimated value, using the property of MLE.

**Proposition 7.2.** Let \( \hat{\alpha} \) be the MLE of \( \alpha \). Then,

\[ \sqrt{n}(\hat{\alpha} - \alpha) \rightarrow_d N \left( 0, \left[ \frac{\partial^2 \log \Gamma_p(\alpha)}{\partial \alpha^2} - \frac{p}{\alpha} \right]^{-1} \right). \]

**Proof.** Because \( \hat{\alpha} \) is the MLE, it follows normal distribution asymptotically, using the property of normality of MLE. To estimate the variance, the Fisher information for \( \alpha \) is calculated from the second derivative of log likelihood, such that

\[ \frac{\partial^2 l}{\partial \alpha^2} = -n \left[ \frac{\partial^2 \log \Gamma_p(\alpha)}{\partial \alpha^2} - \frac{p}{\alpha} \right]. \]

Thus, \( I(\alpha) = -E(\frac{\partial^2 l}{\partial \alpha^2}) = n \left[ \frac{\partial^2 \log \Gamma_p(\alpha)}{\partial \alpha^2} - \frac{p}{\alpha} \right]. \)

Like the estimation of \( \alpha \), the information is also calculated using numerical computation. After this step, the
variance of MLE is obtained by taking inverse of this information. Therefore, the estimated value \( \hat{\alpha} \) follows \( N(\alpha, I(\alpha)^{-1}) \) asymptotically. ■

7.2.2 Estimating shape parameters in the two-sample case

For the two-sample case, \( \alpha_1 \) and \( \alpha_2 \) can be estimated in the same manner as in the one-sample case. We have the likelihood

\[
l(M_1, M_2, \alpha_1^2, \alpha_2^2) \propto n_1 \left[ \log \Gamma_p(\alpha_1) + \alpha_1 \left\{ p \log \alpha_1 - \log |M_1| + \frac{1}{n_1} \sum_{i=1}^{n_1} \log |Y_i| - \text{tr}(M_1^{-T}\bar{Y}_1) \right\} \right] + n_2 \left[ \log \Gamma_p(\alpha_2) + \alpha_2 \left\{ p \log \alpha_2 - \log |M_2| + \frac{1}{n_2} \sum_{j=n_1+1}^{n_1+n_2} \log |Y_j| - \text{tr}(M_2^{-T}\bar{Y}_2) \right\} \right].
\]

Taking the derivative with respect to each \( \alpha_i \), we have

\[
\frac{\partial l}{\partial \alpha_1} = -n_1 \log |M_1| - n_1 \frac{\partial \log \Gamma_p(\alpha_1)}{\partial \alpha_1} + n_1 p \log \alpha_1 + n_1 p + \sum_{i=1}^{n_1} \log |Y_i| - n_1 \text{tr}(M_1^{-1}\bar{Y}_1)
\]

\[
\frac{\partial l}{\partial \alpha_2} = -n_2 \log |M_2| - n_2 \frac{\partial \log \Gamma_p(\alpha_2)}{\partial \alpha_2} + n_2 p \log \alpha_2 + n_2 p + \sum_{i=1}^{n_2} \log |Y_i| - n_2 \text{tr}(M_2^{-1}\bar{Y}_2).
\]

The value of \( \alpha_i \) should satisfy the condition that the first derivative of likelihood evaluated by \( M_i = \hat{M}_i \) is equal to 0. However, since \( \hat{M}_i \) can be a function of both \( \alpha_1 \) and \( \alpha_2 \), depending on the restriction, estimates of \( \alpha_i \) cannot be obtained separately. We can get the solution for \( \alpha_1 \) and \( \alpha_2 \) simultaneously by solving the above systematic equations.

As in the one-sample case, there is no analytical expression for multivariate gamma function. For this reason, no closed-form solution for \( \alpha_i \) is provided. Instead, systematic equations for \( \alpha_i \) presented above can be solved by using the approximation method, such as multivariate Newton’s method. Thus, we estimate \( \hat{\alpha}_i \) using this formula and numerical computation.
Chapter 8

Numerical Study

8.1 Simulation Setting

We perform simulation studies for the scenarios presented in Sections 3, 4, and 6 to confirm the asymptotic distributions of the estimators and the test statistics, and evaluate their statistical power. The dimensions of data matrices are chosen as \( p = 2 \) or \( p = 3 \), depending on the scenarios. In the case of the gamma distribution, the shape parameter \( \alpha \) should be greater than \( p \). Since we do simulations with \( p = 2 \) or \( p = 3 \) in this section, we set \( \alpha = 4 \). In addition, since we ought to use \( M = \alpha V \) for generation of matrix gamma data in R, we use \( M = \alpha V \) for a \( p \times p \) matrix \( V \). We choose the same value of \( M \) for the Gaussian distribution, and pick the covariance parameter as \( \sigma^2 = 4^2 = 16 \).

8.1.1 Distribution of \( \hat{U} \) and \( \hat{D} \) in the one-sample case

To check the distribution of \( \hat{U} \) and \( \hat{D} \) numerically, we generate two different datasets first. Taking the dimension of data matrix as \( p = 2 \), \( n = 200 \) iid samples are generated from the distributions;

1. Gaussian with \( M = \text{Diag}(8,4) \) and \( \text{Cov}(\text{vecd}(Y)) = 4^2 I_3 \).
2. Matrix gamma distribution, with parameters \( \alpha = 4 \) and \( M = \text{Diag}(8,4) \).

With these two datasets, we calculate the MLE of \( \hat{U} \) and \( \hat{D} \) as derived in section 3. This process is performed with \( r = 10000 \) replications, so 10000 MLE sets for each distribution are calculated. Then, with these 10000 MLE sets, we check the asymptotic distributions of off-diagonal elements of \( \hat{A} = \log(U^T \hat{U}) \) and diagonal elements of \( \hat{D} \) by drawing the normal qq plots.

8.1.2 Null distribution of test statistics

For the one-sample case, we perform the four tests listed below:

1. Test of means (Full Matrix test): \( H_0 : M = U_0 D_0 U_0^T \) vs \( H_a : M \) is unrestricted. (sec 4.1)
2. Test of eigenvectors with fixed eigenvalues: \( H_0 : M = U_0 D_0 U_0^T \) vs \( H_a : M \in \mathbb{M}_{D_0} \). (sec 4.2)
3. Test of eigenvalues with fixed eigenvectors: \( H_0 : M = U_0 D_0 U_0^T \) vs \( H_a : M \in \mathbb{M}_{U_0} \). (sec 4.3)
4. Test of multiplicities: \( H_0 : M \in \mathbb{M}_{m_1, \ldots, m_k} \) vs \( H_a : M \) is unrestricted. (sec 4.4)

Taking the dimension of data matrix as \( p = 3 \), \( n = 100 \) iid samples are generated from the distributions:

1. Gaussian with \( M = \text{Diag}(8, 8, 4) \) and \( \text{Cov}(\text{vecd}(Y)) = 4^2 I_6 \).
2. Wishart distribution, with parameters \( \alpha = 4 \) and \( M = \text{Diag}(8, 8, 4) \).

