The Pennsylvania State University
The Graduate School

PROJECTION TEST FOR HIGH-DIMENSIONAL MEAN VECTORS WITH OPTIMAL DIRECTION

A Dissertation in
Statistics
by
Yuan Huang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2015

The dissertation of Yuan Huang was reviewed and approved* by the following:

Runze Li
Distinguished Professor of Statistics
Dissertation Advisor, Chair of Committee

Naomi Altman
Professor of Statistics

David Hunter
Professor of Statistics

Lan Kong
Associate Professor of Public Health Sciences

Aleksandra B. Slavković
Associate Professor of Statistics
Associate Head for Graduate Studies

*Signatures are on file in the Graduate School.

Abstract

Testing the population mean is fundamental in statistical inference. When the dimensionality of a population is high, traditional Hotelling's T^2 test becomes practically infeasible due to the singularity of sample covariance matrix. In this dissertation, we propose a projection-based testing method for the high-dimensional one-sample and two-sample mean problems. Our method projects the original sample to a lower-dimensional space and conducts tests on the projected sample. Different from existing projection-based tests, our approach is based on data-driven estimation of the optimal direction. Meanwhile, our test keeps the equivalence of null hypotheses between the original sample and the projected sample, which is often ignored by previous researches. We show that the test based on projected sample is an exact t-test under the normality assumption and an asymptotic χ^2-test with one degree of freedom without the normality assumption.

In the one-sample problem, we are interested in testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ for a random sample of size N from a p-dimensional population X with finite mean vector $\boldsymbol{\mu}$ and finite positive definite covariance matrix Σ. We derive the theoretical optimal direction with which the test possesses the most power under the alternative. We show that projection to a one-dimensional space with direction $\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$ leads to the optimal power, regardless of the distribution assumption. The null hypothesis with the projected sample is $(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0) = 0$, which holds if and only if $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ for a full rank Σ. A computationally efficient algorithm is developed to implement the new test. Local asymptotic property is studied and we show that under mild conditions the proposed test outperforms the major existing methods. Our numerical comparison shows that the new test retains Type I error rate well and can be more powerful than the existing tests for the high-dimensional data.

In the two-sample problem, we are interested in testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_0 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ for two independent random samples of size N_i from populations

with finite mean vector $\boldsymbol{\mu}_i$ and finite positive definite covariance matrix Σ_i, $i = 1, 2$ respectively. When $\Sigma_1 = \Sigma_2 = \Sigma$, we prove that the optimal direction is $\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, regardless of the distribution assumption. When the covariance matrices are unequal, we show that the optimal projection direction is $\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ for normal population by first taking Bennett's transformation to obtain an one-sample sequence of size N_1 that is distributed from $N(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2)$, assuming $N_1 < N_2$. Both theoretically and empirically, we demonstrate that the proposed test can be much more powerful than the existing ones.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I owe my gratitude to many people who have helped me on the path towards this dissertation.

I am deeply indebted to my advisor Dr. Runze Li for his valuable guidance and constant support, without which this work would never be possible. His insights sharpen this dissertation from all aspects; his way of thinking helps me learn how to really think as an independent researcher; his enthusiasm keeps me move forward in my professional career. I could not have asked for a better advisor and no words can express how thankful I am to him. I would also like to thank the other members of my committee, Dr. Naomi Altman, Dr. David Hunter and Dr. Lan Kong for providing valuable suggestions and improving every detail of this dissertation. Thanks to the Statistic department at the Penn State University for building a world class program, organizing extensive workshops and seminars, and creating an friendly, warming, and encouraging environment. I had a wonderful experience in my graduate studies and a great time living in Happy Valley, which would not be possible without good friends especially the peers in my year. Special words go to Dr. Dennis Lin, who helped me get involved into a new environment and shared the wisdom of life that I am beneficent from. Thanks to the undergraduate students at the Penn State University whom I had the opportunity to teach and learn with.

I also want to extend my gratitude to people outside Penn State. I am indebted to Dr. Shuangge Ma at Yale University for his unwavering encouragement and sincere belief in my work. He is a truly amazing collaborator, mentor, and friend. I would like to thank Dr. Yongjin Jin, my advisor at Renmin University of China, Dr. Hing-po Lo, my supervisor at City University of Hong Kong, and Dr. Ben-Chang Shia, my mentor and collaborator at Fu Jen Catholic University. Without their enlightenment I would never have pursued research on statistics and made it this far.

Finally, thanks to my parents, for all the love and strength they give me. And thanks to my husband Lei, who shares my life and is also my best reader. His

support and influence at every stage in the writing of this dissertation have been essential. I am very blessed to have them in my life.

Chapter 1

Introduction

High-dimensional data are generated by modern technologies at an unprecedented speed and they frequently arise in many research areas including various imaging data applications; different “-omics” disciplines such genomics and proteomics; finance; and sociological study such as political science. In these cases, the number of collected features can be close to, great than or much greater than the sample size. High-dimensionality challenges the classical theory fundamentally and has attracted tremendous research interests in developing new methodologies beyond the classic techniques.

Being one of the most active research areas in statistics (e.g., Donoho, 2000; Fan and Li, 2006; Fan et al., 2014a), much progress has been made so far on the variable selection and sparsity recovery, which select the variables that are truly relevant to the phenomena of interest. However, there are still many challenging unsolved problems that call for the development of new methods and theory, especially on the statistical inference which is less explored. Recently, much attention has been received in literature for hypothesis testing of mean vectors, which is a fundamental question in statistics and the focus of this dissertation as well.

In this introduction, we first provide a very brief review on variable selection, feature screening and statistical inference on high-dimensional regression models. Then we focus on the current development on testing the mean vectors. The last part summarizes our method and its contribution.

1.1 High-dimensional variable selection, feature screening and inference on regressions

The sparsity principle, which assumes that only a small number of predictors contribute to the response, is frequently adopted and deemed useful in regression analysis with high-dimensional predictors. Following this general principle, a large number of variable selection approaches have been developed to estimate a sparse model and select significant variables simultaneously during the last two decades. Fan and Lv (2010) provided a review of variable selection for high-dimensional data. The nonnegative garrote (Breiman, 1996; Yuan and Lin, 2006), the LASSO (Tibshirani, 1996), the smoothly clipped absolute deviation (SCAD) method (Fan and Li, 2001) and the minimax concave penalty (MCP) method (Zhang, 2010) are the most popular approaches for selecting significant variables and estimating regression coefficients simultaneously.

The LASSO method makes use of a penalized least squares with the L_1-penalty, and its solution path can be found by using the LARS algorithm (Efron et al., 2004). Yuan and Lin (2006) proposed the group LASSO for grouped variable selection, and Zou (2006) proposed the adaptive LASSO to reduce estimation bias due to the L_1-penalty. Due to its computation efficiency, the LASSO method has been further extended for various statistical settings by many authors (See, for example, Tibshirani, 1997; Tibshirani et al., 2005; Zou and Hastie, 2005; Meinshausen and Bühlmann, 2006; Park and Hastie, 2007; Rosset and Zhu, 2007; Zhang and Huang, 2008, and among others).

Fan and Li (2001) provided insights into how to choose a penalty function. In particular, Fan and Li (2001) advocated the use of nonconvex penalties such as the SCAD penalty and established the oracle property for nonconvex penalized least squares and nonconcave penalized likelihood methods. In the same spirit of the SCAD, Zhang (2010) proposed the MCP, and Fan and Lv (2011) studied a family of concave penalties that bridge the L_0 and L_1 penalties for model selection and sparse recovery. Regularization parameter controls the model complexity, and therefore plays a critical role in the application of the penalization, in addition to the penalty function. The issue of tuning parameter selection was studied in

Wang et al. (2007), Zhang et al. (2010) and Wang et al. (2013). In employing the nonconcave penalized likelihood, lots of efforts haven been devoted on developing efficient algorithms to compute the nonconcave penalized likelihood estimate. Fan and Li (2001) proposed the local quadratic approximation (LQA) algorithm for nonconcave penalized likelihood. The LQA algorithm was further analyzed by Hunter and Li (2005) by using the MM algorithm techniques (Lange et al., 2000; Hunter and Lange, 2000). Zou and Li (2008) further proposed local linear approximation (LLA) algorithm for the nonconcave penalized likelihood. The LLA algorithm enables us to use LARS algorithm to search the solution of nonconcave penalized likelihood. The nonconcave penalized likelihood approach was applied for survival data analysis (Fan and Li, 2002; Cai et al., 2005), longitudinal data analysis (Fan and Li, 2004), modeling computer experiments (Li and Sudjianto, 2005), and semiparametric regression modeling (Li and Liang, 2008; Liang and Li, 2009; Kai et al., 2011; Liang et al., 2010).

While the aforementioned variable selection methods have been successfully applied in many high-dimensional analysis, modern applications in areas such as genomics and proteomics push the dimensionality of data to an even larger scale, where the dimension of data may grow exponentially with the sample size. This has been called ultrahigh-dimensional data in the literature. Such ultrahigh-dimensional data present simultaneous challenges of computational expediency, statistical accuracy and algorithm stability, which render difficulties in direct applications of the aforementioned variable selection methods (Fan et al., 2009). To address those challenges, Donoho and Elad (2003) derived a characterization of the identifiability of the minimum L_0-norm for undetermined linear equations. Donoho (2004, 2006) showed the individual equivalence of the minimal L_1-norm solution and the minimal L_0-norm solution. Candes and Tao (2007) further extended Donoho (2004, 2006)'s idea and proposed the Dantzig selector for a linear regression model when the number of predictors is much greater than the sample size. Independence learning has been proposed to select significant genes between treatment and control groups for macroarray data by using a two-sample test in Dudoit et al. (2003), Storey and Tibshirani (2003), Fan and Ren (2006), Efron (2007). Fan and Lv (2008) emphasized the importance of feature screening in ultrahigh-dimensional data analysis, and proposed sure independence screening

(SIS) and iterated sure independence screening (ISIS) in the context of a linear regression model. Huang et al. (2008) used the marginal bridge estimators for selecting variables in high-dimensional sparse regression models. Hall and Miller (2009) proposed feature ranking using a generalized correlation. Fan et al. (2009) and Fan and Song (2010) extended SIS and ISIS from a linear model to a generalized linear model. Ravikumar et al. (2009) proposed nonparametric learning under sparse additive models. Fan et al. (2011) proposed a nonparametric marginal screening procedure for additive models based on B-spline expansion. Fan et al. (2014b) further extended the nonparametric B-spline method for varying coefficient models and proposed a marginal sure screening procedure. Liu et al. (2014) proposed a local kernel-based marginal sure screening procedure for varying coefficient models. Aforementioned model-based screening procedures perform well when the underlying models are correctly specified, but their performance may be poor in the presence of model mis-specification. Specifying a correct model for ultrahigh dimensional data may be challenging. Several authors have developed several model-free sure screening procedures, which are particularly appealing for ultrahigh dimensional data (Zhu et al., 2011; Li et al., 2012; He et al., 2013; Cui et al., 2014).

Although variable selection and feature screening have been extensively studied in the literature, there are relative less work on high-dimensional inference including test of hypotheses and confidence interval in linear regression. Dezeure et al. (2014) provided a selective review on this topic. Single sample-splitting and subsequent statistical inference is implicitly contained in Wasserman and Roeder (2009), and further developed to multi-sample-splitting method for multiple testing in Meinshausen et al. (2009). Chatterjee and Lahiri (2013, 2011) developed residual-type bootstrap approaches for adaptive LASSO. Regularized projection was proposed to construct confidence interval for penalized least squares estimates such as the LASSO and the MCP in Zhang and Zhang (2013), van de Geer et al. (2014) and Javanmard and Montanari (2013a,b). Lockhart et al. (2014) proposed using the covariance test to obtain p-values for conditional tests that all relevant variables enter the Lasso solution path first. Efron (2014) considered bootstrap smoothing to tame the erratic discontinuities of selection-based estimators for computing standard errors and confidence intervals that take into account uncertainty due to variable selection using LASSO. Wang et al. (2014b) demonstrated Efron's method works well for other variable selection methods such as SCAD. The dissertation aims to develop new tests for high-dimensional mean vectors.

1.2 Tests on high-dimensional mean vectors

Hypothesis testing for high-dimensional mean vectors has received considerable attention in recent literature. Suppose that $\mathbf{x}_1, \cdots, \mathbf{x}_N$ is a random sample from a p-dimensional population X with finite mean vector $\boldsymbol{\mu}$ and finite positive definite covariance matrix Σ. Of interest is to test the following hypothesis:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \tag{1.1}$$

for a known vector $\boldsymbol{\mu}_0$. This hypothesis testing has been referred to as one-sample problem in multivariate analysis and extensively studied for univariate or multivariate population (i.e. p is fixed). Here we are interested in the large p, small n setting. Let $\bar{\mathbf{x}}$ and $\mathbf{S}$ be the sample mean and sample covariance matrix, respectively. When $N > p$, $\mathbf{S}^{-1}$ is invertible with probability one, and the following Hotelling T^2 test for the one-sample problem is well defined:

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0). \tag{1.2}$$

It is well known that if $\mathbf{x} \sim N_p(\mu, \Sigma)$, $(N-p)T^2/((N-1)p)$ follows $F_{p,N-p}(N\delta)$, the noncentral F-distribution with p and $N-p$ degrees of freedom with noncentrality parameter $N\delta$, where $\delta = \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$. Without normality assumption, T^2 has an asymptotical $\chi^2_p(N\delta)$, the noncentral χ^2-distribution with p degrees of freedom with noncentrality parameter $N\delta$, under mild regularity conditions as $N \to \infty$. The one-sample problem is closely related to the two-sample problem introduced below. In particular, most existing tests for the two-sample problem can be directly applied for the one-sample problem.

Suppose that for $i = 1$ and 2, $\{\mathbf{x}_{ij},\ j = 1, \cdots, N_i\}$ is a random sample from a population with finite mean vector $\boldsymbol{\mu}_i$ and finite positive definite covariance matrix

Σ. The two-sample problem is referred to as testing the following hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{1.3}$$

Let $\bar{\mathbf{x}}_i$ and $\mathbf{S}_i$ be the sample mean and the sample covariance matrix of $\mathbf{x}_{ij}$, respectively, and $\mathbf{S}_0 = \{(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2\}/(N_1 + N_2 - 2)$, the pooled sample covariance matrix. The Hotelling's T^2 test for the two-sample problem (1.3) is

$$T_2^2 = \frac{N_1 N_2}{N_1 + N_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}_0^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \tag{1.4}$$

Testing the hypotheses in (1.1) and (1.3) becomes challenging for high dimensional data. The traditional Hotelling's T^2 test given by (1.2) and (1.4) are not well-defined due to the singularity of $\mathbf{S}$ and $\mathbf{S}_0$ when $p > N$ or $p > N_1 + N_2$. It has been observed in Bai and Saranadasa (1996) that the power of the Hotelling's T^2 test can be adversely affected even when $p < N$ or $p < N_1 + N_2$, if the sample covariance matrix is nearly singular, see also Pan and Zhou (2011). Recently, there has been great interests in the one-sample and two-sample problems in large-dimensional setting with $p/N \to c \in (0, 1)$ (Bai and Saranadasa, 1996; Srivastava and Du, 2008; Srivastava, 2009), and in high-dimensional setting without imposing condition $c \in (0, 1)$ (Lee et al., 2012; Srivastava et al., 2013; Chen and Qin, 2010; Thulin, 2014). Chen et al. (2011) introduced regularized Hotelling's T^2 test by replacing $\mathbf{S}_0$ in (1.4) by $\mathbf{S}_0 + \lambda I_p$, a ridge-type covariance matrix estimator. Based on left-spherical distribution theory, Lauter (1996) and Lauter et al. (1998) proposed exact t and F-test for (1.1) and (1.3) under normality assumption. Power study indicates that the exact t and F-test may have no power under certain alternatives (Frick, 1996). Lopes et al. (2011a,b) suggested applying the Hotelling's T^2 test for random projection samples. Wang et al. (2014a) proposed a high-dimensional spatial sign test for the one-sample problem with heavy-tailed distribution.

1.3 Our method and its contribution

The major goal of this dissertation research is to develop new projection tests for high-dimensional one-sample and two-sample problems. Our work is motivated by searching for a projection direction that maximizes the power of projection test, which distinguishes our work from the existing ones (Lauter, 1996; Lauter et al., 1998; Lopes et al., 2011a,b).

In Chapter 3, we propose a new projection test for the one-sample problem (1.1). Let A be a $p \times k$ matrix with full column-rank $k < N$, and $\mathbf{y}_i$ obtained by projection that $\mathbf{y}_i = A^T \mathbf{x}_i$. The Hotelling's T^2 test is readily applied to the projection sample and the corresponding statistics is $T_A^2 = N\bar{\mathbf{x}}^T A^T (ASA^T)^{-1} A\bar{\mathbf{x}}$. Under normality assumption, we show an inspiring and interesting result that the optimal choice for k is 1 and the optimal direction is $\Sigma^{-1}\boldsymbol{\mu}$ in order to maximize the power of T_A^2. Moreover, the null hypothesis under this particular projection is the same as the original. The resulting test is an exact t-test under the normality assumption and an asymptotic χ^2-test with 1 degree of freedom without the normality assumption.

In practice, however, direction $\Sigma^{-1}\boldsymbol{\mu}$ is unknown and an estimator is required. To ensure the independence of the projected samples, we adopt a random partition strategy such that the original data are randomly partitioned into two separate sets: one set is used to estimate the direction and the other set is used to construct the test based on the estimated direction. We provide a thorough discussion regarding the impact of sample splitting and splitting percentage selection. We conclude that there is still considerable power gain after accounting for the sample loss due to splitting. We also include the asymptotic power comparison with the Hotelling's T^2 test and the other main competitors, and give the conditions under which the proposed method has favorite power.

To estimate the direction, we apply the ridge-like estimator $(\mathbf{S} + \lambda D_{\mathbf{S}})^{-1}$ for $\mathbf{S}^{-1}$, where λ is a turning parameter controlling the amount of penalty and $D_{\mathbf{S}}$ is a diagonal matrix with diagnoal elements equal to the diagonal of $\mathbf{S}$. The proposed ridge-like estimator keeps the test invariant to linear transformation. For implementation, we suggest that $\lambda = N_1^{-0.5}$, where N_1 is the sample size used for estimating the direction. We study the effect of λ and conclude that the performance of the test is fairly robust to the choice of λ. We have conducted several simulation experiments, which empirically show that the proposed exact t-test retains Type I error rate very well and has better power than existing methods under most of the simulation settings.

In Chapter 4, we generalize the one-sample test to its two-sample counterpart (1.3). When the two covariance matrices are equal, we conclude that the $\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is the optimal projection. The null hypothesis under this particular projection is equivalent to the original null. After the projection, the samples are only 1-dimensional such that the classical two-sample t-test can be applied. When the population is normally distributed, the resulting test is exact. The implementation is adapted from the one-sample version with the sample slitting strategy. Asymptotic power comparison is included and we give the conditions when the proposed test would outperform. For the two-sample test, we also discuss the case in which the two covariance matrices are unequal. We approach the optimal projection direction by first apply Bennett transformation to obtain a sample sequence distributed as $N(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2)$, assuming $N_1 < N_2$. In this case, we conclude the optimal direction as $\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

The major contributions of this work can be summarized as follows.

(a) We derive the theoretical optimal directions for projection tests for high-dimensional normal mean problems including both one-sample normal mean problem and two-sample normal mean problem. With the normality assumption, we construct an exact t-test for problem (1.1). The new test retains Type I error rate very well. Both theoretically and empirically, we demonstrate the proposed projection test can be much more powerful than existing ones.

(b) Without the normality assumption, we develop projection-based χ^2-tests for high-dimensional mean problems and derive its optimal direction in an asymptotical sense. The χ^2-tests have an asymptotic χ^2-distribution with 1 degree of freedom. Our numerical comparison shows that the new test retain Type I error rate well and may be more powerful than the existing tests for high-dimensional data.

1.4 Organization of this dissertation

The remainder of this dissertation consists of four parts. In Chapter 2, we provide a review on the hypothesis testing of multivariate one-sample and two-sample mean vectors. In Chapter 3, we introduce the new test procedure for testing the one-sample mean vector. With normality assumption, wee derive the theoretical optimal direction for the projection Hotelling's T^2 test, propose an effective estimation procedure to estimate the optimal direction, and further conduct the asymptotic power analysis of the proposed test statistics. Later we address the issues related to practical implementation of the proposed test and compare its empirical power with existing ones by Monte Carlo simulation study. In the end, a real data example is provided to demonstrate the application of the new test. Without normality assumption, we show that the optimal direction holds for the asymptotic χ^2-tests. In Chapter 4, we generalize the work to test the two-sample mean vectors. We first consider the normal population with equal covariance matrix. Then we generalize the test in two ways. The first is two normal populations with different covariance matrices and the other is the non-normal distribution with equal covariance matrix. We conclude the dissertation and discuss future work in Chapter 5.

Chapter 2

Literature Review

2.1 One-sample test on normal mean vector

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be identically and independently distributed p-dimensional vector from $N_p(\boldsymbol{\mu}, \Sigma)$. Denote $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)'$ such that the i-th row of X is $\mathbf{x}_i'$. The problem of interest is to test the null hypothesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \tag{2.1}$$

versus the alternative hypothesis

$$H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \tag{2.2}$$

where $\boldsymbol{\mu}_0$ is a known vector. The null hypothesis requires equality of the every corresponding element of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_0$, while the alternative indicates that at least one of the elements is different.

2.1.1 Hotelling's T^2 test and Likelihood ratio test

The Hotelling's T^2 test is the multivariate generalization of the student's t-test for one-dimensional case. The T^2 test statistic, as defined in (2.3), takes a quadratic form that evaluates a scaled distance between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}$.

$$T^2 = N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0), \tag{2.3}$$

where $\bar{\mathbf{x}}$ is the sample mean defined by

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{2.4}$$

and $\mathbf{S}$ is the sample covariance matrix defined by

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \tag{2.5}$$

Under H_0, $(N-1)\mathbf{S}$ is distributed as the Wishart distribution $W((N-1), \Sigma)$ and independent from $\bar{\mathbf{x}}$. These lead to the null distribution that

$$\frac{N-p}{(N-1)p}T^2 \sim F_{p,N-p}. \tag{2.6}$$

The Hotelling's T^2 test is uniformly the most powerful among tests with power that depends only on $\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$.

The likelihood ratio test (LRT) is equivalent to the Hotelling's T^2 test. With normality assumption, the likelihood function of observed data is

$$L(\boldsymbol{\mu}, \Sigma) \propto |\Sigma|^{-N/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{N}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T\right)\right\}. \tag{2.7}$$

Denote $A = \sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = (N-1)\mathbf{S}$ and $\mathbf{f} = \sqrt{N}A^{-1/2}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$, we show that the LRT is a monotone function of Hotelling's T^2.

$$\begin{aligned}
\lambda_{LRT} &= \frac{\max_{\Sigma} L(\boldsymbol{\mu}_0, \Sigma)}{\max_{\boldsymbol{\mu},\Sigma} L(\boldsymbol{\mu}, \Sigma)} && (2.8)\\
&= \left(\frac{|A|}{|A + N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T|}\right)^{N/2}\\
&= \left\{\frac{|A|}{|A|^{1/2}\left(|I_N + NA^{-1/2}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T A^{-1/2}\right)|A|^{1/2}}\right\}^{N/2}\\
&= \left|I_N + \mathbf{f}\mathbf{f}^T\right|^{-N/2}\\
&= \left(1 + \mathbf{f}^T\mathbf{f}\right)^{-N/2}\\
&= \left(1 + N(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T A^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)\right)^{-N/2}
\end{aligned}$$

$$= \left(1 + T^2/(N-1)\right)^{-N/2}.$$

Therefore, the Hotelling's T^2 test is equivalent to the LRT. These two tests have the same power.

2.1.2 High-dimensional setting

When $p > N$, the Hotelling's T^2 test becomes invalid due to the rank deficiency of the sample covariance matrix (2.5). Similarly, LRT$= 0$ due to the fact that $|A| = 0$ when $p > N$. Various methods have been proposed in literature for obtaining a valid test under the high-dimensional settings.

The original Dempster test (Dempster, 1958, 1960) was proposed for two-sample mean problem. We present its one-sample version given in Srivastava (2007). The general idea of the test is to evaluate the relative variation of the following two sources: (a)variation of sample mean from the null value and (b) variation of the sample points around the sample mean. To separate these two variations, consider a transformation matrix B of the form

$$B^T = (N^{-1/2}\mathbf{1}_N, \mathbf{b}_2, \ldots, \mathbf{b}_N), \tag{2.9}$$

where $\mathbf{1}_N$ is a vector of N of 1s, and $\mathbf{b}_2, \ldots, \mathbf{b}_N$ are determined such that B is an orthogonal matrix. Denote $Y = BX$ and $\mathbf{y}_i'$ as the i-row of Y. By construction, Y defines a new set of orthogonal axes with $\mathbf{y}_1$ corresponding to the overall mean and following $N(\mathbf{0}, \Sigma)$ distribution under H_0. The rest of axes, $\mathbf{y}_2, \ldots, \mathbf{y}_N$, are distributed as $N(\mathbf{0}, \Sigma)$ and independent from $\mathbf{y}_1$. The variation of the overall mean from the null value is measured by $\mathbf{y}_1^T\mathbf{y}_1$ and $\sum_{i=2}^N \mathbf{y}_i^T\mathbf{y}_i$ defines the variation scattering around the mean. Dempster test approximates $\mathbf{y}_1^T\mathbf{y}_1, \ldots, \mathbf{y}_N^T\mathbf{y}_N$ by χ_r^2, and defines the test statistic as

$$T_D = \frac{\mathbf{y}_1^T\mathbf{y}_1}{\sum_{i=2}^N \mathbf{y}_i^T\mathbf{y}_i/(N-1)}. \tag{2.10}$$

It is clear that under H_0,

$$T_D \sim F_{r,(N-1)r}. \tag{2.11}$$

Srivastava (2007) provided the power function of Dempster test, which had been previously discussed by Bai and Saranadasa (1996) in its two-sample form. Given the parameter r is known,

$$\beta_{T_D}(\boldsymbol{\mu}) \to \Phi\left(-z_\alpha + \frac{N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}\Sigma^2}}\right), \tag{2.12}$$

under conditions that $p/N \to c > 0$, $\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu} = o(\mathrm{tr}\Sigma^2/N)$, and $\lambda_{\max} = o(\sqrt{\mathrm{tr}\Sigma^2})$, where $\lambda_{\max}$ stands for the maximum eigenvalue of Σ, and z_α is upper α quantile of $N(0,1)$ distribution. Practically, the parameter r is unknown. We refer to Section 2.3.2 for the detailed discussion of different estimators. Due to the estimation of r, Dempster test is not exact. Dempster (1958) observed from simulations that the estimated significant level is well balanced around the true and the test is shown to be conservative as r gets larger.

An equivalent form of Dempster test (Srivastava, 2007) is

$$T_D = \frac{N\bar{\mathbf{x}}^T\bar{\mathbf{x}}}{\mathrm{tr}\mathbf{S}}. \tag{2.13}$$

This equivalent formulation shows that the Dempster test utilizes the trace of the sample covariance matrix to standardize the sample mean distance. The test statistic, therefore, is affected by the units of variables.

Srivastava and Du (2008) considered a test based on $\bar{\mathbf{x}}^T D_{\mathbf{S}}^{-1}\bar{\mathbf{x}}$, the standardized distance by $D_{\mathbf{S}}$ to get rid of the unit effect, where $D_{\mathbf{S}}$ is a diagonal matrix with diagonal elements from sample covariance $\mathbf{S}$. The test statistic is defined as

$$T_{SD1} = \frac{N\bar{\mathbf{x}}^T D_{\mathbf{S}}^{-1}\bar{\mathbf{x}} - (N-1)p/(N-3)}{\sqrt{2\left(\mathrm{tr}R^2 - p^2/(N-1)\right)c_{N,p}}}, \tag{2.14}$$

where $c_{N,p}$ is an adjustment coefficient for improving the convergence of T_{SD1}, which approaches 1 in probability as both N and p tend to ∞. The authors suggested

$$c_{N,p} = 1 + \frac{\mathrm{tr}(R^2)}{p^{3/2}}, \tag{2.15}$$

where $R = D_{\mathbf{S}}^{-1/2}\mathbf{S}D_{\mathbf{S}}^{-1/2}$ is the sample correlation matrix.

Given conditions $N = O(p^\zeta)$, $\frac{1}{2} < \zeta \leq 1$, $0 < \lim_{p\to\infty} \mathrm{tr}\mathcal{R}/p < \infty, i = 1,2,3,4$ and $\lim_{p\to\infty} \max_{1\leq i\leq p} \lambda_i/\sqrt{p} = 0$, where $\mathcal{R} = D_\Sigma^{-1/2}\Sigma D_\Sigma^{-1/2}$ with eigenvalues $\lambda_1 \leq$

$\ldots \leq \lambda_p$ and D_Σ is a diagonal matrix with diagonal elements from the covariance matrix Σ, it holds that under H_0,

$$T_{SD1} \to N(0,1). \tag{2.16}$$

Under local alternatives that $\boldsymbol{\mu} = (N(N-1))^{-1/2}\boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a constant vector, if for any p, $\boldsymbol{\delta}^T D_\Sigma^{-1}\boldsymbol{\delta}/p$ is bounded by a constant that does not depend on p, then the asymptotic power is given by

$$\beta_{T_{SD1}}(\boldsymbol{\mu}) \to \Phi\left(-z_\alpha + \frac{N\boldsymbol{\mu}^T D_\Sigma^{-1}\boldsymbol{\mu}}{\sqrt{2\mathrm{tr}\mathcal{R}^2}}\right), \tag{2.17}$$

where z_α is upper α quantile of $N(0,1)$ distribution.

Chen et al. (2011) introduced the regularized Hotelling's T^2 test (RHT) that

$$RHT(\lambda) = N\bar{\mathbf{x}}^T(\mathbf{S}+\lambda I)^{-1}\bar{\mathbf{x}}. \tag{2.18}$$

The following assumptions have been imposed to derive the asymptotic null distribution of RHT: (1)$p/N \to c > 0$, (2) The eigenvalues of Σ, $\lambda_1 \geq \ldots \geq \lambda_p > 0$, satisfy that $\limsup_{p\to\infty} \lambda_1 < \infty$ and $\liminf_{p\to\infty} \lambda_p > 0$; (3) The empirical spectral distribution of Σ, which is defined by $H_p(\tau) = \frac{1}{p}\sum_{j=1}^p \mathbf{1}_{[\tau_j,\infty)(\tau)}$, converges to a probability distribution function $H(\tau)$ at every point of continuity of H as $p \to \infty$. The support of $H \subset [h_1, h_2]$ is compact with $0 < h_1 \leq h_2 < \infty$. For a fixed $\lambda > 0$, the null distribution is derived with the aforementioned assumptions that

$$\frac{\sqrt{p}\,(RHT(\lambda)/p - \Theta_1(\lambda,c))}{(2\Theta_2(\lambda,c))^{1/2}} \to N(0,1), \tag{2.19}$$

where

$$\Theta_1(\lambda,c) = \frac{1-\lambda m_F(-\lambda)}{1-c(1-\lambda m_F(-\lambda))} \tag{2.20}$$

$$\Theta_2(\lambda,c) = \frac{1-\lambda m_F(-\lambda)}{(1-c(1-\lambda m_F(-\lambda)))^3} - \lambda\frac{m_F(-\lambda)-\lambda m_F'(-\lambda)}{(1-c(1-\lambda m_F(-\lambda)))^4} \tag{2.21}$$

and $m_F(z)$ is the solution to the equation that

$$m_F(z) = \int \frac{dH(\tau)}{\tau\,(1 - c - \gamma z m_F(z) - z)}. \tag{2.22}$$

Under local alternative $\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} = O(N^{-\theta})$, where $\theta \in (0, 1/2)$, the power converges to 1 as $N \to \infty$. There is no a simple form for the power function (Chen et al., 2011).

RHT is implemented via a bootstrap procedure which helps improve the performance in scenarios of complicated correlation structure and substantial missingness, but in the cost of increasing computation complexity.

Lauter (1996) and Lauter et al. (1998) proposed a family of score methods for left-spherical distributions. A $n \times p$ matrix X is left-spherically distributed if X and CX follow the same distribution for every fixed $n \times n$ orthogonal matrix C. Recall that the i-th row of X is $\mathbf{x}_i'$, and observation $\mathbf{x}_i$ and $\mathbf{x}_j$ $(i \neq j)$ are independent. If $\mathbf{x}_i \sim N(0, I)$, then X follows a left-spherical distribution by definition. Particularly, the only left-spherical distribution who satisfies the assumption of independent $\mathbf{x}_i$ and $\mathbf{x}_j$ $(i \neq j)$, is multivariate normal distribution. Under normality assumption, Lauter (1996) proposed the following procedure to construct the test for the one-sample problem. Consider linear score $\boldsymbol{z} = X\mathbf{d}(X^T X)$ such that the score weight $\mathbf{d}(X^T X)$ depends only on $X^T X$ and $\mathbf{d}(X^T X) \neq 0$ with probability 1. In this case,

$$\Gamma X \mathbf{d}(X^T X) = \Gamma X \mathbf{d}((\Gamma X)' \Gamma X) \stackrel{d}{=} X \mathbf{d}(X^T X), \tag{2.23}$$

for an orthogonal matrix Γ, where $A \stackrel{d}{=} B$ stands for that A and B follow the same distribution. By definition, $\boldsymbol{z}$ is also left-spherically distributed and

$$\sqrt{N} \bar{z} / s_z \sim t_{N-1}. \tag{2.24}$$

Specifically, Lauter (1996) proposed two ways for obtaining $\mathbf{d}(X^T X)$. The corresponding tests are SS test and PC test. SS test takes

$$\mathbf{d}_{SS} = (\mathrm{diag}(X^T X))^{-1/2}. \tag{2.25}$$

PC test takes $\mathbf{d}_{PC} = \boldsymbol{e}$ for one sided test and $|\boldsymbol{e}|$ for two sided test, where $\boldsymbol{e}$

is the eigenvector corresponding to the largest eigenvalue $\lambda_{\max}$ for the following eigenvalue problem that

$$(X^T X)\boldsymbol{e} = \mathrm{diag}(X^T X)\boldsymbol{e}\lambda_{\max}. \tag{2.26}$$

Power studies (see Logan and Tamhane (2004); Frick (1996); Kropf et al. (1997)) show that SS test attains the highest power in the situation where all variables present nearly the same relative deviation and the same correlation to each other. An especially appropriate application of PC test is with one-factor structure that $\Sigma = K + \boldsymbol{\theta}\boldsymbol{\theta}'$ and $\boldsymbol{\mu} = \boldsymbol{\theta}\delta$, where K is a diagonal matrix, $\boldsymbol{\theta}$ is a vector and δ is nonnegative scalar. Study of Frick (1996) shows that SS test and PC test are power insufficient when the alternative mean vector $\boldsymbol{\mu}$ contains at least one zero element.

PC test is an exact test by following the weights construction rules from Lauter et al. (1998), however, the projection direction associated with the largest eigenvalue may not give the highest power. To improve the performance of PC test, Liang and Tang (2009) proposed the generalized F-statistic (GF) to combine the information from all the projection directions. For each of the eigen-direction $\mathbf{d}_{PC,i}, i = 1, \dots, r$, where r is the number of the non-zero eigenvalues, the $t_{PC,i}$ are computed. GF is proposed to be $\max_{1\le i\le r}\{t^2_{PC,i}\}$. It has been shown that $t_{PC,i}s$ are asymptotically independent and GF is asymptotically distributed as $[F(x; 1, n-1)]^r$. Simulations show that GF test performs at least as well as PC test and does better when the direction is not well chosen.

