A New Approach in Profile Analysis with High-Dimensional Data Using Scores

Cigdem Cengiz

Licentiate Thesis
Swedish University of Agricultural Sciences
Uppsala 2020

A New Approach in Profile Analysis with High-Dimensional Data Using Scores

Cigdem Cengiz

Faculty of Natural Resources and Agricultural Sciences
Department of Energy and Technology
Uppsala

Licentiate Thesis
Swedish University of Agricultural Sciences
Uppsala 2020

Cover illustration by Omer Talha Gulmezer.

Licentiate thesis/Report 113
ISBN (print version): 978-91-576-9783-7
ISBN (electronic version): 978-91-576-9784-4

Print: SLU Service/Repro, Uppsala 2020

A New Approach in Profile Analysis with High-Dimensional Data Using Scores

Abstract

In profile analysis, there exist three tests: test of parallelism, test of levels and test of flatness. In this thesis, these tests have been studied. Firstly, a classical setting, where the sample size is greater than the dimension of the parameter space, is considered. The hypotheses have been established and likelihood ratio tests have been derived. The distributions of these test statistics have been given. In the latter stage, all tests have been derived in a high-dimensional setting, where the number of parameters exceeds the number of sample size. Such settings have become more common due to the advances in computer technologies in the last decades. In high-dimensional data analysis, several issues arise with the dimensionality and different techniques have been developed to deal with these issues. We propose a dimension reduction method using scores that was first proposed by Läuter et al. (1996). To be able to find the specific distributions of the test statistics of profile analysis in this context, the properties of spherical distributions are utilized.

Keywords: High-dimensional data; hypothesis testing; linear scores; multivariate analysis; profile analysis; spherical distributions.

Author's address:
Cigdem Cengiz
SLU, Department of Energy and Technology,
Box 7032, SE-75007 Uppsala, Sweden.
E-mail: cigdem.cengiz@slu.se

Dedication

To my precious mother and the memory of my beloved father

”There are times when I feel like I’m in a big forest and don’t know where I’m going. But then somehow I come to the top of a hill and can see everything more clearly. When that happens it’s really exciting.”

”Of course, the most rewarding part is the ”Aha” moment, the excitement of discovery and enjoyment of understanding something new, the feeling of being on top of a hill, and having a clear view. But most of the time, doing mathematics for me is like being on a long hike with no trail and no end in sight.”

Maryam Mirzakhani, Fields Medalist 2014

Contents

List of Publications

This thesis is based on the work contained in the following report:

I Cengiz, C. and von Rosen, D. (2020). High-dimensional profile analysis. Linköping University Electronic Press, LiTH-MAT-R-2020/07–SE.

1 Introduction

Multivariate statistics focuses on methods for analyzing data which have been collected over time, also called repeated measures data, or for analyzing data which have been taken on different characteristics of a subject. The observations taken from an individual can be written as a column vector and this vector of observations can be considered as taken from a population according to some distribution law. The idea can be illustrated by a general $p \times n$ data matrix with n individuals and p variables:

$$\text{variables}\left\{ \overbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}}^{\text{individuals}} \right. .$$

Many methods that are used for analysing univariate observations can be generalized to multivariate observations. Instead of analysing one dimensional quantities which appear in the univariate case, such as mean, variance, etc., we will have a vector or matrix of these quantities; mean vector which consists of the univariate means and dispersion matrix which consists of univariate variances and bivariate covariances. If we consider the classical problem of comparing means, univariate analysis of variance (ANOVA) will be generalized to the multivariate case which is called multivarite analysis of variance (MANOVA). We should note that the dependency between the variables are important in multivariate analysis and it should be taken into account.

One of the most fundamental resources for our study is the book *Introduction to Multivariate Analysis* by Anderson (2003) which was first published in 1958. Rao's (1973) book *Linear Statistical Inference and its Applications* is another classical resource. Other interesting books have been written by Srivastava and Khatri (1979), Mardia, Kent and Bibby (1979), Muirhead (1982), Bilodeau and Brenner (1999), Rencher (2002) and Morrison (2004). Some of these references focus on the theory, whereas some of them present applications and computer practices. For more advanced studies, the book by Kollo and von Rosen (2005) *Advanced Multivariate Statistics with Matrices* is recommended where the topic is treated with the help of matrix formulations. *Bilinear Regression Analysis: An Introduction* from von Rosen (2018) gives the extension of the theory to bilinear regression models (BRM) which is also known as generalized multivariate analysis of variance (GMANOVA) or the analysis of the growth curve model. These last two references play an important role in the thesis. Technical details, properties, theorems will be frequently used from the first book (Kollo and von Rosen, 2005) and ideas for testing hypotheses from the second book (von Rosen, 2018) give us a crucial insight. Srivastava and Carter (1983) and Srivastava (2002) will be referred to later, in particular for profile analysis which falls into the scope of multivariate analysis. Fang and Zhang's (1990) *Generalized Multivariate Analysis* extends the topic from classical multivariate analysis, which focuses generally on the multivariate normal distributions, to generalized multivariate analysis which is based on elliptically contoured distributions, that involves many multivariate distributions including the multivariate normal distribution. Last but not least, Läuter's (2016) *Multivariate Statistik - drei Manuskripte* gives an introduction and summary of the topic.

1.1 Motivation

This thesis is mainly based on two topics; profile analysis and high-dimensional statistics. Profile analysis is a multivariate technique to compare two or more groups and test for similarity of means. High-dimensional statistics is concerned with the data where the dimension is larger than the sample size which means that the number of variables exceeds the number of subjects. In recent applications of statistics we have started to encounter these type of data sets more often due to the advancing data collection technologies and computing sources. The above mentioned two topics will be treated extensively in the upcoming sections. We mention them briefly here to express our interest. Our motivation lies on the challenges that occur in the high dimensional analysis. Among other problems, there is the problem of singularity. When $p > n$, where p is the number of variables and n is the number of subjects, the sample covariance matrix $\boldsymbol{S}$ becomes singular, so $\boldsymbol{S}^{-1}$ does not exist. Thus, we cannot carry out likelihood ratio tests which contain this inverse or the determinant of this matrix. There have been different approaches proposed to deal with this challenge which we are going to present in the high dimensional section. The idea for our approach includes a dimension reduction method. This method proposes to take linear combinations of variables, which means each individual will have one measure, which also is called score, instead of p separate variables. As Mardia, Kent and Bibby (1979) stated in their book in Section 1.5, taking linear combinations of the varibles is one of the most important tools in multivariate analysis. We should note that the structure of this linear combination has to follow some rules. Useful properties of spherical distributions motivate us to integrate this topic to show the robustness of the methods.

1.2 Aim and outline

Our aim for this thesis is to develop a new method for testing problems in profile analysis within a high-dimensional framework. We list the following specific steps of what is planned to take place:

1. Give a brief introduction to profile analysis, introduce the three tests; test of parallelism, test of levels and test of flatness, give relevant references, introduce mathematical notation and present the hypotheses for two groups;

2. Give a brief introduction to high-dimensional analysis, list appropriate references, discuss common challanges that appear in a high-dimensional setting and give information on common practices, introduce scores and Läuter's ideas;

3. Give a brief introduction for elliptical and spherical distributions, mention relevant references, introduce ideas that will be used in finding distributions of test statistics.

4. Give the results from the report.

The aims of the report which this thesis is based on:

5. Introduce suitable notation;

6. Present related definitons and theorems that are used in the derivations of the test statistics;

7. Present the model of our research problem, derive the likelihood ratio tests for the three hypotheses with q groups in a normal setting where the number of variables is less than the number of subjects, conduct the derivation with matrix notation which is different from, for example Srivastava's approach (1987, 2002);

8. Introduce the high-dimensional problem, scores and spherical distribution, achieve dimension reduction with the help of scores and derive the test statistics of the hypotheses in profile analysis, that are parallelism hypothesis, level hypothesis and flatness hypothesis, find explicit distributions of the test statistics using the theory of spherical distributions.

2 Profile analysis

2.1 Introduction and notation

Profile analysis is a multivariate technique that is used for comparing the patterns of variables between groups. The term arose from social sciences, but, for example, nowadays it is used in many medical applications. We have repeated measurements for each individual and mean levels for each variable are calculated per group. The profile is then obtained by plotting the means for each variable and connecting these points by drawing straight lines.

Example: A survey is conducted within a school to see if there is a difference between classes in terms of success. Say there are n_A students in class A and n_B students in class B. The marks for four subjects, Mathematics, English, Physics and History, will be compared between class A and class B.

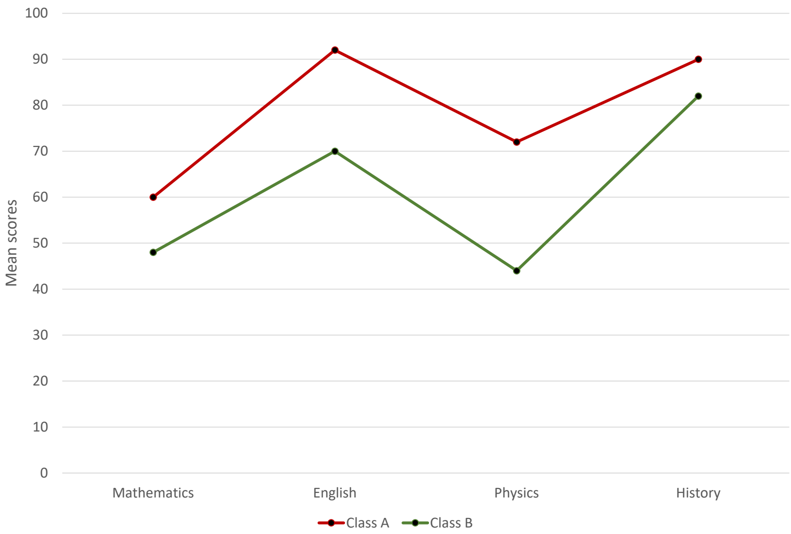

Figure 1: Profiles of class A and class B based on mean level of marks for four subjects; Mathematics, English, Physics and History.

Here we have repeated measurements for each individual and mean levels for each class are plotted in the graph in Figure 1. The main interest is to see if there is a difference between these two profiles or in other words if there is an interaction between classes and tests. Similar profiles will indicate no interaction between subjects and groups.

Note that we are not interested in the mean values of the variables but in the relative difference between the shapes across groups.

Two cases are possible in profile analysis:

1. Different variables can be measured per subject.

2. The same variable can be measured repeatedly over different time points.

The second one is also known as repeated measures case or growth curve case. Note that the response variables are not mutually independent.

There are three types of tests which are commonly used in profile analysis; test of parallelism, test of levels and test of flatness (Srivastava and Carter, 1983; Srivastava, 2002). For simplicity, assume that we have two groups. Then the tests can be visualised as it follows.

1) Test of parallelism

We start testing to see if the profiles are parallel, which is illustrated in Figure 2.

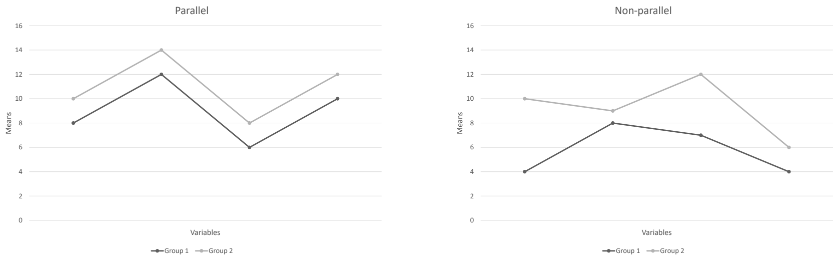

Figure 2: Parallel and non-parallel profiles for two groups.

2) Test of levels

If there exists parallelism between the profiles, then one can check if they coincide or not (see Figure 3).

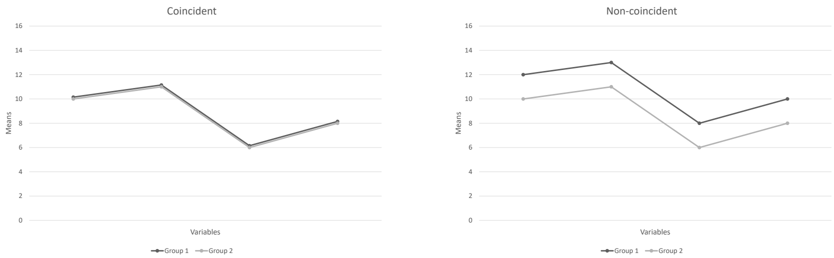

Figure 3: Coincident and non-coincident profiles for two groups.