With the dataset, we estimate the eigenvalues and eigenvectors of the mean parameter matrix using formulas derived in Chapter 3. The likelihood ratio test statistics and p-values are also computed for each hypothesis test. This process is performed with \( r = 10000 \) replications, so 10000 test statistics and p-values are calculated for each scenario.

In the two-sample case, we use the following three tests:

1. Test of means (Full Matrix test): \( H_0 : M_1 = M_2 \) vs \( H_a : M_1 \neq M_2 \). (sec 6.1)
2. Test of equality of two eigenvectors (eigenvector test): \( H_0 : M_1 = M_2 \) vs \( H_a : (M_1, M_2) \in \mathbb{M}_{2,D} \). (sec 6.2)
3. Test of eigenvalue structure (eigenvalue test): \( H_0 : (M_1, M_2) \in \mathbb{M}_{2,D} \) vs \( H_a : (M_1, M_2) \notin \mathbb{M}_{2,D} \). (sec 6.3)

Similar to the one-sample case, taking the dimension of data matrix as \( p = 3 \), \( n_1 = n_2 = 100 \) iid samples are generated from the distributions:

1. Gaussian with \( M = \text{Diag}(12, 8, 4) \) and \( \text{Cov}(\text{vecd}(Y)) = 4^2 I_6 \).
2. Wishart distribution, with parameters \( \alpha = 4 \) and \( M = \text{Diag}(12, 8, 4) \).

With these two groups of data, we estimate eigenvalues and eigenvectors of mean parameter matrices using the formulas derived in Chapter 5. Then, using the estimated eigenvalues and eigenvectors, the likelihood ratio test statistics and p-values are also computed for each hypothesis. This process is performed with \( r = 10000 \) replications, so 10000 test statistics and p-values are calculated for each scenario.

**8.1.3 Statistical Power of Tests**

The simulation scenarios are based on (Schwartzman et al., 2010) with dimension \( p = 3 \). In the one-sample case, under the null hypothesis, \( n = 50 \) iid samples are generated from the distributions:

1. Gaussian with \( M = \text{Diag}(12, 8, 4) \) and \( \text{Cov}(\text{vecd}(Y)) = 4^2 I_6 \).
2. Wishart distribution, with parameters \( \alpha = 4 \) and \( M = \text{Diag}(12, 8, 4) \).

We consider two sets of alternatives to assess the power of the proposed test statistics. The first set of alternatives is defined in terms of changes in the eigenvalues of \( M_0 \). These alternatives are defined as
\[ M_a = M_0 + U_0(\Delta D)U_0^T \] where \( \Delta D \) is a diagonal matrix that represents changes in eigenvalues. This reduces to \( M_a = M_0 + \Delta D \) in our simulation, since \( M_0 \) is diagonal and \( U_0 \) is the identity matrix. Four values of \( \Delta D \) are chosen from easier to harder to detect:

1. \( \Delta D = (0.8,0.8,0.4) \)
2. \( \Delta D = (0.8, -0.8, 0.4) \)
3. \( \Delta D = (-0.4, -0.4, 0.4) \)
4. \( \Delta D = (0.2, 0.2, 0.2) \)

The second set of alternatives is defined by changes in eigenvectors. Using Rodrigues’ rotation formula, a rotation matrix \( Q \in O_3 \) is given as

\[
Q = \exp \left[ \theta \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \right]
\]

where \( a = (a_1, a_2, a_3) \) is a unit vector representing the axis of rotation and \( \theta \) is the rotation angle around this axis. With this rotation matrix, the alternative is formulated as \( M = QM_0Q^T \). Four unit vectors are chosen to make the rotation matrix, with a fixed angle \( \theta = \pi/12 \) for all cases:

1. \( \mathbf{a} = \frac{1}{\sqrt{3}}(-1,-1,-1) \)
2. \( \mathbf{a} = (1,0,0) \)
3. \( \mathbf{a} = \frac{1}{\sqrt{2}}(0,1,1) \)
4. \( \mathbf{a} = \frac{1}{\sqrt{2}}(1,0,-1) \)

With these datasets, we estimate eigenvalues and eigenvectors of the mean parameter matrix using the formulas derived in Chapter 3, and the likelihood ratio test statistics are computed for each of the four scenarios (Full Matrix Test, Eigenvalue test, Eigenvector test, and Multiplicity test). With these test statistics, the statistical power is computed for each set of alternatives. This process is performed with \( r = 10000 \) replications.

In the two-sample case, one sample group is generated from the same process as the null hypothesis setting in the one-sample case, with the same parameters, while the other group is generated under the alternative hypothesis setting as the one-sample case (eigenvalue change, eigenvector change). The sample size is \( n_1 = n_2 = 50 \) for each hypothesis set. With these two groups of data, we estimate eigenvalues and eigenvectors of the mean parameter matrix for each group by using the formula derived in the section 5. These are performed for both Gaussian and Wishart distribution samples. Likelihood ratio test statistics and statistical power are computed for each of the three scenarios (Full Matrix Test, Eigenvalue test, Eigenvector test).
This process is performed with \( r = 10000 \) replications.

### 8.1.4 Estimation of nuisance parameters

As described in Chapter 7, in addition to estimating \( M \) and calculating related test statistics, we need to estimate other parameters such as a variance component and scale parameters. Thus, through the simulation, we need to confirm if our theoretical calculations derived in the previous chapter are appropriate. First, under the Gaussian distribution assumption, we use a spherical covariance matrix to generate the data, which has the form \( \text{Cov}(\text{vecd}(Y)) = \sigma^2 I_q \). Here, we use \( \sigma^2 = 4^2 \). With \( r = 10000 \) replications and the sample size \( n = 100 \), we check the distribution of estimated parameters under the various restrictions. From Proposition 7.1, we have that \( qn\hat{\sigma}^2/\sigma^2 \sim \chi^2(q(n - 1) + df_r) \). Thus, we check the distribution of \( \hat{\sigma}^2 \) with the generated data.