2.2 One-sample test on high-dimensional mean vector

In this section, we review methods that apply to $p \ge N$ framework by imposing a factor-like model structure instead of assuming normality. This structure was first introduced in Bai and Saranadasa (1996) for the two-sample problem. Its corresponding one-sample setup can be formulated as:

$\mathbf{x}_i = \Gamma\boldsymbol{z}_i + \boldsymbol{\mu}$ for $i = 1, \dots, N$, where Γ is a $p \times m$ matrix for some $m \ge p$ such that $\Gamma\Gamma^T = \Sigma$, and $\{\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})'\}_{i=1}^N$ are m-variate independent and identically distributed random vectors satisfying $E(\boldsymbol{z}_i) = \mathbf{0}, Var(\boldsymbol{z}_i) = I_m$, and $E(z_{ik}^4) < \infty$ for $k = 1, \ldots, m$.

Following this factor-like model structure, Srivastava (2009) showed the one-sample version of Bai and Saranadasa (1996)'s test as given in (2.27), which is unscaled distance $\bar{\mathbf{x}}^T\bar{\mathbf{x}}$ with offset tr$\mathbf{S}/N$.

$$T_{BS} = \bar{\mathbf{x}}^T\bar{\mathbf{x}} - \mathrm{tr}\mathbf{S}/N. \tag{2.27}$$

Given $p \to c > 0$ and $\lambda_{\max} = o(\sqrt{\mathrm{tr}\Sigma^2})$, where $\lambda_{\max}$ is the maximum eigenvalue of Σ, if $\{\boldsymbol{z}_i\}_{i=1}^N$ satisfies that $E(\prod_{k=1}^m z_{ik}^{v_k})$ equals 0 when there is at least one $v_k = 1$ and equals 1 when there are two v_k's equal to 2, whenever $\sum_{i=1}^m v_i = 4$, then under local alternative that $\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu} = o(\mathrm{tr}\Sigma^2/N)$,

$$Var(T_{BS}) \to 2\mathrm{tr}(\Sigma^2)/(N(N-1)). \tag{2.28}$$

Under H_0, T_{BS} has mean 0. Therefore, the asymptotic null distribution is

$$\frac{T_{BS}}{\sqrt{2\mathrm{tr}(\Sigma^2)/(N(N-1))}} \to N(0,1). \tag{2.29}$$

The power function under local alternative $\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu} = o(\mathrm{tr}(\Sigma^2)/N)$ is

$$\beta_{T_{BS}}(\boldsymbol{\mu}) = \Phi\left(-z_\alpha + \frac{N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)}}\right), \tag{2.30}$$

where z_α is upper α quantile of $N(0,1)$.

By Bai and Saranadasa (1996), a consistent estimator of $\mathrm{tr}(\Sigma^2)$ is

$$\widehat{\mathrm{tr}(\Sigma^2)} = \frac{(N-1)^2}{(N-2)(N+1)}\left[\mathrm{tr}\mathbf{S}^2 - (\mathrm{tr}\mathbf{S})^2/N\right]. \tag{2.31}$$

The asymptotic power function of T_{BS} is the same as T_D^2 with known r. However, the estimation of r will cause some second order error which will can be seen from the simulation that T_{BS} slightly outperformed T_D^2. Another advantage of T_{BS} over T_D^2 is that T_{BS} does not assume the normality.

Chen and Qin (2010) refined Bai and Saranadasa (1996)'s two-sample test. We present its one-sample version here. The work is inspired by the observation that both the term $\sum_{i=1}^{N} \mathbf{x}_i^T \mathbf{x}_i$ in calculating $\bar{\mathbf{x}}^T \bar{\mathbf{x}}$ and the term tr$\mathbf{S}$ that is used to offset $\bar{\mathbf{x}}^T \bar{\mathbf{x}}$ impose a restricted condition that p and N should be of the same order. The refined statistic takes the simple form that

$$T_{CQ} = \frac{1}{N(N-1)} \sum_{i \neq j}^{N} \mathbf{x}_i^T \mathbf{x}_j. \tag{2.32}$$

Given $\mathrm{tr}(\Sigma^4) = o\{\mathrm{tr}^2(\Sigma^2)\}$, if $\{\boldsymbol{z}_i\}_{i=1}^N$ satisfies that $E(\prod_{k=1}^q z_{il_k}^{\alpha_k}) = \prod_{k=1}^q E(z_{il_k}^{\alpha_k})$ for a positive integer q such that $\sum_{l=1}^q \alpha_l \leq 8$ and $l_1 \neq \ldots \neq l_q$, then under local alternative that $\boldsymbol{\mu}^T \Sigma \boldsymbol{\mu} = o(\mathrm{tr}(\Sigma^2)/N)$,

$$Var(T_{CQ}) \to 2\mathrm{tr}(\Sigma^2)/(N(N-1)). \tag{2.33}$$

Under H_0, T_{CQ} has mean 0. The asymptotic null distribution is

$$\frac{T_{CQ}}{\sqrt{2\mathrm{tr}(\Sigma^2)/(N(N-1))}} \to N(0,1). \tag{2.34}$$

The power function under local alternative $\boldsymbol{\mu}^T \Sigma \boldsymbol{\mu} = o(\mathrm{tr}(\Sigma^2)/N)$ is

$$\beta_{T_{CQ}}(\boldsymbol{\mu}) = \Phi\left(-z_\alpha + \frac{N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)}}\right), \tag{2.35}$$

where z_α is upper α quantile of $N(0,1)$.

The order of p and N is not explicated controlled. Instead, $\mathrm{tr}(\Sigma^4) = o\{\mathrm{tr}^2(\Sigma^2)\}$ is used to regularize the growth of p. A sufficient conditions is $\lambda_p = o\{(p-b)^{1/2}\lambda_1 b^{-1/4}\}$ or $\lambda_p = o\{(p-b)^{1/4}\lambda_1^{1/2}\lambda_{p-b+1}^{1/2}\}$, where $\lambda_1 \leq \ldots, \leq \lambda_p$ are eigenvalues of Σ; b is the number of unbounded eigenvalues such that $(p-b) \to \infty$ and $(p-b)\lambda_1^2 \to \infty$. In this special case, the number of divergent eigenvalues of Σ are not too many and the divergence is not too fast.

The estimator for $\mathrm{tr}(\Sigma^2)$ given in (2.36) is adapted from Bai and Saranadasa (1996) by the same motivation of excluding the term $\sum_{i=1}^{N} \mathbf{x}_i^T \mathbf{x}_i$.

$$\widehat{\mathrm{tr}(\Sigma^2)} = \frac{\mathrm{tr}\left(\sum_{j\neq k}^{N} (\mathbf{x}_j - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_j^T (\mathbf{x}_k - \bar{\mathbf{x}}_{(j,k)})\mathbf{x}_k^T\right)}{N(N-1)}, \tag{2.36}$$

where $\bar{\mathbf{x}}_{(j,k)}$ is the sample mean after excluding $\mathbf{x}_j$ and $\mathbf{x}_k$ and $\bar{\mathbf{x}}_{(l)}$ is the sample mean excluding $\mathbf{x}_l$.

T_{BS} and T_{CQ} are not invariant tests for the units of variables. Srivastava (2009) proposed a test T_{SD2}, which is free of scale by standardizing variables with their corresponding variances that

$$T_{SD2} = \bar{\mathbf{x}}^T (D_{\mathbf{S}}/N)^{-1} \bar{\mathbf{x}} - \frac{(N-1)p}{N-3}. \tag{2.37}$$

Under conditions that $0 < \zeta \leq 1$, $\lim_{p\to\infty} \mathrm{tr}\mathcal{R}^i/p < \infty$ for $i = 1, 2, 3, 4$, where $\mathcal{R} = D_\Sigma^{-1/2} \Sigma D_\Sigma^{-1/2}$ and D_Σ is a diagonal matrix with diagonal elements from the covariance matrix Σ,

$$Var(T_{CQ}) \to 2\mathrm{tr}\mathcal{R}^2. \tag{2.38}$$

Moreover, under H_0,

$$\frac{T_{SD2}}{\sqrt{2\mathrm{tr}\mathcal{R}^2}} \to N(0,1). \tag{2.39}$$

A consistent estimator of $\mathcal{R}^2$ is

$$\mathrm{tr}R^2 - p^2(N-1)^{-1}, \tag{2.40}$$

where $R = D_{\mathbf{S}}^{-1/2} \mathbf{S} D_{\mathbf{S}}^{-1/2}$ is the sample correlation matrix.

The test statistic looks the same as the one proposed by Srivastava and Du (2008) under normality assumption, up to an adjustment coefficient $c_{p,N}$. The condition $\frac{1}{2} < \zeta \leq 1$ is used to guarantee that the adjustment coefficient converge to 1. Srivastava (2009) removed the adjustment coefficient and therefore sets the condition $0 < \zeta \leq 1$.

2.3 Two-sample test on normal mean vectors

Suppose that $\mathbf{x}_{ij} \sim N_p(\boldsymbol{\mu}_i, \Sigma_i), j = 1, \ldots, N_i, i = 1, 2$ are two independent random samples. Denote $X_1 = (\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1N_1})^T$ and $X_2 = (\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2N_2})^T$ such that X_1 and X_2 are independent. For convenience, also denote $N = N_1 + N_2$ and $n = N_1 + N_2 - 2$.

Of interest for the two-sample mean problem is to test the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \tag{2.41}$$

versus the alternative hypothesis

$$H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{2.42}$$

2.3.1 Hotelling's T^2 test

Under classical framework $n > p$, the Hotelling's T^2 test is readily applicable for the case that $\Sigma_1 = \Sigma_2$. The test statistic is defined as

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \left(\mathbf{S} \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \tag{2.43}$$

where S is the pooled sample variance that

$$S = \frac{1}{n} \sum_{i=1}^{2} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T. \tag{2.44}$$

Under H_0, it has been shown that,

$$\frac{n-p+1}{np} T^2 \sim F_{p,n-p+1}. \tag{2.45}$$

If $\Sigma_1 \neq \Sigma_2$, but $N_1 = N_2$, a new sequence of $\boldsymbol{z}_j = \mathbf{x}_{1j} - \mathbf{x}_{2j}$ can be constructed, on which the one-sample Hotelling's T^2 test could be applied. If $\Sigma_1 \neq \Sigma_2$ and $N_1 \neq N_2$, it is known as the Behrens-Fisher problem for univariate case. In this case, the Hotelling's T^2 test is not applicable. Assume $N_1 < N_2$, we apply the method proposed by Scheffe (1943) and generalized to multivariate case by

Bennett (1950) to obtain an one sample sequence of size N_1 that is distributed as $N\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)$. Define

$$\boldsymbol{z}_i = \mathbf{x}_{1i} - \sqrt{\frac{N_1}{N_2}}\mathbf{x}_{2i} + \frac{1}{\sqrt{N_1 N_2}}\sum_{j=1}^{N_1}\mathbf{x}_{2j} - \frac{1}{N_2}\sum_{k=1}^{N_2}\mathbf{x}_{2k}, \quad i = 1, \dots, N_1. \tag{2.46}$$

By construction, $\{\boldsymbol{z}_i\}_{i=1}^{N_1}$ follows the desired $N\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)$ distribution. The one-sample Hotelling's T^2 test could be applied on $\{\boldsymbol{z}_i\}_{i=1}^{N_1}$.

In Section 2.3.2 reviewing the high-dimensional case, we assume $\Sigma_1 = \Sigma_2$ unless otherwise specified.

2.3.2 High-dimensional setting

When $p > n$, the Hotelling's T^2 is not well-defined due to the singular sample covariance matrix (2.44). Even for $p < n$, the Hotelling T^2 could have poor performance if p grows the same order as n. Consider the case that $p/N \to c \in (0,1)$, $N_1/N \to \kappa \in (0,1)$. The asymptotic power of Hotelling's T^2 with local alternative $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(1)$ is,

$$\beta_{T^2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \to \Phi\left(-z_\alpha + \sqrt{\frac{(1-c)}{2c}}\kappa(1-\kappa)\left\{\sqrt{n}\|\Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\|^2\right\}\right), \tag{2.47}$$

where z_α is upper α quantile of $N(0,1)$ distribution (Bai and Saranadasa, 1996). With usual consideration that $\sqrt{n}\|\Sigma^{-\frac{1}{2}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\|^2$ converges to a positive constant, (2.47) reveals that power of the Hotelling's T^2 test increases slowly with increase of the noncentral parameter when the c is close to 1.

Dempster(1958, 1960) proposed T_D test, based on the comparison of variation between two sample mean vectors, and the variation scattering around the mean vectors. Recall that $X_1 = (\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1N_1})^T$, $X_2 = (\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2N_2})^T$ and X_1 is independent from X_2. Denote that

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\boldsymbol{\mu}, I_N \bigotimes \Sigma\right), \quad \boldsymbol{\mu} = \begin{pmatrix} \mathbf{1}_{N_1}\boldsymbol{\mu}_1' \\ \mathbf{1}_{N_2}\boldsymbol{\mu}_2' \end{pmatrix}. \tag{2.48}$$

The sample defines a set of orthogonal axes in a Euclidean space of $N_1 + N_2$ dimensions. Consider an orthogonal transformation for X such that the new set of orthogonal dimensions includes one representing the overall mean, one depicting the mean difference, and the rests as between-sample variation. To achieve this, the transformation matrix B can be constructed in the following way that

$$\begin{aligned} B &= (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N)' \qquad (2.49) \\ \mathbf{b}_1 &= N^{-1/2}\mathbf{1}_N, \\ \mathbf{b}_2 &= \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1/2} \left(\frac{1}{N_1}\mathbf{1}_{N_1}, -\frac{1}{N_2}\mathbf{1}_{N_2}\right), \end{aligned}$$

and the rest of $\mathbf{b}_j, j = 3, \ldots, N$ are chosen such that B is an orthonormal matrix.

Let $Y = BX$ and $\mathbf{y}_i'$ be the ith row of Y. By construction, $\mathbf{y}_1$ represents the grand mean with $E(\mathbf{y}_1) = (N_1\boldsymbol{\mu}_1 + N_2\boldsymbol{\mu}_2)/N$, and $\mathbf{y}_2$ represents the difference between the two sample means with $E(\mathbf{y}_2) = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. It follows that $\mathbf{y}_3, \ldots, \mathbf{y}_N$ are distributed as $N(\mathbf{0}, \Sigma)$ and independent of $\mathbf{y}_1$ and $\mathbf{y}_2$, where $\mathbf{y}_2$ is distributed as $N(\mathbf{0}, \Sigma)$ under H_0. Following this thread, T_D test is defined as

$$T_D = \frac{\mathbf{y}_2^T\mathbf{y}_2}{\sum_{i=3}^N \mathbf{y}_i^T\mathbf{y}_i/n}. \qquad (2.50)$$

As quadratic form of normal distribution, $\mathbf{y}_i^T\mathbf{y}_i$ can be approximated by χ_r^2, which gives $T_D \sim F_{r,nr}$ under H_0.

Dempster (1958) proposes two estimators of r that both apply the method of moment to an approximated distribution obtained by matching moments of asymptotic expansions. The first method approximates

$$n\log\left(\frac{1}{n}\sum_{i=3}^N \mathbf{y}_i^T\mathbf{y}_i\right) - \sum_{i=3}^N \log(\mathbf{y}_i^T\mathbf{y}_i) \qquad (2.51)$$

with

$$\left(\frac{1}{r} + \frac{1 + n^{-1}}{3r^2}\right)\chi_{n-1}^2, \qquad (2.52)$$

which gives $\hat{r}$ as the solution of equation

$$n\log\left(\frac{1}{n}\sum_{i=3}^{N}\mathbf{y}_i^T\mathbf{y}_i\right)-\sum_{i=3}^{N}\log(\mathbf{y}_i^T\mathbf{y}_i)=\left(\frac{1}{r}+\frac{1+n^{-1}}{3r^2}\right)(n-1). \tag{2.53}$$

The second method combines further the information of angles between $\mathbf{y}_i$ for $\mathbf{y}_j$, denoted by θ_{ij} with $3 \le i < j \le N$. It approximates

$$-\log(\sin^2(\theta_{ij})) \tag{2.54}$$

by

$$\left(\frac{1}{r}+\frac{3}{2r^2}\right)\chi_1^2. \tag{2.55}$$

The second estimator can be obtained from

$$\begin{aligned} n\log\left(\frac{1}{n}\sum_{i=3}^{N}\mathbf{y}_i^T\mathbf{y}_i\right)-\sum_{i=3}^{N}\log(\mathbf{y}_i^T\mathbf{y}_i)-\sum_{3\le i<j\le N}\log(\sin^2(\theta_{ij})) \\ = \left(\frac{1}{r}+\frac{1+n^{-1}}{3r^2}\right)(n-1)+\left(\frac{1}{r}+\frac{3}{2r^2}\right)\binom{n}{2}. \end{aligned} \tag{2.56}$$

Bai and Saranadasa (1996) suggested a new estimator of r by showing that

$$r = \frac{(\mathrm{tr}\Sigma)^2}{\mathrm{tr}\Sigma^2}=\frac{p(\mathrm{tr}\Sigma/p)^2}{\mathrm{tr}\Sigma^2/p},$$

and obtaining a consistent estimator of $\mathrm{tr}\Sigma^2$ given in (2.57). See also Srivastava (2007) under conditions that $0<\lim_{p\to\infty}\mathrm{tr}\Sigma/p<\infty, i=1,2,3,4,$

$$\widehat{\mathrm{tr}(\Sigma^2)}=\frac{n^2}{(n-1)(n+2)}\left[\mathrm{tr}\mathbf{S}^2-(\mathrm{tr}\mathbf{S})^2/n\right]. \tag{2.57}$$

The resulting estimator of r is

$$\hat{r}=\frac{(\mathrm{tr}S)^2}{\frac{n^2}{(n-1)(n+2)}\left(\mathrm{tr}S^2-(\mathrm{tr}S)^2/n\right)}. \tag{2.58}$$

Under local alternative $(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\Sigma(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)=o((\frac{1}{N_1}+\frac{1}{N_2})\mathrm{tr}\Sigma^2)$, Bai and Saranadasa (1996) showed that if $p/n\to c>0, N_1/N\to\kappa\in(0,1)$, and $\lambda_{\max}=$

$o(\sqrt{\text{tr}\Sigma^2})$, where $\lambda_{\max}$ is the maximum eigenvalue of Σ, then the asymptotic power function of T_D for a known parameter r is

$$\beta_{T_D}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Phi\left(-\xi_\alpha + \frac{n\kappa(1-\kappa)\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{\sqrt{2\text{tr}\Sigma^2}}\right), \tag{2.59}$$

where z_α is upper α quantile of $N(0,1)$ distribution.

Srivastava and Du (2008) presented an equivalent form for the two-sample T_D test that

$$T_D = \left(\frac{1}{N_1} + \frac{1}{N_2}\right)^{-1} \frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\text{tr}\mathbf{S}}. \tag{2.60}$$

The equivalent formulation shows that the Dempster test utilizes the trace of the sample covariance matrix to standardize the distance between sample means.

Srivastava and Du (2008) proposed a test that standardizes the distance with $D_{\mathbf{S}}$, a diagonal matrix with diagonal elements from the sample covariance $\mathbf{S}$. The test statistic is

$$T_{SD1} = \frac{(\frac{1}{N_1} + \frac{1}{N_2})^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T D_{\mathbf{S}}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - np(n-2)^{-1}}{\sqrt{2\left(trR^2 - p^2n^{-1}\right)c_{p,n}}}, \tag{2.61}$$

where $c_{n,p}$ is an adjustment coefficient for improving the convergence of T_{SD1}. It is suggested that

$$c_{n,p} = 1 + \frac{\text{tr}(R^2)}{p^{3/2}}, \tag{2.62}$$

where $R = D_{\mathbf{S}}^{-1/2}\mathbf{S}D_{\mathbf{S}}^{-1/2}$ is the sample correlation matrix.

Given that $n = O(p^\zeta)$ with $\frac{1}{2} < \zeta \leq 1$ and $N_1/N \to \kappa \in (0,1)$, if $0 < \lim_{p\to\infty} \text{tr}\mathcal{R}^i/p < \infty, i = 1,2,3,4$ and $\lim_{p\to\infty} \max_{1\leq i\leq p} \lambda_i/\sqrt{p} = 0$, where $\mathcal{R} = D_\Sigma^{-1/2}\Sigma D_\Sigma^{-1/2}$ and λ_is are the eigenvalues of $\mathcal{R}$, then under H_0,

$$T_{SD1} \to N(0,1). \tag{2.63}$$

Furthermore, under local alternative that $n(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T D_\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)/(p(\frac{1}{N_1} + \frac{1}{N_2}))$ is bounded by a constant that does not depend on p for all the p, the asymptotic power can be obtained by

$$\beta_{T_{SD1}}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \to \Phi\left(-z_\alpha + \frac{N_1N_2}{N_1+N_2}\frac{(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T D_\Sigma^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)}{\sqrt{2\text{tr}\mathcal{R}^2}}\right), \tag{2.64}$$

where z_α is upper α quantile of $N(0,1)$ distribution.

Chen et al. (2011) introduced regularized Hotelling's T^2 test (RHT) by replacing $\mathbf{S}^{-1}$ with the ridge-like estimator that

$$RHT(\lambda) = N(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T(\mathbf{S} + \lambda I)^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \tag{2.65}$$

The asymptotic properties are derived at the same conditions as one-sample test with one more constraint that $N_1/N \to \kappa \in (0,1)$. For local alternative $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = O(n^{-\theta})$ with $\theta \in (0, 1/2)$, the power converges to 1 as $n \to \infty$.

To better cope with the complicated correlation structure, substantial missingness, and small sample size problem of proteomics data, Chen et al. (2011) proposed the bootstrap procedures to implement RHT. See paper for the detailed procedure. We note that the bootstrap procedure requires very heavy computation.

Lopes et al. (2011a) discussed the idea of reducing the dimension by randomly projecting the sample to a low-dimensional space such that the Hotelling's T^2 test is applicable. Following this idea, Lopes et al. (2011a) proposed a random projection matrix whose entries are randomly drawn from $N(0,1)$ distribution. The original null hypothesis is rejected if the null hypothesis on the projection space is rejected. The procedure results in an exact test for multivariate normal distribution. However, the direction obtained by a single projection generated from a random process may not give desired power. Lopes et al. (2011b) proposed a test which utilizes multiple projections that could potentially increase the chance of getting higher power by combining information from multiple draws. The idea is to project the data to different k-dimensional spaces ($k < n$), perform the classical Hotelling's T^2 test for each projection and build the test statistic based on the average over the ensemble of projection matrices. Let $P_k \in R^{p\times k}$ be a projection matrix that projects data from R^p to R^k. The RP test is defined by

$$T_{RP} = \frac{N_1 N_2}{N_1 + N_2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T E_{P_k}[P_k(P_k^T\hat{\Sigma}P_k)^{-1}P_k^T](\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \tag{2.66}$$

where $E_{P_k}[P_k(P_k^T\hat{\Sigma}P_k)^{-1}P_k^T]$ is a surrogate for $\mathbf{S}^{-1}$ and is estimated by the average of $P_k(P_k^T\hat{\Sigma}P_k)^{-1}P_k^T$ over the ensemble of P_k.

If $N_1/N \to \kappa \in (0,1)$ and $k/n = a + o(1/\sqrt{n})$ for $a \in (0,1)$, then under H_0,

$$\frac{T_{RP} - \bar{\mu}_n}{\bar{\sigma}_n} \to N(0,1), \tag{2.67}$$

where

$$\bar{\mu}_n = \frac{k}{1-k/n}, \quad \bar{\sigma}_n = \sqrt{\frac{2k}{(1-k/n)^3}}. \tag{2.68}$$

With local alternative assumption that $(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = o(1)$, the power function satisfies that

$$\beta_{T_{RP}}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) \to \Phi\left(-z_\alpha + \kappa(1-\kappa)\sqrt{\frac{1-a}{2a}}\bar{\Delta}_k\sqrt{n}\right), \tag{2.69}$$

where z_α is upper α quantile of $N(0,1)$ distribution and

$$\bar{\Delta}_k = E_{P_k}[(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T [P_k(P_k^T \hat{\Sigma} P_k)^{-1} P_k^T](\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)]. \tag{2.70}$$

Two practical issues are discussed for implementation: the choice of k and the number of projections. Lopes et al. (2011b) suggests an optimal choice of k as $k = \lfloor n/2 \rfloor$ under certain conditions. The number of projections can be determined by increasing the copies of projection until the fluctuation of p-value becomes negligible. Simulation studies show that 30 is sufficient to obtain the stable results.

Lauter (1996) and Lauter et al. (1998) proposed a family of score-based tests with weight function designed in the manner such that the resulting score follows a left-spherical distribution. Recall the distribution of X as shown in (2.48) and the transformation matrix B as defined in (2.49). The null hypothesis that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is equivalent to $\boldsymbol{\mu} = \mu \mathbf{b}_1$ where μ is a constant and the deviation from the null hypothesis can be measured by the contrast $\mathbf{b}_2' \boldsymbol{\mu}$. For the weight function $\mathbf{d}$, Lauter (1996) proposed that $\mathbf{d}$ should be determined by $(X - \bar{X})'(X - \bar{X})$ only and $\mathbf{d} \neq 0$ with probability 1. Define linear score $\boldsymbol{z}$ by $\boldsymbol{z} = X\mathbf{d}$, then the test can be carried out by the fact that $\mathbf{b}_2^T \boldsymbol{z}/s_{\boldsymbol{z}}$ follows exact t_n distribution, where

$$s_{\boldsymbol{z}}^2 = \boldsymbol{z}'(I_N - \mathbf{b}_1\mathbf{b}_1' - \mathbf{b}_2\mathbf{b}_2')\boldsymbol{z}/n. \tag{2.71}$$

2.4 Two-sample test on high-dimensional mean vectors

In this section, we review several methods that are applicable to high-dimensional data with normality assumption substituted by a factor-like model structure. The structure is set up in the way that

> $\mathbf{x}_{ij} = \Gamma_i \boldsymbol{z}_{ij} + \boldsymbol{\mu}_i$ for $i = 1, 2, j = 1, \ldots, N_i$, where each Γ_i is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i \Gamma_i' = \Sigma_i$, and $\{\boldsymbol{z}_{ij} = (z_{ij1}, \ldots, z_{ijm})\}_{j=1}^{N_i}$ are m-variate independent and identically distributed random vectors satisfying $E(\boldsymbol{z}_{ij}) = \mathbf{0}, Var(\boldsymbol{z}_{ij}) = I_m$, and $E(z_{ijk}^4) < \infty$, $k = 1, \ldots, m$.

Bai and Saranadasa (1996) proposed a test based on the measure of unscaled distance that

$$T_{BS} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \left(\frac{1}{N_1} + \frac{1}{N_2}\right) \mathrm{tr}\mathbf{S}, \tag{2.72}$$

where T_{BS} has mean 0 under H_0.

Under conditions $p/n \to c > 0$, $N_1/N \to \kappa \in (0, 1)$, $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o((\frac{1}{N_1} + \frac{1}{N_2})\mathrm{tr}\Sigma^2)$ and $\lambda_{\max} = o(\sqrt{\mathrm{tr}\Sigma^2})$, if $\{\boldsymbol{z}_{ij}\}_{j=1}^{N_i}, i = 1, 2$ satisfy that, whenever $\sum_{i=1}^m v_i = 4$, $E(\prod_{k=1}^m z_{ijk}^{v_k})$ equals 0 when there is at least one v_k equals 1 and equals 1 when there are two v_k's equal to 2, then

$$Var(T_{BS}) \to 2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)^2 \left(1 + \frac{1}{n}\right) \mathrm{tr}(\Sigma^2). \tag{2.73}$$

Therefore, under H_0,

$$\frac{T_{BS}}{\sqrt{Var(T_{BS})}} \to N(0, 1). \tag{2.74}$$

Under local alternative $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o((\frac{1}{N_1} + \frac{1}{N_2})\mathrm{tr}\Sigma^2)$, the asymptotic power function of T_{BS} test is

$$\beta(T_{BS}) \to \Phi\left(-z_\alpha + \frac{n\kappa(1-\kappa)\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{\sqrt{2\mathrm{tr}\Sigma^2}}\right), \tag{2.75}$$

where z_α is the upper α quantile of $N(0,1)$.

Bai and Saranadasa (1996) provided a consistent estimator of $\mathrm{tr}(\Sigma^2)$ as

$$\widehat{\mathrm{tr}(\Sigma^2)} = \frac{n^2}{(n-1)(n+2)}\left[\mathrm{tr}\mathbf{S}^2 - (\mathrm{tr}\mathbf{S})^2/n\right]. \tag{2.76}$$

T_{BS} requires $\Sigma_1 = \Sigma_2$ and the p and n should of the same order. Chen and Qin (2010) examined T_{BS} and pointed out that the term $\sum_{j=1}^{N_i}\mathbf{x}_{ij}^T\mathbf{x}_{ij}, i = 1,2$ in $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ is not helpful for the testing and has to offset with $\mathrm{tr}\mathbf{S}$. Both $\sum_{j=1}^{N_i}\mathbf{x}_{ij}^T\mathbf{x}_{ij}$ and $\mathrm{tr}\mathbf{S}$ have to be controlled by imposing the condition that p and n should be of the same order and $\lambda_{\max} = o(\sqrt{\mathrm{tr}\Sigma^2})$. Accordingly, Chen and Qin (2010) proposed to use

$$T_{CQ} = \frac{\sum_{i\neq j}^{N_1}\mathbf{x}_{1i}^T\mathbf{x}_{1j}}{N_1(N_1-1)} + \frac{\sum_{i\neq j}^{N_2}\mathbf{x}_{2i}^T\mathbf{x}_{2j}}{N_2(N_2-1)} - \frac{2\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\mathbf{x}_{1i}^T\mathbf{x}_{2j}}{N_1N_2} - \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2, \tag{2.77}$$

which satisfies that $E(T_{CQ}) = 0$ under H_0.

Under conditions $N_1/N \to k \in (0,1)$, $\mathrm{tr}(\Sigma_i\Sigma_j\Sigma_l\Sigma_h) = o[\mathrm{tr}^2\{(\Sigma_1+\Sigma_2)^2\}]$ for $i,j,l,h = 1,2$ and local alternative $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\Sigma_i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(\mathrm{tr}((\Sigma_1+\Sigma_2)^2)/N)$ for $i = 1,2$, if $\{\boldsymbol{z}_{ij}\}_{j=1}^{N_i}, i = 1,2$ satisfy that $E(\prod_{l=1}^q z_{ijk_l}^{\alpha_l}) = \prod_{l=1}^q E(z_{ijk_l}^{\alpha_l})$ for a positive integer q such that $\sum_{l=1}^q \alpha_l \leq 8$ and $k_1 \neq \ldots \neq k_q$,

$$Var(T_{CQ}) \to \frac{2}{N_1(N_1-1)}\mathrm{tr}(\Sigma_1^2) + \frac{2}{N_2(N_2-1)}\mathrm{tr}(\Sigma_2^2) + \frac{4}{N_1N_2}\mathrm{tr}(\Sigma_1\Sigma_2). \tag{2.78}$$

Therefore, under H_0,

$$\frac{T_{CQ}}{\sqrt{Var(T_{CQ})}} \to N(0,1). \tag{2.79}$$

Under local alternative $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \Sigma_i (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = o(\mathrm{tr}((\Sigma_1 + \Sigma_2)^2)/N)$ for $i = 1, 2$, the asymptotic power function of T_{CQ} is

$$\beta_{T_{CQ}}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Phi\left(-z_\alpha + \frac{nk(1-k)\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2}{\sqrt{2\mathrm{tr}\tilde{\Sigma}(k)^2}}\right), \tag{2.80}$$

where $\tilde{\Sigma}(k) = (1-k)\Sigma_1 + k\Sigma_2$ and z_α is upper α quantile of $N(0,1)$ distribution.

The estimator for the terms $\mathrm{tr}(\Sigma_1^2)$, $\mathrm{tr}(\Sigma_2^2)$ and $\mathrm{tr}(\Sigma_1\Sigma_2)$ are adapted from Bai and Saranadasa (1996) under the same motivation of excluding the term $\sum_{j=1}^{N_i} \mathbf{x}_{ij}^T \mathbf{x}_{ij}, i = 1, 2$. The refined estimators are defined as

$$\begin{aligned} \widehat{\mathrm{tr}(\Sigma_i^2)} &= \frac{\mathrm{tr}\left(\sum_{j\neq k}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{i(j,k)})\mathbf{x}_{ij}^T (\mathbf{x}_{ik} - \bar{\mathbf{x}}_{i(j,k)})\mathbf{x}_{ik}^T\right)}{N_i(N_i - 1)}, \quad i = 1, 2 & (2.81) \\ \mathrm{tr}\widehat{(\Sigma_1\Sigma_2)} &= \frac{\mathrm{tr}\left(\sum_{l=1}^{N_1}\sum_{k=1}^{N_2} (\mathbf{x}_{1l} - \bar{\mathbf{x}}_{1(l)})\mathbf{x}_{1l}^T (\mathbf{x}_{2k} - \bar{\mathbf{x}}_{2(k)})\mathbf{x}_{2k}^T\right)}{N_1 N_2}, & (2.82) \end{aligned}$$

where $\bar{\mathbf{x}}_{i(j,k)}$ is the ith sample mean after excluding $\mathbf{x}_{ij}$ and $\mathbf{x}_{ik}$ and $\bar{\mathbf{x}}_{i(l)}$ is the ith sample mean without $\mathbf{x}_{il}$. All these three estimators are consistent that

$$\frac{\widehat{\mathrm{tr}(\Sigma_1^2)}}{\mathrm{tr}(\Sigma_1^2)} \to 1, \quad \frac{\widehat{\mathrm{tr}(\Sigma_2^2)}}{\mathrm{tr}(\Sigma_2^2)} \to 1, \quad \text{and} \quad \frac{\mathrm{tr}\widehat{(\Sigma_1\Sigma_2)}}{\mathrm{tr}(\Sigma_1\Sigma_2)} \to 1. \tag{2.83}$$

Chapter 3

Projection Test for High-dimensional One-sample Mean Problem

3.1 Background and related work

Let $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be a random sample from a p-variate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$. In the one-sample problem, it is of interest to test

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0. \tag{3.1}$$

Without loss of generality, we assume $\boldsymbol{\mu}_0 = \mathbf{0}$ throughout the rest of this dissertation. The classical solution for the one-sample problem is the Hotelling's T^2 test which is defined as

$$T^2 = N\bar{\mathbf{x}}^T\mathbf{S}^{-1}\bar{\mathbf{x}}, \tag{3.2}$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}^{-1}$ are defined in (2.4) and (2.5), respectively. When the dimension is fixed and small compared to the sample size, the Hotelling's T^2 test is theoretically grounded. Denote $\zeta = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$. It is well known that $(N-p)T^2/(N-1)p$ follows an $F_{p,N-p}(N\zeta)$ distribution, the noncentral F-distribution with $(p, N-p)$ degrees of freedom and noncentrality parameter $N\zeta$.

However, the scaled distance utilized in (3.2) requires taking inverse of the sample covariance matrix, which is rank-deficient when the dimension exceeds the sample size and renders the Hotelling's T^2 test undefined.

Various approaches have been developed to take “inverse” of the sample covariance matrix and provide substitutes for the Hotelling’s T^2 test in scenarios with high-dimensional data. Dempster (1958, 1960) used the trace of sample covariance matrix; Bai and Saranadasa (1996) and Chen and Qin (2010) proposed identity matrix; Srivastava and Du (2008) and Srivastava (2009) suggested the diagnoal matrix of sample covariance matrix; and Chen et al. (2011) adopted the ridge-like estimator for the inverse of $\mathbf{S}$.