3) **Test of flatness**

If the profiles are parallel, one can additionally check if they are flat (see Figure 4).

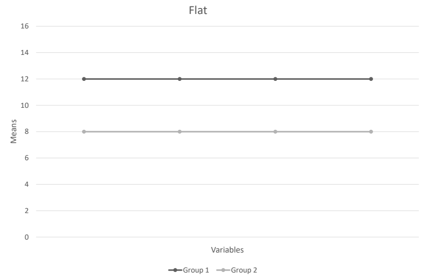

Figure 4: Flat profiles for two groups.

Our interest is to go from the parallelism hypothesis to the level hypothesis or the flatness hypothesis which is following Srivastava's approach of profile analysis. This is illustrated in Figure 5.

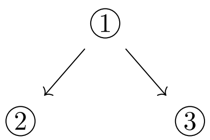

Figure 5: Testing the parallelism hypothesis which is denoted by circle 1 and then moving on to either testing the level hypothesis which is denoted by circle 2 or testing the flatness hypothesis which is denoted by circle 3.

Other types of relationships can also be of interest and one can study if the profiles are parallel, then investigate if the levels are the same, and if the levels are the same, test if they are flat. This is illustrated in Figure 6.

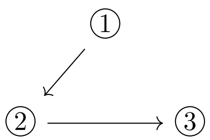

Figure 6: Testing the parallelism hypothesis which is denoted by circle 1 and then moving on to the level hypothesis which is denoted by circle 2 and from there moving on testing the flatness hypothesis which is denoted by circle 3.

The topic has been studied by several authors with different focus. Greenhouse and Geisser (1959) gave a good introduction to the theory and it is one of the early papers on the topic. Srivastava has a substantial contribution to the topic. In his paper, Srivastava (1987) derived the likelihood ratio tests together with their distributions for the three hypotheses. Prior to this paper, Srivastava and Carter (1983) published a book with a chapter on profile analysis. Another reference is the book on multivariate statistics from Srivastava (2002). One of the first papers on the growth curve model is by Potthoff and Roy (1964) and profile analysis with extensions within the framework of growth curve models can be found in Fujikoshi (2009), where he extended profile analysis, especially statistical inference on the parallelism hypothesis. Ohlson and Srivastava (2010) considered profile analysis of several groups where the groups have partly equal means which leads to a profile analysis for a growth curve model. Seo, Sakurai and Fujikoshi (2011) derived likelihood ratio tests for two hypotheses (level and flatness) in profile analysis for growth curve data. Another focus is on profile analysis with random effects covariance structure which has been studied by Srivastava and Singull (2012), Yokoyama (1995) and Yokoyama and Fujikoshi (1993). Singull and Srivastava (2012) constructed tests based on the likelihood ratio, without any restrictions on the parameter space, for testing the covariance matrix for random-effects structure or sphericity. Yokoyama (1995) derived the likelihood ratio criterion with random-effects covariance structure under the parallel profile model. Yokoyama and Fujikoshi (1993) conducted analysis of parallel growth curves of groups where they assumed a random-effects covariance structure. They also gave the asymptotic null distributions of the test. Extension to high-dimensional data are given by Harrar and Kong (2016), Onozawa, Nishiyama and Seo (2016) and Shutoh and Takahashi (2016).

Now we start giving the statistical expressions and the notations for profile analysis. Assume there are p variables of interest and q groups of size n_k, $k = 1, ..., q$. In total, there will be $N = \sum_{k=1}^{q} n_k$ subjects and $N \times p$ observations.

The measurements for the i-th individual in the k-th group can be denoted by

$$(x_{ik1}, x_{ik2}, ..., x_{ikj}, ..., x_{ikp})$$

where there are

- $i = 1, 2, ..., n_k$ individuals in group k,
- $j = 1, 2, ..., p$ variables,
- $k = 1, 2, ..., q$ groups.

The group profile for group k can be written

$$(\bar{x}_{.k1}, \bar{x}_{.k2}, ..., \bar{x}_{.kp}),$$

where $\bar{x}_{.kj}$ represents the group mean (group k) for the variable j.

We visualise all the information in Table 1 given below:

Table 1: Measurements for q groups, each of sample size n_k, where $k = 1, ..., q$, consisting of p variables for each individual and their relevant means.

		Variables					
Group	Individual	x_1	$\cdots$	x_j	$\cdots$	x_p	
1	1	x_{111}		x_{11j}		x_{11p}	
	$\vdots$						
	n_1	$x_{n_1 11}$		$x_{n_1 1j}$		$x_{n_1 1p}$	
Means:	Group 1	$\bar{x}_{.11}$		$\bar{x}_{.1j}$		$\bar{x}_{.1p}$	$\bar{x}_{.1.}$
	$\vdots$						
k	1	x_{1k1}		x_{1kj}		x_{1kp}	
	$\vdots$						
	i	x_{ik1}		x_{ikj}		x_{ikp}	
	$\vdots$						
	n_k	$x_{n_k k1}$		$x_{n_k kj}$		$x_{n_k kp}$	
Means:	Group k	$\bar{x}_{.k1}$		$\bar{x}_{.kj}$		$\bar{x}_{.kp}$	$\bar{x}_{.k.}$
	$\vdots$						
q	1	x_{1q1}		x_{1qj}		x_{1qp}	
	$\vdots$						
	n_q	$x_{n_q q1}$		$x_{n_q qj}$		$x_{n_q qp}$	
Means:	Group q	$\bar{x}_{.q1}$		$\bar{x}_{.qj}$		$\bar{x}_{.qp}$	$\bar{x}_{.q.}$
Means:	All groups	$\bar{x}_{..1}$		$\bar{x}_{..j}$		$\bar{x}_{..p}$	$\bar{x}_{...}$

Since we have multiple measurements for each individual, we will consider it as a multivariate analysis problem. We assume that each individual is randomly sampled from their corresponding groups and

$$E(X_{ikj}) = \mu_{kj}, \quad i = 1, 2, ..., n_k.$$

We have multivariate observations for each individual which means that they cannot be considered as independent. Thus, the observations per individual are correlated which is also called the within-subject correlation. Moreover, it is assumed that the observations between individuals are independent:

$$Cov(X_{i_1 k_1 j_1}, X_{i_2 k_2 j_2}) = \begin{cases} \sigma_{j_1 j_2} & \text{for } i_1 = i_2, k_1 = k_2, \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, it is assumed that we have a p-variate normal distribution with an arbitrary dispersion matrix from which the vector of observations for each individual is sampled:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1p}\sigma_1\sigma_p \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2p}\sigma_2\sigma_p \\ \vdots & \vdots & & \vdots \\ \rho_{1p}\sigma_1\sigma_p & \rho_{2p}\sigma_2\sigma_p & \cdots & \sigma_p^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{pmatrix}.$$

Different covariance structures, such as autocorrelation, intraclass correlation, equicorrelation, etc, can be assumed for the matrix $\boldsymbol{\Sigma}$. In our research, an unstructured variance-covariance matrix is supposed to hold.

2.2 The construction of the three hypotheses for two groups

Let μ_{kj} denote the mean value of variable j of the k-th group, where $k = 1, 2$. Mean vectors can be written

$$\boldsymbol{\mu}_1 = (\mu_{11}, \mu_{12}, ..., \mu_{1p})',$$
$$\boldsymbol{\mu}_2 = (\mu_{21}, \mu_{22}, ..., \mu_{2p})'.$$

Let's introduce a matrix which we are going to use frequently for the hypotheses given below, i.e., the contrast matrix $\boldsymbol{C}$ is a $(p-1) \times p$ matrix which satisfies $\boldsymbol{C}\mathbf{1} = \mathbf{0}$. One possible choice is a matrix satisfying

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}. \tag{1}$$

1. Parallelism hypothesis:

The null hypothesis to test parallel group profiles can be written

$$H_1 = \begin{pmatrix} \mu_{11} - \mu_{12} \\ \vdots \\ \mu_{1,p-1} - \mu_{1,p} \end{pmatrix} = \begin{pmatrix} \mu_{21} - \mu_{22} \\ \vdots \\ \mu_{2,p-1} - \mu_{2,p} \end{pmatrix}$$

which means that for each profile the slopes of line segments are the same, which is equivalent to that there does not exist any interaction between the groups and the responses.

Another way of writing the null hypothesis with the alternative hypothesis is:

$$H_1 : \boldsymbol{C}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = \mathbf{0} \quad \text{versus} \quad A_1 : \boldsymbol{C}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) \neq \mathbf{0},$$

where $\boldsymbol{C}$ is the contrast matrix defined in (1) and A_1 denotes the alternative hypothesis. An equivalent expression is:

$$H_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} = \gamma\mathbf{1} \quad \text{versus} \quad A_1 : \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \neq \gamma\mathbf{1},$$

where γ is an unknown parameter which is called the average difference between two profiles and $\mathbf{1}$ is a vector of ones.

2. Level hypothesis:

The hypotheses for the test of levels can be written

$$H_2|H_1 : \gamma = 0 \quad \text{versus} \quad A_2 : \gamma \neq 0$$

or equivalently

$$H_2|H_1 : \mathbf{1'_p}\boldsymbol{\mu_1} = \mathbf{1'_p}\boldsymbol{\mu_2} \quad \text{versus} \quad A_2 : \mathbf{1'_p}\boldsymbol{\mu_1} \neq \mathbf{1'_p}\boldsymbol{\mu_2},$$

where $H_2|H_1$ denotes the hypothesis H_2 given that H_1 holds.

3. Flatness hypothesis:

Lastly, the hypotheses for testing flatness are given

$$H_3|H_1 : \mu_{11} = \mu_{12} = \cdots = \mu_{1p}, \quad \mu_{21} = \mu_{22} = \cdots = \mu_{2p} \quad \text{versus} \quad A_3 \neq H_3|H_1$$

or equivalently

$$H_3|H_1 : \boldsymbol{C}(\boldsymbol{\mu_1} + \boldsymbol{\mu_2}) = \mathbf{0} \quad \text{versus} \quad A_3 \neq H_3|H_1,$$

where $H_3|H_1$ denotes the hypothesis H_3 given that H_1 holds.

2.3 The test statistics for two groups

The hypotheses have been given in the previous chapter. Now the test statistics for these hypotheses will be presented.

Let the p-dimensional random vectors $\boldsymbol{x}_1^{(i)}, ..., \boldsymbol{x}_{n_i}^{(i)}$, $i = 1, 2$, be independently normally distributed with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}$. The sample mean vectors, the sample covariance matrices and the pooled sample covariance matrix are given by

$$\begin{aligned}
\overline{\boldsymbol{x}}^{(i)} &= \frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{x}_j^{(i)}, \\
\boldsymbol{S}^{(i)} &= \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(\boldsymbol{x}_j^{(i)} - \overline{\boldsymbol{x}}^{(i)})(\boldsymbol{x}_j^{(i)} - \overline{\boldsymbol{x}}^{(i)})', \\
\boldsymbol{S}_p &= \frac{1}{n_1 + n_2 - 2}[(n_1 - 1)\boldsymbol{S}^{(1)} + (n_2 - 1)\boldsymbol{S}^{(2)}].
\end{aligned}$$

Define a $(p-1) \times p$ contrast matrix $\boldsymbol{C}$ which satisfies $\boldsymbol{C}\mathbf{1}_p = \mathbf{0}$ and is of rank $r(\boldsymbol{C}) = p - 1$, where $\mathbf{1}_p$ is a p-vector of ones. Let

$$b = \frac{n_1 n_2}{n_1 + n_2}, \quad f = n_1 + n_2 - 2, \quad \boldsymbol{u} = \overline{\boldsymbol{x}}^{(1)} - \overline{\boldsymbol{x}}^{(2)}.$$

Then the three hypotheses and related test statistics can be written as below (Srivastava and Carter, 1983; Srivastava, 1987, 2002):

(1) *Parallelism hypothesis:* $H_1 : \boldsymbol{C}\boldsymbol{\mu}_1 = \boldsymbol{C}\boldsymbol{\mu}_2$.
The null hypothesis is rejected if

$$\frac{f-(p-1)+1}{f(p-1)} b\boldsymbol{u}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{S}_p\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{u} \geq F_{p-1,f-p+2,\alpha} ,$$

where $F_{p-1,f-p+2,\alpha}$ denotes the α-percentile of the F-distribution with $p-1$ and $f-p+2$ degrees of freedom.