For the two-sample Gaussian case, we have two options: equal or unequal covariance. If we assume that the covariance matrix of two groups are equal, then we need to estimate the pooled covariance matrix \( \Sigma_{12} = \hat{\sigma}^2_{12} I_q \). In Section 7.1, it is said that the pooled covariance parameter \( \hat{\sigma}^2_{12} \) is related to the distribution that \( q(n_1 + n_2)\hat{\sigma}^2_{12}/\sigma^2 \sim \chi^2(q(n_1 + n_2 - 2) + df_{r_2}) \). Thus, we check the distribution of \( \hat{\sigma}^2_{12} \) with the generated data.

Next, for the one-sample matrix gamma case, we have the nuisance parameter \( \alpha \). Thus, to check the distribution of \( \alpha \), we use the Wishart samples since wishart distribution is the special case of matrix gamma distribution. We check if the estimated \( \hat{\alpha} \) is asymptotically normal distributed with mean \( \alpha \) and variance \( \left[ \frac{\partial^2 \log \Gamma_p(\alpha)}{\partial \alpha^2} - \frac{p}{\alpha} \right]^{-1} \). Here, we used \( \alpha = 4 \); therefore, the mean of \( \hat{\alpha} \) should be 4.

In the two-sample matrix gamma problem, we have two nuisance parameters \( \alpha_1 \) and \( \alpha_2 \). Two Wishart samples with different \( \alpha_i, i = 1, 2 \) are generated, and we check if the estimated \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are asymptotically normal distributed. Here, we used \( \alpha_1 = 4 \) and \( \alpha_2 = 2 \), so the mean of \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) are 4 and 2 and variances are \( \left[ \frac{\partial^2 \log \Gamma_p(\alpha_i)}{\partial \alpha_i^2} - \frac{p}{\alpha_i} \right]^{-1} \) for \( i = 1, 2 \).

### 8.2 Simulation Results

#### 8.2.1 Distribution of MLEs

First we evaluate the distribution of the off-diagonal elements of \( \hat{A} = U^T \hat{U} \). Figures 8.1 and 8.2 show the normal Q-Q plots of \( a_{12} \) in the Gaussian and matrix gamma distribution scenarios. Since both figures show that all the points are close to the 45-degree lines, we can confirm that the off-diagonal elements of \( \hat{A} \) in the Gaussian and matrix gamma distributions follow the normal distribution asymptotically. The variances are calculated by evaluating the formula we have derived in Section 3.3 at the true parameter values \( d_1 = 8, \ d_2 = 4, \sigma^2 = 4^2, \) and \( \alpha = 4 \).

Now we move to the distribution of eigenvalues. Figures 8.3 to 8.4 show the normal Q-Q plots of each of the diagonal elements of \( \hat{D} \) in the Gaussian and matrix gamma distributions scenario. As all figures show
that all the points are close to the 45-degree lines, we can confirm that the diagonal elements of $\hat{D}$ in both the Gaussian and matrix gamma distributions follow the normal distribution asymptotically. The variances are
calculated by evaluating the formula we have derived in Section 3.4. at the true parameter values, $d_1 = 8$, $d_2 = 4$, $\sigma^2 = 4^2$ and $\alpha = 4$. It shows that our estimated MLE is appropriate. Moreover, Figure 8.5 shows the gamma Q-Q plot of each of the diagonal elements of $\hat{D}$ in the matrix gamma case, based on the derivation in Section 3.4. It seems that all the points are close to the 45-degree lines, so we can confirm that our derived distribution for diagonal entries of $\hat{D}$ in Example 3.4.2. is accurate.

### 8.2.2 Distribution of P-values

We evaluate the accuracy of the approximate distributions of test statistics by examining the pattern of empirical distribution of plots. Figures 8.6 shows the distribution of the p-values for test statistics of each hypothesis described above in the one-sample case. The closeness of the empirical cumulative distribution function of p-values to the 45-degree line indicates the goodness of the fit. When we see the plot, all tests show good performances since these empirical distributions of p-values are near the 45-degree red line.
Figures 8.6 and 8.7 represent the distribution of p-values for test statistics in the one-sample and two-sample cases, respectively. Like the one-sample case, the closeness of the empirical cdf of p-values to the 45-degree line indicates the goodness of the fit. When we see the plots, all tests show good performances since these empirical distributions of p-values are close to the 45-degree red line. Thus, the derived distributions of test statistics are accurate.
Figure 8.8: ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Gaussian sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$.

8.2.3 Statistical Power of Tests

Figures 8.8 and 8.9 show the results for various alternatives related to eigenvalue changes in the one sample case. When the eigenvalues are closer to each other, it is harder to detect the difference. By design, the eigen-
Figure 8.9: ROC curves for eigenvalue changes: one-sample case, with $n = 50$, Wishart sample. Four eigenvalue-changing matrices are chosen as (a) $\Delta \mathbf{D} = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta \mathbf{D} = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta \mathbf{D} = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta \mathbf{D} = \text{Diag}(0.2, 0.2, 0.2)$. 
Figure 8.10: ROC curves for eigenvector changes: one-sample case, with $n = 50$, Gaussian sample. Four unit vectors for rotation matrices are chosen as (a) $a = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $a = (1, 0, 0)$, (c) $a = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $a = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.
Figure 8.11: ROC curves for eigenvector changes: one-sample case, with $n = 50$, Wishart sample. Four unit vectors for rotation matrices are chosen as (a) $a = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $a = (1, 0, 0)$, (c) $a = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $a = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.
Figure 8.12: ROC curves for eigenvalue changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. 
Figure 8.13: ROC curves for eigenvalue changes: two-sample Wishart case, with $n_1 = n_2 = 50$. Four eigenvalue-changing matrices are chosen as (a) $\Delta D = \text{Diag}(0.8, 0.8, 0.4)$, (b) $\Delta D = \text{Diag}(0.8, -0.8, 0.4)$, (c) $\Delta D = \text{Diag}(-0.4, -0.4, 0.4)$, (d) $\Delta D = \text{Diag}(0.2, 0.2, 0.2)$. 