From a different point of view, Lopes et al. (2011a,b); Lauter (1996); Lauter et al. (1998) approached the problem by considering projecting the sample to a low-dimensional space which makes construction of exact tests possible. Lopes et al. (2011a) proposed a single projection matrix with entries randomly generated from N(0,1) distribution. To achieve higher power, Lopes et al. (2011b) suggested averaging over the multiple random projections. Lauter (1996) designed the projection weight matrix in a delicate way such that the resulting score reaches left-spherical distributions. Exact tests follow by utilizing the properties of left-spherical distributions. Despite intuitive and exact, Lauter’s tests are power insufficient (Frick, 1996) and provide few clues for achieving higher power. Besides, the equivalence of null hypothesis between the original and projected data is often ignored, which tends to weaken the usefulness of this projection method.

The new method is motivated by the aforementioned projection-based tests, in recognition of its capability to obtain an exact test by reducing the dimension to a manageable level. Our new projection tests are distinguished from the existing ones (Lauter, 1996, Lauter, *et al.*, 1998, Lopes, *et al.*, 2011, 2012) in that the proposed tests are based on data-driven estimation of the optimal projection direction. During our search for the optimal projection direction, we take the following aspects into consideration: (a) The test is exact with normality assumption such that the Type I error can be well-controlled. (b) The test should be given an explicit mechanism to achieve optimal power; this ensures the reliability of the projection test so that there is no longer concern about “luck” when conducting the test. (c) The test should do the correct job, meaning that the projection test is working on the same pair of hypotheses as the original data.

3.2 New test based on normal population

In this section, we focus on the high-dimensional one-sample problem (3.1) under the normality assumption, to get insights into the projection test. In Section 3.4, we will discuss extension of the proposed test to the situation without the normality assumption.

3.2.1 Optimal projection direction

To develop the projection test, we consider a matrix $A_{p\times k}$ of rank k such that $k \ll p$ and $k < N$. For each sample, we project the original data $\mathbf{x}_i$ to A to obtain the $\mathbf{y}_i$ with a reduced dimension that $\mathbf{y}_i = A^T\mathbf{x}_i$, for $i \in 1, \ldots, N$.

The projected sample $\mathbf{y}_1, \cdots, \mathbf{y}_N$ are independent and identically distributed according to $N_k(A^T\boldsymbol{\mu}, A^T\Sigma A)$. To test the H_0: $E(\mathbf{y}_i) = \mathbf{0}$, define the projection Hotelling's T^2 test to be

$$T_A^2 = N\bar{\mathbf{x}}^T A(A^T\mathbf{S}A)^{-1}A^T\bar{\mathbf{x}}, \tag{3.3}$$

which is the Hotelling's T^2 test based on the $\mathbf{y}_i$'s. We note that the test statistic in (3.3) is essentially testing

$$H_{0A} : \boldsymbol{\mu}_A = \mathbf{0} \quad \text{vs} \quad H_{0A} : \boldsymbol{\mu}_A \neq \mathbf{0}, \tag{3.4}$$

where $\boldsymbol{\mu}_A = A^T\boldsymbol{\mu}$. Obviously, H_{0A} does not imply H_0 in (3.1) in general. Our search of the optimal direction under a power-maximization framework basically solves two critical issues that how to determine k and how to choose A for a given k so that T_A^2 maximizes the power under H_1 in (3.1). Theorem 3.2.1 provides an elegant solution to these issues.

Theorem 3.2.1. *Suppose that $\mathbf{x}_1, \cdots, \mathbf{x}_N$ is a random sample from $N_p(\boldsymbol{\mu}, \Sigma)$. Let A be a $p \times k$ full column-rank matrix. For a fixed projection rank k, the projection test T_A^2 reaches its best power for H_1 with*

$$A = \Sigma^{-\frac{1}{2}}W, \tag{3.5}$$

where $W = (\mathbf{w}_1, \ldots, \mathbf{w}_k)$ *such that*

$$\mathbf{w}_1 = \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}/\sqrt{\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}}, \tag{3.6}$$

and $\mathbf{w}_2, \ldots, \mathbf{w}_d$ *are taken to be orthogonal to* $\mathbf{w}_1$*. Moreover, the* $k = 1$ *gives the optimal projection rank under which the optimal projection direction is* $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$*.*

Proof. We show proof with unknown covariance Σ. By the property of Hotelling's T^2 test, it holds for the statistic defined in (3.3) that $(N-k)T_A^2/\{k(N-1)\}$ follows the $F_{k,N-k}(N\delta_A)$ distribution, the noncentral F-distribution with degrees of freedom k and $N-k$, and noncentrality parameter $N\delta_A$, where

$$\delta_A = \boldsymbol{\mu}^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu}. \tag{3.7}$$

For a given k, the power of T_A^2 is increasing with δ_A by Lemma 3.6.7. As a result, we need to maximize δ_A with respect to A in order to maximize the power of T_A^2. Denote $W = \Sigma^{1/2}A$ and $P_W = W(W^TW)^{-1}W^T$. Note that P_W is a projection matrix and its eigenvalue is either 0 or 1. Rewrite δ_A in terms of projection matrix P_W as

$$\begin{aligned}
\delta_A &= \boldsymbol{\mu}^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu} \\
&= \boldsymbol{\mu}^T\Sigma^{-\frac{1}{2}}\left(W(W^TW)^{-1}W^T\right)\Sigma^{-\frac{1}{2}}\boldsymbol{\mu} \\
&= \boldsymbol{\mu}^T\Sigma^{-\frac{1}{2}}P_W\Sigma^{-\frac{1}{2}}\boldsymbol{\mu} \\
&\leq \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}.
\end{aligned}$$

Next we construct a matrix A_0 such that δ_{A_0} reaches the upper bound $\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$. Specifically, let $\mathbf{v} = \Sigma^{-1/2}\boldsymbol{\mu}$ and $\mathbf{w}_1 = \mathbf{v}/\|\mathbf{v}\|$. We choose $\mathbf{w}_2, \cdots, \mathbf{w}_p$ so that the matrix $(\mathbf{w}_1, \cdots, \mathbf{w}_p)$ is a $p \times p$ orthogonal matrix. Next we set

$$\begin{aligned}
W_0 &= (\mathbf{w}_1, \cdots, \mathbf{w}_k) \\
A_0 &= \Sigma^{-1/2}W_0.
\end{aligned} \tag{3.8}$$

Then $\delta_{A_0} = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$, which does not depend on the value of k.

By the property of F-distribution (Ghosh, 1973), the power function of T_A^2 with a fixed noncentrality parameter δ_{A_0} is a decreasing function of k. That is, for a given level α, denote by $F_{k,N-k;\alpha}$ the critical value of the F-distribution with k and $N-k$ degrees of freedom, $P\{F_{k,N-k}(N\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}) > F_{k,N-k;\alpha}\}$ is a decreasing function of k when N and $\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$ are fixed. This implies that $k = 1$ with $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$ is the best choice to achieve the optimal power. $\square$

The conclusion from Theorem 3.2.1 is very inspiring in that it not only gives the optimal projection under a very general construction, but also points out that only one dimensional projection space is sufficient to achieve the optimal power. As the result, it provides guideline for how many dimension to select and which space to project to. Based on Theorem 3.2.1, the optimal projection matrix is given by a $p \times 1$ vector $\mathbf{a}$ that

$$\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}. \tag{3.9}$$

Accordingly, denote by $T_{\mathbf{a}}^2$ the projection test T_A^2 with $A = \mathbf{a}$. Clearly, when $A = \mathbf{a}$, H_{0A} becomes $H_{0A} : \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu} = 0$, which is equivalent to $H_0 : \boldsymbol{\mu} = \mathbf{0}$ under the assumption of Σ being full rank. Consequently, $T_{\mathbf{a}}^2$ serves as a feasible tool for testing (3.1). The optimal direction $\mathbf{a}$ is consistent with the results as we directly derive from projection into a one-dimensional space. The proof can be found in Section 3.6.1.

3.2.2 Implementation and practical issues

3.2.2.1 Algorithm

Theorem 3.2.1 lays the groundwork for our method and sheds the light on the theoretical validity. We have shown in Theorem 3.2.1 that the optimal projection $\mathbf{a}$ is $\Sigma^{-1}\boldsymbol{\mu}$, under which the t test of projected sample $X\mathbf{a}$ achieves the optimal power. However, both $\boldsymbol{\mu}$ and Σ are unknown in practice. There arise the natural questions: how to obtain the estimate of the optimal direction, and how to maintain the validity challenged by introducing the estimated weight vector. In this section, we first describe the implementation algorithm, followed by the discussion of how our solutions of the above questions would impact the power of the test.

We propose an intuitive and easy-to-compute algorithm following the single sample-splitting strategy proposed in Wasserman and Roeder (2009). To implement the algorithm, the random sample is first partitioned into two separate sets with a splitting percentage κ such that $\mathcal{S}_1 = \{\mathbf{x}_{11}, \cdots, \mathbf{x}_{1N_1}\}$ and $\mathcal{S}_2 = \{\mathbf{x}_{21}, \cdots, \mathbf{x}_{2N_2}\}$, where $N_1 = \lfloor N\kappa \rfloor, N_1 + N_2 = N$. We propose using $\mathcal{S}_1$ to estimate $\mathbf{a}$ and $\mathcal{S}_2$ to construct $T^2_{\mathbf{a}}$. Specifically, let $\bar{\mathbf{x}}_1$ and $\mathbf{S}_1$ be the sample mean vector and the sample covariance matrix computed from $\mathcal{S}_1$, respectively. Note that $\mathbf{S}_1$ may not be invertible when p is greater than N. A simple way to estimate $\mathbf{a}$ based on $\mathcal{S}_1$ is the ridge-like estimator $\hat{\mathbf{a}} = (\mathbf{S}_1 + \lambda\mathbf{D})^{-1}\bar{\mathbf{x}}_1$, where $\mathbf{D} = \text{diag}(\mathbf{S}_1)$, the diagonal matrix of $\mathbf{S}_1$; and λ is a ridge parameter. We will study how the performance of $T^2_{\mathbf{a}}$ depends on λ in Section 3.3.2. Since $\hat{\mathbf{a}}$ is independent of $\mathcal{S}_2$, $T^2_{\mathbf{a}}$ based on $\hat{\mathbf{a}}^T\mathbf{x}_{21}, \cdots, \hat{\mathbf{a}}^T\mathbf{x}_{2N_2}$ follows exactly a noncentral F-distribution $F_{1,N_2-1}(N_2(\hat{\mathbf{a}}^T\boldsymbol{\mu})^2/(\hat{\mathbf{a}}^T\Sigma\hat{\mathbf{a}}))$. Particularly, under $H_0 : \boldsymbol{\mu} = \mathbf{0}$, $T^2_{\mathbf{a}}$ follows a central F_{1,N_2-1} distribution, which is equivalent to an exact t-test based on $\hat{\mathbf{a}}^T\mathbf{x}_{21}, \cdots, \hat{\mathbf{a}}^T\mathbf{x}_{2N_2}$. We summarize this algorithm into the following diagram.

Algorithm 1: Proposed algorithm for the one-sample test

Input: Data $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, splitting percentage κ, and ridge penalty λ

(a) Randomly partition the sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ into

$$\begin{cases} \text{estimating set } \mathcal{S}_1\text{:} & \{\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N_1}\} \\ \text{testing set } \mathcal{S}_2\text{:} & \{\mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2}\} \end{cases},$$

where $N_1 = \lfloor N\kappa \rfloor$ and $N_2 = N - N_1$.

(b) Obtain weight vector $\hat{\mathbf{a}}$ that

$$\hat{\mathbf{a}} = (\mathbf{S}_1 + \lambda\mathbf{D})^{-1}\bar{\mathbf{x}}_1,$$

where $\bar{\mathbf{x}}_1 = \sum_j^{N_1} \mathbf{x}_{1j}$ and $\mathbf{S}_1 = (N_1 - 1)^{-1} \sum_{j=1}^{N_1}(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1j} - \bar{\mathbf{x}}_1)^T$.

(c) Construct $y_i = \hat{\mathbf{a}}^T\mathbf{x}_i$ for $i = 1, \ldots, N_2$.

(d) Calculate the t statistic and the p-value.

$$p\text{-value} = P\left(\frac{\sqrt{N_2}\bar{y}}{s_y} > t_{N_2-1,\alpha}\right),$$

where $\bar{y} = N_2^{-1} \sum_{j=1}^{N_2} y_j$ and $s_y^2 = (N_2 - 1)^{-1} \sum_{j=1}^{N_2} (y_j - \bar{y})^2$.

The calculation of weight vector $\hat{\mathbf{a}}$ in step (b) involves inverse operation of $p \times p$ matrix $\mathbf{S}_1 + \lambda \mathbf{D}$, which is time consuming. We present a fast calculation method which only requires the inverse operation on a $N_1 \times N_1$ matrix. Note that

$$\begin{aligned} (\mathbf{S}_1 + \lambda \mathbf{D})^{-1} &= \left(\mathbf{D}^{\frac{1}{2}} (\mathbf{D}^{-\frac{1}{2}} \mathbf{S}_1 \mathbf{D}^{-\frac{1}{2}} + \lambda I_p) \mathbf{D}^{\frac{1}{2}} \right)^{-1} \\ &= \lambda^{-1} \mathbf{D}^{-\frac{1}{2}} (I_p + M M^T)^{-1} \mathbf{D}^{-\frac{1}{2}}, \end{aligned} \tag{3.10}$$

where X_1 is the corresponding matrix in X from $\mathcal{S}_1$ and

$$M = (\lambda (N_1 - 1) \mathbf{D})^{-\frac{1}{2}} X_1^T (I_{N_1} - \frac{1}{N_1} 1_{N_1} 1_{N_1}^T). \tag{3.11}$$

For matrix A of $p \times N_1$ and B of $N_1 \times p$, the following identity holds that

$$(I_p + AB)^{-1} = I_p - A(I_{N_1} + BA)^{-1} B,$$

where I_p is $p \times p$ identity matrix with dimension and I_{N_1} is $N_1 \times N_1$ identity matrix. Finally, we obtain the efficient calculation formula by plugging in (3.11) $A = M$.

3.2.2.2 Discussion of sample splitting

Sample splitting strategy plays an important role in the algorithm by avoiding the dependence introduced by estimated weights and therefore maintaining the exactness. Without sample splitting, the numerator and denominator of t test statistic from the projected sample with the estimated weight are correlated, which fails the t-test. We will discuss the selection of splitting parameter in more details in Section 3.3.1.

On the other hand, sample splitting causes loss of power. It is necessary to investigate power gain with the impact of sample splitting taken into account. Here we present a quick (asymptotic) power comparison between the proposed projection $T_{\mathbf{a}}^2$ test and the classical Hotelling's T^2 test with a hypothetical example that N =30, 50, 100, and 200; $p = 10$ and 20; and $N_2 = 0.6N$. In this case with $p < N$, the classical Hotelling's T^2 is applicable with power function $P(F_{p,N-p}(N\zeta) > F_{p,N-p,\alpha})$, where $\zeta = \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu}$. When $N_1 \to \infty$ and $\lambda \to 0$, $(\hat{\mathbf{a}}^T \boldsymbol{\mu})^2 / (\hat{\mathbf{a}}^T \Sigma \hat{\mathbf{a}}) \to \zeta$.

Thus, the noncentrality parameter in $F_{1,N_2-1}(N_2(\hat{\mathbf{a}}^T\boldsymbol{\mu})^2/(\hat{\mathbf{a}}^T\Sigma\hat{\mathbf{a}}))$ approximately equals $N_2\zeta$. In the oracle case, the optimal direction is known and the entire sample is used for testing. The corresponding power function is $P(F_{1,N-1}(N\zeta) > F_{1,N-1,\alpha})$. We include the oracle case in the comparison as a benchmark.

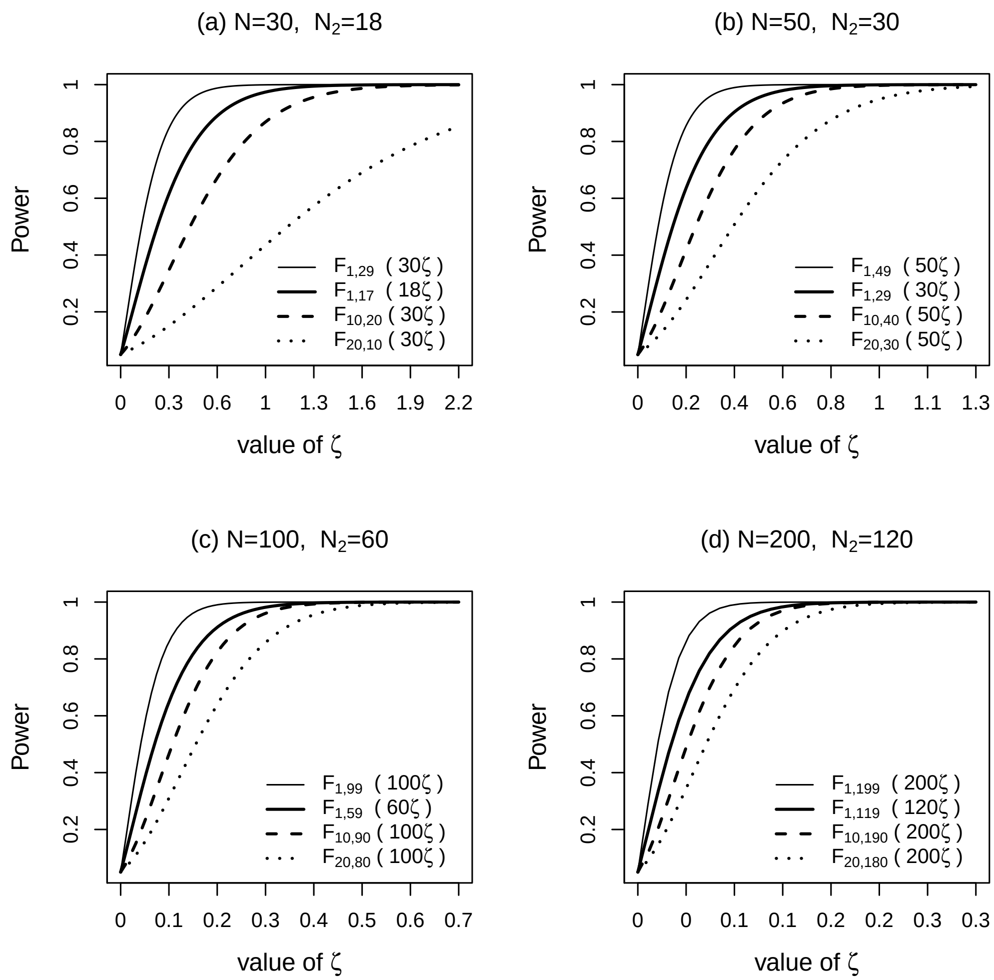


Figure 3.1: Power function of Hotelling's T^2 test and projection Hotelling's T^2 test at level 0.05. Solid line stands for the benchmark (i.e. the power function of the projection test based on the entire sample with **known** optimal direction), the bold solid, dashed and dotted lines stand for the power function of the proposed projection test, Hotelling's T^2 test with $p = 10$ and Hotelling's T^2 test when $p = 20$, respectively.

The following observations are obtained from Figure 3.1, which depicts the power curves under $\alpha = 0.05$: (a) the Hotelling's T^2 test suffers from low power when p is close to N (e.g, the case $N = 30$ and $p = 20$); (b) with a high quality estimate of ζ, the proposed projection test may have higher power than the Hotelling's T^2 when p is close to N, even with the power loss due to sample splitting; and (c) the power function of projection test is close to the benchmark, especially for a large noncentrality parameter.

3.2.2.3 Discussion of ridge-like estimator

In estimating the optimal direction $\Sigma^{-1}\boldsymbol{\mu}$, the calculation of Σ^{-1} confronts the same problem as the classical Hotelling's T^2 when p is greater than N. Estimation of Σ^{-1} is universally challenging in the high-dimensional situations. The discussion is beyond the scope of this study. A simple solution is the ridge-like estimator $(\mathbf{S}_1 + \lambda\mathbf{D})^{-1}$, where λ is a turning parameter that controls the degree of penalty and $\mathbf{D}$ is a diagonal matrix of $\mathbf{S}_1$. A large λ drives the $(\mathbf{S}_1 + \lambda\mathbf{D})^{-1}$ towards an identity matrix while a small λ may not produce stable results. Here the ridge-like estimator is constructed as $(\mathbf{S}_1 + \lambda\mathbf{D})^{-1}$ rather than $(\mathbf{S}_1 + \lambda I)^{-1}$ to keep the test invariant of units. We suggest $\lambda = N_1^{-0.5}$. The intuition is to down-weight the effect of $\lambda\mathbf{D}$ when sample size increases. As will be shown with simulations in Section 3.3.2, the performance of the test is not sensitive to the choice of λ.

3.2.3 Asymptotic power comparison

In Section 3.2.2, we have shown a toy example of power comparison under $N > p$ such that the classical Hotelling's T^2 test is applicable and used as a benchmark in the comparison. In this section, we assume the local alternative hypothesis H_1 : $\boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{N}$. The asymptotic power of the proposed projection test is compared with the Hotelling's T^2 when $N > p$ and compared with some major existing methods in high-dimensional situations.

Recall that the sample is partitioned to estimating set $\mathcal{S}_1$ with sample size N_1 and testing set $\mathcal{S}_2$ with sample size N_2. Assume that $\sqrt{N_2/N} \to b > 0$ as $N \to \infty$, where N_2 is the sample size of $\mathcal{S}_2$. Further assume that $\hat{\mathbf{a}} \to \mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$ in probability as the sample size of $\mathcal{S}_1$ tends to ∞. Let $\Phi(\cdot)$ and z_α denote the cumulative distribution function and upper α quantile of $N(0,1)$, respectively. Then the asymptotic power function of the proposed projection test at a given level α is

$$\beta_{1p}(\eta) = \Phi(-z_{\alpha/2} + b\sqrt{\eta}), \tag{3.12}$$

where $\eta = \boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta}$.

We first derive the asymptotic power function of $T^2 = N\bar{\mathbf{x}}^T \mathbf{S}^{-1}\bar{\mathbf{x}}$. Note that $T^2 \to \chi^2(p)$ in distribution under $H_0 : \boldsymbol{\mu} = \mathbf{0}$, and $T^2 \to \chi_p^2(\eta)$ in distribution under $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{N}$. Since $E\chi_p^2(\eta) = p + \eta$ and $\mathrm{var}\{\chi_p^2(\eta)\} = 2(p + 2\eta)$, its asymptotic power function at level α is

$$\beta_{2p}(\eta|\tau_p) = \Phi\left\{-z_\alpha/\sqrt{1+2\tau_p} + \frac{1}{2}\sqrt{\tau_p/(0.5+\tau_p)} \cdot \sqrt{\eta}\right\}, \tag{3.13}$$

where $\tau_p = \eta/p$. The derivation for $\beta_{2p}(\eta|\tau_p)$ can be found in Section 3.6.3.

Next we give Proposition 3.2.2, which compares $\beta_{1p}(\eta)$ and $\beta_{2p}(\eta|\tau_p)$.

Proposition 3.2.2. *Assume that $\sqrt{N_2/N} \to b$. The following statements are valid.*

(a) If $b > 0.5$ and $\eta \to \infty$ as $p \to \infty$, then $\beta_{1p}(\eta) - \beta_{2p}(\eta|\tau_p) > 0$ for large enough p.

(b) If $\sqrt{\eta}b \geq z_{\alpha/2} - z_\alpha$ and $\tau_p \to 0$ as $p \to \infty$, then $\beta_{1p}(\eta) - \beta_{2p}(\eta|\tau_p) > 0$ for large enough p.

Proof. As to (a), it suffices to show that

$$-z_{\alpha/2} + b\sqrt{\eta} > -z_\alpha/\sqrt{1+2\tau_p} + \frac{1}{2}\sqrt{\tau_p/(0.5+\tau_p)}\sqrt{\eta},$$

which after rearrangement is

$$\sqrt{\eta}\left(b - \frac{1}{2}\sqrt{\tau_p/(0.5+\tau_p)}\right) > z_{\alpha/2} - z_\alpha/\sqrt{1+2\tau_p}.$$

Since $\sqrt{\tau_p/(0.5+\tau_p)} < 1/2$, it holds that

$$\sqrt{\eta}\left(b - \frac{1}{2}\sqrt{\tau_p/(0.5+\tau_p)}\right) > \sqrt{\eta}\left(b - \frac{1}{2}\right).$$

By the assumption that $b > 0.5$ and $\eta \to \infty$, it follows that $\sqrt{\eta}(b - \frac{1}{2}) > z_{\alpha/2}$ for large enough p. Since also $z_\alpha/\sqrt{1+2\tau_p} > 0$ regardless of the value of τ_p, we have $\sqrt{\eta}(b - \frac{1}{2}) > z_{\alpha/2} - z_\alpha/\sqrt{1+2\tau_p}$ for large enough p.

As to (b), if $\tau_p \to 0$, then $\sqrt{1+2\tau_p} \to 1$ and $\sqrt{\tau_p/(0.5+\tau_p)} \to 0$. Thus (b) is valid by the assumption in (b). □

Statement (a) implies that, if $\eta \to \infty$ as $p \to \infty$ and we always set $\mathcal{S}_2$ to include more than 25% of the samples so that $b > 0.5$, then the proposed projection test may be asymptotically more powerful than the traditional T^2. It is also worth noting that $N\bar{\mathbf{x}}^T\Sigma^{-1}\bar{\mathbf{x}}$ and T^2 share the same asymptotic power function $\beta_{2p}(\eta|\tau_p)$. In (b), the condition $\tau_p \to 0$ implies that the signal in $\boldsymbol{\mu}$ is weak or sparse. In such situations, (b) implies that the proposed projection test may be asymptotically more powerful than the traditional T^2 test.

We next derive the asymptotic power of $T_3^2 = N\|\bar{\mathbf{x}}\|^2$, which shares the asymptotic power function of the tests proposed in Bai and Saranadasa (1996), Srivastava and Du (2008), and Chen and Qin (2010) under some mild conditions. When p is close to n, these authors suggested replacing $\mathbf{S}^{-1}$ by I_p, the $p \times p$ identity matrix. Let $\beta_{3p}(\boldsymbol{\delta}, \Sigma)$ be the asymptotic power function of T_3^2 and

$$\beta_{3p}(\eta|\tau_p^*) = \Phi\left\{-z_\alpha/\sqrt{1+2\tau_p^*} + \frac{1}{2}\sqrt{\tau_p^*/(0.5+\tau_p^*)} \cdot \sqrt{\eta}\right\},$$

where $\tau_p^* = \boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}/\mathrm{tr}(\Sigma^2)$. We have the following proposition.

Proposition 3.2.3. *Assume that $\sqrt{N_2/N} \to b$. The following statements are valid.*

(I) Under $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{N}$, $\beta_{3p}(\boldsymbol{\delta}, \Sigma) \leq \beta_{3p}(\eta|\tau_p^*)$

(II) If $b > 0.5$ and $\eta \to \infty$, then $\beta_{1p}(\eta) - \beta_{3p}(\eta|\tau_p^*) > 0$ for large enough p.

(III) If $\sqrt{\eta}b \geq z_{\alpha/2} - z_\alpha$ and $\tau_p^* \to 0$ as $p \to \infty$, then $\beta_{1p}(\eta) - \beta_{3p}(\eta|\tau_p^*) > 0$ for large enough p.

Proof. Since $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, N^{-1}\Sigma)$, by the properties of normal distribution, it holds that $E(T_3^2) = \mathrm{tr}(\Sigma) + N\|\boldsymbol{\mu}\|^2$ and $\mathrm{var}(T_3^2) = 2\mathrm{tr}(\Sigma^2) + 4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}$. Furthermore,

$$\frac{T_3^2 - E(T_3^2)}{\sqrt{\mathrm{var}(T_3^2)}} \to N(0,1)$$

in distribution as $N \to \infty$. Thus under $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{N}$, the asymptotic power function of T_3^2 is

$$\beta_{3p}(\boldsymbol{\mu}, \Sigma) = \Phi\left\{ -z_\alpha \sqrt{\frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}(\Sigma^2) + 2\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}} + \frac{\|\boldsymbol{\delta}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}} \right\}.$$

The derivation of $\beta_{3p}(\boldsymbol{\mu}, \Sigma)$ can be found in Section 3.6.3. Furthermore,

$$\beta_{3p}(\boldsymbol{\mu}, \Sigma) = \Phi\left\{ -z_\alpha / \sqrt{1 + 2\tau_p^*} + \frac{1}{2}\sqrt{\tau_p^*/(0.5 + \tau_p^*)} \cdot \|\boldsymbol{\delta}\|^2 / \sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}} \right\}.$$

By the Cauchy-Schwartz inequality, it holds that $\|\boldsymbol{\delta}\|^2 \le \sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}\sqrt{\boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta}}$. Therefore, $\|\boldsymbol{\delta}\|^2/\sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}} \le \sqrt{\eta}$. Thus, it follows that

$$\beta_{3p}(\boldsymbol{\mu}, \Sigma) \le \Phi\left\{ -z_\alpha / \sqrt{1 + 2\tau_p^*} + \frac{1}{2}\sqrt{\tau_p^*/(0.5 + \tau_p^*)} \cdot \sqrt{\eta} \right\} = \beta_{3p}(\eta|\tau_p^*). \quad (3.14)$$

This completes the proof of Part (I). As a result, $\beta_{2p}(\eta|\tau_p)$ and $\beta_{3p}(\eta|\tau_p^*)$ have the same form. Following the Proposition 3.2.2, Part (II) and (III) hold.

□

In (III), the condition $\tau_p^* \to 0$ means that $\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta} = o\{\mathrm{tr}(\Sigma^2)\}$, which corresponds to the assumption $\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu} = o\{N^{-1}\mathrm{tr}(\Sigma^2)\}$ for the corresponding one-sample test in Bai and Saranadasa (1996). Under this assumption, $\beta_{3p} \approx \Phi\left\{ -z_\alpha + \frac{N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)}} \right\}$, which is exactly the asymptotic power function of the corresponding one-sample test proposed by Bai and Saranadasa (1996) and Chen and Qin (2010), and the test by Srivastava and Du (2008) when all the diagonal elements of Σ equal to 1 (i.e., Σ indeed is a correlation matrix).

In order to better understand the conditions in (II) and (III), we examine two examples with commonly-used correlation structures.

Example 3.1. In this example, we consider the compound symmetry structure, which is defined as $\Sigma = (1-r)I_p + r\mathbf{1}_p\mathbf{1}_p^T$ for $r \in [0, 1)$.

First we check the conditions in (II). Under the compound symmetry assumption,

$$\Sigma^{-1} = (1-r)^{-1}\{I - \{r/(1+(p-1)r)\}\mathbf{1}_p\mathbf{1}_p^T\}.$$

Therefore,

$$\boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta} = (1-r)^{-1}[\boldsymbol{\delta}^T\boldsymbol{\delta} - \{r/(1+(p-1)r)\}(\boldsymbol{\delta}^T\mathbf{1}_p)^2].$$

Since $r/\{1+(p-1)r\} < 1/p$, we have

$$\boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta} > (1-r)^{-1}p\{p^{-1}\boldsymbol{\delta}^T\boldsymbol{\delta} - (p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p)^2\}.$$

As a result, if $\liminf_{p\to\infty} p^{-1}\boldsymbol{\delta}^T\boldsymbol{\delta} - (p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p)^2 \to b_0 > 0$, then $\eta \to \infty$ as $p \to \infty$. Therefore the conditions in (II) hold under some conditions on $\boldsymbol{\delta}$ if $b > 0.5$. Also note that by the Cauchy-Schwartz inequality, it holds that $(\boldsymbol{\delta}^T\mathbf{1}_p)^2 \leq p\boldsymbol{\delta}^T\boldsymbol{\delta}$. Therefore, it is a mild condition to assume that $\liminf_{p\to\infty} p^{-1}\boldsymbol{\delta}^T\boldsymbol{\delta} - (p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p)^2 \to b_0 > 0$ as $p \to \infty$.

We next examine the conditions in (III). Under the compound symmetry assumption,

$$\begin{aligned} \mathrm{tr}(\Sigma^2) &= (1-r^2)p + r^2p^2, \qquad (3.15)\\ \boldsymbol{\delta}^T\Sigma\boldsymbol{\delta} &= (1-r)\boldsymbol{\delta}^T\boldsymbol{\delta} + r(\boldsymbol{\delta}^T\mathbf{1}_p)^2. \end{aligned}$$

Consider $r \in (0,1)$, if $p^{-2}\boldsymbol{\delta}^T\boldsymbol{\delta} \to 0$ and $p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p \to 0$, then $\tau_p^* \to 0$. By Cauchy-Schwartz inequality, $\boldsymbol{\delta}^T\boldsymbol{\delta} \leq \sqrt{(\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta})(\boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta})}$. Consequently, $\sqrt{\eta} > \boldsymbol{\delta}^T\boldsymbol{\delta}/\sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}$. Note that

$$\frac{\boldsymbol{\delta}^T\boldsymbol{\delta}}{\sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}} = \frac{p^{-1}\boldsymbol{\delta}^T\boldsymbol{\delta}}{\sqrt{(1-r)p^{-2}\boldsymbol{\delta}^T\boldsymbol{\delta} + r(p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p)^2}}.$$

Hence, under conditions $p^{-2}\boldsymbol{\delta}^T\boldsymbol{\delta} \to 0$, $p^{-1}\boldsymbol{\delta}^T\mathbf{1}_p \to 0$, and $\liminf_{p\to\infty} p^{-1}\|\boldsymbol{\delta}\|^2 > 0$, we have $\eta \to \infty$ and therefore $\sqrt{\eta}b \geq z_{\alpha/2} - z_\alpha$ for large p. Then, the conditions in (III) hold.

Example 3.2 In this example, we consider the auto-correlation structure that the (i, j)-element of Σ is $r^{|i-j|}$ for $r \in (0, 1)$. Further assume that $\boldsymbol{\delta} = c(\mathbf{1}_s^T, \mathbf{0}_{p-s}^T)^T$ for $c \neq 0$, where $\mathbf{1}_d$ and $\mathbf{0}_d$ stand for the d-dimensional vector with all elements being one and zero, respectively.

First we check the conditions in (II). Since

$$\boldsymbol{\delta}^T \Sigma \boldsymbol{\delta} = c^2[s + 2\{sr(1-r) - r(1-r^s)\}/(1-r)^2,$$

if $s \to \infty$ as $p \to \infty$, then $\boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta} \to \infty$. Hence if we also have $b > 0.5$, then the condition in (II) holds.

Next we examine the conditions in (III). Under the specifications,

$$\begin{aligned} \mathrm{tr}(\Sigma^2) &= p + 2\{p(1-r^2)r^2 - r^2(1-r^{2p})\}/(1-r^2)^2, \\ \boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta} &= c^2\{(1+r^2)s - 2r(s-1) - r^2\}/(1-r^2). \end{aligned} \tag{3.16}$$

If $s/p \to 0$ as $p \to \infty$, then $\tau_p^* \to 0$. Furthermore, if $s \to \infty$ as $p \to \infty$, we have $\sqrt{\eta} \to \infty$. Therefore $\sqrt{\eta} b > z_{\alpha/2} - z_\alpha$ for large p. Thus, the condition in (III) holds for large p if $s/p \to 0$ and $s \to \infty$ as $p \to \infty$.