(2) *Level hypothesis:* $H_2|H_1 : \boldsymbol{1}_p'\boldsymbol{\mu}_1 = \boldsymbol{1}_p'\boldsymbol{\mu}_2$.
The null hypothesis is rejected if

$$\begin{aligned}(\frac{f-p+1}{f})b(\boldsymbol{1}'\boldsymbol{S}_p^{-1}\boldsymbol{u})^2(\boldsymbol{1}'\boldsymbol{S}_p^{-1}\boldsymbol{1})^{-1}(1+f^{-1}T_{p-1}^2)^{-1} &\geq t^2_{f-p+1,\alpha/2} \\ &= F_{1,f-p+1,\alpha} ,\end{aligned}$$

where $T^2_{p-1} = b\boldsymbol{u}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{S}_p\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{u}$ and $t^2_{f-p+1,\alpha/2}$ is the $\alpha/2$-percentile of the t-distribution with $f-p+1$ degrees of freedom.

(3) *Flatness hypothesis:* $H_3|H_1 : \boldsymbol{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = \boldsymbol{0}$.
The null hypothesis is rejected if

$$\frac{n(f-p+3)}{p-1}\boldsymbol{x}'\boldsymbol{C}'(\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}' + b\boldsymbol{C}\boldsymbol{u}\boldsymbol{u}'\boldsymbol{C}')^{-1}\boldsymbol{C}\boldsymbol{x} \geq F_{p-1,n-p+1,\alpha} ,$$

where $\boldsymbol{x} = (n_1\overline{\boldsymbol{x}}^{(1)} + n_2\overline{\boldsymbol{x}}^{(2)})/(n_1+n_2)$ and $\boldsymbol{V} = f\boldsymbol{S}_p$.

As it is mentioned before, the second hypothesis is tested given that H_1 is true. If we fail to reject the first hypothesis, we cannot conclude that the profiles are parallel. It means only that the data do not provide enough evidence for rejecting H_1. Moreover, notice that parallelism hypothesis together with level hypothesis, given that the former holds, gives that the means of the groups are the same, but one cannot calculate the correct level of significance because the levels of significance of these two hypotheses are not additive. If we fail to reject H_1 and H_2 each at 5% significance level, this does not mean that we will fail to reject Hotelling's T^2, which is used to test the equality of mean vectors, at 5% or even 10% significance level. The other way around gives a similar conclusion. Failing to reject that two mean vectors are equal does not necessarily mean that we will fail to reject H_1 and H_2 (Srivastava, 2002, Chapter 7). Thus, Srivastava (2002) suggests that one should perform the tests for profile analysis if similar profiles are expected and one considers the problem of finding the confidence interval for γ. Recall that γ was shown in Section 2.2 and it is also known as the difference in the levels of profiles.

2.4 Hypotheses and test statistics for q groups

The results, which have been presented so far in Chapter 2, have already been given in the literature and they act as informative and supporting bodies for the topic. We will construct the likelihood ratio tests in profile analysis with a different approach and derive the distributions of the test statistics for each hypothesis given in the previous sections of this chapter. This time the tests will be generalized to the setting, where there are q groups. Before we give the motivation of our approach, let's give the definitions of two models which are frequently used in multivariate analysis.

Definition 2.1. (**General Multivariate Linear Model**) A general multivariate linear model is given by

$$\boldsymbol{X} = \boldsymbol{BL} + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{X} : p \times n$ denotes the random matrix, $\boldsymbol{B} : p \times k$ denotes the matrix of unknown parameters, $\boldsymbol{L} : k \times n$ denotes the design matrix which is known and $\boldsymbol{E}$ is the error matrix which satisfies $\boldsymbol{E} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I})$, where $\boldsymbol{\Sigma}$ is an unknown p.d. matrix and $N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I})$ denotes matrix normal distribution.

Definition 2.2. (**Bilinear Regression Model**) A bilinear regression model is given by

$$\boldsymbol{X} = \boldsymbol{KBL} + \boldsymbol{E}, \tag{3}$$

where $\boldsymbol{X}$, $\boldsymbol{B}$, $\boldsymbol{L}$ and $\boldsymbol{E}$ are defined in the same way as in Definition 2.1 and $\boldsymbol{K}$ is a known design matrix. This model is also named as growth curve model or generalized multivariate analysis of variance (GMANOVA) model.

More details on these models can be found in von Rosen (2018). The reason why these models are introduced here is that they are going to be used for the derivation of the test statistics of profile analysis. Our aim is to reformulate the hypotheses given in Section 2.2 as problems in MANOVA and GMANOVA and derive the tests statistics based on this reformulation. The derivations will be based on the scenario where there are q-groups. Reformulation of the problems as problems in MANOVA and GMANOVA is crucial in terms of constructing a structure for later use when we have a high-dimensional setting. The challanges that rise with high-dimensional data will be mentioned in the following chapter, but with the reformulation in MANOVA and GMANOVA that will be introduced in this chapter, it will be easier to propose methods to solve these challanges.

2.4.1 The model

The model for one group, say the k-th group, can be written as

$$\boldsymbol{X}_k = \boldsymbol{M_k D_k} + \boldsymbol{E_k},$$

where $\boldsymbol{X}_k$ represents the matrix of observations, $\boldsymbol{M}_k$ the p-vector of mean parameters, $\boldsymbol{D}_k$ a vector of n_k ones and $\boldsymbol{E}_k$ is an error matrix. The columns of $\boldsymbol{X}_k$ are independently distributed, which means that the columns of $\boldsymbol{E}_k$ are independently distributed. The assumption for the distribution of $\boldsymbol{E}_k$ is that the column vectors of $\boldsymbol{E}_k$ follow a multivariate normal distribution: $\boldsymbol{e}_{jk} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$.

When we have q groups, we have q models:

$$\begin{aligned}(\boldsymbol{X}_1 : \boldsymbol{X}_2 : \cdots : \boldsymbol{X}_q) &= (\boldsymbol{M}_1\boldsymbol{D}_1 : \boldsymbol{M}_2\boldsymbol{D}_2 : \cdots : \boldsymbol{M}_q\boldsymbol{D}_q) + (\boldsymbol{E}_1 : \boldsymbol{E}_2 : \cdots : \boldsymbol{E}_q) \\ &= (\boldsymbol{M}_1 : \boldsymbol{M}_2 : \cdots : \boldsymbol{M}_q)\boldsymbol{D} + (\boldsymbol{E}_1 : \boldsymbol{E}_2 : \cdots : \boldsymbol{E}_q)\,, \end{aligned} \tag{4}$$

where $\boldsymbol{D}$ is a $q \times N$ matrix, $N = \sum_{k=1}^{q} n_k$, which equals

$$\boldsymbol{D} = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

and where $\boldsymbol{E}_k \sim N_{p,n_k}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_{n_k})$. The relation in (4) can be written as

$$\underset{(p\times N)}{\boldsymbol{X}} = \underset{(p\times q)}{\boldsymbol{M}}\underset{(q\times N)}{\boldsymbol{D}} + \underset{(p\times N)}{\boldsymbol{E}}, \qquad \boldsymbol{X} \sim N_{p,N}(\boldsymbol{MD}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)\,,$$

where $\boldsymbol{X} = (\boldsymbol{X}_1 : \boldsymbol{X}_2 : \cdots : \boldsymbol{X}_q)$, $\boldsymbol{M} = (\boldsymbol{M}_1 : \boldsymbol{M}_2 : \cdots : \boldsymbol{M}_q)$ and $\boldsymbol{E} = (\boldsymbol{E}_1 : \boldsymbol{E}_2 : \cdots : \boldsymbol{E}_q)$

Moreover, let $\boldsymbol{F}$ be a $q \times (q-1)$ and $\boldsymbol{C}$ be a $(p-1) \times p$ matrix which satisfies $\boldsymbol{1}'\boldsymbol{F} = \boldsymbol{0}$ and $\boldsymbol{C1} = \boldsymbol{0}$ respectively, i.e.,

$$\boldsymbol{F} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & -1 \end{pmatrix}, \qquad \boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

Since the common $\boldsymbol{F}$ and $\boldsymbol{C}$ are used in each hypothesis, they are introduced here.

2.4.2 Hypotheses and the tests

The model for q groups has been given above by

$$\boldsymbol{X} = \boldsymbol{MD} + \boldsymbol{E}, \tag{5}$$

where $\boldsymbol{X} = (\boldsymbol{X}_1 : \boldsymbol{X}_2 : \cdots : \boldsymbol{X}_q)$. This model is often called the MANOVA model. If we want to deduct any inference from the model, the unknown parameters $\boldsymbol{M}$ and $\boldsymbol{\Sigma}$ need to be estimated.

For the model in (5), the likelihood function equals

$$L(\boldsymbol{M}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{1}{2}pN}|\boldsymbol{\Sigma}|^{-N/2}\exp\Big\{ -\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{MD})(\boldsymbol{X} - \boldsymbol{MD})' \Big\},$$

where $N = \sum_{k=1}^{q} n_k$. One can deduce the following from this likelihood:

$$\boldsymbol{X}\boldsymbol{P}_{D'} = \widehat{\boldsymbol{M}}\boldsymbol{D},$$
$$N\widehat{\boldsymbol{\Sigma}} = \boldsymbol{R}\boldsymbol{R}',$$

where " $\widehat{\cdot}$ " denotes the estimator and $\boldsymbol{R} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{D'})$ with the projection matrix $\boldsymbol{P}_{D'} = \boldsymbol{D}'(\boldsymbol{D}\boldsymbol{D}')^{-1}\boldsymbol{D}$.

This is the general model without any restriction on the mean parameter space. Our aim is to put restrictions on the mean parameter space based on the test and each test (parallelism test, level test or flatness test) requires different formulation of the hypothesis. These restrictions will be given in matrix form.

The structure will be as follows: First, the hypothesis for the parallelism test will be introduced and the test statistic with it's distribution will be given. Then we move on to the level test assuming that the parallelism hypothesis holds. The same structure will be followed, such as, introducing the null and alternative hypotheses initially and then giving the distribution of the test statistic. Alternatively, we move on to the flatness test assuming that the parallelism hypothesis holds.