57
Figure 8.14: ROC curves for eigenvector changes: two-sample Gaussian case, with $n_1 = n_2 = 50$. Four unit vectors for rotation matrices are chosen as (a) $a = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, (b) $a = (1, 0, 0)$, (c) $a = (0, 1/\sqrt{2}, 1/\sqrt{2})$, (d) $a = (1/\sqrt{2}, 0, -1/\sqrt{2})$. The angle $\theta = \pi/12$ is fixed for all cases.
Figure 8.15: ROC curves for eigenvector changes: two-sample Wishart case, with \( n_1 = n_2 = 50 \). Four unit vectors for rotation matrices are chosen as (a) \( \mathbf{a} = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}) \), (b) \( \mathbf{a} = (1, 0, 0) \), (c) \( \mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2}) \), (d) \( \mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2}) \). The angle \( \theta = \pi/12 \) is fixed for all cases.
value tests are the most powerful, with the full matrix tests not far behind. The eigenvector tests have no power since they are not designed to capture changes in eigenvalues. Figures 8.10 and 8.11 show the results for various alternatives related to eigenvector changes in one sample case. For all alternatives, eigenvector tests are the most powerful, while the eigenvalue tests have very little power. Figures 8.12 to 8.13 represent the results for various alternatives related to eigenvalue changes in the two-sample case. Here, similar to the one-sample case, eigenvalue tests are the most powerful for the all alternatives and the full matrix test is also powerful in both distributions. Still, eigenvector tests have no power since they are not designed to capture changes in the eigenvalues. Figures 8.14 to 8.15 present the results for various alternatives related to the eigenvector changes in the two-sample case. Like the one-sample case, eigenvector tests are the most powerful in both distributions, while the eigenvalue tests have very little power.

8.2.4 Multiplicity test

Figure 8.16 and 8.17 show how the multiplicity tests are affected by changes in eigenvalues and eigenvectors. We can see that only case (b) in eigenvalue change scenarios can affect the power of multiplicity test, while the other cases and the change in eigenvectors does not make any difference to the result. The reason is that only the change in eigenvalues for case (b) makes the number of distinct eigenvalues different from the
Figure 8.17: ROC curves for multiplicity test: Gaussian and matrix gamma with \( n = 50 \). Four eigenvector-changing matrices are chosen as (a) \( \mathbf{a} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}) \), (b) \( \mathbf{a} = (1, 0, 0) \), (c) \( \mathbf{a} = (0, 1/\sqrt{2}, 1/\sqrt{2}) \), (d) \( \mathbf{a} = (1/\sqrt{2}, 0, -1/\sqrt{2}) \). Figure (a)

original eigenvalue matrix. Since the multiplicity test is only related to the number of distinct eigenvalues, this result is expected.

8.2.5 Estimation of other parameters

Figures 8.18 to 8.20 show the distribution of the estimated variance parameter under various restrictions in the Gaussian case following the settings in Section 8.1.4. All the relevant distributions are given in Section 7.1. The closeness of the estimated values to the 45-degree line in all the plots indicates the goodness of the fit. Thus, we can confirm that the estimated variance parameters follow the prescribed chi-square distribution.

In addition, we have the nuisance shape parameter in matrix gamma distribution, which is denoted as \( \alpha \). Figure 8.21 is the normal qq plot of \( \hat{\alpha} \), the estimate of the nuisance shape parameter \( \alpha \) in matrix gamma distribution. Since all the points are close to the 45-degree line, we can say that our estimating formula is appropriate. Figure 8.22 shows the distribution of \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \) in the two-sample case under the assumption that \( \alpha_1 \neq \alpha_2 \). For all three tests for the two-sample case, all the points are close to the 45-degree line. Thus, our estimators for \( \alpha_i \) as described in chapter 7 are appropriate.
Figure 8.18: The distribution of $\hat{\sigma}^2$: one-sample Gaussian case, with $n = 100$.

Figure 8.19: The distribution of $\hat{\sigma}_{12}^2$: two-sample Gaussian case, equal covariance assumption, with $n_1 = n_2 = 100$. 
Figure 8.20: The distribution of $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$: two-sample Gaussian case, different covariance assumption, with $n_1 = n_2 = 100$.

Figure 8.21: The distribution of $\hat{\alpha}$: one-sample Wishart case, with $n = 100$. 
Figure 8.22: The distribution of $\hat{\alpha}_1$ and $\hat{\alpha}_2$: two-sample Wishart case, with $n_1 = n_2 = 100$. 
Chapter 9

Real Data Example

9.1 Data description

We analyze a dataset from an observational study of brain images of children (Dougherty et al., 2007; Schwartzman et al., 2010). The dataset consists of 34 brain DTI images of 10-year old children, where 12 images are from boys and 22 images are from girls. The goal of this data analysis is to see whether there exists some regions that represent the anatomic difference between genders.

9.2 Model checking

9.2.1 Model fit

Before we analyze the data, we first applied a test to the data in terms of model checking. Using the property that $\text{vecd}(Y) \sim MVN_q(\text{vecd}(M), \sigma^2 I_q)$ if and only if $Y \sim N_{pp}(M, \sigma^2 I_q)$, we used the vectorized data obtained by applying $\text{vecd}$ operator to check the normality. We performed Mardia’s test (Mardia, 1970) for $p$-dimensional multivariate data at each voxel, which is based on multivariate extensions of skewness $(\gamma_{1,p})$ and kurtosis $(\gamma_{2,p})$ measures as follows:

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}^3$$
$$\hat{\gamma}_{2,p} = \frac{1}{n} \sum_{i=1}^{n} m_{ii}^2$$

where $m_{ij} = (y_i - \bar{y})^T S^{-1} (y_j - \bar{y})$ is the squared Mahalanobis distance, and $p$ is the number of variables. The test statistic for skewness $(n/6)\hat{\gamma}_{1,p}$ is approximately $\chi^2$ distributed with $p(p + 1)(p + 2)/6$ degrees of freedom, and the test statistic for kurtosis $\hat{\gamma}_{2,p}$ is approximately normally distributed with mean $p(p + 2)$ and variance $8p(p + 2)/n$. If at least one of the tests concludes that the null hypothesis is rejected, then multivariate normality assumption is violated. In addition, Mardia (1974) also introduced an correction term
into the skewness test statistic for small samples, usually when \( n < 20 \). The corrected skewness statistic for small samples is \((nk/6)\hat{\gamma}_{1,p}\), where \( k = (p+1)(n+1)(n+3)/(n(n+1)(p+1)6) \), and it also follows \( \chi^2 \) distribution with degrees of freedom \( p(p+1)(p+2)/6 \). Since the sample number of boy group is \( n_1 = 12 \), we used the small sample correction term for skewness to test the normality of boy group at each voxel.