3.3 Simulation studies

In this section, we conduct simulations to evaluate the finite sample performance of the proposed test (Algorithm 1) and the competing ones. In what follows, we conduct three numerical experiments. All three experiments generate samples from multivariate normal populations. In these cases, we study the performance of the exact t-test. Experiments 1 and 2 study the tuning parameter introduced in Algorithm 1. Special attention is given to observing how the tuning parameter would impact power of the proposed test and make suggestions to the selection for latter experiments. Experiment 1 is designed to investigate the effect of splitting percentage. In experiment 2, we study the tuning parameter λ with splitting percentage fixed at 40%. Experiment 3 is designed to compare the proposed method and the alternative tests under multivariate normal distributions.

We consider the sample size $N \in (40, 160)$ and the dimension $p \in (400, 1600)$, which together give four pairs of (N, p). The mean vector $\boldsymbol{\mu}$ takes the form of $(c\mathbf{1}'_{10}, \mathbf{0}'_{p-10})^T$, where $\mathbf{1}_d$ and $\mathbf{0}_d$ stand for d-dimensional vector with all elements being one and zero, respectively. The constant c takes values in $(0, 0.5, 1)$. When $c = 0$, we obtain the observed Type I error and $c = 0.5$ or 1 lead to power under different strength of the means. For the covariance matrix, we set all the marginal variances to 1, and consider the following three types of off-diagonal setups:

- Compound symmetry structure Σ_1 of which $\Sigma_1(i, j) = \rho$,
- Autocorrelation structure Σ_2 of which $\Sigma_2(i, j) = \rho^{|i-j|}$,
- Combined structure Σ_3 as a weighted matrix of Σ_1 and Σ_2 that $\Sigma_3 = 0.5\Sigma_1 + 0.5\Sigma_2$.

We set $\rho \in (0.25, 0.5, 0.75, 0.95)$ to examine the influence of correlation on the power. Table 3.1 summarizes the aforementioned settings. Not all the settings will be used in each experiment. We refer to the introduction before each experiment for the detailed settings. We set $\alpha = 0.05$ and our results are obtained based on 10000 replicates.

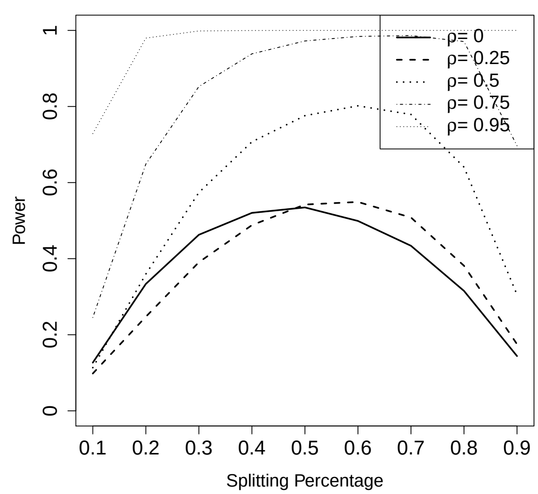
Table 3.1: Simulation setting specifications

Parameter	Choices	More details
Σ	$\Sigma_1(i, j) = \rho$ $\Sigma_2(i, j) = \rho^{\vert i-j\vert}$ $\Sigma_3 = 0.5\Sigma_1 + 0.5\Sigma_2$	$\rho = 0.25, 0.5, 0.75, 0.95$
$\boldsymbol{\mu}$	$(c\mathbf{1}'_{10}, \mathbf{0}'_{p-10})^T$	$c = 0, 0.5, 1$
N	40, 160	
p	400, 1600	

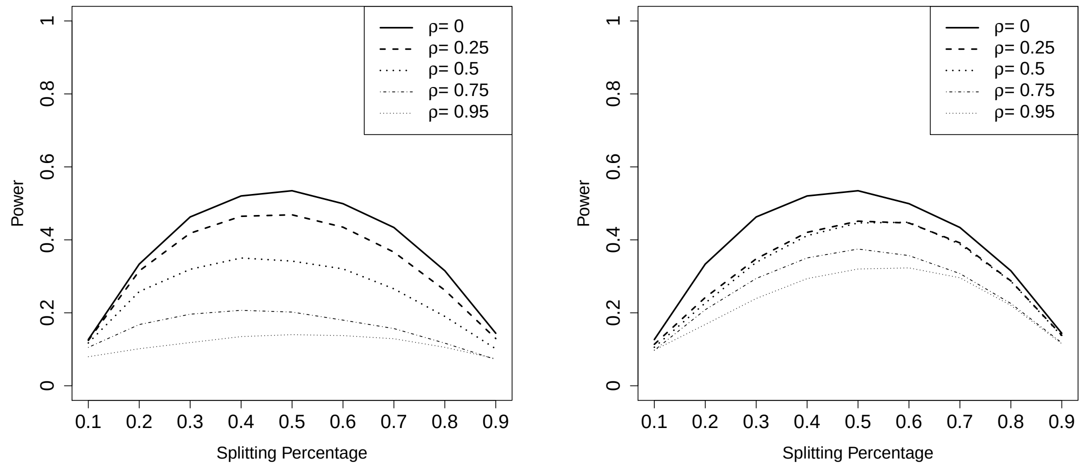
3.3.1 Experiment 1

With splitting percentage κ, the sample size used for estimation and the test construction are $N_1 = \lfloor N\kappa \rfloor$ and $N_2 = N - \lfloor N\kappa \rfloor$, respectively. As a pedagogical argument, large κ allocates more sample for estimation and therefore produces a more accurate estimation of the optimal direction. Meanwhile, large κ also reduces the sample size used for constructing the t statistic and inevitably leads to a certain level of power loss. In this section, we explore this trade-off in the simulations by taking a grid of κ over $(0, 1)$ as $10\%, 20\%, \ldots, 90\%$ and compare the power of each gird value. During the implementation, λ is set to $N_1^{-0.5}$ and its effect will be studied in experiment 2.

Figure 3.2 presents the power curves with setting $(N, p, c) = (40, 400, 0.5)$ and different covariance matrices as Σ_1, Σ_2 and Σ_3 respectively. All power curves display an approximate quadratic pattern. As the percentage increases, the power first improves as the estimation of the projection direction is more accurate. However, exceedingly increase of the percentage adversely affects the power, as the power is also an increasing function of sample size used in the testing procedure. The optimal splitting percentage varies by situations, with most peaks occurring at grid value 40%, 50% and 60% under our setting that $(N, p) = (40, 400)$. It is difficult in practice to determine a splitting percentage that works universally best, due to unknown covariance structure, relative size of N and p, and the concern of data quality. However, we argue that $40\% - 60\%$ is a reasonable range. In the following simulation studies, we set the splitting percentage to 40%.



(a) Σ_1 (Compound symmetry)

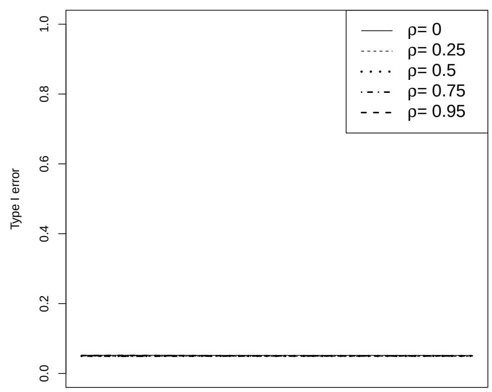


(b) Σ_2 (Autocorrelation) (c) Σ_3 (Combined correlation)

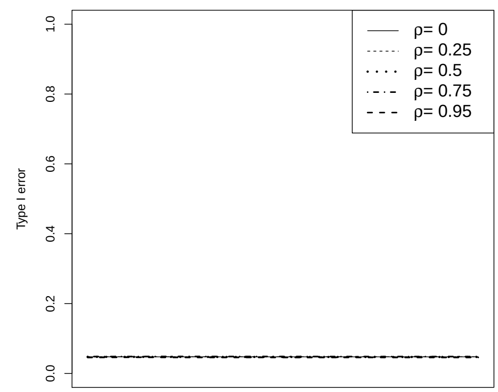
Figure 3.2: Effect of splitting percentage for $(N, p, c) = (40, 400, 0.5)$

3.3.2 Experiment 2

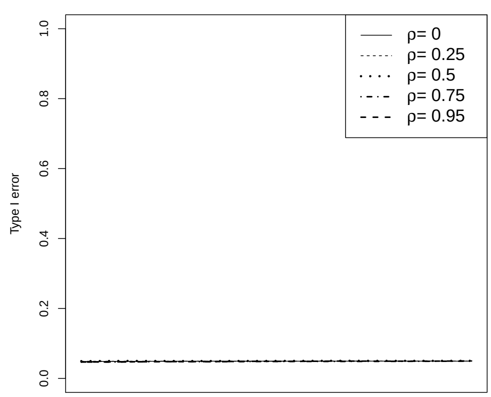
Tuning parameter λ is used in the construction of ridge-like estimator $S^{-1}(\lambda) = (S + \lambda D_S)^{-1}$ for Σ^{-1}. In Section 3.2.2, we suggest $\lambda = N_1^{-0.5}$, which down-weights λ when the sample size increases. In this experiment, we study the effect of tuning parameter λ by comparing power under various λ values. The splitting percentage is fixed with 40% and $N_1 = 0.4N$. We take $\lambda = N_1^{-\tau}$ and $\tau \in (0.1, 0.2, \ldots, 1)$. Under this specification, λ takes value between $(0.063, 0.758)$ for $N = 40$ and takes value between $(0.016, 0.660)$ for $N = 160$. The data is generated from multivariate normal distribution with sample size, dimension and the mean vector as specified in Table 3.1. For covariance matrix, we only include the compound symmetry correlation structure R_1. The pattern under autocorrelation and the combined correlation structures are similar and therefore omitted from here. Power plots presented by Figure 3.3 - Figure 3.5 show that the proposed test is fairly robust for a wide range of λ.



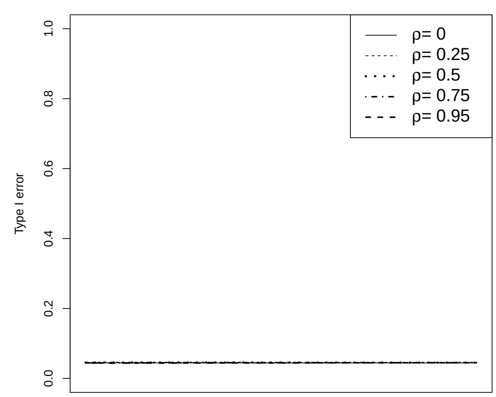
(a) N = 40, p = 400



(b) N = 40, p=1600

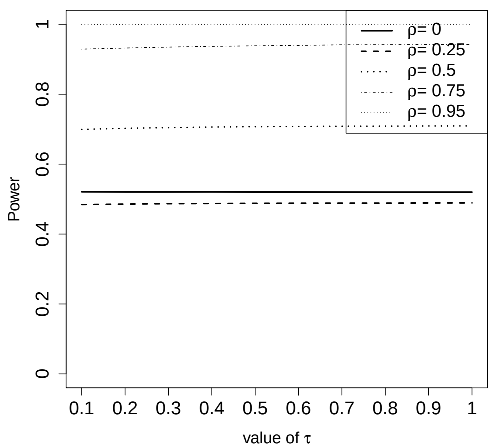


(c) N=160, p = 400

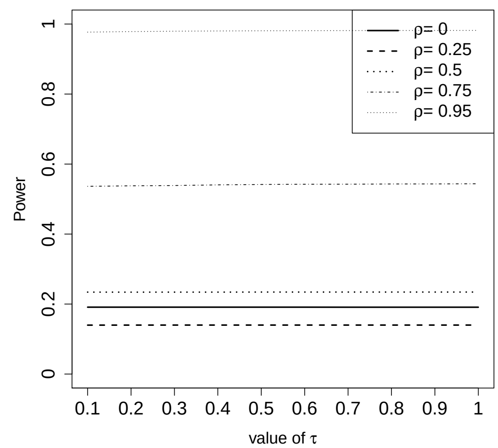


(d) N=160, p=1600

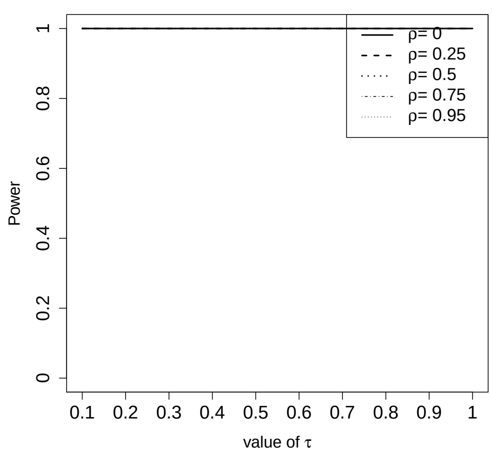
Figure 3.3: Type I error under various values of λ with $\lambda = N_1^{-\tau}$ and $N_1 = 0.4N$. The mean vector is set to $\mathbf{0}$.



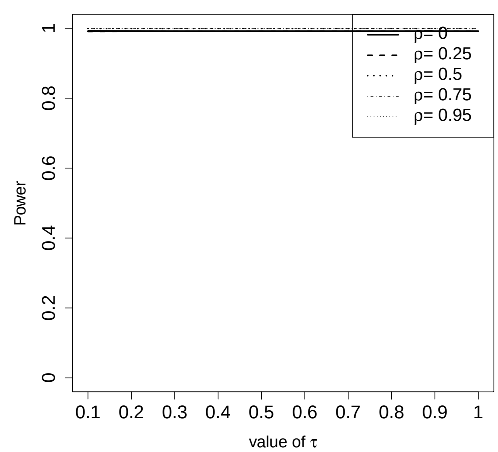
(a) N = 40, p = 400



(b) N = 40, p=1600

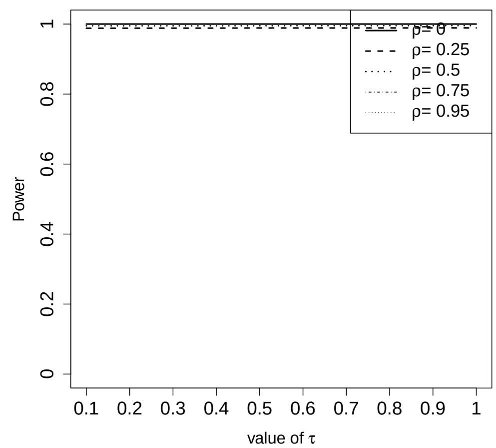


(c) N=160, p = 400

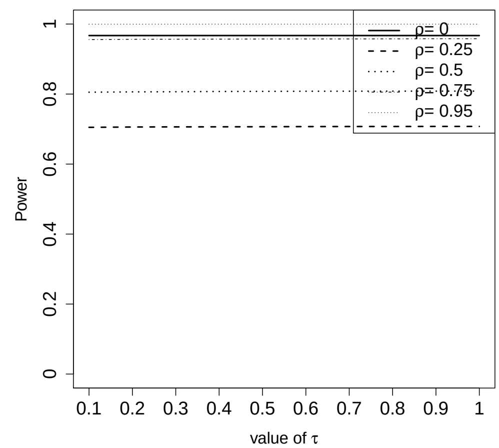


(d) N=160, p=1600

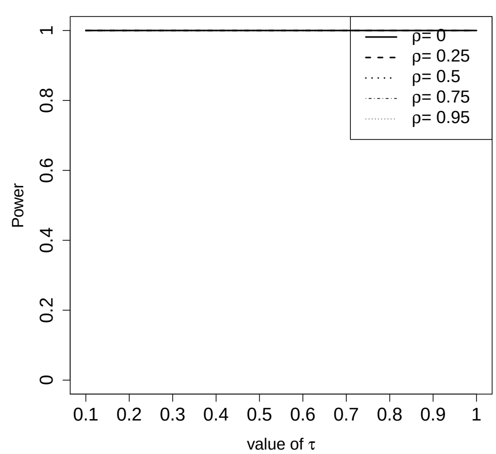
Figure 3.4: Power under various values of λ with $\lambda = N_1^{-\tau}$ and $N_1 = 0.4N$. The mean vector is set to $(0.5\mathbf{1}'_{10}, \mathbf{0}'_{p-10})^T$.



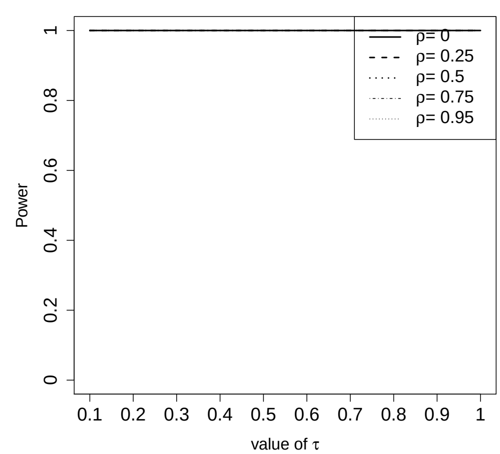
(a) N = 40, p = 400



(b) N = 40, p=1600



(c) N=160, p = 400



(d) N=160, p=1600

Figure 3.5: Power under various values of λ with $\lambda = N_1^{-\tau}$ and $N_1 = 0.4N$. The mean vector is set to $(1\mathbf{1}'_{10}, \mathbf{0}'_{p-10})^T$.

3.3.3 Experiment 3

In Experiment 3, we compare the proposed test and several alternative tests on random samples from multivariate normal distributions with configurations specified in Table 3.1. The competing algorithms include Dempster test (Dempster, 1958), BS test (Bai and Saranadasa, 1996), CQ test (Chen and Qin, 2010), SD test with adjusting factor (Srivastava and Du, 2008), SD test without adjusting factor (Srivastava, 2009), Lauter's PC test (Lauter, 1996) and the two versions of the RP test (Lopes et al., 2011a,b). The parameter r for Dempster test is estimated by (2.58)'s one-sample form. For the RP test with average of multiple projections, we set the number of projection to 30 as Lopes et al. (2011b) suggested that the number of projections over 30 does not make considerable difference. Tables 3.2-3.4 show the percentage of rejection based on 10000 replicates for three correlation structures with various configurations of (N, p, c).

First we examine the empirical Type I errors, which correspond to the columns with $c = 0$. For significance level $\alpha = 0.05$, the Monte Carlo error equals to $1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$. Therefore, well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%. Across all configurations considered, we observe that the proposed test, Lauter's PC test and RP test with single projection have good control of Type I error, with the false positive rate close to the pre-assigned significance level $\alpha = 0.05$. The version of RP test with average over multiple projections is conservative and tends to produce very low Type I error. In contrast, the alternatives BS test, CQ test and SD test, tend to overestimate the Type I error. These observations line well with the theoretical results as the proposed test, Lauter's PC test and RP test with single projection are all exact tests, while the other tests are developed based on the asymptotic approximations. We note that Dempster's test also performs well in terms of controlling Type I errors. Dempster's test is exact up to estimation of a parameter for use in the testing.

Inspection over Tables 3.2-3.4 reveals that power depends strongly on signal strength, as well as the correlation structures. It is clear that the increase of signal strength c helps improve the power as a larger signal is easier to detect. Specification of correlation structure can also affect the power. For example, Dempster test, BS test, CQ test and SD test are built up on $\|\bar{\mathbf{x}}\|^2$. In other words, they substitute the $\mathbf{S}^{-1}$ with identity matrix and ignore the off-diagonal information. Therefore, they are expected to have favorable performance when the true covariance matrix is closer to identity matrix, such as autocorrelation structure with small ρ. Next we discuss each of the Tables 3.2-3.4 in details.

Table 3.2 shows the results with compound symmetry correlation Σ_1, of which all entries equal to ρ. With Σ_1, our test outperforms all the other alternatives. We observe a dramatic increase of the power as c increases from 0.5 to 1. When $c = 0.5$, the power of the proposed test increases significantly with the increase of ρ. As dimension increases, there is a downward trend. However, even in the extreme case with $(N, p) = (40, 1600)$, the proposed test can manage to produce a high power when $c = 1$, and when $c = 0.5$ with a large value of ρ. The two versions of RP tests present similar pattern as the proposed test and outperform the rest alternatives. It is very interesting to compare the performance of the proposed test, and the two versions of the RP tests more closely. Our proposed test is an exact test constructed with a designed projection direction for optimal power. The RP test with single projection is also an exact test, but with the projection direction randomly generated. As expected, the proposed test outperforms RP test with single projection. The RP test with average over multiple projections is developed to improve performance, by increasing the chance of catching a better projection, in the cost of losing the exactness property. However, our numerical results show that the RP test with multiple projections cannot guarantee the improvement over the single projection version, at least under our simulation settings. For example, under configuration $(N, p, c) = (40, 1600, 0.5)$, the power of RP test with multiple projections is lower than that of the RP test with single projection for most of the ρ values. Lauter's PC test is known to be power deficient when the true mean vector contains 0 elements. This is well observed from our simulation studies that increase of sample size or the value of c cannot help improve the power for PC test. The other alternative tests, Dempster test, BS test, CQ test and SD test, tend to be adversely affected by the value of ρ and have power decreasing dramatically with even a slight increase of ρ. Their overall performance is unsatisfactory for the compound symmetry case. When $c = 0.5$, these tests almost have no power. Except for the cases with ρ as small as 0.25, the other alternative tests tend to have fairly small amount of improvement in both $c = 0.5$ and 1 when N increases from 40 to 160. As discussed, this could be due to the fact that they do not consider the off-diagonal information of the covariance matrix.

Table 3.3 shows the results with autocorrelation structure Σ_2, of which the (i, j) entry is $\rho^{|i-j|}$. With Σ_2 and especially under small ρ, the Dempster test, BS test, CQ test and SD test have satisfactory performance and produce higher power over the proposed test and the two versions of RP tests. We observe a large gap in power when $(N, c) = (40, 0.5)$. However, the power of the proposed test increases significantly to a comparable level as the competing test with either increase of c from 0.5 to 1, or the sample size N from 40 to 160. Similar increasing pattern is observed for the two versions of RP tests, except that the level of increase is considerably slower than our proposed test. Again, there is no guarantee that the RP test with multiple projections would outperform the one with single projection. We observe that there are a few cases where the RP test with multiple projections outperforms the proposed test. This is due to the sample splitting procedure of the proposed algorithm, which reduces the sample used in testing and causes a certain amount of loss in power.

Apart from the compound symmetry Σ_1 and the autocorrelation structure Σ_2, we also consider a more complex correlation structure Σ_3 as a combination of Σ_1 and Σ_2 under equal weights. The corresponding simulation results are shown in Table 3.4. Under this combined correlation structure, the proposed test outperforms all the competing tests. The power is generally lower than the case of Σ_1, but higher than the case of Σ_2. Another difference from the case of R_1 is that the power of the proposed test is decreasing with ρ. For $c = 0.5$, Dempster test, BS test, CQ test and the SD test have very low power except for $(N, p) = (160, 400)$. In general, the power of these test decreases with ρ.

Table 3.2: Comparison for one-sample tests: multivariate normal with Σ_1

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N = 40, p = 400}$							
New	5.16	5.04	4.86	5.05	48.60	70.34	94.10	99.98	98.81	99.54	99.96	100.00
Dempster	6.77	6.22	5.71	5.49	12.63	8.13	6.98	6.46	80.44	22.71	13.06	10.23
BS	7.73	7.80	7.79	7.80	14.64	10.55	9.39	9.11	88.21	30.28	18.51	15.33
CQ	7.72	7.82	7.79	7.77	14.64	10.50	9.41	9.11	88.18	30.22	18.50	15.32
SD_withAdjust	4.20	1.71	0.52	0.15	7.97	2.29	0.63	0.22	54.21	6.41	1.29	0.36
SD_noAdjust	8.48	8.21	7.87	7.71	16.34	11.15	9.53	8.96	90.25	32.69	18.93	15.06
Lauter's PC	5.18	5.18	5.17	5.15	5.66	5.21	5.17	5.11	6.25	5.59	5.31	5.22
RP_single	5.01	4.99	4.86	5.03	13.80	20.65	40.58	98.34	54.05	74.46	95.94	100.00
RP_average	0.06	0.06	0.12	0.09	4.58	20.42	94.41	100.00	99.87	100.00	100.00	100.00
					$\mathbf{N = 40, p = 1600}$							
New	4.91	4.82	4.85	4.69	13.92	23.54	53.93	98.13	70.54	80.97	95.99	99.99
Dempster	6.93	6.19	5.73	5.48	7.81	6.71	5.96	5.73	12.45	8.12	6.96	6.45
BS	7.74	7.79	7.78	7.79	8.85	8.30	8.12	8.06	14.47	10.42	9.35	8.92
CQ	7.76	7.80	7.77	7.77	8.88	8.32	8.14	8.05	14.49	10.41	9.34	8.92
SD_withAdjust	2.76	0.69	0.14	0.00	3.22	0.73	0.17	0.00	5.26	0.97	0.20	0.01
SD_noAdjust	8.37	8.12	7.86	7.71	9.67	8.70	8.24	7.94	15.79	11.14	9.52	8.78
Lauter's PC	5.15	5.15	5.15	5.15	5.30	5.22	5.19	5.18	5.50	5.29	5.23	5.18
RP_single	4.65	5.08	4.99	4.95	6.77	7.68	11.52	52.16	14.49	20.55	42.17	98.29
RP_average	0.04	0.04	0.06	0.05	0.13	0.24	1.21	99.90	3.10	19.18	97.23	100.00
					$\mathbf{N = 160, p = 400}$							
New	4.93	4.98	4.99	4.88	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.74	5.26	4.89	4.77	87.20	19.91	12.09	9.73	100.00	100.00	99.95	89.50
BS	6.66	6.71	6.69	6.71	94.00	26.42	16.69	13.78	100.00	100.00	100.00	99.41
CQ	6.66	6.71	6.69	6.71	94.02	26.45	16.69	13.76	100.00	100.00	100.00	99.39
SD_withAdjust	3.11	0.99	0.34	0.07	50.59	3.63	0.72	0.17	100.00	92.93	7.69	1.27
SD_noAdjust	6.87	6.83	6.71	6.65	94.39	26.76	16.83	13.72	100.00	100.00	100.00	99.36
Lauter's PC	4.76	4.74	4.74	4.73	5.98	5.23	4.95	4.87	7.08	5.88	5.32	5.11
RP_single	4.81	5.15	4.99	4.84	98.07	99.92	100.00	100.00	100.00	100.00	100.00	100.00
RP_average	0.01	0.02	0.01	0.02	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
					$\mathbf{N = 160, p = 1600}$							
New	4.54	4.48	4.44	4.50	98.94	99.96	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.76	5.22	4.87	4.77	11.15	7.32	6.11	5.68	93.07	19.49	11.60	9.41
BS	6.71	6.69	6.69	6.69	13.09	9.46	8.37	8.11	97.90	26.09	16.49	13.60
CQ	6.71	6.69	6.70	6.70	13.10	9.47	8.37	8.11	97.91	26.11	16.48	13.61
SD_withAdjust	2.10	0.40	0.05	0.02	3.82	0.53	0.05	0.02	29.18	1.19	0.15	0.03
SD_noAdjust	6.90	6.82	6.71	6.66	13.48	9.46	8.43	8.09	98.05	26.51	16.39	13.53
Lauter's PC	4.72	4.73	4.73	4.74	4.76	4.71	4.69	4.71	4.93	4.73	4.73	4.71
RP_single	5.23	4.83	4.80	4.70	34.28	55.48	91.85	100.00	98.27	99.95	100.00	100.00
RP_average	0.00	0.00	0.00	0.00	41.05	97.30	100.00	100.00	100.00	100.00	100.00	100.00

Table 3.3: Comparison for one-sample tests: multivariate normal with Σ_2

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N = 40, p = 400}$							
New	5.01	5.04	5.02	5.02	46.81	34.83	20.93	13.46	99.99	99.52	91.13	68.01
Dempster	5.06	4.97	4.75	5.30	89.47	77.24	51.45	17.29	100.00	100.00	99.96	84.57
BS	5.57	5.57	5.46	6.86	90.19	78.40	53.88	20.81	100.00	100.00	99.99	88.16
CQ	5.59	5.57	5.44	6.85	90.16	78.39	53.83	20.81	100.00	100.00	99.99	88.15
SD_withAdjust	3.75	3.68	3.30	2.72	84.86	70.93	44.71	9.94	100.00	100.00	99.85	68.93
SD_noAdjust	7.25	7.28	7.61	8.52	90.57	80.54	57.97	23.86	100.00	100.00	99.96	87.61
Lauter's PC	4.69	4.67	4.93	4.96	36.78	28.64	16.49	6.63	92.38	77.29	40.81	9.26
RP_single	5.52	5.11	5.00	4.97	12.71	12.15	11.51	15.28	44.17	43.04	42.40	60.42
RP_average	0.02	0.02	0.13	1.80	1.36	2.35	4.14	18.45	96.36	93.38	86.98	97.08
					$\mathbf{N = 40, p = 1600}$							
New	4.93	5.07	4.83	5.16	17.36	13.68	9.80	6.53	94.73	84.77	58.18	22.55
Dempster	4.91	5.14	4.88	4.74	48.45	37.63	23.47	9.96	99.99	99.91	94.58	42.28
BS	5.05	5.46	5.36	5.49	49.13	38.40	24.63	11.40	99.99	99.91	94.96	45.81
CQ	5.08	5.48	5.29	5.50	49.26	38.35	24.60	11.44	99.99	99.91	94.94	45.74
SD_withAdjust	1.77	1.91	2.04	1.81	30.97	22.82	12.73	3.66	99.92	99.03	86.28	23.53
SD_noAdjust	7.04	7.13	7.19	7.55	53.38	43.79	29.11	14.45	99.98	99.79	95.07	50.80
Lauter's PC	4.92	5.11	5.08	4.99	15.69	13.57	9.77	6.17	45.99	34.47	19.55	7.99
RP_single	4.61	4.99	4.87	4.89	6.04	6.47	6.17	6.68	11.71	12.12	11.46	13.14
RP_average	0.00	0.00	0.00	0.23	0.00	0.00	0.00	0.55	0.59	0.46	0.99	6.55
					$\mathbf{N = 160, p = 400}$							
New	4.77	4.95	4.76	4.91	100.00	99.50	90.60	96.54	100.00	100.00	100.00	100.00
Dempster	4.61	4.97	5.12	5.34	100.00	100.00	100.00	85.83	100.00	100.00	100.00	100.00
BS	5.03	5.50	5.83	6.61	100.00	100.00	100.00	89.10	100.00	100.00	100.00	100.00
CQ	5.03	5.49	5.83	6.62	100.00	100.00	100.00	89.10	100.00	100.00	100.00	100.00
SD_withAdjust	4.20	4.42	4.17	2.73	100.00	100.00	100.00	72.60	100.00	100.00	100.00	100.00
SD_noAdjust	5.41	5.78	6.19	6.93	100.00	100.00	100.00	88.85	100.00	100.00	100.00	100.00
Lauter's PC	4.87	4.71	4.70	5.00	89.99	71.70	34.04	7.34	100.00	100.00	71.60	10.28
RP_single	4.65	4.95	4.75	5.27	89.44	85.36	80.43	98.54	100.00	100.00	100.00	100.00
RP_average	0.00	0.08	0.66	1.66	100.00	99.84	97.95	100.00	100.00	100.00	100.00	100.00
					$\mathbf{N = 160, p = 1600}$							
New	4.50	4.82	4.60	4.84	97.50	89.89	61.44	36.12	100.00	100.00	100.00	99.54
Dempster	4.73	4.72	4.99	5.11	99.99	99.89	95.03	42.55	100.00	100.00	100.00	100.00
BS	4.86	5.00	5.30	5.98	100.00	99.90	95.35	45.67	100.00	100.00	100.00	100.00
CQ	4.86	4.98	5.29	5.99	100.00	99.90	95.35	45.66	100.00	100.00	100.00	100.00
SD_withAdjust	3.47	3.46	3.57	2.70	100.00	99.83	93.02	29.91	100.00	100.00	100.00	99.99
SD_noAdjust	5.40	5.48	5.65	6.33	100.00	99.87	95.47	46.65	100.00	100.00	100.00	100.00
Lauter's PC	5.27	5.08	4.84	4.78	42.34	31.61	16.88	6.42	97.49	83.24	39.61	9.04
RP_single	4.86	4.67	4.56	5.45	25.35	24.48	23.41	37.24	92.06	90.94	90.47	98.77
RP_average	0.00	0.00	0.00	0.58	8.88	9.63	12.68	49.05	100.00	100.00	99.99	100.00

Table 3.4: Comparison for one-sample tests: multivariate normal with Σ_3

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N = 40, p = 400}$							
New	5.04	4.97	5.24	5.19	42.03	41.23	34.91	29.71	99.04	98.27	97.16	93.76
Dempster	6.82	6.68	6.38	6.02	26.75	12.64	9.38	8.05	99.93	78.81	38.79	21.82
BS	7.37	7.73	7.75	7.86	29.27	14.57	11.73	10.39	99.95	86.13	48.62	29.47
CQ	7.40	7.71	7.71	7.89	29.30	14.54	11.68	10.35	99.96	86.09	48.67	29.42
SD_withAdjust	5.39	4.27	2.71	1.32	20.68	8.06	4.18	1.88	99.61	52.80	15.70	5.47
SD_noAdjust	8.15	8.36	8.34	8.27	33.33	15.93	12.72	10.85	99.96	88.48	52.41	32.24
Lauter's PC	5.07	5.17	5.13	5.19	6.12	5.57	5.33	5.22	8.10	6.30	5.80	5.57
RP_single	4.86	5.20	4.95	5.06	12.83	14.60	15.63	26.31	48.15	53.40	60.76	86.04
RP_average	0.00	0.03	0.17	1.45	2.52	5.21	12.10	45.20	98.58	99.56	99.78	99.99
					$\mathbf{N = 40, p = 1600}$							
New	4.76	4.89	5.20	5.42	13.05	12.51	11.88	8.75	70.52	67.60	65.63	45.33
Dempster	7.17	6.90	6.48	6.19	9.58	7.78	7.18	6.81	27.30	12.45	9.31	8.28
BS	7.74	7.80	7.80	7.78	10.34	9.01	8.46	8.30	30.00	14.44	11.59	10.42
CQ	7.76	7.81	7.78	7.72	10.28	8.95	8.46	8.29	30.05	14.45	11.54	10.40
SD_withAdjust	4.19	2.70	1.44	0.73	5.79	3.18	1.61	0.77	15.53	5.26	2.19	1.04
SD_noAdjust	8.48	8.43	8.32	8.19	11.70	9.76	9.05	8.82	34.82	15.94	12.59	11.17
Lauter's PC	5.10	5.14	5.19	5.20	5.39	5.27	5.21	5.21	5.71	5.44	5.30	5.31
RP_single	5.00	4.84	4.79	5.23	6.68	6.81	7.05	8.80	13.07	14.15	16.65	23.86
RP_average	0.00	0.01	0.04	0.42	0.04	0.10	0.19	1.35	1.45	3.11	8.49	32.33
					$\mathbf{N = 160, p = 400}$							
New	4.93	4.68	4.60	4.99	100.00	100.00	99.93	99.99	100.00	100.00	100.00	100.00
Dempster	5.98	5.73	5.44	5.04	100.00	83.26	33.73	18.91	100.00	100.00	100.00	100.00
BS	6.45	6.67	6.73	6.72	100.00	90.99	43.97	25.76	100.00	100.00	100.00	100.00
CQ	6.47	6.67	6.72	6.72	100.00	91.00	43.98	25.74	100.00	100.00	100.00	100.00
SD_withAdjust	4.92	3.04	1.81	0.82	99.99	48.74	10.03	2.98	100.00	100.00	99.99	78.09
SD_noAdjust	6.70	6.80	6.89	6.85	100.00	91.47	45.28	26.38	100.00	100.00	100.00	100.00
Lauter's PC	4.70	4.75	4.75	4.74	7.09	6.01	5.56	5.20	10.27	7.09	6.21	5.83
RP_single	5.36	5.28	4.97	4.91	94.31	96.07	97.07	100.00	100.00	100.00	100.00	100.00
RP_average	0.00	0.01	0.25	0.96	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
					$\mathbf{N = 160, p = 1600}$							
New	4.54	4.44	4.41	4.72	97.44	97.59	94.35	79.76	100.00	100.00	100.00	100.00
Dempster	6.12	5.77	5.45	5.25	23.85	11.17	8.51	7.42	100.00	91.94	33.55	20.31
BS	6.66	6.68	6.70	6.77	26.24	13.24	10.44	9.46	100.00	97.40	44.83	27.33
CQ	6.67	6.68	6.70	6.77	26.26	13.22	10.41	9.44	100.00	97.42	44.81	27.32
SD_withAdjust	4.13	2.06	0.84	0.39	15.66	3.70	1.26	0.58	100.00	29.28	4.52	1.28
SD_noAdjust	6.83	6.80	6.85	6.86	27.30	13.49	10.70	9.61	100.00	97.42	46.28	28.33
Lauter's PC	4.69	4.71	4.75	4.74	4.93	4.78	4.71	4.70	5.44	4.91	4.79	4.74
RP_single	4.97	5.30	5.41	5.20	28.38	33.95	40.01	68.67	94.88	97.86	99.36	99.99
RP_average	0.00	0.00	0.00	0.47	18.49	35.78	58.69	95.91	100.00	100.00	100.00	100.00

3.4 Extension to non-normal distributions

3.4.1 Optimal projection direction

In this section, we further investigate the proposed projection test with extension to non-normal distributions. It is known that under mild conditions, T_A^2 has an asymptotic noncentral χ^2-distribution with k degrees of freedom and noncentrality parameter $N\delta_A$. The optimal direction can be obtained using the similar techniques as the normal case. Thus, we have the following theorem.