1) Parallelism hypothesis

For the parallelism hypothesis, the restrictions on the mean parameter space will be introduced with two matrices; $\boldsymbol{C}$ and $\boldsymbol{F}$ which have been defined in Section 2.4.1. Then the null hypothesis and the alternative hypothesis for parallelism can be written

$$\begin{aligned} H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{M}\boldsymbol{D}, \quad \boldsymbol{C}\boldsymbol{M}\boldsymbol{F} = \boldsymbol{0}, \\ A_1: &\quad E(\boldsymbol{X}) = \boldsymbol{M}\boldsymbol{D}, \quad \boldsymbol{C}\boldsymbol{M}\boldsymbol{F} \neq \boldsymbol{0}. \end{aligned} \tag{6}$$

Now the aim is to find the maximum of the likelihoods, more precisely proportional expressions, for both hypotheses. These results are deducted from the estimation of $\boldsymbol{\Sigma}$. Denote the maximum likelihood estimator of $\boldsymbol{\Sigma}$ under the null hypothesis by $\widehat{\boldsymbol{\Sigma}}_{H_1}$ and under the alternative hypothesis by $\widehat{\boldsymbol{\Sigma}}_{A_1}$. Then we can give the following theorem:

Theorem 2.1. *The likelihood ratio statistic for the parallelism hypothesis presented in* (6) *can be given as*

$$\lambda^{2/N} = \frac{|N\widehat{\boldsymbol{\Sigma}}_{A_1}|}{|N\widehat{\boldsymbol{\Sigma}}_{H_1}|} = \frac{|\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}'|}{|\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{X}\boldsymbol{P}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'|}, \tag{7}$$

where $\boldsymbol{S} = \boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'$ *and* $\boldsymbol{K}$ *is any matrix satisfying* $\mathcal{C}(\boldsymbol{K}) = \mathcal{C}(\boldsymbol{D}) \cap \mathcal{C}(\boldsymbol{F})$,

$$\boldsymbol{C}\boldsymbol{S}\boldsymbol{C}' \sim W_{p-1}(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}', N - r(\boldsymbol{D})),$$
$$\boldsymbol{C}\boldsymbol{X}\boldsymbol{P}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' \sim W_{p-1}(\boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}', r(\boldsymbol{K})).$$

Then

$$\lambda^{2/N} = \frac{|\boldsymbol{CSC}'|}{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'|} \sim \Lambda(p-1, N - r(\boldsymbol{D}), r(\boldsymbol{K})),$$

where $\Lambda(\cdot,\cdot,\cdot)$ *denotes the Wilks' lambda distribution.*

2) Level Hypothesis

Assuming that the profiles are parallel, that is $\boldsymbol{CMF} = \boldsymbol{0}$, we will construct a test to check if they have equals level, in other words if they coincide. In this case, the restrictions on the mean parameter space will be introduced only with the matrix $\boldsymbol{F}$. Then the null hypothesis and the alternative hypothesis for the level test can be written

$$\begin{aligned} H_2|H_1 : &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{MF} = \boldsymbol{0}, \\ A_2|H_1 : &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} = \boldsymbol{0}. \end{aligned}$$

The same routine as with parallelism hypothesis will be followed; find the maximum of the likelihoods for both hypotheses and investigate the ratio of these two. Then the following theorem can be given:

Theorem 2.2. *The likelihood ratio statistic for the level hypothesis can be expressed as*

$$\begin{aligned} \lambda^{2/N} &= \frac{|N\widehat{\boldsymbol{\Sigma}}_{A_2}|}{|N\widehat{\boldsymbol{\Sigma}}_{H_2}|} \\ &= \frac{|(\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ}|^{-1}}{\left|\begin{array}{l}((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1} + ((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}\boldsymbol{XD}' \\ \quad \times (\boldsymbol{DD}')^{-1}\boldsymbol{KQ}^{-1}\boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{DX}'\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ}((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1}\end{array}\right|}, \end{aligned} \tag{8}$$

where $\boldsymbol{Q} = \boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{K} + \boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{DX}'\boldsymbol{C}'(\boldsymbol{CSC}')^{-1}\boldsymbol{CXD}'(\boldsymbol{DD}')^{-1}\boldsymbol{K}$ *and* $\mathcal{C}(\boldsymbol{K}) = \mathcal{C}(\boldsymbol{D}) \cap \mathcal{C}(\boldsymbol{F})$,

$$\begin{aligned} &((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1} \sim W_1(((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, N - r(\boldsymbol{D}) - p + 1), \\ &((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}\boldsymbol{XD}'(\boldsymbol{DD}')^{-1}\boldsymbol{KQ}^{-1}\boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{DX}'\boldsymbol{S}^{-1} \\ &\qquad \times (\boldsymbol{C}')^{\circ}((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1} \sim W_1(((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, r(\boldsymbol{K})). \end{aligned}$$

Then

$$\lambda^{2/N} \sim \Lambda(1, N - r(\boldsymbol{D}) - p + 1, r(\boldsymbol{K})).$$

3) Flatness Hypothesis

Assuming that the profiles are parallel, we will test if they are flat or not. The restrictions on the mean parameter space will be introduced only with the matrix $\boldsymbol{C}$:

$$\begin{aligned} H_3|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CM} = \boldsymbol{0}, \\ A_3|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} = \boldsymbol{0}. \end{aligned}$$

Theorem 2.3. *The likelihood ratio statistic for the flatness hypothesis is given by*

$$\lambda^{2/N} = \frac{|N\widehat{\boldsymbol{\Sigma}}_{A_3}|}{|N\widehat{\boldsymbol{\Sigma}}_{H_3}|} = \frac{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'|}{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' + \boldsymbol{CXP}_{D'F^\circ}\boldsymbol{X}'\boldsymbol{C}'|}, \tag{9}$$

where

$$\begin{aligned} \boldsymbol{CXP}_{D'F^\circ}\boldsymbol{X}'\boldsymbol{C}' &\sim W_{p-1}(\boldsymbol{C\Sigma C}', r(\boldsymbol{D}'\boldsymbol{F}^\circ)), \\ \boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' &\sim W_{p-1}(\boldsymbol{C\Sigma C}', N - r(\boldsymbol{D}) + r(\boldsymbol{K})). \end{aligned}$$

Then

$$\lambda^{2/N} \sim \Lambda(p-1, N - r(\boldsymbol{D}) + r(\boldsymbol{K}), r(\boldsymbol{D}'\boldsymbol{F}^\circ)).$$

The details of the derivations can be found in the report.

3 Elliptical and spherical distributions

3.1 Introduction

The multivariate normal distribution has been in the centre of the statistical analysis of multivariate observations. In general, the normal distribution:

1. describes the process of data generation (as an underlying distribution);

2. approximates the sampling distribution of a statistic (as a limiting distribution).

In reality, the assumption that the data follows the normal distribution is often not satisfied. Even if there exist some optimal tests which may work efficiently under non-normality, one needs to consider the situations where this is violated and investigate the robustness of the procedures. This has led to a search for extending the theory to cover not just the normal distribution, but a wider class of distributions. The focus has been reflected on the elliptically contoured distributions (or elliptical distributions). Note that the multivariate normal distribution is in this class and moreover, several properties for normal distribution can be transmitted to the elliptical distributions. Thus, the class of elliptical distributions can be considered as an extension of the class of multivariate normal distributions. For example, in addition to the multivariate normal distribution, there exist the multivariate t, the multivariate Cauchy, the multivariate Laplace, the multivariate uniform and mixtures of normal distributions which all belong to the class of elliptically contoured distributions (Anderson and Fang, 1990).

The first detailed paper on spherical distributions was written by Kelker (1970), in which he studied the distribution theory of spherical distributions. Dawid (1977, 1978) worked on spherical matrix distributions. An early reference on the theory of elliptically contoured distributions is Cambanis, Huang and Simons (1981). Anderson and Fang (1982, 1990) presented technical reports on the topic. Two book references are Kariya and Sinha (1989) and Fang, Kotz and Ng (1990), where the former presents the development of a general theory of the robustness of tests and its application to a wide variety of multiparameter hypotheses testing problems and the latter presents symmetric multivariate and related distributions. We will focus mainly on the book by Fang and Zhang (1990) which we consider as an important reference. Some classical results on elliptical distributions can also be found in Kollo and von Rosen (2005).

In the following sections, the definitions of both elliptical and spherical distribution will be given and the difference between the two will be pointed out.

3.2 Spherical distributions

Definition 3.1. An n-dimensional random vector $\boldsymbol{x}$ is distributed according to an *elliptically contoured distribution* with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ϕ if the characteristic function of $\boldsymbol{x}$ has the form $\exp(i\boldsymbol{t}'\boldsymbol{\mu})\phi(\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t})$, where $\boldsymbol{\mu} : n \times 1$, $\boldsymbol{\Sigma} : n \times n$ and $\boldsymbol{\Sigma} \geq \boldsymbol{0}$ and this is denoted by $\boldsymbol{x} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

Definition 3.2. If in Definition 3.1 $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_n$, $EC_n(\boldsymbol{0}, \boldsymbol{I}_n, \phi)$ is called a *spherical distribution* and denoted by $\mathcal{S}_n(\phi)$. Then the characteristic function of $\boldsymbol{x}$ has the form $\phi(\boldsymbol{t}'\boldsymbol{t})$.

$\phi(.)$ is a function of a scalar variable and it is called the *characteristic generator* of the spherical distribution. This means, if $\boldsymbol{x}$ has a spherical distribution, there exists a function $\phi(.)$ of a scalar variable such that $\psi(\boldsymbol{t}) = \phi(\boldsymbol{t}'\boldsymbol{t})$, where $\psi(\boldsymbol{t})$ is the characteristic function of the random vector $\boldsymbol{x}$. The reverse holds as well.

One important spherical distribution is the uniform distribution. Let $\boldsymbol{u}^{(n)}$ be a random vector distributed uniformly on the unit sphere in $\mathbb{R}^n$. The characteristic function of $\boldsymbol{u}^{(n)}$ is of the form $\phi(\boldsymbol{t}'\boldsymbol{t})$. The proof can be found in Fang and Zhang (1990), which will be omitted here. The uniform distribution will appear in the following theorems that will be later used in the report and in the proofs of the results which will be presented in the thesis. Another example is the multivariate normal distribution. Let $\boldsymbol{x}' = (x_1, ..., x_n)$ be distributed as $N_n(\boldsymbol{0}, \boldsymbol{I}_n)$. The characteristic function of $\boldsymbol{x}$ is

$$\exp\left\{-\frac{1}{2}(t_1^2 + \cdots t_n^2)\right\} = \exp\left\{-\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right\}. \tag{10}$$

We begin with a theorem which denotes a relation that we will come across in Corollary 3.1.

Theorem 3.1. *A function $\phi(.) \in \Phi_n$, where Φ_n is the set of all possible ϕ's, that is $\Phi_n = \{\phi(.)|\phi(t_1^2 + \cdots + t_n^2)$ is a characteristic function$\}$, if and only if*

$$\phi(\boldsymbol{x}) = \int_0^\infty \Omega_n(\boldsymbol{x}r^2)dF(r), \tag{11}$$

where $F(.)$ is a cumulative distribution function (c.d.f.) over $[0, \infty)$ and

$$\Omega_n(\boldsymbol{y}'\boldsymbol{y}) = \int_{S:\boldsymbol{x}'\boldsymbol{x}=1} e^{i\boldsymbol{y}'\boldsymbol{x}} dS/S_n,$$

where S_n is the area of unit sphere surface in $\mathbb{R}^n$, i.e., $\Omega_n(\boldsymbol{t}'\boldsymbol{t})$ is the characteristic function of $\boldsymbol{u}^{(n)}$.

Now let's give two crucial corollaries.

Corollary 3.1. *Assume that the characteristic function of a* $n \times 1$ *random vector* $\boldsymbol{x}$ *is* $\phi(\boldsymbol{t}'\boldsymbol{t})$ *and* $\phi \in \Phi_n$*. Then* $\boldsymbol{x}$ *has a stochastic representation*

$$\boldsymbol{x} \stackrel{d}{=} R\boldsymbol{u}^{(n)}, \tag{12}$$

where $R \sim F(x)$ *is related to* ϕ *as in* (11) *and is independent of* $\mathbf{u}^{(n)}$*. This random variable* R *can be thought of as a radius (Kollo and von Rosen, 2005).*

Corollary 3.2. *An* $n \times 1$ *random vector* $\boldsymbol{x} \sim \mathcal{S}_n(\phi)$ *if and only if for every* $\boldsymbol{\Gamma} \in \mathcal{O}(n)$*,*

$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{\Gamma}\boldsymbol{x},$$

where $\mathcal{O}(n)$ *is the set of* $n \times n$ *orthogonal matrices.*

Corollary 3.2 can sometimes be used as a definition for spherical distributions (Fang, Kotz and Ng, 1990). As one can see, there are several definitions for spherical distibution or theorems which indicates sphericity. If all information given so far is gathered in one theorem, the following can be written.