For this data, the null hypothesis that the two groups of observed data are from multivariate normal distribution was not rejected at the 0.05 significance level for 76.40% of voxels. In addition, we applied the same test to the log-transformed data and get the result that the null hypothesis was not rejected at the 0.05 significance level for 73.56% of voxels.

Figure 9.1 and 9.2 show histograms of two test statistics for each group, to check the normality and log-normality. The red lines represent the critical value of test statistics to reject the null hypothesis at \( \alpha = 0.05 \). For the skewness test, the left side of the red line contains voxels where the null hypothesis is not rejected. And the area between two red lines represents the voxels where the null hypothesis of kurtosis test is not rejected.

There are approximately 80000 voxels (76.43% approximately) where the skewness test concludes that the null hypothesis is not rejected, while the kurtosis test does not reject the null hypothesis at the 97.51% of voxels. Since we can assess the normality or log-normality of the data only if both two tests are not rejected in two groups at each voxel, the number of voxels where the normality assumption holds is lower than the number of voxels where one of the tests does not reject the null hypothesis.
We also draw some chi-square q-q plots for multivariate data, using the property that the squared mahalanobis distance $m_{ii}$ defined above follows approximately chi-square distribution with the number of degrees of freedom $p$ if the data is from the multivariate normal distribution. This is because the squared mahalanobis distance is the sum of $p$ squared standard normal variables if the data follows multivariate normal distribution. Figure 9.3 shows chi-square q-q plots at a voxel where the null hypothesis is not rejected, and Figure 9.4 shows chi-square q-q plots at a voxel where the normality assumption is violated. We can see that even though the null hypothesis is rejected, the pattern of data points at this voxel is not so different from the voxel where the normality assumption is valid. Thus, based on the results of above tests and figures,
it seems to be possible that the assumption of Gaussian distribution and log-normal distribution can be set for the DTs at every voxel.

In addition to checking the normality, we also performed a tests to check the multivariate gamma distribution. Using the property that the diagonal elements of matrix gamma distribution are independently gamma distributed as mentioned in Example 3.4.2, we performed elementwise Kolmogorov-Smirnov test to check whether the possibility of gamma distribution exists or not. The Kolmogorov-Smirnov test is a nonparametric test of the equality of probability distributions which can be used as a goodness of fit test (Smirnov, 1948). The Kolmogorov-Smirnov statistic is calculated as a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. Figure 9.5 and 9.6 show histograms of p-values for three diagonal elements. The red lines represent the significance level $\alpha = 0.05$. As we can see, there are relatively small amounts of voxels where the null hypothesis is rejected. All three
diagonal elements show similar patterns. Thus, the matrix gamma distribution assumption can be considered in most voxel, for both boy and girl groups.

Based on the previous results, we assumed three distributions for the DTs; Gaussian distribution, log-normal distribution, and matrix gamma distribution. Since our data consists of diffusion tensors and the DTs in each voxel are all positive definite by definition, the matrix gamma distribution was also considered. The positive definite constraint for the data is satisfied in the log-normal and the matrix gamma distributions automatically. For the Gaussian distribution, we assume that the probability that matrices are not positive definite is small.

At each voxel, the two-sample comparisons are performed, by computing the test statistics for each hypothesis test (eigenvector test, eigenvalue test, full matrix test) under each distribution assumption. With these test statistics, the corresponding p-values are also computed using their approximate null distributions that were described in previous chapters. For our data, \( n_1 = 12 \), \( n_2 = 22 \), and \( p = 3 \) at each voxel, and we have \( v = 105,822 \) voxels in total inside the brain.

### 9.2.2 Model checking for the Gaussian model

Figure 9.7 presents the p-value distribution of LRT and Box’s M test, which tests whether the covariance matrices of the two groups are equal. The LRT statistic for this hypothesis test is

\[
T = (n_1 + n_2) \log |S| - n_1 \log |S_1| - n_2 \log |S_2|
\]
where $S_i$ is a covariance matrix of each group and $S = \frac{(n_1 \Sigma_1 + n_2 \Sigma_2)}{(n_1 + n_2)}$ is a pooled covariance matrix (Mardia, Kent, and Bibby, 1979, page 140).

For a case where the sample size is small, Box’s M test (Box, 1949) is also provided;

$$
M = \gamma \left( \sum_{i=1}^{2} S_{ui} \right)
$$

where

$$
\gamma = 1 - \frac{(2p^2 + 3p - 1)}{6(p+1)} \left( (n_1 - 1)^{-1} + (n_2 - 1)^{-1} - 2(n_1 + n_2 - 2)^{-1} \right)
$$

$$
S_{ui} = (n_i - 1) \log \left( \left( \frac{n_i}{n_i - 1} S_i \right)^{-1} \frac{(n_1 + n_2)}{(n_1 + n_2 - 2)} S \right).
$$

Both LRT and Box’s M test statistics follow $\chi^2$ distribution with degrees of freedom $\frac{q(q+1)}{2}$ asymptotically. Since the plot shows that two lines of p-value distributions are far from 45-degree lines, we need to make an assumption that the covariance matrices are different in most voxels.

Figure 9.8 shows the test of sphericity of covariance matrices for each voxel, as described in Section 7.1.1. We can see that the distribution of p-values for boys and girls are far from 45-degree line. Thus, in most voxels, we reject the null hypothesis that covariance matrices are spherical. Based on this result, we need to make an arbitrary covariance assumption for the model.

The next step is to check the structure of eigenvalues and eigenvectors. We performed two multiplicity tests to check the number of distinct eigenvalues;
The distribution of p-values for these tests are shown in Figure 9.9 and 9.10. We can see that both two p-value distributions are far from the 45-degree line. In addition, the p-value level plots for these tests shown in Figure 9.11 and 9.12 represent that almost all of the voxel are highlighted. These plots show that the two multiplicity hypotheses are rejected in most voxels. Thus, we can make the assumption that three eigenvalues of mean parameter are distinct. Note that most of the highlighted voxels where the eigenvalues are different correspond to the brain’s white matter, conforming to the known fact that diffusion in the white matter is anisotropic.