Theorem 3.4.1. *Suppose that* $\mathbf{x}_1, \cdots, \mathbf{x}_N$ *is a random sample from a population with mean* $\boldsymbol{\mu}$ *and covariance matrix* Σ. *Further assume that for any nonzero constant* $p \times k$ *matrix* A *with a fixed* k, $A^T\bar{\mathbf{x}} \to N(A^T\boldsymbol{\mu}, A^T\Sigma A)$ *in distribution and* $A^T\mathbf{S}A - A^T\Sigma A \to \mathbf{0}$ *in probability as* $N \to \infty$. *The projection test* T_A^2 *defined in (3.3) for the one-sample problem (3.1) reaches its asymptotic best power for* H_1 *in (3.1) at* $k = 1$ *and* $A = \Sigma^{-1}\boldsymbol{\mu}$.

Proof. Under some mild regularity conditions T_A^2 follows the $\chi_k^2(N\delta_A)$ distribution, the noncentral χ^2-distribution with k degrees of freedom and noncentrality parameter $N\delta_A$, where $\delta_A = \boldsymbol{\mu}^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu}$. Using the property of χ^2-distribution as Lemma 3.6.6 (shown in Kallenberg, 1990) , the power of T_A^2 increases with δ_A for given k and N. As a result, we want to maximize δ_A with respect to A in order to maximize the power of T_A^2. Denote by $\chi_{k;\alpha}^2$ the critical value of the χ^2-distribution with k degrees of freedom. By Lemma 3.6.5 (Theorem 2 in Ghosh, 1973), for a given level α, $P\{\chi_k^2(N\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}) > \chi_{k;\alpha}^2\}$ is a decreasing function of k when $N\boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}$ is fixed. Using this property of χ^2-distribution and exact the same argument as those used in the proof of Theorem 3.2.1, it can be shown that $k = 1$ with $A = \Sigma^{-1}\boldsymbol{\mu}$ is the best choice to achieve the optimal power. This completes the proof of Theorem 3.4.1.

□

Theorem 3.4.1 implies that we may construct a projection χ^2-test by projecting the original sample along the direction $\hat{\Sigma}^{-1}\hat{\boldsymbol{\mu}}$ for the one-sample problem. Thus, the single sample-splitting strategy used for the projection Hotelling's T^2 test can be used to construct a projection χ^2-test.

3.4.2 Simulation results

In this section, we conduct simulations to investigate the performance of the algorithm for non-normal populations. To this end, we generate random samples from the multivariate t-distribution. Denote by $t_\nu(\boldsymbol{\mu}, \Sigma)$ the multivariate-t distribution with mean vector $\boldsymbol{\mu}$, covariance matrix Σ, and the degrees of freedom ν. The construction of $t_\nu(\boldsymbol{\mu}, \Sigma)$ is based on the fact that given $\boldsymbol{z} \sim N(\mathbf{0}, \Sigma)$, $u \sim \chi^2_\nu$, and u is independent from $\boldsymbol{z}$, then $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{z}/\sqrt{u/\nu}$ is distributed as $t_\nu(\boldsymbol{\mu}, \Sigma)$. The other simulation settings are the same as used in experiment 3 in Section 3.3.3. Percentage of rejection based on 10000 replicates for three correlation structures with various configurations of (N, p, c) are shown in Tables 3.5-3.7. Since the sample sizes in some of the settings are small, we use the small sample correction which applies the t-test. In this case, we can also examine the robustness of the tests on the normality assumption.

We first take a look at the Type I error at the columns with $c = 0$. We note that the proposed test still keeps the Type I error well. The other projection tests seem to be problematic in different ways. The RP test with single projection tends to be a little conservative in general. The Lauter test is well behaved in the compound symmetry case and the combined case, but turns extremely conservative for the autocorrelation case. Similar extreme situation for autocorrelation structure is also observed for BS test and SD test. On the other hand, CQ test tends to have an improved Type I error with autocorrelation structure.

Generally, the power performance is less satisfied compared with cases of multivariate normal as expected. The patterns of power under compound symmetry structure and the combined structure are very similar to their multivariate normal counterpart. For autocorrelation, we recall that Dempster test, BS test, CQ test and SD test are powerful in the multivariate normal cases. In Table 3.6, however, we observe that Dempster test, BS test and SD test break down in cases where $(N, p, c) = (40, 1600)$ or $(N, p, c) = (40, 400, 0.5)$. The CQ test keeps momentum and generally outperforms the proposed test. However, the power of the proposed test increases dramatically with the signal strength c and the sample size N.

Table 3.5: Comparison for one-sample tests: multivariate t with Σ_1

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N=40,p=400}$							
New	5.09	4.94	4.88	4.71	37.02	57.48	89.13	99.94	96.51	98.59	99.86	100.00
Dempster	5.07	5.44	5.16	5.05	8.52	7.12	6.27	5.84	38.09	14.37	9.86	8.29
BS	6.25	7.41	7.66	7.76	10.01	9.24	8.91	8.71	45.99	19.11	13.95	12.35
CQ	7.85	7.96	7.93	7.94	12.19	9.75	9.17	8.92	55.22	20.18	14.47	12.58
SD_withAdjust	2.97	1.45	0.45	0.15	4.80	1.83	0.50	0.18	22.74	4.06	0.84	0.27
SD_noAdjust	7.14	7.86	7.74	7.57	11.36	9.75	9.06	8.52	54.58	21.00	14.34	12.11
Lauter	4.66	4.80	4.81	4.82	5.17	4.95	4.85	4.79	5.89	5.34	5.00	4.90
RP_single	4.33	4.15	4.36	4.23	11.65	16.81	34.72	96.43	45.75	65.28	91.90	100.00
RP_average	0.01	0.04	0.03	0.06	1.81	9.90	79.92	100.00	98.10	99.99	100.00	100.00
					$\mathbf{N=40,p=1600}$							
New	4.72	4.62	4.69	4.71	11.88	18.94	44.98	96.04	61.34	74.40	92.50	99.97
Dempster	5.44	5.81	5.57	5.37	6.04	6.09	5.76	5.54	8.52	7.16	6.36	5.99
BS	6.53	7.45	7.74	7.82	7.25	7.89	8.03	8.03	10.05	9.41	9.08	8.93
CQ	7.86	7.96	7.95	7.96	8.92	8.35	8.25	8.24	12.24	9.87	9.30	9.10
SD_withAdjust	1.69	0.63	0.12	0.00	1.92	0.69	0.12	0.00	2.76	0.79	0.15	0.00
SD_noAdjust	7.31	7.97	7.85	7.64	8.16	8.42	8.12	7.88	11.55	9.90	9.14	8.68
Lauter	5.12	5.18	5.20	5.21	5.18	5.14	5.16	5.19	5.28	5.17	5.14	5.15
RP_single	4.35	4.36	4.39	4.29	5.93	7.02	9.91	45.00	12.36	17.23	36.15	96.25
RP_average	0.01	0.02	0.02	0.02	0.03	0.11	0.52	97.53	1.16	8.83	84.39	100.00
					$\mathbf{N=160,p=400}$							
New	4.86	4.77	4.90	4.76	99.99	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.46	5.18	4.97	4.89	36.38	12.65	8.92	7.70	99.99	98.68	63.94	34.77
BS	6.43	6.73	6.80	6.84	44.60	16.77	12.38	11.03	99.99	99.73	85.32	54.53
CQ	6.82	6.86	6.87	6.89	48.30	17.04	12.51	11.13	99.99	99.78	86.04	55.09
SD_withAdjust	2.99	0.98	0.28	0.14	19.04	2.45	0.43	0.21	99.91	37.91	2.91	0.50
SD_noAdjust	6.57	6.83	6.84	6.83	47.67	17.29	12.49	10.95	99.99	99.76	85.83	54.02
Lauter	4.80	4.84	4.84	4.83	5.65	5.18	5.05	4.95	6.73	5.74	5.38	5.15
RP_single	4.31	4.21	4.17	4.36	94.90	99.49	100.00	100.00	100.00	100.00	100.00	100.00
RP_average	0.01	0.00	0.00	0.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
					$\mathbf{N=160,p=1600}$							
New	4.27	4.64	4.69	4.69	96.30	99.74	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.28	4.90	4.66	4.60	8.34	6.32	5.51	5.29	36.94	12.33	8.68	7.47
BS	6.29	6.59	6.66	6.69	9.75	8.20	7.76	7.56	45.75	16.45	12.13	10.80
CQ	6.74	6.71	6.69	6.71	10.46	8.36	7.87	7.61	49.87	16.73	12.27	10.90
SD_withAdjust	1.89	0.35	0.09	0.01	2.71	0.38	0.10	0.02	10.66	0.81	0.10	0.04
SD_noAdjust	6.54	6.68	6.70	6.67	10.18	8.42	7.83	7.51	49.57	16.80	12.13	10.68
Lauter	4.54	4.54	4.57	4.56	4.86	4.71	4.64	4.61	5.04	4.83	4.78	4.71
RP_single	4.41	4.35	4.23	4.35	28.75	46.19	85.36	100.00	95.08	99.65	100.00	100.00
RP_average	0.00	0.00	0.00	0.00	16.62	82.52	100.00	100.00	100.00	100.00	100.00	100.00

Table 3.6: Comparison for one-sample tests: multivariate t with Σ_2

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N=40, p=400}$							
New	4.92	4.72	5.15	4.55	34.49	26.51	16.07	10.81	99.32	96.11	79.58	55.31
Dempster	0.05	0.17	0.80	3.27	10.04	11.00	11.39	8.56	92.77	91.15	83.76	48.47
BS	0.08	0.25	1.02	4.65	12.42	13.43	14.05	11.29	94.45	92.97	86.68	55.35
CQ	5.49	5.71	5.83	6.77	68.47	55.52	36.29	15.44	100.00	99.97	97.97	64.15
SD_withAdjust	0.04	0.05	0.37	1.34	7.13	8.03	8.00	3.84	91.59	89.34	79.00	33.06
SD_noAdjust	0.16	0.48	1.57	5.99	20.35	20.70	19.99	14.67	97.78	96.67	91.34	61.40
Lauter	0.46	0.76	1.66	4.00	1.53	2.31	3.49	4.98	5.90	6.81	7.87	6.73
RP_single	3.85	4.40	4.34	4.12	10.16	10.22	10.12	13.08	37.12	36.14	35.35	51.57
RP_average	0.00	0.00	0.03	1.26	0.39	0.73	2.20	12.52	84.30	79.70	73.11	91.40
					$\mathbf{N=40, p=1600}$							
New	4.54	4.64	4.91	4.78	13.20	11.24	8.16	6.00	82.95	70.16	44.58	17.93
Dempster	0.00	0.00	0.02	1.13	0.00	0.00	0.22	2.09	7.62	7.79	9.38	9.44
BS	0.00	0.00	0.06	1.58	0.00	0.00	0.30	2.72	9.59	9.90	11.96	11.86
CQ	5.07	5.16	5.23	5.93	30.83	24.57	16.62	9.58	98.44	94.23	75.89	29.10
SD_withAdjust	0.00	0.00	0.00	0.11	0.00	0.00	0.02	0.30	1.74	2.05	2.91	2.41
SD_noAdjust	0.00	0.00	0.13	2.51	0.00	0.05	0.57	4.55	18.33	18.81	19.84	17.45
Lauter	0.05	0.12	0.40	2.08	0.10	0.18	0.52	2.30	0.21	0.40	0.80	2.87
RP_single	4.26	4.24	4.22	4.18	5.60	5.31	5.49	6.01	10.32	9.83	9.97	11.53
RP_average	0.00	0.00	0.00	0.11	0.00	0.00	0.01	0.29	0.13	0.19	0.26	3.61
					$\mathbf{N=160, p=400}$							
New	4.66	4.86	4.90	4.95	99.69	96.53	78.24	88.90	100.00	100.00	100.00	100.00
Dempster	0.41	1.03	2.41	4.33	99.52	99.02	95.86	53.53	99.99	99.99	99.99	99.96
BS	0.52	1.30	2.97	5.57	99.61	99.27	96.64	59.70	99.99	100.00	99.99	99.98
CQ	5.32	5.68	5.75	6.38	99.99	99.92	98.91	62.70	100.00	100.00	100.00	100.00
SD_withAdjust	0.31	0.81	1.70	2.14	99.59	99.00	94.47	37.44	99.99	100.00	99.99	99.93
SD_noAdjust	0.77	1.48	3.28	6.09	99.80	99.46	96.90	61.76	100.00	100.00	100.00	99.99
Lauter	0.92	1.60	2.87	4.35	6.34	8.40	9.53	5.81	24.25	25.56	22.83	8.63
RP_single	4.55	4.42	4.36	4.38	82.42	76.72	70.40	95.73	100.00	100.00	100.00	100.00
RP_average	0.00	0.02	0.23	0.88	99.94	98.87	93.69	99.97	100.00	100.00	100.00	100.00
					$\mathbf{N=160, p=1600}$							
New	4.50	4.78	4.74	4.83	89.31	74.86	46.17	27.09	100.00	100.00	99.98	97.38
Dempster	0.00	0.03	0.40	3.04	43.74	40.56	33.32	17.30	99.72	99.67	99.64	95.47
BS	0.00	0.07	0.47	3.80	47.26	43.91	36.50	19.82	99.78	99.74	99.75	96.40
CQ	4.98	4.93	5.16	6.08	98.83	95.34	75.76	27.93	100.00	100.00	100.00	98.82
SD_withAdjust	0.00	0.00	0.16	1.22	33.20	30.69	24.45	9.06	99.59	99.57	99.37	88.00
SD_noAdjust	0.00	0.05	0.57	4.16	53.10	49.61	40.86	21.79	99.92	99.86	99.90	96.36
Lauter	0.15	0.26	0.65	3.31	0.29	0.66	1.22	4.07	0.95	1.56	2.51	5.19
RP_single	4.48	3.98	4.32	4.05	19.83	19.19	18.86	30.07	84.29	84.13	82.51	96.48
RP_average	0.00	0.00	0.00	0.44	2.43	3.08	5.69	33.96	100.00	100.00	100.00	100.00

Table 3.7: Comparison for one-sample tests: multivariate t with Σ_3

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{N=40, p=400}$							
New	5.01	5.19	5.24	4.88	31.76	31.21	27.25	22.56	96.29	95.28	92.42	85.93
Dempster	3.39	5.02	5.51	5.40	8.88	8.38	7.51	6.95	78.22	37.80	20.45	13.58
BS	3.95	6.21	7.09	7.31	10.61	9.87	9.48	9.14	82.84	45.72	26.91	18.78
CQ	7.39	7.79	7.86	7.73	19.60	12.02	10.35	9.67	95.36	54.45	29.48	20.02
SD_withAdjust	2.32	2.99	2.12	1.11	6.30	4.75	3.04	1.46	68.31	22.40	8.13	3.36
SD_noAdjust	4.89	7.07	7.71	7.74	13.44	11.53	10.47	9.61	90.14	54.13	30.77	20.85
Lauter	4.07	4.66	4.86	4.83	4.88	5.13	5.05	4.92	6.48	5.94	5.58	5.30
RP_single	4.11	4.41	3.92	4.70	10.21	11.67	12.90	22.49	40.10	45.59	52.01	78.51
RP_average	0.01	0.01	0.10	1.05	0.99	2.53	6.07	33.46	91.26	95.85	97.57	99.89
					$\mathbf{N=40, p=1600}$							
New	4.97	4.68	5.08	5.06	10.21	10.33	9.94	7.48	59.66	57.14	54.29	36.00
Dempster	3.55	5.45	5.70	5.73	4.40	6.01	6.18	6.11	8.72	8.47	7.65	7.19
BS	4.09	6.49	7.17	7.42	5.15	7.21	7.76	7.86	10.40	10.02	9.63	9.46
CQ	7.78	7.89	7.92	7.82	9.68	8.82	8.59	8.28	20.33	12.28	10.64	9.88
SD_withAdjust	1.59	1.63	1.03	0.65	1.89	1.88	1.11	0.68	4.16	2.74	1.48	0.84
SD_noAdjust	5.04	7.22	7.82	7.85	6.30	8.12	8.46	8.37	13.12	11.46	10.54	9.99
Lauter	4.45	5.11	5.15	5.18	4.57	5.11	5.18	5.14	4.76	5.25	5.24	5.20
RP_single	4.27	4.09	4.21	4.36	5.68	5.62	6.36	7.28	11.38	11.96	14.38	19.23
RP_average	0.00	0.00	0.00	0.17	0.00	0.03	0.07	0.75	0.64	1.28	3.62	19.59
					$\mathbf{N=160, p=400}$							
New	4.69	4.83	4.66	5.02	99.96	99.82	98.91	99.73	100.00	100.00	100.00	100.00
Dempster	4.85	5.47	5.34	4.99	93.01	36.48	17.99	12.35	99.98	99.96	99.84	95.43
BS	5.55	6.51	6.72	6.82	95.22	44.56	22.75	16.47	99.99	99.99	99.92	98.69
CQ	6.78	6.93	6.93	6.91	97.98	47.73	23.55	16.82	100.00	99.99	99.96	98.93
SD_withAdjust	3.72	2.94	1.68	0.78	87.30	19.33	6.15	2.02	99.99	99.89	88.11	28.52
SD_noAdjust	5.89	6.76	6.88	6.90	96.54	46.91	23.64	16.91	100.00	99.99	99.93	98.25
Lauter	4.68	4.86	4.83	4.83	6.51	5.72	5.33	5.20	9.13	6.63	6.08	5.66
RP_single	4.46	4.39	4.29	4.38	87.88	90.63	92.57	99.90	100.00	100.00	100.00	100.00
RP_average	0.01	0.01	0.10	0.51	99.99	100.00	99.99	100.00	100.00	100.00	100.00	100.00
					$\mathbf{N=160, p=1600}$							
New	4.51	4.64	4.72	4.81	91.20	91.58	83.69	64.67	100.00	100.00	100.00	100.00
Dempster	4.67	5.30	5.09	4.90	11.60	8.30	6.98	6.23	97.17	36.84	17.48	12.69
BS	5.28	6.38	6.49	6.60	13.14	9.78	8.83	8.37	98.13	46.23	22.37	16.87
CQ	6.70	6.71	6.68	6.70	16.74	10.56	9.17	8.48	99.67	50.00	23.19	17.27
SD_withAdjust	2.94	1.87	0.79	0.35	6.89	2.67	1.12	0.39	85.63	10.63	2.55	0.85
SD_noAdjust	5.57	6.52	6.64	6.73	14.08	10.26	9.10	8.60	98.82	49.48	23.29	17.56
Lauter	4.46	4.54	4.57	4.52	4.91	4.81	4.74	4.69	5.20	5.01	4.88	4.80
RP_single	4.34	4.45	3.97	4.47	23.40	27.21	31.95	58.09	90.09	94.17	97.28	99.94
RP_average	0.00	0.00	0.00	0.27	6.12	15.53	34.56	87.79	100.00	100.00	100.00	100.00

3.5 Real data example

In this section, we demonstrate the proposed projection test with an empirical analysis of a high resolution micro-computed tomography dataset which contains the bone volume measured at many different bone density levels in a genetic mutation study. Data were collected at Center for Quantitative X-Ray Imaging at Pennsylvania State University. See Percival et al. (2014) for detailed description of protocols. In this empirical analysis, we compare the performance of the proposed projection test with several competing ones, on the comparative study for the bone density pattern of two adjacent bones in mice's skull.

We first normalize bone volume by dividing the bone volume at each bone density level by the total bone volume across all the density levels to get rid of the total bone size effect such that the two bones are comparable. Then we prepare the data for one-sample test by taking difference of the two normalized bone volume at the corresponding density level for each subject. The volume difference at bone intensity levels from 10 to 100 are used in this analysis. There are total 22 samples available for this analysis. Thus, the dimension and the sample size of this data set are $p = 91$ and $N = 22$, respectively.

Table 3.8: Bone volume dataset: p-values of one-sample tests

δ	New	D1958	BS1996	CQ2010	SD2008w	SD2008wo	L1996	LJW2011	LJW2012
1	0	0	0	0	0	0	0	0	0
0.8	0	0	0	0	0	0	0	0.0005	0
0.6	0	0	0	0	0	0	0	0.0007	0
0.4	0.001	.0001	0	0	0	0	0.0008	0.0713	0.0002
0.2	0.003	0.098	0.0771	0.0782	0.2482	0.0583	0.1003	0.5337	0.6499

We first apply the testing procedures that are included in experiments 3 and 4 to this set. To implement the proposed method, we set the value of λ to $(22 \times 0.4)^{-0.5} = 0.34$ and the splitting percentage as 40%. The p-values are reported in Table 3.8. All the p-values are very small, implying that the growth pattern is significantly different. To compare the powers of different tests under this specific setting, we demonstrate the how p-values of the tests change as the signal becomes weaker. To this end, we set $\bar{\mathbf{x}}$ to be the same mean, and $\boldsymbol{r}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, the residual vector of the i-th subject, and then construct new data $\boldsymbol{z}_i = \delta\bar{\mathbf{x}}+\boldsymbol{r}_i$ for $i = 1, \cdots, N$. By the construction, smaller δ results in weaker signals, and would make the test more challenging. Table 3.8 depicts p-values of all tests on new data with $\delta = 1, 0.8, \cdots, 0.2$. As expected, the p-value of each test increases as δ decreases.

The single random projection test fails to reject H_0 at significant level 0.05 when $\delta = 0.4$. All tests but the newly proposed projection test fail to reject the null hypothesis at significant level 0.05 when $\delta = 0.2$. We plot the absolute values of the upper triangular elements of the sample correlation matrix in Figure 3.6, which indicates that there exist high correlations among variables. This may be the reason why the proposed projection test is more powerful than Dempster test, BS test and SD test as these tests ignore the correlation among variables.

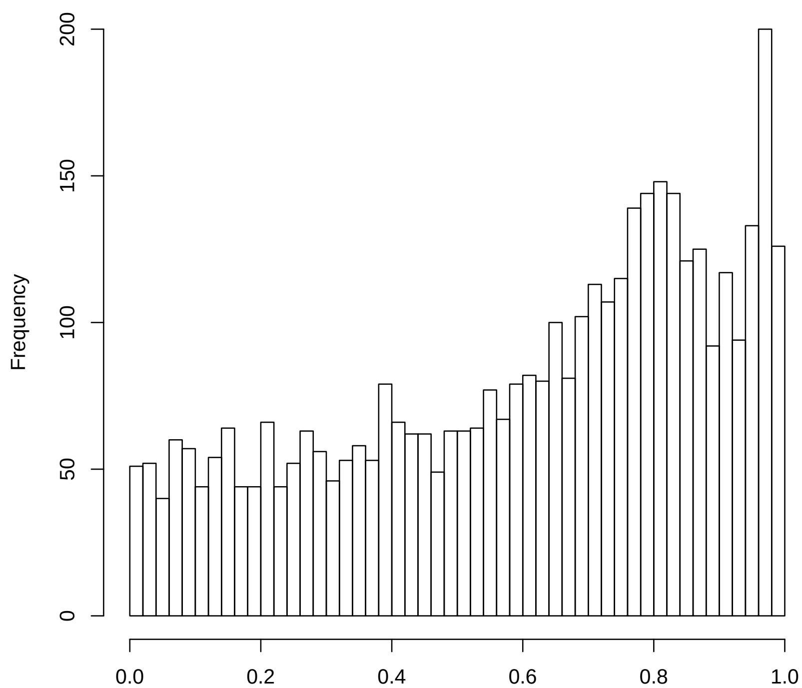


Figure 3.6: Histogram of absolute values of paired sample correlations among bone volumes at all different bone density levels.

3.6 Proofs

3.6.1 Projection to one dimensional space

In this section, we consider the projection to a one-dimensional space and show that the corresponding optimal direction is consistent with the findings in Theorem 3.2.1. In this special case, we search for a vector $\mathbf{a}$ of length d such that the test based on linear score of $X\mathbf{a}$ gives highest power and is equivalent to the original test.

Denote $\mathbf{y} = X\mathbf{a}$ such that $\mathbf{y}$ is a vector of length N. Denote the i-th element as y_i and they are identically and independently distributed as $N(\mu_Y, \sigma_Y^2)$, where $\mu_Y = \mathbf{a}'\boldsymbol{\mu}, \sigma_Y^2 = \mathbf{a}'\Sigma\mathbf{a}$. For testing the univariate mean $H_0^{\mathbf{a}} : \mu_Y = 0$, we use z-test with known Σ and t-test unknown Σ. In each case, we show that the power function is increasing with noncentrality parameter and the maximum power can be attained by taking the projection direction to be $\Sigma^{-1}\boldsymbol{\mu}$.

Lemma 3.6.1. *Let $y_1, \ldots, y_N$ be random sample from $N(\mu_Y, \sigma_Y^2)$, where σ_Y^2 is known. The power of the z-test to $H_0 : \mu_Y = 0$ increases with $|\mu_Y|/\sigma_Y$.*

Proof. z-test is applied to test $H_0 : \mu_Y = 0$ with the test statistic $Z = \sqrt{N}\bar{y}/\sigma_Y$. Denote $\delta = \sqrt{N}\mu_Y/\sigma_Y$. Consider

$$
\begin{aligned}
P(|Z| < z_{\alpha/2}) &= P(|Z| < z_{\alpha/2}) \\
&= P\left(-z_{\alpha/2} < \frac{\sqrt{N}\bar{y}}{\sigma} < z_{\alpha/2}\right) \\
&= P\left(-z_{\alpha/2} - \frac{\sqrt{N}\mu_Y}{\sigma_Y} < \frac{\sqrt{N}(\bar{y} - \mu_Y)}{\sigma_Y} < z_{\alpha/2} - \frac{\sqrt{N}\mu_Y}{\sigma_Y}\right) \\
&= \Phi\left(z_{\alpha/2} - \delta\right) - \Phi\left(-z_{\alpha/2} - \delta\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}-\delta}^{z_{\alpha/2}-\delta} e^{-x^2/2} dx.
\end{aligned}
$$

Denote the power function as

$$\beta(\delta) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-z_{\alpha/2}-\delta}^{z_{\alpha/2}-\delta} e^{-x^2/2} dx.$$

By Leibniz's rule,

$$\frac{\partial\beta(\delta)}{\partial\delta} = \frac{1}{\sqrt{2\pi}}\left\{\exp\left(-\frac{1}{2}(z_{\alpha/2}-\delta)^2\right) - \exp\left(-\frac{1}{2}(z_{\alpha/2}+\delta)^2\right)\right\}.$$

It follows that

$$\frac{\partial\beta(\delta)}{\partial\delta}\begin{cases} <0 & \text{for} \quad \delta<0 \\ >0 & \text{for} \quad \delta>0 \end{cases}$$

Hence, $\beta(\delta)$ is a monotone increasing function of $|\delta|$. □

Lemma 3.6.2. *Let $y_1,\ldots,y_N$ be random sample from $N(\mu_Y,\sigma_Y^2)$, where σ_Y^2 is unknown. The power of the t-test to $H_0:\mu_Y=0$ increases with $|\mu_Y|/\sigma_Y$.*

Proof. t-test is applied to test $H_0:\mu_Y=0$ with the test statistic $t=\sqrt{N}\bar{y}/s_Y$, $s_Y^2=\sum(y_i-\bar{y})^2/(N-1)$. Denote $\delta=\sqrt{N}\mu_Y/\sigma_Y$. Consider

$$\begin{aligned} P(|t|<t_{\alpha/2,N-1}) &= P\left(|t|<t_{\alpha/2,N-1}\right) \\ &= P\left(-t_{\alpha/2,N-1}<\frac{\sqrt{N}\bar{y}}{s_Y}<t_{\alpha/2,N-1}\right) \\ &= P\left(-t_{\alpha/2,N-1}<\frac{\bar{y}-\mu_Y+\mu_Y}{\sqrt{\sigma_Y^2/N}\sqrt{\frac{s_Y^2/N}{\sigma_Y^2/N}}}<t_{\alpha/2,N-1}\right) \\ &= P\left(-t_{\alpha/2,N-1}<\frac{\frac{\bar{y}-\mu_Y}{\sqrt{\sigma_Y^2/N}}+\delta}{\sqrt{\frac{s_Y^2/N}{\sigma_Y^2/N}}}<t_{\alpha/2,N-1}\right). \end{aligned}$$

Denote

$$Z=\frac{\bar{y}-\mu_Y}{\sqrt{\sigma_Y^2/N}}, \quad Q=\frac{s_Y^2/N}{\sigma_Y^2/N}.$$

It follows that

$$Z\sim N(0,1), \quad Q\sim\chi_{N-1}^2/(N-1),$$

and Z is independent with Q. Hence the power function can be written as

$$\begin{aligned} P(|t|>t_{\alpha/2,N-1}) &= 1-P\left(-t_{\alpha/2,N-1}\sqrt{Q}-\delta<Z<t_{\alpha/2,N-1}\sqrt{Q}-\delta\right) \\ &= 1-E\left(P\left(-t_{\alpha/2,N-1}\sqrt{Q}-\delta<Z<t_{\alpha/2,N-1}\sqrt{Q}-\delta\right)\Big|Q\right) \end{aligned}$$

$$= \quad 1 - \int_0^\infty \left(\int_{-t_{\alpha/2,N-1}\sqrt{q}-\delta}^{t_{\alpha/2,N-1}\sqrt{q}-\delta} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right) f(q) dq.$$

For $\forall q$, denote

$$g(\delta) = \int_{-t_{\alpha/2,N-1}\sqrt{q}-\delta}^{t_{\alpha/2,N-1}\sqrt{q}-\delta} e^{-x^2/2} dx.$$

By Leibniz's rule,

$$\frac{\partial g(\delta)}{\partial \delta} = \exp\left(-\frac{1}{2} \left(t_{\alpha/2,N-1}\sqrt{q} + \delta \right)^2 \right) - \exp\left(-\frac{1}{2} \left(t_{\alpha/2,N-1}\sqrt{q} - \delta \right)^2 \right).$$

It follows that

$$\frac{\partial \beta(\delta)}{\partial \delta} \begin{cases} < 0 & \text{for} \quad \delta < 0, \\ > 0 & \text{for} \quad \delta > 0. \end{cases}$$

Hence, $g(\delta)$ is a decreasing function of $|\delta|$ and $\beta(\delta)$ is a monotone increasing function of $|\delta|$.

□

As Lemma 3.6.1 and 3.6.2 show, the power function is an increasing function of noncentrality parameter $|\mu_Y|/\sigma_Y$, i.e., $|\mathbf{a}'\boldsymbol{\mu}|/\sqrt{\mathbf{a}'\Sigma\mathbf{a}}$, for both cases with known and unknown Σ. Next we will show that the maximum can be attained with $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}$.

Lemma 3.6.3.

$$arg \max_{\mathbf{a},\mathbf{a}\neq\mathbf{0}} \left| \frac{\mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\Sigma\mathbf{a}}} \right| = \Sigma^{-1}\boldsymbol{\mu}.$$

Proof. Let $\boldsymbol{w} = \Sigma^{\frac{1}{2}}\mathbf{a}$ such that $\mathbf{a} = \Sigma^{-\frac{1}{2}}\boldsymbol{w}$,

$$\begin{aligned} \arg\max_{\mathbf{a},\mathbf{a}\neq\mathbf{0}} \left| \frac{\mathbf{a}'\boldsymbol{\mu}}{\sqrt{\mathbf{a}'\Sigma\mathbf{a}}} \right| &= \arg\max_{\mathbf{a},\mathbf{a}\neq\mathbf{0}} \frac{(\mathbf{a}'\boldsymbol{\mu})^2}{\mathbf{a}'\Sigma\mathbf{a}} \\ &= \arg\max_{\mathbf{a},\mathbf{a}\neq\mathbf{0}} \frac{\mathbf{a}'\boldsymbol{\mu}\boldsymbol{\mu}'\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} \\ &= \arg\max_{\mathbf{a},\mathbf{a}\neq\mathbf{0}} \frac{\boldsymbol{w}'\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\Sigma^{-\frac{1}{2}}\boldsymbol{w}}{\boldsymbol{w}'\boldsymbol{w}}. \end{aligned}$$

Let $f(\boldsymbol{w}) = \boldsymbol{w}'\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\Sigma^{-\frac{1}{2}}\boldsymbol{w}/\boldsymbol{w}'\boldsymbol{w}$. By matrix algebra, $\max f(\boldsymbol{w}) = \lambda_1$, where λ_1 is the maximum eigenvalue of $\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\Sigma^{-\frac{1}{2}}$ and it is attained when $\boldsymbol{w}$ equals to the eigenvector corresponding to λ_1. Denote this eigenvector as $\boldsymbol{w}_1$.

By the definition of eigenvector, $\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}\boldsymbol{\mu}'\Sigma^{-\frac{1}{2}}\boldsymbol{w}_1 = \lambda_1\boldsymbol{w}_1$. By using $\boldsymbol{\mu}'\Sigma^{-\frac{1}{2}}\boldsymbol{w}_1$ is a scalar, it gives that $\boldsymbol{w}_1 \propto \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}$ and $\lambda_1 = \boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$.

□

Therefore, if we project the data into a one-dimensional space, the optimal direction is $\Sigma^{-1}\boldsymbol{\mu}$, which is consistent with the results in Theorem 3.2.1.

3.6.2 Projection to multi-dimensional space

Here we introduce lemmas that facilitate the proof of Theorem 3.2.1 with the unknown Σ. Proof of these lemmas can be found in the reference paper if not presented here.

Lemma 3.6.4. *For any fixed $0 < \alpha < 1$, $P(F_{k,n-k}(\delta) > F_{k,n-k,\alpha})$ is a decreasing function of k with fixed noncentrality parameter δ.(Theorem 7 of Ghosh (1973))*

Lemma 3.6.5. *For any fixed $0 < \alpha < 1$, $P(\chi^2_k(\delta) > \chi^2_{k,\alpha})$ is a decreasing function of k with fixed noncentrality parameter δ.(Theorem 2 of Ghosh (1973))*

Lemma 3.6.6. *Let $Q \sim \chi^2_k(\delta)(k \geq 1)$, for all $c \in R$, $P(Q \geq c)$ is an increasing function of δ.(Part of Theorem 1.1 of Kallenberg (1990))*

Proof. The proof is given in Kallenberg (1990). We outline it here for completeness.