Theorem 3.2. *Let* $\boldsymbol{x}$ *be an* $n \times 1$ *random vector. Then the following statements are equivalent:*

(i) The characteristic function of $\boldsymbol{x}$ *has the form* $\phi(\boldsymbol{t}'\boldsymbol{t})$*, where* $\phi \in \Phi_n$*;*

(ii) $\boldsymbol{x}$ *has a stochastic representation* $\boldsymbol{x} \stackrel{d}{=} R\boldsymbol{u}^{(n)}$*, where* $R \geq 0$ *is independent of* $\boldsymbol{u}^{(n)}$*;*

(iii) $\boldsymbol{x} \stackrel{d}{=} \boldsymbol{\Gamma}\boldsymbol{x}$ *for every* $\boldsymbol{\Gamma} \in \mathcal{O}(n)$*.*

Corollary 3.3. *Suppose* $\boldsymbol{x} \stackrel{d}{=} R\boldsymbol{u}^{(n)} \sim \mathcal{S}_n(\phi)$ *and* $P(\boldsymbol{x} = \boldsymbol{0}) = 0$*. Then*

$$\|\boldsymbol{x}\| \stackrel{d}{=} R, \qquad \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \stackrel{d}{=} \boldsymbol{u}^{(n)},$$

which are independent. $\|\boldsymbol{x}\|$ *denotes the Euclidean norm, that is* $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{n} x_i^2}$*.*

If $\boldsymbol{x} \sim \mathcal{S}_n(\phi)$ and $P(\boldsymbol{x} = \boldsymbol{0}) = 0$, this is denoted by $\boldsymbol{x} \sim \mathcal{S}_n^+(\phi)$. A very important fact is that the distribution of $\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ does not depend on any special element of the class of $\mathcal{S}_n^+(\phi)$'s. Thus, one can assume $\boldsymbol{x} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$ (Fang and Zhang, 1990). This connection between the standard normal distribution and the spherical distrbution will be the core of some important results which are going to be introduced below.

Corollary 3.4. *Let* $\boldsymbol{x} \sim \mathcal{S}_n(\phi)$ *with* $\phi \in \Phi_\infty$*, where* $\Phi_\infty = \bigcap_{n=1}^{\infty} \Phi_n$*, if and only if*

$$\boldsymbol{x} \stackrel{d}{=} R\boldsymbol{z},$$

where $\boldsymbol{z} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$ *is independent of* $R \geq 0$*.*

The results that have been given so far and that will be introduced after Definition 3.3 are based on a random vector $\boldsymbol{x}$. The only property which is based on a random matrix is the following definition.

Definition 3.3. Let $\boldsymbol{X}$ be a $p \times n$ random matrix. If $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X\Gamma}$ for every $\boldsymbol{\Gamma} \in \mathcal{O}(n)$, $\boldsymbol{X}$ is called *right-spherical.*

All theorems and corollaries given so far form a basis for the understanding why spherical distributions are of interest and how they can be used in our research. As mentioned before, the multivariate normal distibution lies at the centre of multivariate statistics and an appropriate extension of this distribution will provide broader usage of properties which will be valid also in a wider class of distributions. Let's demonstrate the idea of invariant distribution with two examples.

Let $\boldsymbol{x} = (x_1, ..., x_n)'$ be a random vector and

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}_n'\boldsymbol{x}, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\boldsymbol{x}'\boldsymbol{J}\boldsymbol{x},$$

where $\mathbf{1}_n = (1, ..., 1)'$ and $\boldsymbol{J} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$.

Example 3.1. Define

$$t = \sqrt{n}\frac{\bar{x}}{s}.$$

It is known that if $\boldsymbol{x} \sim N_n(\mathbf{0}, \boldsymbol{I}_n)$, $t \sim t_{n-1}$, where t_{n-1} denotes the t-distribution with $n-1$ degrees of freedom. To show that this holds whenever $\boldsymbol{x} \sim \mathcal{S}_n^+(\phi)$, first define

$$f(\boldsymbol{x}) = \sqrt{n}\frac{\frac{1}{n}\mathbf{1}_n'\boldsymbol{x}}{\left(\frac{1}{n-1}\boldsymbol{x}'\boldsymbol{J}\boldsymbol{x}\right)^{\frac{1}{2}}}.$$

Then

$$t = f(\boldsymbol{x}) \stackrel{d}{=} f(R\boldsymbol{u}^{(n)}) = \sqrt{n}\frac{\frac{R}{n}\mathbf{1}_n'\boldsymbol{u}^{(n)}}{\left(\frac{R^2}{n-1}\boldsymbol{u}^{(n)\prime}\boldsymbol{J}\boldsymbol{u}^{(n)}\right)^{\frac{1}{2}}} = \sqrt{n}\frac{\frac{1}{n}\mathbf{1}_n'\boldsymbol{u}^{(n)}}{\left(\frac{1}{n-1}\boldsymbol{u}^{(n)\prime}\boldsymbol{J}\boldsymbol{u}^{(n)}\right)^{\frac{1}{2}}}.$$

Notice that t's distribution is independent of R. Recall that the normal distribution $N_n(\mathbf{0}, \boldsymbol{I}_n)$ is $\mathcal{S}_n^+(\phi)$ with the characteristic generator ϕ given in (10). Thus, t follows the same distribution t_{n-1} for the whole class $\{\mathcal{S}_n^+(\phi)\}$.

Example 3.2. Define

$$F = \frac{\boldsymbol{x}'\boldsymbol{P}_1\boldsymbol{x}/r_1}{\boldsymbol{x}'\boldsymbol{P}_2\boldsymbol{x}/r_2}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are two orthogonal projection matrices of ranks r_1 and r_2 respectively, such that $\boldsymbol{P}_1\boldsymbol{P}_2 = \mathbf{0}$. It is known that if $\boldsymbol{x} \sim N_n(\mathbf{0}, \boldsymbol{I}_n)$, $F \sim F(r_1, r_2)$, where $F(r_1, r_2)$ denotes the F-distribution with r_1 and r_2

degrees of freedom. However, F follows the same distribution $F(r_1, r_2)$ in the class $\{\mathcal{S}_n^+(\phi)\}$. The proof can be given similarly to the way it was given for the t-distribution in Example 3.1.

The idea of invariant distribution can be generalized with the following theorem.

Theorem 3.3. *A statistic $t(\boldsymbol{x})$'s distribution remains the same whenever $\boldsymbol{x} \sim \mathcal{S}_n^+(\phi)$ if*

$$t(\alpha \boldsymbol{x}) \stackrel{d}{=} t(\boldsymbol{x}) \tag{13}$$

for each $\alpha > 0$ and each $\boldsymbol{x} \sim \mathcal{S}_n^+(.)$.

Theorem 3.3 provides a very useful connection between the normal distribution and other members of the class of spherical distributions. One will be able to determine the distribution of a statistic even if the random quantity does not have a normal distibution as long as the relation given by (13) is satisfied. But first, one needs to show that the random quantity is spherically distributed. This theorem will be used in the derivations of the test statistics in a high-dimensional setting, particularly when we try to attain the distributions of the ratios that are derived for the three hypotheses.

The definitions and theorems presented so far in this chapter have been taken mainly from Fang and Zhang (1990) and also from Fang, Kotz and Ng (1990). More have been presented in these references and also in the other references mentioned in the introduction, such as properties of elliptical distributions, spherical matrix distributions, estimation of parameters etc. If one is interested further in the topic, see the reference list in Section 3.1.

4 High-dimensional analysis

The methods and theories that have been introduced and developed so far consider the setting where the number of independent observations is greater than the number of dependent variables. In classical multivariate methods, the number of dependent variables, which is denoted by p, needs to be less than the number of independent observations, which is denoted by n, for estimability reasons. Several multivariate techniques are based on large sample approximations, where p is fixed and $n \to \infty$. For example, central limit theorems and the law of large numbers are often based on these assumptions.

However, due to the substantial advances in computer technologies, data storage capacity and computing speed have improved significantly. Therefore, recently one encounters high-dimensional data more frequently in several applications of statistics. Examples include genetic data, finance, brain imaging, climate data, signal processing, etc. In this setting, p is larger than n, which is also called "large p, small n" paradigm.

In the high-dimensional setting, where the dimension (p) is larger than the sample size (n), the conventional testing methods, for instance likelihod ratio tests, are not applicable. One important problem is that $\boldsymbol{S}$, which is the sample covariance matrix, becomes singular, so $\boldsymbol{S}^{-1}$, which is used as an unbiased estimator of $\boldsymbol{\Sigma}^{-1}$, does not exist. The following three principles can be used to overcome the problem:

- Shrinking: Use $\boldsymbol{S}^{+}$ instead of $\boldsymbol{\Sigma}^{-1}$.
- Tikhonov regularization: $(\boldsymbol{S} + \lambda \boldsymbol{I})^{-1}$ instead of $\boldsymbol{\Sigma}^{-1}$.
- Krylov space method: Based on the Cayley-Hamilton

  $$\boldsymbol{\Sigma}^{-1} = \sum_{i=1}^{p} c_i \boldsymbol{\Sigma}^{i-1},$$

  where $\boldsymbol{\Sigma}$ is of the size $p \times p$ and since $\boldsymbol{\Sigma}$ is unknown, the constants c_i are also unknown. Then an approximation of $\boldsymbol{\Sigma}^{-1}$ is given by

  $$\boldsymbol{\Sigma}^{-1} \approx \sum_{i=1}^{a} c_i \boldsymbol{\Sigma}^{i-1}, \qquad a \leq p,$$

  and an estimator is found via $\widehat{\boldsymbol{\Sigma}}^{-1} \approx \sum_{i=1}^{a} \hat{c}_i \boldsymbol{S}^{i-1}$. When determining c_i, a Krylov space method, partial least squares (PLS), is used.

These challenges have led to a search for re-examining the classical methods and extending them to high-dimensional settings. Ledoit and Wolf (2002) derived the hypothesis tests for the covariance matrix in a high dimensional setting. Srivastava (2005) also developed tests for certain hypotheses on the covariance matrix in high dimension. Srivastava and Fujikoshi (2006), Srivastava (2007), Srivastava and Du (2008) are other examples in the multivariate area. Kollo, von Rosen and von Rosen (2011) focused on estimating the parameters describing the mean structure in the Growth Curve model. Testing for the mean matrix in a Growth Curve model for high dimensions was studied by Srivastava and Singull (2017) as well. Fujikoshi, Ulyanov and Shimizu (2010) focused on high dimensional and large-sample approximations for multivariate statistics.

The attention in this thesis is on high dimensional profile analysis. Onozawa, Nishiyama and Seo (2016) derived test statistics for profile analysis with unequal covariance matrices in high dimension. Similarly, Harrar and Kong (2016) worked on this topic. Shutoh and Takahashi (2016) proposed new test statistics in profile analysis with high-dimensional data by using the Cauchy-Schwarz inequality. All these references study the asymptotic distributions of the test statistics. They introduce different high-dimensional asymptotic frameworks and derive the test statistics in profile analysis under these frameworks. In this context, such asymptotics can be of interest as $n \to \infty$:

(i) $\frac{p}{n} \to c$, where $c \in (a, b)$,

(ii) $\frac{p}{n} \to \infty$.

Our approach will be different than the approaches mentioned above. We will not focus on the asymptotic distributions of the test statistics. In this thesis, fixed p and n are of interest and the following scenario is considered:

$$p > n \qquad \text{or} \qquad p \gg n. \tag{14}$$

4.1 Dimension reduction using scores and spherical distributions

As stated before, many classical tests are not feasible in high-dimensional situations. Läuter (1996, 2016) and Läuter, Glimm and Kropf (1996, 1998) proposed a new method for dealing with the problem that arises in high dimensional settings. The tests they proposed are based on linear scores which have been obtained by using score coefficients that are determined from data via sums of products matrices. These scores are basically linear combinations of the repeated measures and the coefficients of the linear combinations are also called vectors of weights. With this approach high-dimensional observations are compressed into low-dimensional scores. Then they use these scores for the analysis instead of the original data. This approach can be useful in many situations because we often do not have the knowledge on the effect of each single variable or one may want to investigate the joint effect of several variables.

Let's give the mathematical representation of the theory. Suppose

$$\boldsymbol{x} = (x_i) \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

and n individual p-dimensional vectors form the $p \times n$ matrix $\boldsymbol{X}$ which satisfies

$$\boldsymbol{X} = (x_{ij}) \sim N_{p,n}(\boldsymbol{\mu}\mathbf{1}'_n, \boldsymbol{\Sigma}, \boldsymbol{I}_n).$$

Consider a single score

$$\boldsymbol{z}' = (z_1, z_2, \cdots, z_n) = (d_1, d_2, \cdots, d_p)\boldsymbol{X} = \boldsymbol{d}'\boldsymbol{X},$$

where $\boldsymbol{d}$ is the vector of weights and z_j's, $j = 1, ..., n$ are the individual scores. The rule for choosing the vector $\boldsymbol{d}$ of the coefficients is that it has to be a unique function of $\boldsymbol{X}\boldsymbol{X}'$ which is the $p \times p$ matrix of the sums of the products. Moreover, the condition $\boldsymbol{d}'\boldsymbol{X} \neq \mathbf{0}$ with probability 1 needs to be satisfied. The total sums of product matrix $\boldsymbol{X}\boldsymbol{X}'$ corresponds to the hypothesis $\boldsymbol{\mu} = \mathbf{0}$. Consequently, the structure of the function can change based on the hypothesis. We will try to illustrate the idea with two primary theorems presented in Läuter, Glimm and Kropf (1996).