9.3 Analysis

9.3.1 Gaussian distribution

Under the Gaussian distribution assumption, we used arbitrary covariance assumption, based on the result of equality and the sphericity test in the previous section. The form of the test statistic and its distribution are described in Schwartzman et al. (2010). We used a sample covariance matrix, which has the form of

\[
S_i = \hat{\text{Cov}}(\text{vecd}(Y_i)) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \text{vecd}(Y_{ij} - \bar{Y}_i)\{\text{vecd}(Y_{ij} - \bar{Y}_i)\}^T, \quad i = 1, 2.
\]
9.3.2 Log-normal distribution

Now we took the logarithm of the original data to eliminate the positive definite constraints, referring to the data analysis in (Schwartzman et al. 2010). Thus, we assume that the original data follows the log-normal distribution. Figure 9.14 shows the empirical distribution of the p-values for the three tests with the log-normal distribution assumption. Like the Gaussian distribution case, the full matrix test and eigenvector test are not so different from each other, but both are better than the eigenvalue test. The results appear to be
Figure 9.12: The p-value level plot in scale $-\log_{10}(p)$ for multiplicity test at slice $z = 49$ : $k = 2$, matrix gamma distribution

Figure 9.13: The distribution of p-values: two-sample Gaussian distribution with arbitrary covariance, using weighted chi-square distribution

slightly better than when modeling the data according to the normal distribution, but the difference is small.

9.3.3 Matrix Gamma distribution

The matrix gamma distribution is defined for positive definite matrix data. Thus, we can also assume that the original data follows the matrix gamma distribution. Before the analysis, we first check whether $\alpha_1$ and $\alpha_2$ are different or not. Figure 9.15 shows the distribution of the log of the relative difference between $\alpha_1$ and $\alpha_2$. 
Figure 9.14: The distribution of p-values: two-sample log-normal distribution with arbitrary covariance, using weighted chi-square distribution.

Figure 9.15: Histogram of $\log \frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\alpha_1}$.

$\alpha_2$ across voxels, and Figure 9.16 presents the scatterplot of $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Based on the shape of the histogram and scatterplot, we can assume that $\alpha_i$s are different between groups.

Figure 9.17 shows the empirical distribution of the p-values for the three tests with the matrix gamma distribution assumption. Three tests are entangled at the high p-value area, but the eigenvector test has the closest to the 45-degree line and the eigenvalue test is better than the other two tests. Therefore, to detect the difference between these two groups, the eigenvalue test seems to be better.
Figure 9.16: Scatterplot of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

Figure 9.17: The distribution of p-values: two-sample Wishart distribution

### 9.3.4 P-value Maps

Figure 9.18 and 9.19 show maps of the p-values in $-\log_{10}$ scale under the Gaussian distribution assumption. Some of the most significant differences between boys and girls are presented as white in the plot. We can see that the eigenvector test shows the largest area of differences between the two groups, highlighted in white. Figure 9.20 and 9.21 show maps of the p-values in $-\log_{10}$ scale under the log-normal distribution assumption. Like the gaussian distribution, the eigenvector test shows the largest highlighted area.
Figure 9.18: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 40$: Gaussian distribution

Figure 9.19: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 49$: Gaussian distribution

Figure 9.22 and 9.23 show maps of the p-values in $-\log_{10}$ scale under the matrix gamma distribution assumption. We can see that eigenvalue test and full matrix test show similar size of highlighted area. While the eigenvalue and full matrix tests in the matrix gamma case detect more voxels that are different between the two groups, the eigenvector test in the matrix gamma case shows the same anatomical pattern as in the Gaussian and log-normal cases. Therefore, the various modeling assumptions do not give contradictory results in terms of the eigenvector test. Yet, the eigenvalue test in the matrix gamma case is sensitive to differences in eigenvalues in the gray matter that are not picked up by the Gaussian model.
Figure 9.20: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 40$: log-normal distribution

Figure 9.21: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 49$: log-normal distribution

Figure 9.22: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 40$: matrix gamma distribution
Figure 9.23: The p-value level plot in scale $-\log_{10}(p)$ at slice $z = 49$: matrix gamma distribution
Chapter 10

Summary and Discussion

10.1 Summary

This dissertation has described inferences for the eigenstructure of the mean matrix of symmetric positive definite matrix-variate data. We have studied an exponential family of positive definite symmetric matrices, including the derivation of MLEs and LLR test statistics for the eigenstructure of the mean parameter $M$ under the exponential family assumption. These procedures were applied to both one-sample and two-sample problems. Through simulations and data analysis, we showed that the exponential family assumption could be appropriate to analyze the matrix-variate data.

10.2 Discussion

We defined an exponential family model for symmetric matrix-variate data to find a general way to make inferences for the mean matrix. This model flexibly included both the spherical Gaussian and matrix-variate Gamma distributions. The matrix-variate normal distribution with orthogonally invariant covariance (OI covariance) as defined in Schwartzman et al. (2008) does not conform to this model because it requires another parameter $\tau$ in addition to the variance parameter $\sigma^2$, which affects the norm of the matrix $Y - M$, and therefore it cannot be written in the exponential family form we defined. Our research could be extended to find a more general form of distribution family that handles this problem.

As our primary interest was the inference for the mean parameter $M$, all other parameters were treated as known or fixed parameters. However, since these nuisance parameters can affect the estimation procedure and test statistics in data analysis, we needed to estimate them in an appropriate way. However, here we used the maximum likelihood estimation to estimate these parameters, because of their efficiency (Pearson, 1936). Nevertheless, the method of moments can yield consistent estimators under weak assumptions quickly and easily. Using other methods to estimate nuisance parameters can be an extension of our research.

In the data analysis, we found that the sphericity of covariance matrices assumption was invalid. To handle
this problem in the Gaussian case, we used the approximation in Schwartzman et al. (2010) for the distribution of LRT statistics assuming an arbitrary covariance structure. The problem did not occur in the case of the matrix-Gamma distribution because the covariance matrix depends on the mean matrix, which is arbitrary. However, the covariance cannot be specified separately, as in the Gaussian case.

Our research could be extended to generalized linear models (GLMs) for matrix-variate data. In this dissertation, we focused on looking for differences in eigenvalues and eigenvectors between two groups under the exponential family assumption. In a GLM formulation it could be assumed that each outcome of the dependent variables $Y$ is generated from a particular distribution in the exponential family as $Y_i = \psi^{-1}(XB) + E_i$, where $X$ is an independent variable matrix, $B$ is an (unknown) coefficient matrix, and $\psi(*)$ is the link function such that $E(Y) = M = \psi^{-1}(XB)$. Such a GLM could be used to look for differences among several groups or to include covariates that can affect the eigenstructure of mean parameter.
REFERENCES


Appendix A

Properties needed for the proof of Theorems

A.1 Derivation of derivative rules in Example 2.2.2.