Q can be decomposed to the sum of two independent random variables V and X^2, where

$$V \sim \chi^2_{k-1}(0), \quad \text{and} \quad X \sim N(\sqrt{\delta}, 1).$$

Let g denote the density for V and let Φ and ϕ be the probability distribution function and density function for $Z \sim N(0,1)$.

Let $Q_1 \sim \chi^2_k(\delta_1)$ and $Q_2 \sim \chi^2_k(\delta_2)$, where $\delta_1 \leq \delta_2$.

$$\begin{aligned}
& P(Q_2 \geq c) - P(Q_1 \geq c) \\
= \; & P(Q_1 \leq c) - P(Q_2 \leq c) \\
= \; & \int_0^c \Big[\Big\{ \Phi(\delta_1^{1/2} + (c-v)^{1/2}) - \Phi(\delta_1^{1/2} - (c-v)^{1/2}) \Big\} \\
& - \Big\{ \Phi(\delta_2^{1/2} + (c-v)^{1/2}) - \Phi(\delta_2^{1/2} - (c-v)^{1/2}) \Big\} \Big] \, g(v) dv
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^c \left[\int_{\delta_1^{1/2}-(c-v)^{1/2}}^{\delta_2^{1/2}-(c-v)^{1/2}} \phi(z)dz - \int_{\delta_1^{1/2}+(c-v)^{1/2}}^{\delta_2^{1/2}+(c-v)^{1/2}} \phi(z)dz \right] g(v)dv \\
&= \int_0^c \left\{ \int_{\delta_1^{1/2}-(c-v)^{1/2}}^{\delta_2^{1/2}-(c-v)^{1/2}} \phi(z) \left[1 - \exp\left(-2(c-v)^{1/2}(z+(c-v)^{1/2})\right)\right] \right\} g(v)dv.
\end{aligned}
$$

It holds that for all $u, h \in R$,

$$0 \le \phi(z)\left[1 - \exp\left(-2h(z+h)\right)\right] \le (2\pi)^{-1/2}.$$

Hence the conclusion follows. □

Lemma 3.6.7. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a random sample of size N from $N_p(\boldsymbol{\mu}, \Sigma), p < N$ and Σ is unknown. The power of the Hotelling's T^2 test to $H_0 : \boldsymbol{\mu} = \mathbf{0}$ increases with $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$.*

Proof. In this case of unknown Σ and $p < N$, the Hotelling's T^2 can be applied with the test statistic

$$T^2 = N\bar{\mathbf{x}}'\mathbf{S}^{-1}\bar{\mathbf{x}},$$

where $\mathbf{S} = \frac{1}{N-1}\sum_i(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. The power function is

$$P\left((N-p)T^2/(p(N-1)) > F_{p,N-p,\alpha}\right),$$

where $(N-p)T^2/(p(N-1))$ is a noncentral F with the noncentrality parameter $\delta = N\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$.

By construction, denote $F = \frac{V/p}{W/(n-p)}$, where $V \sim \chi_p^2(\delta)$, $W \sim \chi_{n-p}^2$, and V is independent with W. The power function is

$$
\begin{aligned}
P(F > F_{p,n-p,\alpha}) &= P\left(\frac{V/p}{W/(n-p)} > F_{p,n-p,\alpha}\right) \\
&= P\left(V > \frac{p}{n-p} W F_{p,n-p,\alpha}\right) \\
&= E\left[P\left(V > \frac{p}{n-p} W F_{p,n-p,\alpha}\right) \Big| W\right] \\
&= \int_0^\infty P\left(V > \frac{p}{n-p} w F_{p,n-p,\alpha}\right) f(w)dw.
\end{aligned}
$$

For each each w, denote

$$g(\delta) = P\left(V > \frac{p}{n-p} w F_{p,n-p,\alpha}\right).$$

By lemma 3.6.6, $g(\delta)$ is an increasing function of δ for each w. Hence the power of Hotelling's T^2 test is also an increasing function of $\boldsymbol{\mu}'\Sigma^{-1}\boldsymbol{\mu}$.

□

3.6.3 Derivation for the $\beta_{2p}(\eta|\tau_p)$ and $\beta_{3p}(\boldsymbol{\mu}, \Sigma)$

3.6.3.1 Derivation for the $\beta_{2p}(\eta|\tau_p)$

The statistic $T^2 = N\bar{\mathbf{x}}^T\mathbf{S}^{-1}\bar{\mathbf{x}}$ is distributed as $\chi^2(p)$ with mean p and variance $2p$ under the null hypothesis, and distributed as $\chi^2_p(\eta)$ with mean $p+\eta$ and variance $2(p+2\eta)$ under the alternative. Therefore, the power function is

$$\begin{aligned}
\beta_{2p}(\eta|\tau_p) &= P\left(\frac{T^2 - p}{\sqrt{2p}} > z_\alpha\right) \\
&= P\left(\frac{T^2 - (p+\eta)}{\sqrt{2(p+2\eta)}} > \frac{\sqrt{2p}z_\alpha + p - (p+\eta)}{\sqrt{2(p+2\eta)}}\right) \\
&= \Phi\left(-\frac{\sqrt{2p}z_\alpha + p - (p+\eta)}{\sqrt{2(p+2\eta)}}\right) \\
&= \Phi\left(-\frac{z_\alpha}{\sqrt{1+2\eta/p}} + \frac{1}{2}\sqrt{\frac{\eta}{0.5p+\eta}}\sqrt{\eta}\right).
\end{aligned} \tag{3.17}$$

Plugging $\tau_p = \eta/p$ will give (3.13).

3.6.3.2 Derivation for the $\beta_{3p}(\boldsymbol{\mu}, \Sigma)$

Because $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, N^{-1}\Sigma)$, the statistic $T_3^2 = N\|\bar{\mathbf{x}}\|^2$ has mean $\mathrm{tr}(\Sigma)$ and variance $2\mathrm{tr}(\Sigma^2)$ under null hypothesis, and mean $\mathrm{tr}(\Sigma) + N\|\boldsymbol{\mu}\|^2$ and variance $2\mathrm{tr}(\Sigma^2) + 4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}$ under alternative hypothesis. Also by the asymptotic normality that

$$\frac{T_3^2 - E(T_3^2)}{\sqrt{\mathrm{var}(T_3^2)}} \to N(0,1),$$

it holds that

$$\begin{aligned}\beta_{3p}(\boldsymbol{\mu},\Sigma) &= P\left(\frac{T_3^2-\mathrm{tr}(\Sigma)}{\sqrt{2\mathrm{tr}(\Sigma^2)}}>z_\alpha\right) \qquad (3.18)\\ &= P\left(\frac{T_3^2-(\mathrm{tr}(\Sigma)+N\|\boldsymbol{\mu}\|^2)}{\sqrt{2\mathrm{tr}(\Sigma^2)+4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}}}>\frac{\sqrt{2\mathrm{tr}(\Sigma^2)}z_\alpha-N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)+4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}}}\right)\\ &= \Phi\left(-\frac{\sqrt{2\mathrm{tr}(\Sigma^2)}z_\alpha-N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)+4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}}}\right)\\ &= \Phi\left(-z_\alpha\sqrt{\frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}(\Sigma^2)+2N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}}}+\frac{N\|\boldsymbol{\mu}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)+4N\boldsymbol{\mu}^T\Sigma\boldsymbol{\mu}}}\right).\end{aligned}$$

$$\beta_{3p}(\boldsymbol{\mu},\Sigma)=\Phi\left\{-z_\alpha\sqrt{\frac{\mathrm{tr}(\Sigma^2)}{\mathrm{tr}(\Sigma^2)+2\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}}+\frac{\|\boldsymbol{\delta}\|^2}{\sqrt{2\mathrm{tr}(\Sigma^2)+4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}}\right\}.$$

Plug in $\boldsymbol{\mu}=\boldsymbol{\delta}/\sqrt{N}$ will complete the derivation.

Chapter 4

Projection Test for High-dimensional Two-sample Mean Problem

4.1 Introduction

In Chapter 3, we have developed a projection test for the one-sample mean problem under high-dimensional scenarios in which the dimension of data can be greater than the sample size. For a random sample $\mathbf{x}_1, \ldots, \mathbf{x}_N$ from a p-dimensional population $\mathbf{x}$ with finite mean $E(\mathbf{x}) = \boldsymbol{\mu}$ and finite positive definite covariance matrix $\text{cov}(\mathbf{x}) = \Sigma$, we show that the optimal direction for this projection test is $\Sigma^{-1}\boldsymbol{\mu}$ and the test under this projection has equivalent hypotheses as the original data.

In this chapter, we extend this idea of optimal projection from the one-sample mean problem to the two-sample case, in which two independent random samples $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N_1}$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2}$ are available. Classical settings often assume normal population and the equal covariance matrix. Let $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N_1}$ be a random sample from the population $N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2}$ be another random sample from the population $N_p(\boldsymbol{\mu}_2, \Sigma)$. Denote $n = N_1 + N_2 - 2$. Under the high-dimensional specification, $n \ll p$. We are interested in testing

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{4.1}$$

Let $\bar{\mathbf{x}}_i$ and $\mathbf{S}_i$ be the sample mean vector and sample covariance matrix with respect to the ith sample, respectively. In this case, the two-sample Hotelling's T^2

test can be applied with the test statistic

$$\tilde{T}^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \left(\mathbf{S}_0 \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \right)^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \tag{4.2}$$

where $\mathbf{S}_0$ is the pooled sample covariance $\mathbf{S}_0 = \{(N_1 - 1)\mathbf{S}_1 + (N_2 - 1)\mathbf{S}_2\}/n$.

For high-dimensional data in which $n < p$, $\mathbf{S}_0$ is singular and as a result the Hotelling's T^2 test is undefined. In literature, there are two families of approaches developed to deal with this high-dimensionality: one is to obtain an appropriate substitute for the inverse of the $\mathbf{S}$ and the other is to project the original data to a smaller space for dimension reduction such that the classical methods can still be applied. Our work follows the projection scheme that we first obtain a projection direction and then conduct the test on the projected sample. The major difference with the previous methods is that the proposed direction is derived explicitly by maximizing the power of the projection test. When $\Sigma_1 = \Sigma_2 = \Sigma$, the projection test can be naturally extended from the one-sample test. A piece of complexity arises when the two covariance matrices differ. This unequal covariance matrices situation needs special attention and we will discuss it separately.

The remains of this chapter is organized into three main sections. Firstly, we discuss the classic case with the normal population and equal covariance assumption. Secondly, we discuss the extension of equal covariance case without normality assumption. Thirdly, we further study the situation of normal population but with unequal covariances.

4.2 New test based on normal populations with equal covariance

In this section, we assume normal populations and equal covariance matrix. Let $\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N_1}$ be a random sample from the population $N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{x}_{21}, \ldots, \mathbf{x}_{2N_2}$ be another from the population $N_p(\boldsymbol{\mu}_2, \Sigma)$. The two samples are independent with each other. The problem of interest is to test the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus the alternative $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. Denote $\boldsymbol{\mu}_d = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, we can rewrite the hypotheses as $H_0 : \boldsymbol{\mu}_d = \mathbf{0}$ versus $H_0 : \boldsymbol{\mu}_d \neq \mathbf{0}$.

4.2.1 Optimal projection direction

To develop the projection test, we consider a matrix $A_{p\times k}$ of rank k such that $k \ll p$ and $k < N$. For each sample, we project the original data to A to obtain a sequence of new sample with reduced dimension. That is $\mathbf{y}_{ij} = A^T\mathbf{x}_{ij}$, for $i = 1, 2$ and $j = 1, \ldots, N_i$. Here we use the common projection matrix A in order to maintain the equality of the covariance matrices of two projected samples. The distribution of the projected sample can be easily obtained that

$$\mathbf{y}_{1j} \overset{\text{i.i.d.}}{\sim} N(A^T\boldsymbol{\mu}_1, A^T\Sigma A), \quad j = 1, \ldots, N_1, \tag{4.3}$$

$$\mathbf{y}_{2j} \overset{\text{i.i.d.}}{\sim} N(A^T\boldsymbol{\mu}_2, A^T\Sigma A), \quad j = 1, \ldots, N_2. \tag{4.4}$$

Denote $\boldsymbol{\mu}_{1A} = A^T\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_{2A} = A^T\boldsymbol{\mu}_2$. On these projected samples, the Hotelling's T^2 test can be applied to test

$$H_{0A} : \boldsymbol{\mu}_{1A} = \boldsymbol{\mu}_{2A} \quad \text{versus} \quad H_{1A} : \boldsymbol{\mu}_{1A} \neq \boldsymbol{\mu}_{2A}. \tag{4.5}$$

The corresponding test statistic is

$$\tilde{T}_A^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T A \left(A^T\mathbf{S}_0 \left(\frac{1}{N_1} + \frac{1}{N_2} \right) A \right)^{-1} A^T(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2), \tag{4.6}$$

where $\bar{\mathbf{x}}_i$ and $\mathbf{S}_0$ are as defined in (4.2) .

As we have discussed in Chapter 3, the search for the optimal projection involves two steps. The first step is to determine the optimal projection with a given projection dimension k. The second step is to search the dimension k and the corresponding projection direction that results in the best power among the all. As a direct result of Theorem 3.2.1, we have the following corollary which addresses both issues. In the end, it is also very important to be able to conclude that the projected test is testing the same hypothesis as the original. We will confirm the equivalence after the corollary.

Corollary 4.2.1. *Suppose that for $i = 1$ and 2, $\mathbf{x}_{ij}$, $j = 1, \cdots, N_i$, is a random sample from $N(\boldsymbol{\mu}_i, \Sigma)$. Let A be a $p \times k$ full column-rank matrix. For a fixed* *projection rank k, the projection test $\tilde{T}_A^2$ reaches its best power for H_1 with*

$$A = \Sigma^{-\frac{1}{2}}\tilde{W}, \tag{4.7}$$

where $\tilde{W} = (\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_k)$ with

$$\tilde{\mathbf{w}}_1 = \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}_d / \sqrt{\boldsymbol{\mu}_d^T \Sigma^{-1}\boldsymbol{\mu}_d}, \tag{4.8}$$

and $\tilde{\mathbf{w}}_2, \ldots, \tilde{\mathbf{w}}_d$ are taken to be orthogonal to $\tilde{\mathbf{w}}_1$. Moreover, the $k = 1$ gives the optimal projection rank under which the optimal projection direction is $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d$.

Proof. By the property of the Hotelling's T^2 test, it holds for the statistic defined in (4.6) that $((n-k+1)/kn)\tilde{T}_A^2$ follows the $F_{k,n-k+1}(N_1N_2/(N_1+N_2)\tilde{\delta}_A)$ distribution, the noncentral F-distribution with degrees of freedom k and $n-k+1$, and noncentrality parameter $N_1N_2/(N_1+N_2)\tilde{\delta}_A$, where

$$\tilde{\delta}_A = \boldsymbol{\mu}_d^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu}_d. \tag{4.9}$$

By Lemma 3.6.7, the power function of the Hotelling's T^2 test depends only on (4.9) and is an increasing function of (4.9). Therefore, the goal of maximizing the power is equivalent to optimizing (4.9), and the conclusion follows by the similar argument as used in Theorem 3.2.1. Here we include the rest of the steps for completeness of presentation.

To proceed, we first introduce some notations. Denote

$$\begin{aligned} \tilde{\mathbf{v}} &= \Sigma^{-\frac{1}{2}}\boldsymbol{\mu}_d, \\ \tilde{W} &= \Sigma^{\frac{1}{2}}A, \\ P_{\tilde{W}} &= \tilde{W}(\tilde{W}'\tilde{W})^{-1}\tilde{W}'. \end{aligned} \tag{4.10}$$

The projection matrix $P_{\tilde{W}}$ connects the dots and plays an important role in finding the optimal direction. With a little algebra, we can rewrite $\tilde{\delta}_A$ as follows.

$$\begin{aligned} \tilde{\delta}_A &= \boldsymbol{\mu}_d^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu}_d \qquad (4.11) \\ &= \boldsymbol{\mu}_d^T \Sigma^{-\frac{1}{2}}(\tilde{W}(\tilde{W}^T\tilde{W})^{-1}\tilde{W}^T)\Sigma^{-\frac{1}{2}}\boldsymbol{\mu}_d \\ &= \tilde{\mathbf{v}}^T P_{\tilde{W}}\tilde{\mathbf{v}}. \end{aligned}$$

Recall that our goal is to maximize $\tilde{\delta}_A$, we next show that $\tilde{\delta}_A$ has an attainable maximum $\|\tilde{\mathbf{v}}\|_2^2$. By the property of projection matrix, there is a $p \times p$ orthogonal matrix $\tilde{H}$ such that

$$P_{\tilde{W}} = \tilde{H}\tilde{M}\tilde{H}^T, \tag{4.12}$$

where

$$\tilde{M} = \begin{pmatrix} I_k & 0_{k\times(p-k)} \\ 0_{(p-k)\times k} & 0_{(p-k)\times(p-k)} \end{pmatrix}. \tag{4.13}$$

Therefore, it follows that

$$\tilde{\delta}_A \leq \tilde{\mathbf{v}}^T\tilde{H}^T I_{p\times p}\tilde{H}\tilde{\mathbf{v}} = \|\tilde{\mathbf{v}}\|^2. \tag{4.14}$$

To show that $\|\tilde{\mathbf{v}}\|^2$ is attainable, we give a constructed case that reaches $\|\tilde{\mathbf{v}}\|^2$. Denote the i-th column of $\tilde{H}$ as $\tilde{\mathbf{h}}_i$. Specifically, take the first column of $\tilde{H}$ as

$$\tilde{\mathbf{h}}_1 = \tilde{\mathbf{v}}/\|\tilde{\mathbf{v}}\|,$$

and the rest columns $\tilde{\mathbf{h}}_2, \ldots, \tilde{\mathbf{h}}_d$ are taken so that $\tilde{H}$ is orthonormal. With this constructed example, it is clear that $\tilde{\delta}_A$ attains the maximum. To see this,

$$\begin{aligned} \tilde{\mathbf{v}}^T\tilde{H} &= (\|\tilde{\mathbf{v}}\|, \underbrace{0, \ldots, 0}_{p-1}), \\ \tilde{\delta}_A &= \tilde{\mathbf{v}}^T\tilde{H}\tilde{M}\tilde{H}^T\tilde{\mathbf{v}} = \|\tilde{\mathbf{v}}\|^2. \end{aligned} \tag{4.15}$$

The matrix A is still unknown so far. To obtain the expression for A, we trace back from $\tilde{H}$ to $\tilde{W}$, and finally land to A by the relationship of A and W defined by (4.10).

$$\begin{aligned} P_{\tilde{W}} &= \tilde{H}\tilde{M}\tilde{H}^T \\ &= (\tilde{H}\tilde{M})(\tilde{M}\tilde{H}^T) \\ &= (\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_k, \underbrace{0, \ldots, 0}_{p-k})(\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_k, \underbrace{0, \ldots, 0}_{p-k})^T \\ &= \sum_1^k \tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^T. \end{aligned} \tag{4.16}$$

For $\tilde{W} = (\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_k)$, it can be easily verified that $\tilde{\mathbf{w}}_i = \tilde{\mathbf{h}}_i, i \in 1, \dots, k$ satisfies (4.16). Accordingly, the projection direction can be obtained by $A = \Sigma^{-\frac{1}{2}}\tilde{W}$.

We note that the definition of $\|\tilde{\mathbf{v}}\|^2$ does not involve the projection dimension k. Therefore, we can conclude that with the selection of the projection direction, the projection schemes under different projection dimension k reach the same noncentrality parameter value. By Theorem 7 of Ghosh (1973), $P(F_{k,n-k}(\delta) > F_{k,n-k,\alpha})$ is a decreasing function of k with a fixed noncentrality parameter δ (see Lemma 3.6.4). Hence the test with $k = 1$ achieves the optimal power with the projection direction $\Sigma^{-1}\boldsymbol{\mu}_d$.

□

Based on Corollary 4.2.1, the optimal projection direction for the two-sample projection test with equal covariance matrix is given by a $p \times 1$ vector $\mathbf{a}$ that

$$\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d. \tag{4.17}$$

The hypothesis based on this $\mathbf{a}$, $H_0^{\mathbf{a}}$, is equivalent to the original test by recognizing that Σ is positive definite. The projected mean $\boldsymbol{\mu}_d^T\Sigma^{-1}\boldsymbol{\mu}_d = 0$ if and only if $\boldsymbol{\mu}_d = \mathbf{0}$.

If the projection direction is a known quantity, then we can easily obtain two projected samples on which the traditional two-sample t-test can be applied due to the 1-dimensional nature. However, the projection direction is rarely known in practice. In what follows, we propose an implementation for the algorithm.

4.2.2 Implementation

In this section, we discuss the implementation of the test, especially the estimation of the optimal projection direction proposed in Corollary 4.2.1. There are two underlying difficulties: first, the estimation of $\Sigma^{-\frac{1}{2}}$ is challenging under high-dimensional situation; second, the estimation of the direction may introduce dependency to the projected sample, which will break down the exactness of the test.

Similar to Algorithm 1, we consider partitioning each of the two samples into two separate data sets. One set of the data is used to estimate the direction, and the other set of the data is used to construct the projected sample and execute the test. The purpose of partition is to get the direction vector $\mathbf{a}$ independent of the test data, so that the exact level can be maintained. The discussion of estimating Σ^{-1} is beyond the scope of this dissertation. We use the ridge-like estimator to estimate Σ^{-1}.

Algorithm 2: Proposed algorithm for the two-sample test with equal covariance.

Input: Two independent random samples $\mathcal{S}_1 = \{\mathbf{x}_{1i}\}_{i=1}^{N_1}$ and $\mathcal{S}_2 = \{\mathbf{x}_{2i}\}_{i=1}^{N_2}$, splitting percentage κ, and ridge penalty λ.

(a) Conduct the random partition to both samples. Sample $\mathcal{S}_1$ is partitioned to estimating set $\mathcal{S}_{11}$ of size $N_{11} = \lfloor N_1\kappa \rfloor$ and testing set $\mathcal{S}_{12}$ of size $N_{12} = N_1 - N_{11}$. Likewise, sample $\mathcal{S}_2$ is partitioned to estimating set $\mathcal{S}_{21}$ of size $N_{21} = \lfloor N_2\kappa \rfloor$ and testing set $\mathcal{S}_{22}$ of size $N_{22} = N_2 - N_{21}$.

(b) Obtain weight vector $\hat{\mathbf{a}}$ based on estimating sets $\mathcal{S}_{11}$ and $\mathcal{S}_{21}$.

$$\begin{aligned}\hat{\mathbf{a}} &= (\mathbf{S}_{01} + \lambda \mathbf{D}_{\mathbf{S}_{01}})^{-1}(\bar{\mathbf{x}}_{11} - \bar{\mathbf{x}}_{21}),\\ \mathbf{S}_{01} &= ((N_{11}-1)\mathbf{S}_{11} + (N_{21}-1)\mathbf{S}_{21})/(N_{11}+N_{21}-2)\end{aligned}$$

where $\bar{\mathbf{x}}_{i1}$ and $\mathbf{S}_{i1}$ are the sample mean vector and sample covariance matrix of the corresponding estimating sets of the i-th sample, respectively.

(c) Construct the projected sample that

$$\begin{aligned} y_{1j} &= \hat{\mathbf{a}}^T \mathbf{x}_{1j}, \quad \mathbf{x}_{1j} \in \mathcal{S}_{12}, \\ y_{2j} &= \hat{\mathbf{a}}^T \mathbf{x}_{2j}, \quad \mathbf{x}_{2j} \in \mathcal{S}_{22}. \end{aligned} \tag{4.18}$$

(d) Denote the $\bar{y}_i$ and s_i^2 as the sample mean and variance with respect to the $\{y_{ij}\}_{j=1}^{N_{i2}}$, respectively. Calculate the t statistic,

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_y\sqrt{\frac{1}{N_{12}} + \frac{1}{N_{22}}}},$$

where s_y^2 is the pooled sample mean $s_y^2 = ((N_{12}-1)s_1^2 + (N_{22}-1)s_2^2)/(N_{12} + N_{22} - 2)$.

Reject the null hypothesis if $|t| > t_{df,\alpha/2}, df = N_{12} + N_{22} - 2$.

In the numerical study, we use $\lambda = \min\{N_{11}, N_{21}\}^{-0.5}$. Meanwhile, the splitting percentage is set to be 0.4. The calculation of $(\mathbf{S}_{01} + \lambda \mathbf{D}_{\mathbf{S}_{01}})^{-1}$ uses the same trick as (3.10).

4.2.3 Asymptotic power comparison

In this section, we study the asymptotic power of the proposed two-sample test. We consider the local alternative that

$$H_1 : \boldsymbol{\mu}_d = \boldsymbol{\delta}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}. \tag{4.19}$$

Without loss of generality, we assume that $\sqrt{N_{12}/N_1} \to b > 0$ as $N_1 \to \infty$ and $\sqrt{N_{22}/N_2} \to b > 0$ as $N_2 \to \infty$, where N_{12} and N_{22} are the sample sizes of the testing sets $\mathcal{S}_{12}$ and $\mathcal{S}_{22}$ for the first and second sample, respectively. Further assume that $\hat{\mathbf{a}} \to \mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d$ in probability as the sample size of estimating sets $\mathcal{S}_{11}$ and $\mathcal{S}_{21}$ tend to ∞. Let $\Phi(\cdot)$ and z_α denote the cumulative distribution function and $1-\alpha$ quantile of $N(0,1)$, respectively. Then the asymptotic power function of the proposed projection test at a given level α is

$$\tilde{\beta}_{1p}(\eta) = \Phi(-z_{\alpha/2} + b\sqrt{\eta}), \tag{4.20}$$

where $\eta = \boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta}$. Note that the power function(4.20) takes the same form as its one-sample counterpart.

We would like to compare the asymptotic power function $\tilde{\beta}_{1p}(\eta)$ with that of the test statistic $\tilde{T}_3^2 = (1/N_1 + 1/N_2)^{-1}\|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\|^2$, which is enlightened by the ideas to replace $\mathbf{S}^{-1}$ by I_p, the $p \times p$ identity matrix. The tests proposed in Bai and Saranadasa (1996), Srivastava and Du (2008), and Chen and Qin (2010) share the asymptotic power as $\tilde{T}_3^2$. Let $\tilde{\beta}_{3p}(\boldsymbol{\delta}, \Sigma)$ be the asymptotic power function of $\tilde{T}_3^2$. Define that

$$\tilde{\beta}_{3p}(\eta|\tilde{\tau}_p) = \Phi\left\{-z_\alpha/\sqrt{1+2\tilde{\tau}_p} + \frac{1}{2}\sqrt{\tilde{\tau}_p/(0.5+\tilde{\tau}_p)}\cdot\sqrt{\eta}\right\},$$

where $\tilde{\tau}_p = \boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}/\mathrm{tr}(\Sigma^2)$. We have the following proposition.

Proposition 4.2.2. *Assume that* $\sqrt{N_{12}/N_1} \to b$ *as* $N_1 \to \infty$ *and* $\sqrt{N_{22}/N_2} \to b$ *as* $N_2 \to \infty$. *The following statements are valid.*

(I) Under $H_1 : \boldsymbol{\mu}_d = \boldsymbol{\delta}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$, $\tilde{\beta}_{3p}(\boldsymbol{\delta}, \Sigma) \leq \tilde{\beta}_{3p}(\eta|\tilde{\tau}_p)$

(II) If $b > 0.5$ *and* $\eta \to \infty$, *then* $\tilde{\beta}_{1p}(\eta) - \tilde{\beta}_{3p}(\eta|\tilde{\tau}_p) > 0$ *for large enough* p.

(III) If $\sqrt{\eta}b \geq z_{\alpha/2} - z_\alpha$ *and* $\tilde{\tau}_p \to 0$ *as* $p \to \infty$, *then* $\tilde{\beta}_{1p}(\eta) - \tilde{\beta}_{3p}(\eta|\tilde{\tau}_p) > 0$ *for large enough* p.

Proof. Since $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 \sim N(\boldsymbol{\mu}_d, (\frac{1}{N_1} + \frac{1}{N_2})\Sigma)$, by the properties of normal distribution, it holds that $E(\tilde{T}_3^2) = \text{tr}(\Sigma) + \|\boldsymbol{\delta}\|^2$ and $\text{var}(\tilde{T}_3^2) = 2\text{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}$. Furthermore,

$$\frac{\tilde{T}_3^2 - E(\tilde{T}_3^2)}{\sqrt{\text{var}(\tilde{T}_3^2)}} \to N(0,1)$$

in distribution as $N \to \infty$. Thus under $H_1 : \boldsymbol{\mu}_d = \boldsymbol{\delta}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$, the asymptotic power function of $\tilde{T}_3^2$ can be obtained by the standard process as below.

$$\begin{aligned}
\tilde{\beta}_{3p}(\boldsymbol{\delta}, \Sigma) &= P\left(\frac{\tilde{T}_3^2 - \text{tr}(\Sigma)}{\sqrt{2\text{tr}(\Sigma^2)}} > z_\alpha\right) \qquad (4.21) \\
&= P\left(\frac{\tilde{T}_3^2 - (\text{tr}(\Sigma) + \|\boldsymbol{\delta}\|^2)}{\sqrt{2\text{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}} > \frac{\sqrt{2\text{tr}(\Sigma^2)}z_\alpha - \|\boldsymbol{\delta}\|^2}{\sqrt{2\text{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}}\right) \\
&= \Phi\left(-\frac{\sqrt{2\text{tr}(\Sigma^2)}z_\alpha - \|\boldsymbol{\delta}\|^2}{\sqrt{2\text{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}}\right) \\
&= \Phi\left(-z_\alpha\sqrt{\frac{\text{tr}(\Sigma^2)}{\text{tr}(\Sigma^2) + 2\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}} + \frac{\|\boldsymbol{\delta}\|^2}{\sqrt{2\text{tr}(\Sigma^2) + 4\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}}\right)
\end{aligned}$$

Furthermore, by plugging in the $\tilde{\tau}_p$,

$$\tilde{\beta}_{3p}(\boldsymbol{\delta}, \Sigma) = \Phi\left\{-z_\alpha/\sqrt{1 + 2\tilde{\tau}_p} + \frac{1}{2}\sqrt{\tilde{\tau}_p/(0.5 + \tilde{\tau}_p)} \cdot \|\boldsymbol{\delta}\|^2/\sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}\right\}.$$

By the Cauchy-Schwartz inequality, it holds that $\|\boldsymbol{\delta}\|^2 \leq \sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}}\sqrt{\boldsymbol{\delta}^T\Sigma^{-1}\boldsymbol{\delta}}$. Therefore, $\|\boldsymbol{\delta}\|^2/\sqrt{\boldsymbol{\delta}^T\Sigma\boldsymbol{\delta}} \leq \sqrt{\eta}$. Thus, it follows that

$$\tilde{\beta}_{3p}(\boldsymbol{\delta}, \Sigma) \leq \Phi\left\{-z_\alpha/\sqrt{1+2\tilde{\tau}_p} + \frac{1}{2}\sqrt{\tilde{\tau}_p/(0.5+\tilde{\tau}_p)} \cdot \sqrt{\eta}\right\} = \tilde{\beta}_{3p}(\eta|\tilde{\tau}_p). \tag{4.22}$$

This completes the proof of Part (I). Following the similar proof of Proposition 3.2.3, Part (II) and (III) hold.

□

4.2.4 Simulation study

In this section, we perform a simulation study to evaluate the finite sample performance of the two-sample tests with equal covariance matrix. In this case, we set the sample sizes of both samples to be equal. We note that for the equal covariance case, unequal sample size should not have much effect in general. Hence $N_1 = N_2 = \tilde{N}$ and both are selected from the set $\{40, 160\}$. For each of of the sample size, the dimension p can take values of 400 or 1600, which sets different levels of high dimensionality. In the first sample, the mean of the first 10 random variables are selected from $c \in \{0, 0.5, 1\}$, and the rest random variables have mean 0. For the second sample, all the random variables have mean 0. We configure the covariance matrices in three ways: compound symmetry $\Sigma_1(i,j) = \rho$, autocorrelation $\Sigma_2 = \rho^{|i-j|}$, and the combined structure $\Sigma_3 = 0.5\Sigma_1 + 0.5\Sigma_2$. The ρ takes value from $\{0.25, 0.5, 0.75, 0.95\}$, indicating different levels of correlation.

The empirical powers (i.e., rejection rate in percentage at level 0.05) of these tests are reported in Tables 4.1-4.3 for three different covariance structures and various levels of $(\tilde{N}, p, c)$. Examination reveals the similar patterns as those being discovered in the one-sample situations. As expected, the proposed test, Lauter's test and the RP test with single projection keep the Type I error rate very well. The other tests fail on this because their critical values are based on their asymptotic distributions. Also seen from Tables 4.1-4.3, the power of the tests strongly relies on the covariance structures as well as the value of p and c. We note that the overall patterns are very similar to the one-sample test situation, as shown in Section 3.3.

The proposed test is a clear winner for the compound symmetry structure case in Table 4.1. When the sample sizes are small at 40, the power of the proposed test grows dramatically as the increase of ρ, or the c. For example, it goes from 8.72% to 55.79% when c goes from 0.5 to 1 under the case $(\tilde{N}, p, \rho) = (40, 1600, 0.25)$. Instead, the performance of the other alternatives is not that satisfactory. Under the same $(\tilde{N}, p, \rho) = (40, 1600, 0.25)$ case, the powers are barely changed. In general, the two RP methods reveal similar patterns as the proposed method when c, N or ρ increases, and outperform the rest tests.

Table 4.2 summarizes the simulation results for the autocorrelation covariance structure. In this scenario, the Dempster test, BS test, CQ test, and SD test have more satisfactory performance than the proposed test when $(\tilde{N}, c) = (40, 0.5)$. It is also observed that these tests quickly lose their powers as the correlation increases. This is expected, because all these four tests ignore the correlations among variables. Therefore, their performance is better when the true correlations between variables are weak. Nevertheless, the power of our test is at a comparable level as the competing tests with the increase of c or the sample size. Similar increasing pattern is also observed for the two versions of random projection tests, but their increasing rates are considerably slower than our test.

To make a fair comparison, we consider the composite covariance matrix Σ_3. The corresponding simulation results are presented in Table 4.3, which clearly shows that the proposed test outperforms all the alternative tests in general. Occasionally there are cases that one version of the random projection test reaches higher power than ours. In this case, the power of our test shows a slight decrease as the increase of ρ, while the other two versions of the random projection tests preserve the increasing trend. This observation is interesting so we know that the pattern could be different for our test and the random projection tests, even though they share some similarity in the design. The reason could be that the random projection is designed from the general idea to utilize the correlation information by randomly projected to a smaller space so that the projected sample covariance is invertible. In our case, the power is tightly correlated with the quantity of $\eta = \boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta}$, so it may not always increase with ρ.