Theorem 4.1. ***(Läuter, Glimm and Kropf, 1996)*** *Assume that $\boldsymbol{X}$ is a $p \times n$ matrix consisting of n p-dimensional observations ($p \geq 1$, $n \geq 2$) that follows the normal distribution $\boldsymbol{X} \sim N_{p,n}(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$. Define a p-dimensional vector of weights $\boldsymbol{d}$ which is a function of $\boldsymbol{X}\boldsymbol{X}'$ and assume $\boldsymbol{d}'\boldsymbol{X} \neq \mathbf{0}$ with probability* 1*. Then*

$$t = \frac{\sqrt{n}\bar{z}}{s_z} \tag{15}$$

has the exact t distribution with $n-1$ degrees of freedom, where

$$\boldsymbol{z}' = (z_j)' = \boldsymbol{d}'\boldsymbol{X}, \quad \bar{z} = \frac{1}{n}\boldsymbol{z}'\mathbf{1}_n, \quad s_z^2 = \frac{1}{n-1}(\boldsymbol{z}'\boldsymbol{z} - n\bar{z}^2).$$

Proof. Define the orthogonal transformation of the random matrix $\boldsymbol{X}$:

$$\boldsymbol{X}_\Gamma = \boldsymbol{X}\boldsymbol{\Gamma},$$

where $\boldsymbol{\Gamma}$ is an $n \times n$ orthogonal matrix. Take a linear combination of both $\boldsymbol{X}$ and $\boldsymbol{X}_\Gamma$ in the following way:

$$\boldsymbol{z}' = \boldsymbol{d}'\boldsymbol{X}, \qquad \boldsymbol{z}'_\Gamma = \boldsymbol{d}'_\Gamma\boldsymbol{X}_\Gamma.$$

Since the coefficients $\boldsymbol{d}$ and $\boldsymbol{d}'_\Gamma$ are derived from the same matrix, they are equal, $\boldsymbol{d} = \boldsymbol{d}_\Gamma$:

$$\boldsymbol{X}_\Gamma\boldsymbol{X}'_\Gamma = (\boldsymbol{X}\boldsymbol{\Gamma})(\boldsymbol{X}\boldsymbol{\Gamma})' = \boldsymbol{X}\boldsymbol{\Gamma}\boldsymbol{\Gamma}'\boldsymbol{X}' = \boldsymbol{X}\boldsymbol{X}'.$$

Thus,

$$\boldsymbol{z}'_\Gamma = \boldsymbol{d}'_\Gamma\boldsymbol{X}_\Gamma = \boldsymbol{d}'\boldsymbol{X}\boldsymbol{\Gamma} = \boldsymbol{z}'\boldsymbol{\Gamma}. \tag{16}$$

Then we need to show that $\boldsymbol{X}$ and $\boldsymbol{X}_\Gamma$ have the same normal distribution:

$$\begin{aligned} E[\boldsymbol{X}_\Gamma] &= E[\boldsymbol{X}\boldsymbol{\Gamma}] = E[\boldsymbol{X}]\boldsymbol{\Gamma} = \mathbf{0}, \\ D[\boldsymbol{X}_\Gamma] &= D[\boldsymbol{X}\boldsymbol{\Gamma}] = \boldsymbol{\Gamma}'\boldsymbol{\Gamma} \otimes \boldsymbol{\Sigma} = \boldsymbol{I} \otimes \boldsymbol{\Sigma}. \end{aligned}$$

It is proved that $\boldsymbol{X}$ and $\boldsymbol{X}_\Gamma$ have the same distribution, consequently $\boldsymbol{X}$ is right-spherically distributed. This means that the transformed random vectors $\boldsymbol{z}'$ and $\boldsymbol{z}'_\Gamma$ have the same distribution which is spherical due to (16). From Theorem 3.3, the statistic t is distibuted according to t-distribution with $n-1$ degrees of freedom.

□

Theorem 4.2. ***(Läuter, Glimm and Kropf, 1996)*** *Assume that* $\boldsymbol{H} \sim W_p(\boldsymbol{\Sigma}, m)$ *and* $\boldsymbol{G} \sim W_p(\boldsymbol{\Sigma}, f)$ *and they are independently distributed. Define a p-dimensional vector of weights* $\boldsymbol{d}$ *which is a function of* $\boldsymbol{H}+\boldsymbol{G}$ *and assume* $\boldsymbol{d}'(\boldsymbol{H}+\boldsymbol{G})\boldsymbol{d} \neq \boldsymbol{0}$ *with probability* 1*. Then*

$$F = \frac{f}{m}\frac{\boldsymbol{d}'\boldsymbol{H}\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{G}\boldsymbol{d}}$$

follows an F-distibution with m and f degrees of freedom.

Proof. From the definition of the Wishart distribution, $\boldsymbol{H}$ and $\boldsymbol{G}$ can be written

$$\boldsymbol{H} = \sum_{j=1}^{m} \boldsymbol{h}_j\boldsymbol{h}'_j, \qquad \boldsymbol{G} = \sum_{j=1}^{f} \boldsymbol{g}_j\boldsymbol{g}'_j,$$

where $\boldsymbol{h}_1, \boldsymbol{h}_2, ..., \boldsymbol{h}_m$ and $\boldsymbol{g}_1, \boldsymbol{g}_2, ..., \boldsymbol{g}_f$ are independent vectors, each with the normal distribution $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Define the matrix

$$\boldsymbol{X} = (\boldsymbol{h}_1 \;\; \boldsymbol{h}_2 \;\; ... \;\; \boldsymbol{h}_m \;\; \boldsymbol{g}_1 \;\; \boldsymbol{g}_2 \;\; ... \;\; \boldsymbol{g}_f)$$

of size $p \times n$ and notice that $n = m + f$. Then $\boldsymbol{X} \sim N_{p,n}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{I}_n)$ and $\boldsymbol{H}+\boldsymbol{G} = \boldsymbol{X}\boldsymbol{X}'$. One can show that $\boldsymbol{z}' = \boldsymbol{d}'\boldsymbol{X}$ has a spherical distribution in the same way that is shown in Theorem 4.1. Then again from Theorem 3.3, the F-statistic has an F-distribution with m and n degrees of freedom.

□

These theorems have been given to illustrate the idea of the approach we are going to use in the report.

5 Results from the report

5.1 Motivation

The three hypotheses of profile analysis have been given and test statistics with their distributions have been presented in Chapter 2, Section 2.4 in this thesis. The details for the calculations can be found in Section 3 of the report. These calculations have been carried out by the authors. The results which were given in the literature have also been presented (see Section 2.3). Recall that our approach was an extension of the current methods to problems in MANOVA and GMANOVA model. The motivation for this reformulation comes from intention to see the tests in compact forms and to be able to detect the effect of dimensionality on the test statistics. For instance. let's have a look at the parallelism hypothesis. The aim is to investigate if the differences of the mean values across groups are the same. Instead of writing individual differences, such as $\mu_{11} - \mu_{12} = \mu_{12} - \mu_{13} = \cdots = \mu_{1,p-1} - \mu_{1,p} = \ldots = \mu_{q1} - \mu_{q2} = \mu_{q2} - \mu_{q3} = \cdots = \mu_{q,p-1} - \mu_{q,p}$, the test can be summarized with $\boldsymbol{CMF} = \boldsymbol{0}$ (see (6)). The likelihood ratio has been derived based on this matrix formulation. However, when we have high-dimensional data, $\boldsymbol{S}$, which is equal to $\boldsymbol{X}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'$, becomes singular, consequently the determinants given in (7) will become zero and in this case, the likelihood ratio is not defined. The same complication arises in the other two hypotheses. Thus, we propose a dimension reduction method to solve the problem. The idea of dimension reduction using linear combination of data by Läuter (1996, 2016) and Läuter, Glimm and Kropf (1996, 1998) is given in Section 4.1. By this way, each p-dimensional observation will be compressed into one-dimensional observation. The same linear combination will be taken for each individual. But our approach is slightly different than Läuter's and Läuter, Glimm and Kropf's in terms of when and where this reduction takes place.

According to the present theories, one applies the weight vector $\boldsymbol{d}$ to the matrix $\boldsymbol{X}$ and obtain scores which are denoted by $\boldsymbol{z}$:

$$\boldsymbol{z}' = \boldsymbol{d}'\boldsymbol{X} = [z_1, z_2, \cdots, z_n] = [d_1, d_2, \cdots, d_p] \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}.$$

Here, the illustration is with one group of size n. If this approach, which is applying the vector $\boldsymbol{d}$ from the beginning to the matrix $\boldsymbol{X}$, is followed in our research problem (high-dimensional profile analysis), we cannot continue to conduct the tests in profile analysis, because the profiles will be reduced to one dimensional quantities:

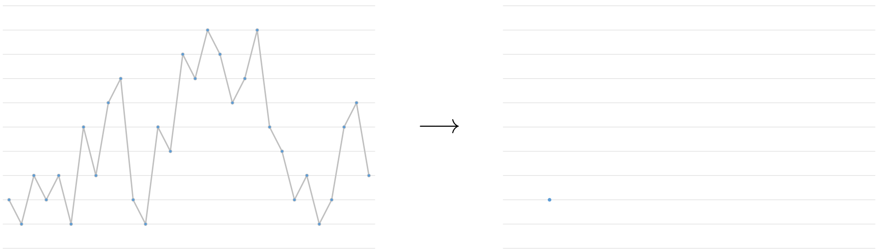

Figure 7: First figure corresponds to the profile of one group. After the dimension reduction with $\boldsymbol{d}$, the profile information will be summarised with one point, shown with the second figure.

The solution that is proposed is that dimension reduction should be implemented to the ratio for each hypothesis which is given by (7), (8) and (9). The detailed derivation of these ratios, which have been given in the report in Section 3, will be useful when we implement the methods in the high-dimensional setting since these derivations give clear intuitions where dimension reduction is needed and how the calculations should progress. There will be a special case with the level hypothesis. The specifics on this matter will be given in the following chapter. Now let's give the test statistics and distributions.

5.2 Test statistics and distributions of profile analysis with high-dimensional data

Consider the high-dimensional setting where $p > N$. The first step is to construct scores and the latter step is to derive the likelihood ratio tests based on these scores.