We will derive the derivative of vectorized inverse matrix, which is

\[
\begin{align*}
\frac{d\text{vec}(A^{-1})}{d\text{vec}(A)} &= -D^T(A^{-1} \otimes A^{-1})D.
\end{align*}
\]

Starting with \(AA^{-1} = I\), we get \(dAA^{-1} + AdA^{-1} = 0\) by taking the derivative. Simplifying this equation, we have

\[
\begin{align*}
dAA^{-1} &= -AdA^{-1} \\
dA^{-1} &= -A^{-1}dAA^{-1}.
\end{align*}
\]

Then, applying \(\text{vec}\) operator on both sides, we can get this term using the product rule:

\[
\begin{align*}
\text{vec}(dA^{-1}) &= -(A^{-1} \otimes A^{-1})\text{vec}(dA) \\
d\text{vec}(A^{-1}) &= -(A^{-1} \otimes A^{-1})d\text{vec}(A).
\end{align*}
\]

The, using the duplicate matrix \(D\), \(\text{vecd}\) operator is applied to above equation:

\[
\begin{align*}
Dd\text{vecd}(A^{-1}) &= -(A^{-1} \otimes A^{-1})Dd\text{vecd}(A) \\
d\text{vecd}(A^{-1}) &= -D^T(A^{-1} \otimes A^{-1})Dd\text{vecd}(A)
\end{align*}
\]

Then, we can get the derivative of inverse matrix as follows:

\[
\begin{align*}
\frac{d\text{vecd}(A^{-1})}{d\text{vecd}(A)} &= -D^T(A^{-1} \otimes A^{-1})D.
\end{align*}
\]
A.2 Derivative of a scalar function with respect to a diagonal matrix

Before we started the estimation in section 3.4., we needed a definition for the derivative of a scalar function by a diagonal matrix. In general, the derivative of a scalar function by a matrix is defined as a matrix with the elements \((\frac{\partial g(X)}{\partial X})_{ij} = \frac{\partial g(X)}{\partial x_{ij}})\). However, in the case of a diagonal matrix, all off-diagonal elements are 0, so we cannot take a derivative with respect to 0. Thus, using the concept of directional derivative, we define the partial derivative in an arbitrary \(ij\)th entry.

Let \(Y_{(ij)}\) be a matrix such that \(y_{ij} = 1\) and 0 elsewhere. Then, using the directional derivative definition to matrices, we have the directional derivative of \(g(X)\) in the direction of \(Y_{(ij)}\) as follows:

\[
\nabla_{Y_{(ij)}} g(X) = \lim_{t \to 0} \frac{g(X + tY_{(ij)}) - g(X)}{t}.
\]

This is the partial derivative of \(g(X)\) with respect to \(x_{ij}\). Note that for any \(i \neq j\), the elements of \(X\) are all 0. That means \(g(X)\) is originally the function of “diagonal” elements of \(X\) only. Thus, changes in off-diagonal elements of \(X\) cannot affect to the value of \(g(X)\), i.e. \(g(X + tY_{(ij)}) = g(X)\) for all \(i \neq j\), so the directional derivative of \(g(X)\) in the direction of \(Y_{(ij)}\) are 0 for all \(i \neq j\). Using this fact, we can conclude that the partial derivative of \(g(X)\) with respect to \(x_{ij}\) is 0 for all \(i \neq j\). Thus, we can use the given definition for the derivative of a scalar function by a diagonal matrix.

A.3 Other algebraic properties

**Proposition A.1.** If \(D \in \mathbb{D}_p\) and \(A\) is a general \(p \times p\) matrix, \(\text{Diag}(DA) = D \times \text{Diag}(A)\).

**Proof.** Let \(D = \text{Diag}(d_1, d_2, ..., d_p)\) and \(A = (a_1, ..., a_p)\) where \(a_j\) is the \(j\)th column of \(A\). Then, the \(ij\)th element of \(DA\) is \((DA)_{ij} = d_i e^T_i a_j\) where \(e_i\) is a vector whose \(i\)th element is 1 and 0 otherwise. Therefore, \((DA)_{ij} = d_i e^T_i a_j = d_i a_{ij}\) for all \(i, j\). Thus, the diagonal of \(DA\) is

\[
\text{Diag}(DA) = \text{Diag}((DA)_{11}, (DA)_{22}, ..., (DA)_{pp}) = \text{Diag}(d_1 a_{11}, ..., d_p a_{pp}).
\]

On the other hand, we know that \(\text{Diag}(A) = \text{Diag}(a_{11}, ..., a_{pp})\). Therefore, we can obtain that

\[
D \times \text{Diag}(A) = \text{Diag}(d_1, d_2, ..., d_p) \times \text{Diag}(a_{11}, ..., a_{pp}) = \text{Diag}(d_1 a_{11}, ..., d_p a_{pp})
\]

The first and second results are the same. Thus, this proposition has been proven.

**Proposition A.2.** Let \(\text{diag}(A)\) be a vector whose elements are diagonal elements of \(A\) and \(\text{offdiag}(A)\) be a vector whose elements are all off-diagonal elements of \(A\). Define the Frobenius inner product of two
\( p \times p \) matrices as \(< A, B >= \{ \text{vec}_d(A) \}^T \text{vec}(B) \) where \( \text{vec}_d(A) = (\text{diag}(A), \sqrt{2} \text{offdiag}(A)) \) is a \( \frac{p(p+1)}{2} \times 1 \) vector. If we have \(< A, B >= 0 \) for any \( B \in \mathbb{D}_p \), all diagonal elements of \( A \) are 0.

**Proof.** By the definition of inner product, we have

\[
< A, B >= \{ \text{vec}_d(A) \}^T \text{vec}(B) = (\text{diag}(A), \sqrt{2} \text{offdiag}(A))^T (\text{diag}(B), \sqrt{2} \text{offdiag}(B)).
\]

As we assume that \( B \) is a diagonal matrix, \( \text{offdiag}(B) = (0, ..., 0) \). Thus,

\[
< A, B >= (\text{diag}(A), \sqrt{2} \text{offdiag}(A))^T (\text{diag}(B), 0, ..., 0) = (\text{diag}(A))^T (\text{diag}(B)).
\]

From above, \(< A, B >= 0 \) is equivalent to \((\text{diag}(A))^T (\text{diag}(B)) = 0\). To make this inner product equal to zero for any \( B \), it is evident that \( \text{diag}(A) = (0, ..., 0) \).

**Proposition A.3.** If a general matrix \( A \) is positive definite, all the diagonal elements of \( A \) are positive. Thus, \( \text{Diag}(A) \in \mathbb{D}_p^+ \).