Table 4.1: Comparison for two-sample tests (equal covariance): multivariate normal with Σ_1

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\tilde{\mathbf{N}} = \mathbf{40}, \mathbf{p} = \mathbf{400}$							
New	4.90	4.89	5.00	4.95	24.53	44.27	88.86	100.00	97.23	99.48	100.00	100.00
Dempster	5.80	5.25	4.84	4.64	8.11	6.33	5.55	5.27	23.44	10.52	7.78	6.90
BS	6.97	7.09	7.05	7.03	9.76	8.33	7.89	7.70	27.85	14.04	11.18	10.01
CQ	6.96	7.08	7.03	7.05	9.74	8.33	7.90	7.69	27.86	14.04	11.21	10.00
SD_with_adjust	3.39	1.32	0.25	0.08	4.81	1.50	0.31	0.08	13.59	2.38	0.49	0.09
SD_no_adjust	7.33	7.30	7.11	6.97	10.27	8.62	7.92	7.60	29.59	14.42	11.26	9.88
Lauter	4.57	4.50	4.49	4.49	5.03	4.80	4.73	4.64	5.54	5.01	4.93	4.87
RP_single	5.18	4.90	5.16	4.96	11.51	16.21	32.55	98.24	45.28	67.39	95.39	100.00
RP_average	0.01	0.01	0.00	0.02	0.44	2.53	44.83	100.00	83.83	99.79	100.00	100.00
					$\tilde{\mathbf{N}} = \mathbf{40}, \mathbf{p} = \mathbf{1600}$							
New	4.88	5.06	4.95	4.86	8.72	13.71	37.53	98.54	55.79	78.54	96.77	100.00
Dempster	6.37	5.69	5.18	5.05	6.81	5.97	5.40	5.18	8.64	6.58	5.89	5.57
BS	7.37	7.41	7.39	7.39	7.92	7.65	7.53	7.50	10.06	8.64	8.22	8.04
CQ	7.36	7.44	7.43	7.39	7.92	7.69	7.55	7.50	10.08	8.67	8.23	8.03
SD_with_adjust	2.30	0.48	0.10	0.02	2.44	0.49	0.11	0.03	3.20	0.55	0.12	0.03
SD_no_adjust	7.78	7.56	7.45	7.28	8.41	7.90	7.61	7.43	10.51	8.85	8.29	7.93
Lauter	4.95	4.94	4.94	4.94	5.02	4.96	4.94	4.93	5.18	5.04	5.01	4.96
RP_single	4.96	5.25	5.36	5.31	6.08	7.43	9.89	42.82	11.91	16.77	34.41	98.60
RP_average	0.00	0.00	0.00	0.00	0.00	0.03	0.09	85.18	0.10	0.92	47.12	100.00
					$\tilde{\mathbf{N}} = \mathbf{160}, \mathbf{p} = \mathbf{400}$							
New	4.90	4.96	4.90	5.00	99.13	99.99	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.92	5.49	5.24	5.19	21.07	9.74	7.63	6.96	100.00	87.81	28.09	18.84
BS	6.76	6.87	6.89	6.87	25.36	13.05	10.15	9.31	100.00	98.82	41.68	27.19
CQ	6.76	6.87	6.89	6.88	25.36	13.04	10.15	9.33	100.00	98.83	41.66	27.17
SD_with_adjust	3.53	1.00	0.24	0.08	11.74	2.04	0.37	0.09	100.00	11.66	1.46	0.32
SD_no_adjust	6.81	6.91	6.89	6.87	25.84	13.13	10.14	9.26	100.00	98.70	41.83	27.09
Lauter	5.19	5.18	5.18	5.18	5.55	5.19	5.13	5.10	6.43	5.67	5.40	5.24
RP_single	4.72	5.05	5.11	4.79	93.42	99.66	100.00	100.00	100.00	100.00	100.00	100.00
RP_average	0.25	0.24	0.25	0.21	99.97	100.00	100.00	100.00	100.00	100.00	100.00	100.00
					$\tilde{\mathbf{N}} = \mathbf{160}, \mathbf{p} = \mathbf{1600}$							
New	5.00	5.05	4.75	5.09	81.12	98.32	99.99	100.00	100.00	100.00	100.00	100.00
Dempster	5.87	5.28	5.00	4.95	8.07	6.43	5.59	5.40	21.32	9.86	7.64	7.07
BS	6.95	6.94	6.87	6.86	9.45	8.04	7.71	7.51	25.49	12.64	10.46	9.64
CQ	6.95	6.94	6.88	6.85	9.46	8.05	7.71	7.51	25.51	12.65	10.45	9.64
SD_with_adjust	2.17	0.36	0.03	0.00	2.87	0.42	0.03	0.00	7.33	0.64	0.04	0.00
SD_no_adjust	7.04	6.98	6.88	6.85	9.54	8.13	7.71	7.50	25.94	12.70	10.48	9.64
Lauter	4.91	4.93	4.94	4.93	4.98	4.89	4.89	4.92	5.17	5.00	4.93	4.90
RP_single	5.20	5.09	4.86	4.96	25.28	40.57	81.37	100.00	94.24	99.57	100.00	100.00
RP_average	0.00	0.00	0.00	0.00	5.82	44.86	99.99	100.00	100.00	100.00	100.00	100.00

Table 4.2: Comparison for two-sample tests (equal covariance): multivariate normal with Σ_2

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{\tilde{N}=40, p=400}$							
New	5.28	5.04	5.08	5.18	18.06	13.33	9.49	9.42	93.08	79.98	53.04	44.97
Dempster	5.07	5.21	5.20	5.36	46.73	36.37	24.12	9.97	99.96	99.35	91.83	39.01
BS	5.47	5.82	5.83	6.41	48.33	37.97	25.88	12.16	99.96	99.44	92.61	44.27
CQ	5.48	5.83	5.82	6.45	48.33	37.98	25.91	12.10	99.96	99.44	92.60	44.24
SD_with_adjust	4.23	4.20	4.03	2.70	42.93	32.64	19.93	5.30	99.94	99.10	88.95	26.01
SD_no_adjust	6.32	6.47	6.60	7.23	50.60	40.45	27.79	13.49	99.95	99.52	92.62	46.27
Lauter	4.94	4.95	5.05	5.22	16.79	13.93	9.34	5.81	63.46	47.61	24.40	7.26
RP_single	5.05	4.52	4.87	5.16	9.67	9.00	9.12	12.60	33.55	30.23	30.09	51.58
RP_average	0.00	0.01	0.04	1.36	0.15	0.39	1.28	8.63	45.41	39.57	37.47	75.17
					$\mathbf{\tilde{N}=40, p=1600}$							
New	4.94	4.89	4.71	4.72	8.72	7.57	6.07	4.93	54.84	40.84	22.98	10.99
Dempster	4.84	4.88	5.06	4.93	21.01	16.99	12.02	7.26	92.49	82.12	54.97	18.05
BS	5.02	5.13	5.43	5.75	21.68	17.71	12.75	8.31	92.75	82.73	56.59	20.21
CQ	5.02	5.15	5.44	5.73	21.70	17.75	12.75	8.28	92.79	82.75	56.41	20.20
SD_with_adjust	2.74	2.72	2.64	2.32	14.25	11.01	7.43	3.39	87.32	73.48	44.52	9.84
SD_no_adjust	6.06	6.12	6.25	6.57	24.06	19.41	14.66	9.41	93.19	83.46	58.60	21.92
Lauter	5.14	5.08	4.76	5.25	9.53	8.47	6.89	5.57	25.81	20.69	12.77	6.67
RP_single	5.06	4.93	5.18	5.11	6.22	6.02	6.35	6.34	10.40	9.96	9.88	11.38
RP_average	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.57	0.01	0.00	0.06	3.15
					$\mathbf{\tilde{N}=160, p=400}$							
New	5.11	4.79	4.94	5.05	84.85	65.17	41.14	77.79	100.00	100.00	99.82	100.00
Dempster	4.95	4.98	5.17	5.20	99.99	99.45	91.48	39.72	100.00	100.00	100.00	99.99
BS	5.34	5.54	5.78	6.55	99.99	99.49	92.12	44.81	100.00	100.00	100.00	100.00
CQ	5.34	5.55	5.78	6.55	99.99	99.49	92.12	44.79	100.00	100.00	100.00	100.00
SD_with_adjust	4.50	4.67	4.49	2.53	99.97	99.40	90.12	25.68	100.00	100.00	100.00	99.93
SD_no_adjust	5.51	5.57	6.10	6.68	99.99	99.47	92.44	45.25	100.00	100.00	100.00	99.99
Lauter	5.65	5.13	4.75	5.21	46.98	36.55	15.96	5.84	99.76	92.57	44.28	7.78
RP_single	4.93	4.94	4.99	5.09	71.71	58.51	46.96	88.94	100.00	99.99	99.94	100.00
RP_average	0.31	0.56	1.11	1.31	91.64	75.31	55.32	98.57	100.00	100.00	100.00	100.00
					$\mathbf{\tilde{N}=160, p=1600}$							
New	4.84	4.73	4.59	5.26	54.85	38.91	21.90	19.69	100.00	100.00	97.36	94.23
Dempster	5.28	5.12	5.38	5.69	92.57	81.36	55.32	18.54	100.00	100.00	100.00	89.77
BS	5.44	5.34	5.82	6.37	92.85	82.00	56.76	20.60	100.00	100.00	100.00	91.16
CQ	5.43	5.34	5.82	6.37	92.84	81.99	56.77	20.61	100.00	100.00	100.00	91.15
SD_with_adjust	4.50	4.37	4.41	3.18	91.45	79.11	51.43	12.30	100.00	100.00	100.00	82.78
SD_no_adjust	5.61	5.46	5.95	6.53	92.82	82.22	57.37	21.10	100.00	100.00	100.00	91.09
Lauter	5.43	5.08	5.00	5.07	18.07	15.23	9.62	5.46	69.95	50.90	23.63	6.89
RP_single	5.33	5.17	5.17	5.02	18.17	16.89	16.75	26.88	79.29	76.30	74.97	96.62
RP_average	0.00	0.00	0.04	0.72	0.92	1.58	3.60	23.16	99.96	99.72	97.24	99.98

Table 4.3: Comparison for two-sample tests (equal covariance): multivariate normal with Σ_3

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{\tilde{N} = 40, p = 400}$							
New	4.86	4.74	5.07	5.00	18.92	19.48	16.43	17.11	93.76	92.73	85.87	83.38
Dempster	6.03	5.90	5.47	5.09	12.17	8.19	7.04	6.23	77.93	23.57	13.87	9.93
BS	6.64	6.97	6.99	6.94	13.32	9.68	8.75	8.09	82.45	28.04	17.34	13.52
CQ	6.62	6.97	6.99	6.93	13.33	9.71	8.74	8.09	82.42	28.08	17.32	13.51
SD_with_adjust	4.87	3.33	2.06	1.06	9.75	4.77	2.53	1.22	66.61	13.53	4.98	1.92
SD_no_adjust	7.20	7.25	7.25	7.13	14.39	10.21	9.10	8.38	84.68	29.94	18.11	14.07
Lauter	4.65	4.51	4.45	4.50	5.18	4.98	4.85	4.80	6.48	5.48	5.22	5.03
RP_single	4.86	4.52	4.88	5.17	9.92	10.74	11.82	20.15	37.07	42.24	48.03	79.80
RP_average	0.00	0.04	0.01	1.00	0.29	0.63	2.41	18.33	60.50	72.30	79.87	99.00
					$\mathbf{\tilde{N} = 40, p = 1600}$							
New	5.05	4.91	5.11	4.73	8.11	7.91	8.05	6.12	47.15	49.04	42.75	23.79
Dempster	6.64	6.29	5.96	5.58	7.76	6.79	6.23	5.78	12.62	8.76	7.33	6.61
BS	7.26	7.38	7.38	7.35	8.39	8.02	7.70	7.65	13.57	10.01	9.07	8.58
CQ	7.26	7.38	7.38	7.35	8.41	8.07	7.72	7.68	13.59	10.03	9.09	8.59
SD_with_adjust	4.21	2.31	0.99	0.49	4.95	2.45	1.06	0.51	7.92	3.21	1.30	0.58
SD_no_adjust	7.61	7.77	7.65	7.57	8.98	8.38	8.09	7.86	14.54	10.54	9.40	8.81
Lauter	4.94	4.92	4.90	4.95	5.08	5.05	4.98	4.97	5.34	5.19	5.11	5.06
RP_single	4.95	4.86	4.92	4.61	6.18	6.24	6.56	7.06	10.35	11.66	12.82	18.11
RP_average	0.00	0.00	0.00	0.25	0.00	0.02	0.01	0.69	0.07	0.20	0.53	11.21
					$\mathbf{\tilde{N} = 160, p = 400}$							
New	4.93	4.98	4.88	5.09	93.92	91.14	81.09	98.81	100.00	100.00	100.00	100.00
Dempster	6.17	6.02	5.72	5.47	79.91	21.32	12.60	9.52	100.00	100.00	99.42	73.33
BS	6.63	6.75	6.77	6.76	84.04	25.44	15.91	12.53	100.00	100.00	99.97	88.38
CQ	6.63	6.77	6.77	6.77	84.05	25.42	15.91	12.53	100.00	100.00	99.96	88.37
SD_with_adjust	5.29	3.46	1.80	0.78	70.33	11.49	4.15	1.53	100.00	99.99	54.66	9.33
SD_no_adjust	6.70	6.87	6.87	6.82	84.51	25.85	16.02	12.59	100.00	100.00	99.92	87.85
Lauter	5.11	5.17	5.19	5.15	6.06	5.53	5.29	5.16	8.59	6.36	5.87	5.68
RP_single	4.91	4.80	4.93	5.23	82.78	80.63	75.25	98.83	100.00	100.00	100.00	100.00
RP_average	0.23	0.42	0.72	0.91	97.96	96.21	91.46	99.98	100.00	100.00	100.00	100.00
					$\mathbf{\tilde{N} = 160, p = 1600}$							
New	4.96	4.87	4.72	5.26	65.75	67.20	53.07	49.92	100.00	100.00	100.00	99.99
Dempster	6.30	5.84	5.52	5.27	11.71	7.98	7.06	6.37	93.51	21.38	12.62	10.14
BS	6.87	6.92	6.95	6.85	12.76	9.40	8.42	8.17	96.54	25.66	15.94	12.94
CQ	6.87	6.92	6.94	6.85	12.76	9.40	8.42	8.17	96.52	25.64	15.92	12.94
SD_with_adjust	4.36	2.14	0.80	0.42	8.28	2.85	1.07	0.48	70.67	7.17	1.93	0.67
SD_no_adjust	7.03	7.01	7.00	6.92	13.04	9.50	8.53	8.24	96.62	25.96	16.11	13.06
Lauter	4.85	4.94	4.93	4.92	5.07	4.97	4.90	4.88	5.62	5.17	5.01	5.03
RP_single	5.39	5.08	4.78	4.92	20.14	22.14	24.87	51.11	86.78	91.47	95.65	99.98
RP_average	0.00	0.00	0.01	0.40	2.00	5.08	12.65	66.27	100.00	100.00	100.00	100.00

4.2.5 Real data example

In this section, we present a real data application of the proposed method. Transcription factor p53 is a well-known tumor-suppressor, which encodes proteins that bind to DNA and regulates gene expression to prevent mutations of the genome. Subramanian et al. (2005) examined gene expression patterns in response of mutation of p53 in NCI-60 collection of cancer cell lines to identify targets of the transcription factor p53. In this session, we apply the proposed method to test the difference of the expression level of gene sets grouped in KEGG pathways in cancer for normal p53 and mutated p53 samples. The KEGG pathway data can be downloaded from URL http://www.broadinstitute.org/gsea/msigdb/cards/ KEGG_PATHWAYS_IN_CANCER. The cancer cell line data can be downloaded from http://www.broadinstitute.org/gsea/datasets.jsp.

KEGG pathway of cancer contains 328 genes, among which 263 are available in the cancer cell line data. Hence we test the gene expression level between p53 mutant group which has sample size 33 and p53 normal group which has sample size 17, on a gene set of length 263. We apply the log2 transformation to the expression data before the analysis.

With such sample size, it is difficult to justify whether the covariance matrices are equal. In this section, we show the results with the assumption of equal covariance matrix. The results without this assumption is presented in Section 4.4.4.

Table 4.4: Gene pathway dataset: p-values of the two-sample tests

New	Dempster	BS	CQ	SD_with	SD_no	Lauter	RP_s	RP_m
0.03	0.38	0.42	0.43	0.10	0.05	0.55	0.73	0.16

From Table 4.4, we observe that both our test and the SD test without adjustment identify the significance under the assumption of equal covariance matrix. This implies an interesting result that the pathway in cancer is a possible target of the p53, which may explain expression level difference between the p53 normal and mutant group. Experiments are needed for the confirmatory conclusion.

To further explore the data and obtain the significant genes that contribute to the pathway, we conduct a follow-up analysis with the two-sample t-test to each individual gene. Figure 4.1 shows the histogram of the p-values.

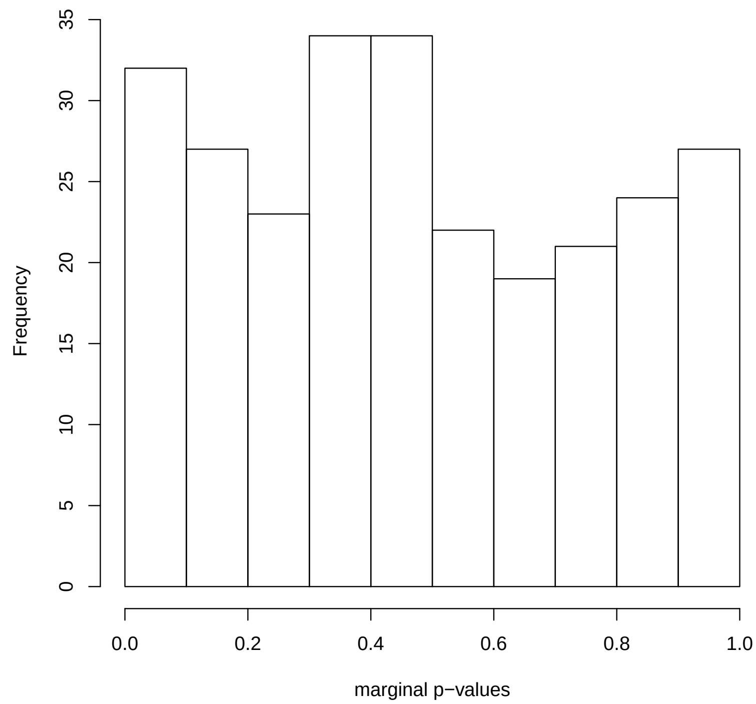


Figure 4.1: Histogram of the marginal p-values from the two-sample t-tests.

As shown in Table 4.1, marginal p-values are well-scattered in the (0,1) interval. Therefore, majority of the genes have no significant signals. Among the 263 genes, 20 of them have p-value less than 0.05. With Bonferroni correction, the number of significant genes drops down to 2. The two left genes are BAX and MDM2, with p-values 4.92×10^{-6} and 1.55×10^{-4}, respectively. We search the NCBI website for further information of those two genes, especially their relationship with p53. NCBI stands for National Center for Biotechnology Information. It serves an integrated, one-stop, genomic information infrastructure, and can be accessed with URL: http://www.ncbi.nlm.nih.gov/. By the time of May 2015, the information of BAX can be found in http://www.ncbi.nlm.nih.gov/gene/581. BAX is a known apoptosis regulator. Its relationship with p53 is well-documented in literature that it has been shown to be involved in p53-mediated apoptosis and its expression is regulated by p53. For example, Chipuk et al. (2004) claims a direct activation of Bax by p53 in a well cited science paper. Therefore, it is reasonable that the expression level of BAX differs for normal and mutated p53. The search of MDM2 reveals interesting two-way interaction relationship between MDM2 and p53. The information of MDM2 can be found in http://www.ncbi.nlm.nih.gov/gene/4193. On one side, MDM2 is transcriptionally-regulated by p53; on the other side, a protein encoded by MDM2 can promote tumor formation by targeting tumor suppressor proteins, such as p53. Therefore, is is not surprising to observe the significant difference between p53 normal and mutated groups.

4.3 Extension to non-normal distributions

4.3.1 Optimal projection direction

In this section, we further investigate the proposed projection test without the normality assumption. Here we still assume the equal covariance matrix. In this case, the optimal direction can be derived as a direct generalization of the one-sample case as shown in Theorem 3.4.1. We show that the optimal direction for non-normal populations and equal covariance matrix is still $\Sigma^{-1}\boldsymbol{\mu}_d$. We include the proof here for completeness.

Corollary 4.3.1. *Suppose that for $i = 1$ and 2, $\mathbf{x}_{ij}$, $j = 1, \cdots, N_i$, is a random sample from a population X_i with finite mean $\boldsymbol{\mu}_i$ and covariance matrix Σ. Further assume that for any nonzero constant $p \times k$ matrix A with a fixed k, $A^T\bar{\mathbf{x}}_i \to N(A^T\boldsymbol{\mu}_i, A^T\Sigma A)$ in distribution and $A^T\mathbf{S}_0A - A^T\Sigma A \to \mathbf{0}$ in probability as $n \to \infty$. The projection test $\tilde{T}_A^2$ for the two-sample problem (4.1) reaches its asymptotic best power for H_1 in (4.1) at $k = 1$ and $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d$.*

Proof. Under some mild regularity conditions, $\tilde{T}_A^2$ follows the $\chi_k^2(N_1N_2/(N_1 + N_2)\tilde{\delta}_A)$ distribution, the noncentral χ^2-distribution with k degrees of freedom and noncentrality parameter $N_1N_2/(N_1+N_2)\tilde{\delta}_A$, where $\tilde{\delta}_A = \boldsymbol{\mu}_d^T A(A^T\Sigma A)^{-1}A^T\boldsymbol{\mu}_d$. Using the property of χ^2-distribution (Kallenberg, 1990) , the power of $\tilde{T}_A^2$ increases with $\tilde{\delta}_A$ for given k. As a result, we only need to maximize $\tilde{\delta}_A^2$ with respect to A

in order to maximize the power of $\tilde{T}_A^2$. Denote by $\chi^2_{k;\alpha}$ the critical value of the χ^2-distribution with k degrees of freedom. By Lemma 3.6.5 (shown in Ghosh, 1973), for a given level α, $P\{\chi^2_k(c) > \chi^2_{k;\alpha}\}$ is a decreasing function of k when c are fixed. Using this property of χ^2-distribution and exact the same argument as those in the proof of Theorem 3.2.1, it can be shown that $k = 1$ with $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d$ is the best choice to achieve the optimal power.

□

4.3.2 Simulation study

Corollary 4.3.1 shows that when the two populations have equal covariance matrix, the optimal direction $\mathbf{a} = \Sigma^{-1}\boldsymbol{\mu}_d$ still applies even without normality assumption. In this section, we conduct simulations to investigate the performance of the algorithm for non-normal populations. To this end, we generate random samples from the multivariate t-distribution. Denote by $t_\nu(\boldsymbol{\mu}, \Sigma)$ the multivariate t-distribution with mean vector $\boldsymbol{\mu}$, covariance matrix Σ, and the degrees of freedom ν. The two samples are simulated from $t_6(\boldsymbol{\mu}_1, \Sigma)$ and $t_6(\boldsymbol{\mu}_2, \Sigma)$. The other settings are the same as used in Section 4.2.4. Since the sample sizes in some of the settings are small, we use the small sample correction which applies the t-test instead of χ^2-test. Percentage of rejection based on 10000 replicates for three correlation structures with various configurations of $(\tilde{N}, p, c)$ are shown in Tables 4.5-4.7.

In general, we observe similar patterns as from non-normal distributions in the one-sample test. For Type I errors, the proposed test still maintains it well around 0.05. RP test, however, becomes slightly conservative. Lauter's test is well-behaved except for the autocorrelation case. For the rest of the tests, CQ test has an improved Type I error with autocorrelation structure compared with the compound symmetry and the combined cases. The power of the tests is generally inferior than that of their multivariate normal counterparts, even though the comparative patterns are similar. In compound symmetry case, the proposed test significantly outperforms the rest alternatives and the power increases dramatically with ρ, c and $\tilde{N}$. For autocorrelation structure, CQ test shows robust performance and outperforms the proposed method in general. Dempster test, BS test and SD test have deteriorated power for certain scenarios.

Table 4.5: Comparison for two-sample tests (equal covariance): multivariate t with Σ_1

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{\tilde{N} = 40, p = 400}$							
New	4.91	4.88	4.75	4.72	17.90	32.01	76.26	99.99	91.18	98.23	99.97	100.00
Dempster	5.49	5.44	5.23	5.14	7.00	6.12	5.62	5.43	13.67	8.85	7.05	6.47
BS	6.57	7.15	7.21	7.19	8.22	8.15	7.88	7.67	16.22	11.36	10.08	9.38
CQ	7.30	7.38	7.32	7.31	9.42	8.34	8.04	7.81	18.00	11.66	10.25	9.55
SD_with_adjust	3.05	1.23	0.33	0.13	3.91	1.35	0.34	0.14	7.91	1.99	0.42	0.17
SD_no_adjust	7.00	7.32	7.27	7.17	8.83	8.37	7.94	7.57	17.41	11.72	10.15	9.28
Lauter	5.00	5.04	5.01	5.01	5.15	5.12	5.11	5.12	5.50	5.19	5.09	5.02
RP_single	4.15	4.40	4.17	4.23	9.76	13.24	27.52	95.76	37.57	57.23	90.44	100.00
RP_average	0.00	0.00	0.01	0.01	0.14	0.74	24.04	100.00	61.60	97.73	100.00	100.00
					$\mathbf{\tilde{N} = 40, p = 400}$							
New	5.12	5.27	4.94	5.16	7.67	10.70	28.07	96.48	42.25	66.05	93.55	100.00
Dempster	5.51	5.35	5.13	5.04	5.79	5.55	5.22	5.15	6.62	6.04	5.50	5.31
BS	6.45	6.99	7.09	7.14	6.80	7.10	7.16	7.19	7.87	7.80	7.69	7.54
CQ	7.11	7.14	7.17	7.18	7.50	7.36	7.32	7.32	8.96	8.04	7.73	7.63
SD_with_adjust	1.75	0.41	0.09	0.02	1.84	0.41	0.10	0.02	2.22	0.42	0.10	0.02
SD_no_adjust	6.86	7.15	7.14	7.06	7.12	7.31	7.21	7.05	8.50	8.00	7.70	7.46
Lauter	4.90	4.97	4.95	4.96	4.92	4.99	4.98	4.98	5.01	5.02	4.99	4.99
RP_single	4.59	4.73	4.18	4.44	5.64	6.45	7.86	36.43	9.99	14.16	27.18	95.95
RP_average	0.00	0.00	0.00	0.01	0.00	0.00	0.01	54.92	0.05	0.26	19.84	100.00
					$\mathbf{\tilde{N} = 160, p = 400}$							
New	5.02	4.95	4.97	4.86	95.78	99.90	100.00	100.00	100.00	100.00	100.00	100.00
Dempster	5.61	5.20	4.90	4.86	13.36	8.01	6.52	5.98	99.34	31.08	15.97	12.23
BS	6.51	6.71	6.77	6.80	16.07	10.53	9.16	8.54	99.74	44.42	22.03	17.26
CQ	6.80	6.80	6.78	6.80	16.74	10.63	9.21	8.60	99.87	45.07	22.21	17.36
SD_with_adjust	3.14	1.17	0.22	0.04	7.51	1.71	0.37	0.06	88.62	5.45	0.99	0.20
SD_no_adjust	6.65	6.79	6.78	6.79	16.42	10.55	9.16	8.55	99.86	45.88	22.12	17.28
Lauter	4.82	4.82	4.84	4.83	5.16	4.98	4.90	4.87	6.19	5.46	5.15	5.06
RP_single	4.07	4.30	4.35	4.42	87.25	98.59	100.00	100.00	100.00	100.00	100.00	100.00
RP_average	0.12	0.13	0.07	0.12	99.64	100.00	100.00	100.00	100.00	100.00	100.00	100.00
					$\mathbf{\tilde{N} = 160, p = 1600}$							
New	4.84	4.90	4.93	4.79	64.03	91.40	99.98	100.00	100.00	100.00	100.00	100.00
Dempster	5.79	5.51	5.20	5.09	7.07	5.99	5.53	5.40	13.68	8.04	6.61	6.27
BS	6.67	6.80	6.81	6.86	8.23	7.58	7.33	7.22	15.92	10.62	9.19	8.67
CQ	6.87	6.87	6.88	6.90	8.51	7.68	7.41	7.25	16.48	10.76	9.23	8.70
SD_with_adjust	1.98	0.37	0.05	0.01	2.46	0.39	0.04	0.01	4.50	0.46	0.07	0.01
SD_no_adjust	6.75	6.88	6.84	6.82	8.36	7.68	7.37	7.18	16.23	10.74	9.20	8.65
Lauter	5.13	5.09	5.08	5.07	5.08	5.12	5.15	5.13	5.14	5.09	5.05	5.08
RP_single	4.18	4.23	4.50	4.46	19.18	32.89	70.48	100.00	88.06	98.55	100.00	100.00
RP_average	0.00	0.00	0.00	0.00	1.45	19.73	99.65	100.00	100.00	100.00	100.00	100.00

Table 4.6: Comparison for two-sample tests (equal covariance): multivariate t with Σ_2

	$c=0$				$c=0.5$				$c=1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\tilde{\mathbf{N}}=\mathbf{40}, \mathbf{p}=\mathbf{400}$							
New	5.04	5.10	5.08	4.68	13.38	10.90	7.75	7.46	80.51	64.01	39.52	34.36
Dempster	0.09	0.29	1.42	3.92	3.37	4.67	6.08	6.66	70.00	62.35	48.70	20.05
BS	0.15	0.50	1.84	5.16	4.24	6.01	7.27	8.62	73.79	66.36	53.13	24.32
CQ	5.44	5.40	5.78	6.57	31.40	25.32	17.16	10.48	97.77	92.09	73.19	28.26
SD_with_adjust	0.07	0.17	0.77	1.64	2.46	3.48	4.36	2.83	66.85	58.39	42.85	10.98
SD_no_adjust	0.32	0.80	2.25	6.13	6.02	7.73	9.01	9.77	79.91	72.58	58.98	27.25
Lauter	0.89	1.18	2.11	4.71	1.25	1.95	3.16	4.86	3.21	4.37	5.76	5.76
RP_single	4.13	4.23	4.01	4.43	8.17	7.85	7.29	10.25	27.11	25.63	24.47	42.71
RP_average	0.00	0.00	0.06	0.95	0.02	0.15	0.62	5.85	24.79	22.91	24.10	61.47
					$\tilde{\mathbf{N}}=\mathbf{40}, \mathbf{p}=\mathbf{1600}$							
New	5.13	4.92	4.89	5.06	7.81	6.66	6.02	4.95	38.90	29.74	16.68	8.55
Dempster	0.00	0.00	0.08	1.98	0.00	0.04	0.24	2.74	0.51	1.20	3.01	5.89
BS	0.00	0.00	0.13	2.68	0.00	0.06	0.35	3.57	0.81	1.63	3.66	7.33
CQ	4.91	5.24	5.43	6.30	14.31	12.57	10.22	7.85	68.72	56.03	35.45	14.09
SD_with_adjust	0.00	0.00	0.00	0.46	0.00	0.00	0.06	0.75	0.07	0.28	0.90	1.82
SD_no_adjust	0.00	0.01	0.23	3.22	0.02	0.09	0.56	4.36	1.55	2.67	5.41	8.80
Lauter	0.06	0.15	0.46	2.92	0.09	0.18	0.62	3.01	0.15	0.27	0.86	3.31
RP_single	4.13	4.26	4.42	4.12	5.00	4.96	5.43	5.21	8.26	8.08	8.17	9.49
RP_average	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.20	0.00	0.01	0.01	1.55
					$\tilde{\mathbf{N}}=\mathbf{160}, \mathbf{p}=\mathbf{400}$							
New	4.99	5.03	5.28	5.06	71.00	50.11	30.79	60.74	100.00	99.98	98.92	100.00
Dempster	1.29	2.09	3.29	5.01	92.02	83.80	62.53	22.30	99.99	99.98	99.97	96.86
BS	1.61	2.43	3.93	6.43	92.90	85.13	65.17	26.29	100.00	99.99	99.98	97.86
CQ	5.48	5.86	6.07	6.86	98.18	92.34	71.95	27.74	100.00	100.00	100.00	98.21
SD_with_adjust	1.15	1.77	2.66	2.23	91.46	83.11	59.66	12.83	100.00	99.99	99.99	91.12
SD_no_adjust	1.67	2.72	4.15	6.57	93.79	86.45	66.57	27.01	100.00	100.00	99.99	97.79
Lauter	1.60	2.36	3.52	4.72	3.85	5.82	7.20	5.12	13.70	17.66	17.39	6.39
RP_single	4.42	4.19	4.80	4.32	61.58	48.45	37.46	80.79	100.00	99.98	99.68	100.00
RP_average	0.13	0.33	0.67	0.93	80.63	60.66	42.38	95.20	100.00	100.00	100.00	100.00
					$\tilde{\mathbf{N}}=\mathbf{160}, \mathbf{p}=\mathbf{1600}$							
New	4.41	4.40	4.63	4.93	39.23	27.39	15.91	14.29	99.99	99.41	90.02	83.68
Dempster	0.03	0.10	1.19	3.75	12.97	13.83	13.79	9.77	99.06	98.44	95.22	55.31
BS	0.03	0.17	1.36	4.39	14.52	15.45	15.22	11.14	99.21	98.71	96.02	59.15
CQ	5.18	5.41	5.55	5.92	70.50	55.80	35.26	14.12	100.00	100.00	99.58	65.32
SD_with_adjust	0.01	0.08	0.72	1.68	10.05	11.05	10.58	5.21	98.94	98.21	93.73	41.70
SD_no_adjust	0.05	0.19	1.43	4.79	16.30	17.05	16.65	11.51	99.47	99.15	96.45	60.32
Lauter	0.25	0.37	0.95	3.98	0.27	0.62	1.40	4.38	0.61	1.12	2.59	5.00
RP_single	4.30	4.24	4.25	3.97	14.12	13.28	12.58	21.07	69.82	66.38	62.81	91.42
RP_average	0.00	0.00	0.00	0.45	0.21	0.47	1.41	14.51	99.01	96.42	89.90	99.70

Table 4.7: Comparison for two-sample tests (equal covariance): multivariate t with Σ_3

	$c = 0$				$c = 0.5$				$c = 1$			
Methods	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
					$\mathbf{\tilde{N} = 40, p = 400}$							
New	4.93	5.01	4.90	4.83	13.99	14.83	12.72	13.03	82.71	81.46	71.78	69.60
Dempster	4.39	5.47	5.57	5.31	6.81	6.90	6.50	5.97	27.86	13.57	10.26	8.35
BS	4.99	6.53	6.94	7.04	7.63	8.18	8.14	7.86	32.10	16.03	12.60	10.98
CQ	7.17	7.33	7.26	7.18	11.25	9.32	8.50	8.05	47.11	18.06	13.23	11.28
SD_with_adjust	3.23	2.96	2.07	0.95	5.01	3.85	2.42	1.15	21.16	7.76	3.75	1.60
SD_no_adjust	5.34	6.92	7.23	7.23	8.59	8.72	8.42	7.97	37.46	17.57	13.40	11.39
Lauter	4.65	4.89	5.03	5.01	5.10	5.21	5.14	5.07	5.90	5.56	5.25	5.18
RP_single	4.10	4.30	4.60	4.36	8.56	8.84	10.34	16.59	30.99	34.13	39.78	71.04
RP_average	0.00	0.00	0.00	0.65	0.06	0.16	1.12	12.45	36.62	50.00	62.66	96.40
					$\mathbf{\tilde{N} = 40, p = 1600}$							
New	5.14	5.09	5.36	4.90	7.08	7.06	7.26	5.74	35.19	36.14	31.04	17.83
Dempster	4.37	5.46	5.50	5.30	4.88	5.74	5.71	5.48	6.51	6.62	6.35	6.08
BS	5.01	6.50	6.87	6.97	5.43	6.85	7.06	7.17	7.32	7.89	7.85	7.73
CQ	7.05	7.18	7.18	7.23	7.82	7.52	7.43	7.32	11.19	8.92	8.30	8.01
SD_with_adjust	2.18	1.72	0.90	0.40	2.49	1.83	0.95	0.41	3.59	2.23	1.10	0.44
SD_no_adjust	5.42	6.80	7.09	7.18	5.97	7.09	7.32	7.32	8.21	8.49	8.17	7.96
Lauter	4.66	4.94	4.98	4.96	4.75	4.96	5.02	5.02	4.88	5.00	5.04	5.03
RP_single	4.45	4.25	4.15	4.17	5.43	5.34	5.39	6.25	9.21	9.44	11.00	15.23
RP_average	0.00	0.00	0.01	0.07	0.00	0.01	0.01	0.31	0.01	0.02	0.23	6.03
					$\mathbf{\tilde{N} = 160, p = 400}$							
New	5.11	4.94	5.10	4.92	83.61	79.63	66.40	93.42	100.00	100.00	100.00	100.00
Dempster	5.27	5.57	5.30	4.99	33.35	13.43	9.67	7.86	99.97	98.65	64.09	30.21
BS	5.86	6.66	6.83	6.69	38.18	15.74	12.16	10.40	99.99	99.50	78.48	42.45
CQ	6.64	6.82	6.94	6.73	43.47	16.26	12.32	10.52	100.00	99.63	79.42	42.93
SD_with_adjust	4.24	3.15	2.00	0.95	26.34	7.34	3.18	1.43	99.97	85.02	17.74	4.41
SD_no_adjust	5.98	6.70	6.83	6.71	39.84	16.07	12.19	10.38	100.00	99.61	79.48	43.56
Lauter	4.86	4.84	4.82	4.86	5.59	5.16	5.07	4.93	7.67	6.19	5.64	5.34
RP_single	4.57	4.57	4.50	4.43	72.28	70.50	64.16	95.87	100.00	100.00	100.00	100.00
RP_average	0.13	0.17	0.30	0.54	93.39	90.03	81.47	99.89	100.00	100.00	100.00	100.00
					$\mathbf{\tilde{N} = 160, p = 1600}$							
New	4.77	4.47	4.69	5.01	48.37	49.24	37.79	36.38	99.99	99.99	99.96	99.78
Dempster	5.60	5.77	5.62	5.39	8.47	7.04	6.42	6.03	34.66	13.47	9.74	8.25
BS	6.16	6.61	6.74	6.84	9.53	8.21	7.81	7.67	39.56	15.78	12.07	10.57
CQ	6.89	6.87	6.89	6.90	10.64	8.56	7.97	7.81	46.56	16.52	12.30	10.67
SD_with_adjust	3.80	1.94	0.69	0.38	5.60	2.41	0.80	0.38	21.05	4.50	1.34	0.50
SD_no_adjust	6.28	6.73	6.81	6.86	9.66	8.29	7.96	7.74	41.54	16.25	12.26	10.67
Lauter	5.07	5.15	5.10	5.11	5.06	5.07	5.10	5.14	5.29	5.17	5.09	5.10
RP_single	4.65	4.12	4.42	4.38	16.65	17.85	21.38	42.50	78.06	84.39	89.74	99.91
RP_average	0.00	0.00	0.00	0.23	0.37	1.37	5.67	49.92	99.92	99.98	99.95	100.00

4.4 Extension to normal populations with unequal covariances

In this section, we discuss the situation in which the two random samples have unequal covariance matrices. Under this setup, we consider two independent normal random samples, $\{\mathbf{x}_{1i}\}_{i=1}^{N_1}$ from $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $\{\mathbf{x}_{2i}\}_{i=1}^{N_2}$ from $N_p(\boldsymbol{\mu}_2, \Sigma_2)$. The problem of interest is to test the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus the alternative $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.