1) Parallelism hypothesis

Recall from Section 2.4.2

$$\begin{aligned} H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} = \boldsymbol{0}, \\ A_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} \neq \boldsymbol{0} \end{aligned}$$

and

$$LR = \frac{|\boldsymbol{\Sigma}_{A_1}|}{|\boldsymbol{\Sigma}_{H_1}|} = \frac{|\boldsymbol{CSC}'|}{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'|},$$

where

$$\begin{aligned} \boldsymbol{CSC}' &\sim W_{p-1}(\boldsymbol{C\Sigma C}', N - r(\boldsymbol{D})), \\ \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' &\sim W_{p-1}(\boldsymbol{C\Sigma C}', r(\boldsymbol{K})). \end{aligned} \quad (17)$$

It is assumed that $\boldsymbol{X} \sim N_{p,N}(\boldsymbol{MD}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$. From the properties of matrix normal distribution (see Theorem 2.5 in the report), $\boldsymbol{CX}$ is also normally distributed with $\boldsymbol{CX} \sim N_{(p-1),N}(\boldsymbol{CMD}, \boldsymbol{C\Sigma C}', \boldsymbol{I}_N)$. One can easily notice that $\boldsymbol{X}$ appears with $\boldsymbol{C}$ in the statistics $\boldsymbol{CSC}'$ and $\boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'$. Say $\boldsymbol{CX} = \boldsymbol{Y}$. Then

$$\boldsymbol{CSC}' = \boldsymbol{CX}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}' = \boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{Y}' \sim W_{p-1}(\boldsymbol{C\Sigma C}', N - r(\boldsymbol{D})),$$
$$\boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' = \boldsymbol{YP}_{D'(DD')^{-1}K}\boldsymbol{Y}' \sim W_{p-1}(\boldsymbol{C\Sigma C}', r(\boldsymbol{K})).$$

We propose to apply the vector $\boldsymbol{d}$ to $\boldsymbol{Y}$ instead of applying it to $\boldsymbol{X}$. This means we are taking a linear combination of the restricted $\boldsymbol{X}$. Here $\boldsymbol{d}$ is a $(p-1) \times 1$ vector. When $\boldsymbol{Y}$ is multiplied with $\boldsymbol{d}'$ from left (and $\boldsymbol{Y}'$ with $\boldsymbol{d}$ from right), it will be reduced to a vector and we call this new vector the score vector and denote it by $\boldsymbol{z}$, that is $\boldsymbol{z}' = \boldsymbol{d}'\boldsymbol{Y}$, which was obtained in the report:

$$\begin{aligned}\lambda^{2/N} &= \frac{\boldsymbol{d}'\boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{Y}'\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{Y}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{YP}_{D'(DD')^{-1}K}\boldsymbol{Y}'\boldsymbol{d}} \\ &= \frac{\boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{z}}{\boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{P}_{D'(DD')^{-1}K}\boldsymbol{z}}. \end{aligned} \tag{18}$$

Notice that this ratio is not a likelihood ratio anymore. Now let's give the distribution of this ratio.

Theorem 5.1. *The ratio given in* (18) *follows Wilks' lambda distribution with parameters* 1*,* $N - r(\boldsymbol{D})$ *and* $r(\boldsymbol{K})$ *that is denoted by* $\Lambda(1, N - r(\boldsymbol{D}), r(\boldsymbol{K}))$ *which is equivalent to* $B\left(\frac{N-r(\boldsymbol{D})}{2}, \frac{r(\boldsymbol{K})}{2}\right)$*, where* $B(\cdot,\cdot)$ *denotes the Beta-distribution.*

2) Level hypothesis

The null and the alternative hypotheses in the normal setting have been given in Section 2.4.2 as:

$$\begin{aligned} H_2|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{MF} = \boldsymbol{0}, \\ A_2|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} = \boldsymbol{0} \end{aligned}$$

and the likelihood ratio equals

$$LR = \frac{|(\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ}|^{-1}}{\left|\begin{matrix}((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1} + ((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}\boldsymbol{XD}'(\boldsymbol{DD}')^{-1}\boldsymbol{K} \\ \times \boldsymbol{Q}^{-1}\boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{DX}'\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ}((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1}\end{matrix}\right|}. \tag{19}$$

Here we have a special case, because both expressions in the likelihood ratio are already one dimensional. If LR is imagined as $LR = \frac{|\boldsymbol{U}|}{|\boldsymbol{U}+\boldsymbol{V}|}$, the expressions in question are $\boldsymbol{U}$ and $\boldsymbol{V}$. To see this, one should calculate the dimension of $(\boldsymbol{C}')^{\circ'}$:

$$\boldsymbol{C} : (p-1) \times p \;\Rightarrow\; \boldsymbol{C}' : p \times (p-1) \;\Rightarrow\; (\boldsymbol{C}')^{\circ} : p \times 1 \;\Rightarrow\; (\boldsymbol{C}')^{\circ'} : 1 \times p.$$

The issue for this hypothesis lies within the degrees of freedom in the Wishart distribution. Recall

$$((\boldsymbol{C}')^{\circ'}\boldsymbol{S}^{-1}(\boldsymbol{C}')^{\circ})^{-1} \sim W_1(((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, N - r(\boldsymbol{D}) - p + 1).$$

When p exceeds N, that is $p > N$, the degrees of freedom will become negative which cannot take place. In addition to this, $\boldsymbol{S}^{-1}$ does not exist. To handle these problems we propose a dimension reduction by using the weight vector $\boldsymbol{d}$ which has been defined previously. In the beginning, it is not very clear where to apply the dimension reduction in the ratio given by (19). After using some properties given in Appendix, we have found equivalent expressions for $\boldsymbol{U}$ and $\boldsymbol{V}$ and then $\boldsymbol{d}$ is applied to $\boldsymbol{CX}$. Denote these new statistics with $\widetilde{\boldsymbol{U}}_L$ and $\widetilde{\boldsymbol{V}}_L$. Then

$$\begin{aligned}\widetilde{\boldsymbol{U}}_L &= ((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}[(\boldsymbol{I} - \boldsymbol{P}_{D'})(\boldsymbol{I} - \boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d}(\boldsymbol{d}'\boldsymbol{CX} \\ &\quad \times (\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{CX}(\boldsymbol{I} - \boldsymbol{P}_{D'}))]\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ}((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, \\ \widetilde{\boldsymbol{V}}_L &= ((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}[\boldsymbol{I} - (\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d}(\boldsymbol{d}'\boldsymbol{CX} \\ &\quad \times (\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{CX}]\boldsymbol{D}'(\boldsymbol{DD}')^{-1}\boldsymbol{K}\widetilde{\boldsymbol{Q}}^{-1}\boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{D} \\ &\quad \times [\boldsymbol{I} - \boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d}(\boldsymbol{d}'\boldsymbol{CX}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{d})^{-1}\boldsymbol{d}'\boldsymbol{CX}(\boldsymbol{I} - \boldsymbol{P}_{D'})]\boldsymbol{X}' \\ &\quad \times \boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ}((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}.\end{aligned}$$

After simplification,

$$\begin{aligned}\widetilde{\boldsymbol{U}}_L &= ((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}[\boldsymbol{S}(\boldsymbol{I} - \boldsymbol{P}_{C'd,S^{-1}})]\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ} \\ &\quad \times ((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1} \sim W_1(((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, N - r(\boldsymbol{D}) - 1), \\ \widetilde{\boldsymbol{V}}_L &= ((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}(\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{P}'_{C'd,S^{-1}})\boldsymbol{XD}'(\boldsymbol{DD}')^{-1}\boldsymbol{K}\widetilde{\boldsymbol{Q}}^{-1} \\ &\quad \times \boldsymbol{K}'(\boldsymbol{DD}')^{-1}\boldsymbol{DX}'(\boldsymbol{I} - \boldsymbol{P}_{C'd,S^{-1}}) \sim W_1(((\boldsymbol{C}')^{\circ'}\boldsymbol{\Sigma}^{-1}(\boldsymbol{C}')^{\circ})^{-1}, r(\boldsymbol{K})).\end{aligned}$$

The following theorem was presented in the report.

Theorem 5.2.

$$\begin{aligned}\lambda^{2/N} = \frac{\widetilde{\boldsymbol{U}}_L}{\widetilde{\boldsymbol{U}}_L + \widetilde{\boldsymbol{V}}_L} &\sim \Lambda(1, N - r(\boldsymbol{D}) - 1, r(\boldsymbol{K})) \\ &\equiv B\left(\frac{N - r(\boldsymbol{D}) - 1}{2}, \frac{r(\boldsymbol{K})}{2}\right).\end{aligned}$$

3) Flatness Hypothesis

The test statistics in the likelihood ratio for the parallelism hypothesis and the flatness hypothesis have similar structures, therefore a similar approach to the parallelism hypothesis will be followed. Recall the null and the alternative hypotheses from Section 2.4.2:

$$\begin{aligned}H_3|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CM} = \boldsymbol{0}, \\ A_3|H_1: &\quad E(\boldsymbol{X}) = \boldsymbol{MD}, \quad \boldsymbol{CMF} = \boldsymbol{0}\end{aligned}$$

and

$$LR = \frac{|\widehat{\boldsymbol{\Sigma}}_{A_3}|}{|\widehat{\boldsymbol{\Sigma}}_{H_3}|} = \frac{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}'|}{|\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' + \boldsymbol{CXP}_{D'F^\circ}\boldsymbol{X}'\boldsymbol{C}'|},$$

where

$$\boldsymbol{CSC}' + \boldsymbol{CXP}_{D'(DD')^{-1}K}\boldsymbol{X}'\boldsymbol{C}' \sim W_{p-1}(\boldsymbol{C\Sigma C}', N - r(\boldsymbol{D}) + r(\boldsymbol{K})), \quad (20)$$

$$\boldsymbol{CXP}_{D'F^\circ}\boldsymbol{X}'\boldsymbol{C}' \sim W_{p-1}(\boldsymbol{C\Sigma C}', r(\boldsymbol{D}'\boldsymbol{F}^\circ)). \quad (21)$$

We investigate $\boldsymbol{CSC}'$ first. $\boldsymbol{CSC}' = \boldsymbol{CX}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{X}'\boldsymbol{C}'$ has been given several times before. Put $\boldsymbol{Y} = \boldsymbol{CX}$ and since $\boldsymbol{X} \sim N_{p,N}(\boldsymbol{MD}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$, $\boldsymbol{Y} \sim N_{(p-1),N}(\boldsymbol{CMD}, \boldsymbol{C\Sigma C}', \boldsymbol{I}_N)$. To reduce the dimension, the weight vector $\boldsymbol{d}$ will be applied to $\boldsymbol{Y}$ instead of the data matrix $\boldsymbol{X}$ like we did for the parallelism hypothesis and denote the new vector by $\boldsymbol{z}$, where $\boldsymbol{z}' = \boldsymbol{d}'\boldsymbol{Y}$. Then

$$\begin{aligned}\lambda^{2/N} &= \frac{\boldsymbol{d}'\boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{Y}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{YP}_{D'(DD')^{-1}K}\boldsymbol{Y}'\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{Y}(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{Y}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{YP}_{D'(DD')^{-1}K}\boldsymbol{Y}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{YP}_{D'F^\circ}\boldsymbol{Y}'\boldsymbol{d}} \\ &= \frac{\boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{P}_{D'(DD')^{-1}K}\boldsymbol{z}}{\boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{P}_{D'})\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{P}_{D'(DD')^{-1}K}\boldsymbol{z} + \boldsymbol{z}'\boldsymbol{P}_{D'F^\circ}\boldsymbol{z}}. \end{aligned} \quad (22)$$

The distribution of this ratio is presented in the next theorem.

Theorem 5.3. *The ratio given in* (22) *follows Wilks' lambda distribution with parameters* 1, $N - r(\boldsymbol{D}) + r(\boldsymbol{K})$ *and* $r(\boldsymbol{D}'\boldsymbol{F}^\circ)$ *that is denoted by* $\Lambda(1, N - r(\boldsymbol{D}) + r(\boldsymbol{K}), r(\boldsymbol{D}'\boldsymbol{F}^\circ))$ *which equals* $B\big(\frac{N-r(\boldsymbol{D})+r(\boldsymbol{K})}{2}, \frac{r(\boldsymbol{D}'\boldsymbol{F}^\circ)}{2}\big)$.

6 Conclusion

6.1 Discussion

Our aim in this thesis was to develop a method which can be applied to compare profiles in a high-dimensional setting. The three hypotheses of profile analysis were investigated. To be able to work in high-dimensions, we first conducted the derivations in the normal setting ($N > p$). The calculations were directed to form a ratio $\frac{|\boldsymbol{U}|}{|\boldsymbol{U}+\boldsymbol{V}|}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent and Wishart distributed. The distribution of this ratio is well-known, i.e., the Wilks' lambda distribution. Then the crucial question was how to implement a dimension reduction method using scores which was inspired by Läuter (1996, 2016) and Läuter, Glimm and Kropf (1996, 1998). Applying a vector $\boldsymbol{d}$ solely to $\boldsymbol{X}$ would not work in our problem since it would not be possible to apply the restrictions on the mean parameter space.

Note that the level hypothesis in high-dimensions was treated differently than the other two hypotheses, because the matrices in the last statistic were already one-dimensional. Since the degrees of freedom of the Wishart distribution for one of the statistics that appears in the likelihood ratio became negative when $p > N$, a dimension reduction was necessary.