**Proof.** A matrix \( A \) is positive definite if and only if it satisfies these two conditions; All the diagonal elements of \( A \) are positive and \( A \) is strictly diagonally dominant (\(|a_{ii}| > \sum_{j \neq i} |a_{ij}|\)). As we assume that \( A \) is a positive definite, it naturally satisfies these two conditions. Thus, all the diagonal elements of \( A \) are positive, so \( \text{diag}(A) \in \mathbb{D}_p^+ \).

**Proposition A.4.** If two \( p \times p \) matrices \( A \) and \( B \) commute and \( A \) is a diagonal matrix, then \( B \) is also a diagonal matrix.

**Proof.** As \( A \) and \( B \) commute, \( AB = BA \). Thus,

\[
AB = \begin{pmatrix}
    a_{11} & 0 & \cdots & 0 \\
    0 & a_{22} & \cdots & 0 \\
    \vdots & \ddots & \ddots & \vdots \\
    0 & \cdots & 0 & a_{kk}
\end{pmatrix} \begin{pmatrix}
    b_{11} & \cdots & b_{1k} \\
    b_{21} & \cdots & b_{2k} \\
    \vdots & \ddots & \ddots \\
    b_{k1} & \cdots & b_{kk}
\end{pmatrix} = \begin{pmatrix}
    a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1k} \\
    a_{22}b_{12} & a_{22}b_{22} & \cdots & a_{22}b_{2k} \\
    \vdots & \ddots & \ddots & \vdots \\
    a_{kk}b_{1k} & \cdots & \cdots & a_{kk}b_{kk}
\end{pmatrix}
\]

\[
BA = \begin{pmatrix}
    b_{11} & b_{12} & \cdots & b_{1k} \\
    b_{21} & b_{22} & \cdots & b_{2k} \\
    \vdots & \ddots & \ddots & \vdots \\
    b_{k1} & \cdots & b_{kk}
\end{pmatrix} \begin{pmatrix}
    a_{11} & 0 & \cdots & 0 \\
    0 & a_{22} & \cdots & 0 \\
    \vdots & \ddots & \ddots & \vdots \\
    0 & \cdots & 0 & a_{kk}
\end{pmatrix} = \begin{pmatrix}
    b_{11}a_{11} & b_{12}a_{22} & \cdots & b_{1k}a_{kk} \\
    b_{21}a_{11} & b_{22}a_{22} & \cdots & b_{2k}a_{kk} \\
    \vdots & \ddots & \ddots & \vdots \\
    b_{k1}a_{11} & \cdots & \cdots & b_{kk}a_{kk}
\end{pmatrix}
\]

Now, we assume that \( AB = BA \), so we have
As all the off-diagonal elements in $AB$ and $BA$ are the same, all the off-diagonal elements of $B$ should satisfy all these equations:

\[
\begin{align*}
    a_{11}b_{12} &= b_{12}a_{22} \\
    a_{22}b_{12} &= b_{12}a_{11} \\
    \vdots \\
    a_{kk}b_{1k} &= b_{1k}a_{11} \\
    \vdots \\
    a_{kk}b_{kk} &= b_{kk}a_{kk}
\end{align*}
\]

These equations are satisfied only if all the off-diagonal elements of $B$ are 0. Thus, $B$ should be also a diagonal matrix.

**Proposition A.5.** The projection of some matrix $A$ to the space of antisymmetric $p \times p$ matrices $\mathbb{A}_p$, denoted as $A_{\text{proj}}$, is the antisymmetric part of $A$, which is $\text{Asymm}(A) = \frac{1}{2}(A - A^T)$.

**Proof.** From the definition of the projection, we need to find $\hat{A}$ which minimizes the distance between $A$ and $\hat{A}$. That is, $\hat{A} = \arg\min_{\hat{A} \in \mathbb{A}_p} ||A - \hat{A}||^2$. Separating this norm into symmetric and antisymmetric parts, we have

\[
||A - \hat{A}||^2 = ||A - \frac{A^T}{2} + \frac{A^T}{2} - \hat{A}||^2
\]

\[
= ||\frac{A - A^T}{2} + \frac{A + A^T}{2} - \hat{A}||^2
\]

\[
= ||\frac{A + A^T}{2} + (\frac{A - A^T}{2} - \hat{A})||^2
\]

\[
= ||\frac{A + A^T}{2}||^2 + ||\frac{A - A^T}{2} - \hat{A}||^2 + 2 < \frac{A + A^T}{2}, \frac{A - A^T}{2} - \hat{A}>
\]

The last term is equal to zero, since the symmetric matrix is orthogonal to the antisymmetric matrix. Thus, the problem is simplified as $\hat{A} = \arg\min_{\hat{A} \in \mathbb{A}_p} ||A - \frac{A - A^T}{2}||^2$. To minimize this, we should take $\hat{A} = \frac{A - A^T}{2}$, which is the antisymmetric part of $A$.

**Proposition A.6.** The projection of some matrix $A$ to the space of diagonal $p \times p$ matrices $\mathbb{D}_p$, denoted as...
\( \hat{D} \), is the diagonal matrix whose diagonal elements are the diagonal part of \( A \). That is, \( \hat{D} = \text{Diag}(A) \).

**Proof.** From the definition of the projection, we need to find \( \hat{D} \) which minimizes the distance between \( A \) and \( \hat{D} \). That is, \( \hat{D} = \text{argmin}_{\hat{D} \in D_p} \| A - \hat{D} \|^2 \). Separating this norm into diagonal and off-diagonal parts using the definition of trace, we have

\[
\| A - \hat{D} \|^2 = \text{tr}((A - \hat{D})^2) = \sum_{i,j} (A - \hat{D})_{ij}^2 = \sum_{i} (A - \hat{D})_{ii}^2 + \sum_{i \neq j} (A - \hat{D})_{ij}^2 = \| \text{diag}(A - \hat{D}) \|^2 + \| \text{offdiag}(A - \hat{D}) \|^2
\]

The second term is equal to \( \| \text{offdiag}(A) \| ^2 \), since \( \hat{D} \) is a diagonal matrix. Thus, to minimize the distance \( \| A - \hat{D} \|^2 \), we need to choose \( \hat{D} \) which minimizes \( \| \text{Diag}(A - \hat{D}) \|^2 \). Thus, we should take \( \hat{D} = \text{Diag}(A) \) since this \( \hat{D} \) makes the first term equal to zero. Therefore, \( \hat{D} = \text{Diag}(A) \) is the projection of \( A \) to the space of diagonal matrices.