Multivariate normal distributions are frequently used in the multivariate analysis and the results that were derived in the first part of the thesis (when $N > p$) are based on the assumption of normality. We assumed that $\boldsymbol{X} \sim N_{p,N}(\boldsymbol{MD}, \boldsymbol{\Sigma}, \boldsymbol{I}_N)$, but normality will not hold when the dimension reduction is applied because $\boldsymbol{d}$ depends on $\boldsymbol{X}$. This is where we utilized the properties of spherical distributions. In the end, for all three hypotheses, the ratios follow Wilks' lambda distribution after properly done dimension reductions.

6.2 Future work

Besides the results that have been given in this thesis, there are still several questions which require attention. There are mainly two problems which we are interested in:

1. How should we choose the vector $\boldsymbol{d}$?

   As Läuter, Glimm and Kropf (1996) suggested, $\boldsymbol{d}$ has to be a unique function of the $p \times p$ matrix of the sums of the products $\boldsymbol{XX}'$. This is the matrix which corresponds to the hypothesis $\boldsymbol{\mu} = \boldsymbol{0}$. If one compares the equality of group means, i.e. $H : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_q$, the weight vector $\boldsymbol{d}$ may be some arbitrary function of the total sum of squares matrix under H, that is $(\boldsymbol{X} - \bar{\boldsymbol{X}})(\boldsymbol{X} - \bar{\boldsymbol{X}})'$. For the standardized sum test, $\boldsymbol{d}$ is chosen based on $\boldsymbol{d} = [\mathrm{Diag}(\boldsymbol{XX}')]^{-1/2}\mathbf{1}_p$ and for the principle component test, $\boldsymbol{d}$ is the solution to the eigenvalue problem $(\boldsymbol{XX}')\boldsymbol{d} = \mathrm{Diag}(\boldsymbol{XX}')\boldsymbol{d}\lambda$ (Läuter, Glimm and Kropf, 1996, 1998). In this thesis, $\boldsymbol{d}$ is a function of $\boldsymbol{CX}$, where $\boldsymbol{C}$ is defined in Section 2.4.1, but it is still unclear how to choose $\boldsymbol{d}$ exactly. This requires further investigations.

2. Instead of using one linear combination of the variables, which is denoted by the vector $\boldsymbol{d}$, what will happen if one considers several, say r linear scores, which is denoted by the $p \times r$ weight matrix $\widetilde{\boldsymbol{D}}$ where $r \leq p$?

   In this case, the scores are obtained via $\boldsymbol{Z} = \widetilde{\boldsymbol{D}}'\boldsymbol{X}$ (Läuter, Glimm and Kropf, 1996, 1998). For our research, it will be $\boldsymbol{Z} = \widetilde{\boldsymbol{D}}'\boldsymbol{CX}$. Part of the research will be to investigate how the distributions of the likelohood ratios will be effected.

References

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed). Wiley, New York.

[2] Anderson, T. W. and Fang, K. T. (1982). On the theory of multivariate elliptically contoured distributions and their applications. Technical Report No. 54, Stanford University, California.

[3] Anderson, T. W. and Fang, K. T. (1990). Theory and applications of elliptically contoured and related distributions. Technical Report No. 24, Stanford University, California.

[4] Bilodeau, M. and Brenner, D. (1999). *Theory of Multivariate Statistics.* Springer-Verlag, New York.

[5] Cambanis, S., Huang, S., and Simons, G. (1981). On the theory of elliptically contoured distribution. *Journal of Multivariate Analysis*, 11, 368-385.

[6] Dawid, A. P. (1977). Spherical matrix distributions and a multivariate model. *Journal of the Royal Statistical Society, Series B*, 39, 254-261.

[7] Dawid, A. P. (1978). Extendibility of spherical matrix distributions. *Journal of Multivariate Analysis*, 8, 559-566.

[8] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions.* Springer-Science+Business Media, B.V.

[9] Fang, K. T. and Zhang, Y. T. (1990). *Generalized Multivariate Analysis.* Springer-Verlag, Berlin.

[10] Fujikoshi, Y. (2009). Statistical inference for parallelism hypothesis in growth curve model. *SUT Journal of Mathematics*, 45, 137-148.

[11] Fujikoshi, Y., Ulyanov, V. V., and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations.* Wiley, Hoboken, New Jersey.

[12] Geisser, S. (2003). The analysis of profile data–revisited. *Statistics in Medicine*, 22, 3337-3346.

[13] Greenhouse, S. W., Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika* 24, 95-112.

[14] Harrar, S. W. and Kong, X. (2016). High-dimensional multivariate repeated measures analysis with unequal covariance matrices. *Journal of Multivariate Analysis* 145, 1-21.

[15] Kariya, T. and Sinha, B. K. (1989). *Robustness of Statistical Tests.* Academic Press, Boston.

[16] Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* 32, 419-430.

[17] Kollo, T. and von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices*. Springer Dordrecht.

[18] Kollo, T., von Rosen, T. and von Rosen, D. (2011). Estimation in high-himensional analysis and multivariate linear models. *Communications in Statistics - Theory and Methods*, 40, 1241-1253.

[19] Kropf, S. (2000). *Hochdimensionale Multivariate Verfahren In Der Medizinischen Statistik*. Shaker Verlag, Aachen.

[20] Läuter, J. (1996). Exact t and F tests for analyzing studies with multiple endpoints. *Biometrics*, 52, 964-970.

[21] Läuter, J. (2016). *Multivariate Statistik - drei Manuskripte*. Shaker Verlag, Aachen.

[22] Läuter, J., Glimm, E., and Kropf, S. (1996). New multivariate tests for data with an inherent structure. *Biometrical Journal*, 38, 5-23.

[23] Läuter, J., Glimm, E. and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *The Annals of Statistics* 26, 1972-1988.

[24] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics* 30, 1081-1102.

[25] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, New York.

[26] Morrison, D. F. (2004). *Multivariate Statistical Methods*, 4th edition. Duxbury Press, CA.

[27] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

[28] O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, 40, 1079-1087.

[29] Ohlson, M. and Srivastava, M. S. (2010). Profile analysis for a growth curve model. *Journal of the Japan Statistical Society*, 40, 1-21.

[30] Onozawa, M., Nishiyama, T. and Seo, T. (2016). On test statistics in profile analysis with high-dimensional data. *Communications in Statistics - Simulation and Computation*, 45, 3716-3743.

[31] Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 51, 313-326.

[32] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edition. Wiley, New York.

[33] Rencher, A. C. (2002). *Methods of Multivariate Analysis*, 2nd edition. Wiley, New York.

[34] Seo, T., Sakurai, T. and Fujikoshi, Y. (2011). LR tests for two hypotheses in profile analysis of growth curve data. *SUT Journal of Mathematics*, 47, 105-118.

[35] Shutoh, N. and Takahashi, S. (2016). Tests for parallelism and flatness hypotheses of two mean vectors in high-dimensional settings. *Journal of Statistical Computation and Simulation*, 86, 1150-1165.

[36] Srivastava, M. S. (1987). Profile analysis of several groups. *Communications in Statistics - Theory and Methods*, 16, 909-926.

[37] Srivastava, M. S. (2002). *Methods of Multivariate Statistics*. Wiley, New York.

[38] Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society*, 35, 251–272.

[39] Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 37, 53-86.

[40] Srivastava, M. S. and Carter, E. M.(1983). *An Introduction to Applied Multivariate Statistics*. North Holland, New York.

[41] Srivastava, M. S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99, 386–402.

[42] Srivastava, M. S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *Journal of Multivariate Analysis*, 97, 1927–1940.

[43] Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*. North Holland, New York.

[44] Srivastava, M. S. and Singull, M. (2012). Profile analysis with random-effects covariance structure. *Journal of the Japan Statistical Society*, 42, 145-164.

[45] Srivastava, M. S. and Singull, M. (2017). Test for the mean matrix in a growth curve model for high dimensions. *Communications in Statistics - Theory and Methods*, 46, 6668-6683.

[46] von Rosen, D. (2018). *Bilinear Regression Analysis: An Introduction*. Springer International Publishing, New York.

[47] Yokoyama, T. (1995). LR test for random-effects covariance structure in a parallel profile model. *Annals of the Institute of Statistical Mathematics*, 47, 309-320.

[48] Yokoyama, T. and Fujikoshi, Y. (1993). A parallel profile model with random-effects covariance structure. *Journal of the Japan Statistical Society*, 23, 83-89.

Acknowledgements

A very simple causal diagram can be written as $\textcircled{A} \longrightarrow \textcircled{B}$, meaning event A causes event B. If we consider the event B as completing my licentiate degree, I can list many A's pointing to B. Below, I would like to mention the people who influence the A's and accordingly the B.

First on the very long list below, I would like to express my gratitude to my main supervisor, Prof. Dietrich von Rosen. I have been very lucky to work under your supervision and I could not have asked for a better supervisor. You were always available and gave constructive criticism, but most importantly taught me how to look at things from a different perspective and to like and enjoy my work. You have been very kind and supportive. To my co-supervisor, Dr. Martin Singull, thank you very much for your feedback on the manuscripts, which has helped a lot to improve things. I am looking forward to working with you in Linköping. To my co-supervisor, Prof. Julia Volaufova, thank you for carefully reading my work and pointing out important details.

I want to show my appreciation to Prof. Mujgan Tez from Marmara University, who informed me about this PhD position in the first place. I would not be here if it were not for you.

I would like to thank the good people of the Department of Energy and Technology for providing such a friendly environment. Eva-Lena, you have been very helpful with all the questions and issues I had. What would I have done without you? Sven, many thanks for the technical help, especially during the last period. Claudia and Johannes, it was a pleasure to work with you with teaching which was a rewarding experience. Tomas and Jesper, I enjoyed our lunch times together very much when we talked about many different things. Razaw, you always asked how I was and if I was doing fine. My previous office mate Harimurti, we stood by each other during this tough journey and thank you for all your help. Good luck in Stuttgart my friend, I believe in you. My current office mate Mahlet, this snack girl is going to miss you. Sahar and Hjalmar, I cannot thank enough for your kind invitations and great gatherings. They were some of the nicest moments I will remember. Prithvi, we had so much fun together and I will miss our conversations. Sweet Shan, I wish we could have met earlier. You have become such a good friend in a short time, but we will meet again in the der Kuss by Klimt. Torun, my Swedish teacher, sorry for not being a very good student, but I appreciate your effort. Jennifer, thank you for your help with the proof-reading. My PhD fellows from Linköping University, Denise and Emelyne, we had quite nice adventures taking courses together, going to conferences and attending seminars. Lastly, Behnaz, we shared not only the same working space, but also the same living space for 2.5 years. You treated me with the kindness of a mother. We supported each other during this time. Zor septas, with all my heart.

Going to conferences or doing research visits is an important part of the program. Bedlewo, Poland has become like a second home which I had the chance to visit three times. I would like to express my sincere gratitude to Augustyn Markiewicz, Katarzyna Filipiak and the rest of the team. Another gratitude goes to Shinpei Imori from Hiroshima University who accepted to host me for JSPS Summer Program. I am looking forward to our collaboration. Prof. Siegfried Kropf, thank you for sending me your book and the papers and sharing ideas.

The biggest emotional support comes from the family, especially when it is needed the most. My cousin, Elif, you are like a sister to me. Thank you for being one of the greatest joys in my life. You cheer me up when I feel low, you bring a smile to my face when I feel sad. My brother, Kubilay, you called me almost everyday to check if I am doing OK, if everything is fine. I am one lucky sister. And my dear mum, there is not enough vocabulary to tell you how much I love you. You are the strongest person I have seen and you keep inspiring me everyday. Being far away is hard, but please know that you are always in my heart. Hope that father is watching and keeping an eye on us.

The three tests of profile analysis: test of parallelism, test of level and test of flatness have been studied. Likelihood ratio tests have been derived. Firstly, a traditional setting, where the sample size is greater than the dimension of the parameter space, is considered. Then, all tests have been derived in a high-dimensional setting. A dimension reduction method using scores has been proposed.

Cigdem Cengiz, the author of this thesis, received her postgraduate education at the Department of Energy and Technology, SLU, Uppsala. She obtained her Master Degree in Statistics with Biostatistics pathway from The University of Manchester, the UK.

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

Online publication of the thesis summary: http://pub.epsilon.slu.se/

ISBN (print version): 978-91-576-9783-7
ISBN (electronic version): 978-91-576-9784-4