4.4.1 Optimal projection direction

Recall that the search for the optimal direction requires expressing the power function of the projected data analytically at the first place, which provides an access to identify factors that contribute to the increase of the power. In the case of unequal covariance matrices, the Hotelling's T^2 test can not be directly applied, which means that the power function can hardly be established. To see this, consider a projection matrix $A_{p\times k}$ of rank k. The previously-adopted projection scheme gives

$$\mathbf{y}_{1j} = A^T\mathbf{x}_{1j}, j = 1, \ldots, N_1, \tag{4.23}$$

$$\mathbf{y}_{2j} = A^T\mathbf{x}_{2j}, j = 1, \ldots, N_2. \tag{4.24}$$

By the property of normal distribution,

$$\mathbf{y}_{1j} \overset{\text{i.i.d.}}{\sim} N(A^T\boldsymbol{\mu}_1, A'\Sigma_1 A), j = 1, \ldots, N_1, \tag{4.25}$$

$$\mathbf{y}_{2j} \overset{\text{i.i.d.}}{\sim} N(A^T\boldsymbol{\mu}_2, A'\Sigma_2 A), j = 1, \ldots, N_2. \tag{4.26}$$

It follows that

$$\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2 \sim N\left(A^T\boldsymbol{\mu}_d, A^T\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)A\right). \tag{4.27}$$

The Hotelling's T^2 test can not be applied to (4.27). In fact, this case is well-known as the Behrens-Fisher problem for the univariate case. Assume $N_1 < N_2$, we apply the method proposed by Scheffe (1943) and generalized to multivariate case by Bennett (1950) to obtain an one-sample sequence of size N_1 that is distributed as $N(\boldsymbol{\mu}_d, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2)$. Define

$$\mathbf{z}_i = \mathbf{x}_{1i} - \sqrt{\frac{N_1}{N_2}}\mathbf{x}_{2i} + \frac{1}{\sqrt{N_1 N_2}}\sum_{j=1}^{N_1}\mathbf{x}_{2j} - \frac{1}{N_2}\sum_{k=1}^{N_2}\mathbf{x}_{2k}, \quad i = 1, \ldots, N_1 \tag{4.28}$$

It follows that

$$\mathbf{z}_i \overset{\text{i.i.d.}}{\sim} N\left(\boldsymbol{\mu}_d, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right), \quad i = 1, \ldots, N_1. \tag{4.29}$$

Then the one-sample test technique could be applied to $\{\mathbf{z}_i\}_{i=1}^{N_1}$. Following the results of the one-sample test in Theorem 3.2.1, the optimal projection direction is

$$\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)^{-1}\boldsymbol{\mu}_d. \tag{4.30}$$

4.4.2 Asymptotic power comparison

Following previous discussion, in the situation where $\Sigma_1 \neq \Sigma_2$, a one-sample sequence of size N_1 is constructed with distribution $N(\boldsymbol{\mu}_d, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2)$, assuming $N_1 < N_2$. With the constructed one-sample sequence, the tests are just one-sample tests whose power have already been compared in Section 3.2.3. We note that among the alternatives, CQ test can be applied to the unequal covariance case without modification of the data. Therefore, we need to compare the power of our test on the constructed sample with the power of the CQ test on the original sample, before we can complete the picture.

For our projection test, the implementation is taken on the constructed data and the procedures are the same as the one-sample test. That is, the data is partitioned into an estimating set of size N_{11}, and a testing set of size N_{12}. Assume that $\sqrt{N_{12}/N_1} \to b > 0$ as $N_1 \to \infty$, Further assume that $\hat{\mathbf{a}} \to \mathbf{a} = (\Sigma_1 + (N_1/N_2)\Sigma_2)^{-1}\boldsymbol{\mu}_d$ in probability as N_1 tends to ∞. Let $\Phi(\cdot)$ and z_α denote the cumulative distribution function and upper α quantile of $N(0,1)$, respectively. Then the asymptotic power function of the proposed projection test at a given level α is

$$\beta_{1p}^*(\eta^*) = \Phi(-z_{\alpha/2} + b\sqrt{\eta^*}), \tag{4.31}$$

where

$$\eta^* = \boldsymbol{\mu}_d^T \left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2 \right)^{-1} \boldsymbol{\mu}_d. \tag{4.32}$$

Denote $\beta^*(T_{CQ2}^2)$ as the power function of CQ test under this setting and

$$\beta_{3p}^*(\eta^*|\tau_p^*) = \Phi\left\{ -z_\alpha/\sqrt{1+2\tau_p^*} + \frac{1}{2}\sqrt{\tau_p^*/(0.5+\tau_p^*)} \cdot \sqrt{\eta^*} \right\}, \tag{4.33}$$

where τ_p^* is defined in (4.38).

Proposition 4.4.1. *Assume $\sqrt{N_{12}/N_1} \to b$. The following statements are valid.*

(I) Under H_1, $\beta^*(T_{CQ2}^2) \leq \beta_{3p}^*(\eta^*|\tau_p^*)$

(II) If $b > 0.5$ and $\eta^* \to \infty$, then $\beta_{1p}^*(\eta^*) - \beta_{3p}^*(\eta^*|\tau_p^*) > 0$ for large enough p.

(III) If $\sqrt{\eta^*}b \geq z_{\alpha/2} - z_\alpha$ and $\tau_p^* \to 0$ as $p \to \infty$, then $\beta_{1p}^*(\eta^*) - \beta_{3p}^*(\eta^*|\tau_p^*) > 0$ for large enough p.

Proof. First derive the power function for CQ test. Let T_{CQ2}^2 denote the test statistic of the CQ test. Then

$$\begin{aligned} \mathrm{E}(T_{CQ2}^2) &= \|\boldsymbol{\mu}_d\|^2, & (4.34)\\ \mathrm{var}(T_{CQ2:H1}^2) &= \frac{2}{N_1(N_1-1)}\mathrm{tr}(\Sigma_1^2) + \frac{2}{N_2(N_2-1)}\mathrm{tr}(\Sigma_2^2) & (4.35)\\ &\quad + \frac{4}{N_1N_2}\mathrm{tr}(\Sigma_1\Sigma_2) + 4\boldsymbol{\mu}_d^T\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)\boldsymbol{\mu}_d. \end{aligned}$$

Under null hypothesis, the mean of T_{CQ2}^2 is 0 and the variance is

$$\mathrm{var}(T_{CQ2:H0}^2) = \frac{2}{N_1(N_1-1)}\mathrm{tr}(\Sigma_1^2) + \frac{2}{N_2(N_2-1)}\mathrm{tr}(\Sigma_2^2) + \frac{4}{N_1N_2}\mathrm{tr}(\Sigma_1\Sigma_2). \tag{4.36}$$

The power function of CQ test can be derived as follows.

$$\beta^*(T_{CQ2}^2) = P\left(\frac{T_{CQ2}^2}{\sqrt{\mathrm{var}(T_{CQ2:H0}^2)}} > z_\alpha \right) \tag{4.37}$$

$$= \quad P\left(\frac{T_{cq}-\|\boldsymbol{\mu}_d\|^2}{\sqrt{\mathrm{var}(T^2_{CQ2:H1})}} > z_\alpha \frac{\sqrt{\mathrm{var}(T^2_{CQ2:H0})}}{\sqrt{\mathrm{var}(T^2_{CQ2:H1})}} - \frac{\|\boldsymbol{\mu}_d\|^2}{\sqrt{\mathrm{var}(T^2_{CQ2:H1})}}\right)$$

$$= \quad \Phi\left(-z_\alpha \frac{\sqrt{\mathrm{var}(T^2_{CQ2:H0})}}{\sqrt{\mathrm{var}(T^2_{CQ2:H1})}} + \frac{\|\boldsymbol{\mu}_d\|^2}{\sqrt{\mathrm{var}(T^2_{CQ2:H1})}}\right)$$

Define

$$\tau_p^* = \frac{\boldsymbol{\mu}_d^T\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)\boldsymbol{\mu}_d}{\frac{1}{N_1(N_1-1)}\mathrm{tr}(\Sigma_1^2) + \frac{1}{N_2(N_2-1)}\mathrm{tr}(\Sigma_2^2) + \frac{2}{N_1N_2}\mathrm{tr}(\Sigma_1\Sigma_2)} \tag{4.38}$$

Then $\beta^*(T^2_{CQ2})$ can be formulated as

$$\beta^*(T^2_{CQ2}) = \Phi\left\{-z_\alpha/\sqrt{1+2\tau_p^*} + \frac{1}{2}\sqrt{\tau_p^*/(0.5+\tau_p^*)w}\right\}, \tag{4.39}$$

where

$$w = \frac{\|\boldsymbol{\mu}_d\|^2}{\boldsymbol{\mu}_d^T\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)\boldsymbol{\mu}_d} \tag{4.40}$$

By Cauchy-Schwartz inequality,

$$\|\boldsymbol{\mu}_d\|^2 \le \left(\boldsymbol{\mu}_d^T\left(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2\right)\boldsymbol{\mu}_d\right)^{1/2}\sqrt{\eta^*}$$

Therefore we have $\beta^*(T^2_{CQ2}) \le \beta^*_{3p}(\eta^*|\tau_p^*)$. This completes the proof for part (I). The proof of part (II) and (III) follows the similar arguments in Proposition 3.2.3.

□

4.4.3 Simulation study

In this section, we conduct simulations to compare the finite sample performance under the unequal covariance. In this case, we set the two sample sizes to be different that N_1 is selected from $(40, 160)$ and N_2 is 1.25 times of N_1. The dimension p is still set to be 400 or 1600. The settings of the mean vectors are the same as the equal covariance cases. The first 10 random variables in the first sample have mean c, which takes value in $\{0, 0.5, 1\}$, and the rest random variables have mean

0. All the random variables in the second sample have mean 0. We consider the following scenarios for the different covariance matrices.

1. Both Σ_1 and Σ_2 have autocorrelation structure but with different levels of correlation. That is, $\Sigma_1(i,j) = \rho_1$, $\Sigma_2(i,j) = \rho_2$, and ρ_1 and ρ_2 take different values. We consider three pairs of (ρ_1, ρ_2) in the simulation study as $(\rho_1, \rho_2) = (0.25, 0.5), (0.5, 0.75)$ and $(0.25, 0.75)$.

2. Both Σ_1 and Σ_2 have compound symmetry structure but with different levels of correlation. That is, $\Sigma_1(i,j) = \rho_1^{|i-j|}$, $\Sigma_2(i,j) = \rho_2^{|i-j|}$, and ρ_1 and ρ_2 take different values. We consider three pairs of (ρ_1, ρ_2) in the simulation study as $(\rho_1, \rho_2) = (0.25, 0.5), (0.5, 0.75)$ and $(0.25, 0.75)$.

3. Σ_1 and Σ_2 take different correlation structures. Σ_1 is autocorrelation and the Σ_2 is compound symmetry. We consider $\Sigma_1(i,j) = \rho_1$ and $\Sigma_2(i,j) = \rho_2^{|i-j|}$ with $\rho_1 = \rho_2 = \rho \in \{0.25, 0.5, 0.75\}$.

In implementation of CQ test, we use the two-sample version which can handle the unequal covariances. The other tests are applied to the constructed one-sample sequence with the corresponding one-sample tests. The simulation results are presented in Tables 4.8-4.10. Generally, the results line well with the equal variance scenarios. The proposed projection test, Lauter test and Random single projection test keep the Type I error well around α. When both covariance matrices are compound symmetry, Table 4.8 shows that our proposed test significantly outperforms all the other alternatives, and preserves strong power even under small value of c. Even though CQ test can handle the unequal covariance, the power of CQ test is no higher than BS test. Table 4.9 shows that Dempster test, BS test, CQ test and SD test have advantages in the autocorrelation setting. The proposed test can reach to a comparable level with the increase of c or the sample size. Under this case, we observe a slight increase in power of CQ test over BS test. In Table 4.10, the projection test is the most powerful in the great majority of settings. With ρ small as 0.25, the Dempster test, BS test, CQ test and SD test may show comparable performance. However, their power drops significantly with the increase of ρ.

Table 4.8: Comparison for two-sample tests (unequal covariances): multivariate normal with $\Sigma_1(i,j) = \rho_1$ and $\Sigma_2(i,j) = \rho_2$

	$c = 0$			$c = 0.5$			$c = 1$		
ρ_1	0.25	0.25	0.50	0.25	0.25	0.50	0.25	0.25	0.50
ρ_2	0.50	0.75	0.75	0.50	0.75	0.75	0.50	0.75	0.75
$\mathbf{N_1 = 40, N_2 = 50, p = 400}$									
New	4.94	5.15	4.81	26.91	34.24	51.75	92.42	95.06	97.97
Dempster	6.12	5.87	5.59	7.94	7.11	6.55	17.48	12.91	10.29
BS	7.50	7.56	7.56	9.78	9.28	8.87	21.93	16.99	14.00
CQ	7.19	7.07	7.13	9.01	8.46	8.11	20.11	15.75	12.80
SD_with_adjust	2.62	1.76	1.02	3.47	2.10	1.17	7.80	3.92	1.79
SD_no_adjust	8.14	8.05	7.79	10.59	9.85	9.23	23.77	18.32	14.48
Lauter	4.83	4.70	4.90	4.97	4.79	5.00	5.35	5.12	5.15
RP_single	4.94	4.62	4.91	10.63	11.55	15.94	36.28	44.43	58.04
RP_average	0.08	0.13	0.17	1.13	2.00	6.31	84.43	96.77	99.98
$\mathbf{N_1 = 40, N_2 = 50, p = 1600}$									
New	4.71	4.62	4.76	8.67	10.14	14.16	48.43	56.35	69.13
Dempster	6.22	5.93	5.57	6.51	6.22	5.73	8.09	7.18	6.44
BS	7.66	7.68	7.71	8.22	8.02	7.99	9.76	9.18	8.91
CQ	7.34	7.25	7.31	7.70	7.61	7.52	9.20	8.81	8.26
SD_with_adjust	1.79	0.90	0.40	1.90	0.98	0.42	2.22	1.15	0.45
SD_no_adjust	8.34	8.15	8.05	8.93	8.48	8.38	10.65	9.72	9.26
Lauter	4.83	4.95	4.97	4.85	4.97	4.94	4.90	4.99	4.94
RP_single	5.08	5.09	5.06	6.18	6.62	7.19	10.97	11.88	15.05
RP_average	0.05	0.05	0.08	0.12	0.12	0.23	0.73	1.48	4.74
$\mathbf{N_1 = 160, N_2 = 200, p = 400}$									
New	4.81	5.30	5.05	99.97	99.99	100.00	100.00	100.00	100.00
Dempster	5.80	5.43	5.42	15.32	11.45	9.14	100.00	98.29	66.73
BS	6.95	6.93	6.98	18.83	14.86	12.47	100.00	99.88	89.75
CQ	6.88	6.88	6.97	18.53	14.63	12.00	100.00	100.00	93.35
SD_with_adjust	2.28	1.28	0.66	5.64	2.68	1.14	87.04	25.35	5.99
SD_no_adjust	7.04	7.08	7.02	19.25	15.04	12.67	100.00	99.89	89.95
Lauter	5.09	5.09	5.18	5.17	5.10	5.15	5.96	5.63	5.46
RP_single	5.09	5.07	5.05	86.60	93.52	98.56	100.00	100.00	100.00
RP_average	0.04	0.01	0.01	100.00	100.00	100.00	100.00	100.00	100.00
$\mathbf{N_1 = 160, N_2 = 200, p = 1600}$									
New	4.53	4.89	4.34	89.59	96.09	99.41	100.00	100.00	100.00
Dempster	5.77	5.48	5.35	7.21	6.56	6.18	15.07	11.39	9.12
BS	6.93	7.00	6.97	8.94	8.45	7.96	19.03	14.83	12.45
CQ	6.76	6.90	6.96	8.86	8.27	7.90	18.58	14.64	12.20
SD_with_adjust	1.03	0.47	0.10	1.31	0.63	0.13	2.77	0.97	0.25
SD_no_adjust	7.12	7.18	7.07	9.09	8.52	8.07	19.55	15.14	12.55
Lauter	5.08	4.99	5.04	5.10	5.04	5.01	5.08	5.09	5.01
RP_single	5.20	4.91	4.71	22.62	27.08	37.10	87.32	93.92	98.69
RP_average	0.00	0.00	0.00	3.30	12.09	53.26	100.00	100.00	100.00

Table 4.9: Comparison for two-sample tests (unequal covariances): multivariate normal with $\Sigma_1(i,j) = \rho_1^{|i-j|}$ and $\Sigma_2(i,j) = \rho_2^{|i-j|}$

	$c = 0$			$c = 0.5$			$c = 1$		
ρ_1	0.25	0.25	0.50	0.25	0.25	0.50	0.25	0.25	0.50
ρ_2	0.50	0.75	0.75	0.50	0.75	0.75	0.50	0.75	0.75
$\mathbf{N_1 = 40, N_2 = 50, p = 400}$									
New	4.90	5.06	5.11	17.28	14.67	13.07	91.25	81.54	73.40
Dempster	4.88	5.15	5.03	48.00	40.26	34.12	99.97	99.67	98.70
BS	5.18	5.80	5.71	49.46	41.87	35.74	99.97	99.71	98.93
CQ	5.33	5.63	5.56	50.03	42.06	36.13	99.98	99.74	98.99
SD_with_adjust	3.52	3.78	3.59	41.11	34.27	28.35	99.93	99.33	97.52
SD_no_adjust	6.99	7.21	7.57	54.01	45.82	40.10	100.00	99.74	98.91
Lauter	4.62	4.71	4.87	19.29	16.07	14.54	65.13	50.35	41.71
RP_single	5.11	4.80	4.72	8.32	8.07	8.27	23.20	23.22	22.83
RP_average	0.00	0.01	0.03	0.19	0.38	0.54	33.08	32.46	32.70
$\mathbf{N_1 = 40, N_2 = 50, p = 1600}$									
Ours	4.78	4.96	4.57	8.49	7.54	7.04	52.54	42.42	35.29
Dempster	4.69	4.95	5.03	21.60	18.18	16.49	92.68	84.31	76.47
BS	4.96	5.27	5.36	22.43	19.09	17.33	92.92	84.85	77.32
CQ	5.11	5.29	5.35	22.29	19.13	17.23	93.29	85.17	77.92
SD_with_adjust	1.77	1.87	2.02	11.22	9.16	8.22	83.34	71.17	61.06
SD_no_adjust	7.01	7.14	7.44	26.68	22.75	20.84	93.85	86.61	80.17
Lauter	4.81	5.11	5.07	10.46	9.19	8.98	28.41	21.94	19.11
RP_single	4.38	5.31	5.08	5.21	6.05	5.90	8.38	8.62	8.59
RP_average	0.00	0.00	0.00	0.00	0.01	0.00	0.05	0.04	0.03
$\mathbf{N_1 = 160, N_2 = 200, p = 400}$									
Ours	4.99	5.15	4.93	90.55	79.53	69.84	100.00	100.00	100.00
Dempster	5.02	5.10	5.39	99.96	99.70	98.84	100.00	100.00	100.00
BS	5.40	5.57	6.00	99.96	99.75	99.00	100.00	100.00	100.00
CQ	5.29	5.65	5.93	99.96	99.79	98.98	100.00	100.00	100.00
SD_with_adjust	4.32	4.39	4.71	99.93	99.60	98.52	100.00	100.00	100.00
SD_no_adjust	5.85	6.20	6.58	99.96	99.71	98.97	100.00	100.00	100.00
Lauter	5.20	4.92	5.21	55.41	39.42	32.05	99.86	92.39	81.74
RP_single	5.02	4.74	5.20	53.43	47.32	47.14	99.94	99.91	99.88
RP_average	0.02	0.11	0.16	83.23	72.46	69.07	100.00	100.00	100.00
$\mathbf{N_1 = 160, N_2 = 200, p = 1600}$									
Ours	4.62	4.67	4.99	57.29	45.20	36.66	100.00	100.00	99.96
Dempster	5.35	4.93	4.99	93.56	85.85	77.42	100.00	100.00	100.00
BS	5.68	5.23	5.25	93.71	86.36	78.21	100.00	100.00	100.00
CQ	5.60	5.12	5.19	93.75	86.33	78.48	100.00	100.00	100.00
SD_with_adjust	3.86	3.78	3.85	91.12	82.22	72.45	100.00	100.00	100.00
SD_no_adjust	5.98	5.72	5.66	93.93	86.69	78.96	100.00	100.00	100.00
Lauter	5.13	4.95	5.09	22.27	17.76	15.66	73.04	53.18	43.04
RP_single	5.06	5.50	4.83	13.85	14.05	13.73	59.87	59.91	59.25
RP_average	0.00	0.00	0.00	0.20	0.27	0.62	98.71	97.21	96.49

Table 4.10: Comparison for two-sample tests (unequal covariances): multivariate normal with $\Sigma_1(i,j) = \rho_1^{|i-j|}$ and $\Sigma_2(i,j) = \rho_2$

	$c = 0$			$c = 0.5$			$c = 1$		
ρ_1	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
ρ_2	0.25	0.50	0.75	0.25	0.50	0.75	0.25	0.50	0.75
$\mathbf{N_1 = 40, N_2 = 50, p = 400}$									
New	5.07	5.20	5.11	17.59	17.25	15.15	88.68	84.44	77.70
Dempster	6.70	6.29	6.01	15.84	10.00	8.03	91.59	37.68	19.56
BS	7.35	7.39	7.47	17.32	11.56	10.06	93.67	43.58	24.03
CQ	7.29	7.12	7.15	16.95	11.17	9.71	94.62	42.60	22.96
SD_with_adjust	5.10	4.11	2.88	12.78	6.48	3.79	83.56	24.04	8.89
SD_no_adjust	8.53	8.20	8.15	19.82	12.88	10.97	95.01	48.82	26.84
Lauter	4.72	4.69	4.63	5.60	5.17	4.93	7.48	5.91	5.28
RP_single	4.98	4.79	5.06	8.66	8.88	10.20	26.14	28.17	32.18
RP_average	0.05	0.07	0.21	0.29	0.80	2.27	42.82	54.86	67.10
$\mathbf{N_1 = 40, N_2 = 50, p = 1600}$									
New	5.04	4.70	4.68	7.46	7.23	6.69	38.91	36.40	30.60
Dempster	7.08	6.78	6.33	8.62	7.57	6.79	16.81	10.37	8.38
BS	7.71	7.63	7.72	9.28	8.44	8.28	17.98	11.64	10.26
CQ	7.56	7.61	7.70	8.99	8.44	8.22	17.56	11.05	9.76
SD_with_adjust	4.25	2.89	1.66	5.28	3.28	1.75	9.88	4.44	2.46
SD_no_adjust	8.54	8.48	8.43	10.40	9.51	9.05	20.51	12.82	11.12
Lauter	4.84	4.91	4.99	5.01	5.04	5.07	5.43	5.19	5.15
RP_single	4.87	4.94	4.72	5.77	5.69	5.91	8.94	9.13	10.04
RP_average	0.02	0.05	0.07	0.03	0.07	0.07	0.20	0.39	0.75
$\mathbf{N_1 = 160, N_2 = 200, p = 400}$									
New	4.92	5.30	4.59	97.79	96.27	86.13	100.00	100.00	100.00
Dempster	6.30	6.09	5.79	94.78	33.49	16.86	100.00	100.00	99.98
BS	6.83	7.05	7.14	96.11	40.32	20.92	100.00	100.00	99.99
CQ	6.83	7.07	7.17	96.45	39.72	20.82	100.00	100.00	99.99
SD_with_adjust	5.21	3.89	2.32	90.31	19.12	6.46	100.00	100.00	90.79
SD_no_adjust	7.01	7.26	7.32	96.33	42.05	21.41	100.00	100.00	100.00
Lauter	5.04	5.04	5.01	6.30	5.63	5.50	8.99	6.55	6.20
RP_single	4.93	4.91	5.04	64.17	66.05	67.09	99.99	100.00	100.00
RP_average	0.05	0.06	0.33	96.16	95.90	92.95	100.00	100.00	100.00
$\mathbf{N_1 = 160, N_2 = 200, p = 1600}$									
New	4.73	4.68	4.79	67.08	65.46	52.38	100.00	100.00	100.00
Dempster	6.11	5.77	5.60	14.31	8.81	7.24	99.66	31.98	16.75
BS	6.60	6.75	6.56	15.54	10.15	8.79	99.91	38.65	20.60
CQ	6.59	6.71	6.65	15.31	10.00	8.69	99.94	38.14	20.51
SD_with_adjust	4.48	2.47	1.31	9.93	3.75	1.59	94.85	12.04	3.27
SD_no_adjust	6.79	6.99	6.70	16.03	10.43	9.06	99.81	40.20	21.09
Lauter	4.76	4.77	4.84	4.93	4.87	4.84	5.26	5.12	4.93
RP_single	4.90	5.31	5.09	15.32	17.30	19.86	68.64	74.51	81.02
RP_average	0.00	0.00	0.01	0.38	1.26	4.69	99.90	99.96	99.98

4.4.4 Real data example

In this section, we revisit the example that have been analyzed in the Section 4.2.5. Recall that two groups of data, a mutant group of size 33 and a normal group of size 17, are available with dimension $p = 263$. We have applied the two-sample test methods to this dataset and found that the proposed method and the SD test without adjustment can identify the difference. A follow up analysis with two-sample t-test and Bonferroni correction identifies two genes BAX and MDM2, both of which have well-established relation with the p53 factor. As mentioned, it is difficult to justify whether the covariance matrix are equal with a small sample size. Therefore, it is good to construct a transformed one-sample sequence and apply the one-sample test which does not require the equal covariance matrix assumption. By doing so, we could obtain a better understanding of the data with test results from both versions.

In this section, we apply the one-sample t-test to the constructed one-sample sequence which has 17 samples and the dimension is 263. Table 4.11 shows the p-values of different tests. In Table 4.11, the p-value of the CQ test is also obtained from the constructed sequence. Its p-value for the direct two-sample test is 0.43 as shown in Table 4.4.

Table 4.11: Gene pathway dataset: p-values of the one-sample tests on the constructed sequence

New	Dempster	BS	CQ	SD_with	SD_no	Lauter	RP_s	RP_m
0.02	0.46	0.50	0.50	0.18	0.08	0.94	0.25	0.22

As shown in Table 4.11, our test has p-value less than 0.05 and is able to conclude the different expression levels for the pathway in cancer between the p53 normal and mutant groups. The other alternatives, including the SD test without adjustment, however, do not show the significant evidence. Our test indicates that the pathway in cancer is a possible target of the p53. Again experiments are needed for the confirmatory conclusion.

We conduct a follow-up analysis with marginal one-sample t-tests. The histogram of the corresponding p-values is shown in Figure 4.2. Similar to Figure 4.1 of the two-sample marginal test, the p-values are well-scattered in the (0,1) interval, indicating weak signals for the majority of the individual genes. There are still

20 genes that have the marginal significance under $\alpha = 0.05$, of which 17 genes are overlapped with the marginally significant genes in the two-sample t-test. Here only gene BAX survives after Bonferroni correction, with p-value 9.23×10^{-6}.

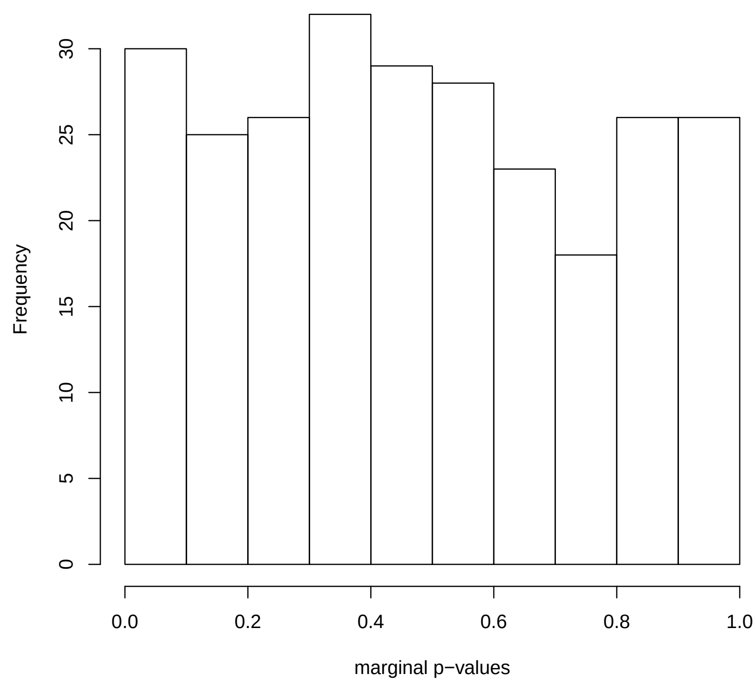


Figure 4.2: Histogram of the marginal p-values from the one-sample t-tests.

Chapter 5

Conclusions and Future Work

In the analysis of high-dimensional data, considerable attention has been devoted to sparse recovery and feature selection in the past decades. However, much less is known about the hypothesis testing. In classical settings for the one-sample and two-sample mean problems, the Hotelling's T^2 test can be applied and is theoretically grounded. Hotelling's T^2 evaluates the weighted distance between mean vectors by scaling with inverse of sample covariance matrix. In the high-dimensional scenarios, the sample covariance matrix is singular which renders the Hotelling's test invalid. Recently in literature, hypothesis testing for high-dimensional mean vectors has received considerable attention. Existing methods can be classified in two categories. One main focus is to substitute the inverse sample covariance matrix and obtain an asymptotic normal distribution to determine the rejection region. The other projects the data into a lower dimensional space such that the classical methods could be applied which leads to an exact test.

Our work is motived by the idea of projection test by noting that the current projection based methods lack an explicit mechanism to promote the power. Under this unique power-maximizing framework, we determine the optimal projection direction and develop an algorithm to obtain the estimate. In this section, we conclude this dissertation by first summarizing our work in the one-sample and two-sample tests; and then discussing the potential ways to extend the study.

5.1 Conclusions

This dissertation has been focused on testing mean vectors in the one-sample and two-sample problems. These problems are fundamental in statistics and well established in the classical settings where the sample size is larger than the dimension. However, as dimension grows, new methods are in demand to handle the high-dimensionality. We propose a novel projection test that is readily applicable to the high-dimensional settings with a computationally efficient algorithm for implementation. Moreover, we derive the asymptotic power and give conditions in which the proposed methods have higher power. Finite sample performance is studied with extensive simulations.

Chapter 3 focuses on the high-dimensional one-sample mean problem, which tests $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ against $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ for a random sample of size N from a p-dimensional ($p > N$) population $\mathbf{x}$ with finite mean $E(\mathbf{x}) = \boldsymbol{\mu}$ and finite positive definite covariance matrix $\mathrm{cov}(\mathbf{x}) = \Sigma$. We show the optimal direction for the projection test is $\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_0)$. The implementation follows a random partition scheme such that the sample is separated into an estimating set from which the direction is estimated and a testing set in which the test is conducted. Ridge-like estimator $(\mathbf{S} + \lambda\mathbf{D})^{-1}$ is used to estimate Σ^{-1} where $\mathbf{S}$ is the sample covariance matrix and $\mathbf{D}$ is the diagonal matrix of $\mathbf{S}$. Here the λ is the ridge penalty that tunes the estimator. With normality assumption, the proposed test is an exact test and we study the asymptotic power function of the proposed test under local alternative $H_1 : \boldsymbol{\mu} = \boldsymbol{\delta}/\sqrt{N}$. We show that under mild conditions, our algorithm can achieve higher power and discuss in details the two commonly-used covariance matrix structures. For the implementation, we carefully study the two tuning parameters, the splitting percentage and the ridge penalty, in extensive simulation experiments. We conclude that the splitting percentage is flexible in a range from 40% - 60% and suggest the λ take value of the $N_1^{-0.5}$, where N_1 is the sample size in the estimating set. Without normality assumption, we show that the optimal projection direction still applies.

Chapter 4 focuses on the two-sample mean problem, which tests $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ against $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ for two independent random samples, a size N_1 sample from a population with finite mean $\boldsymbol{\mu}_1$ and finite positive definite covariance matrix Σ_1, and a size N_2 sample from a population with finite mean $\boldsymbol{\mu}_2$ and finite positive definite covariance matrix Σ_2. The equal covariance case $\Sigma_1 = \Sigma_2 = \Sigma$ follows a natural extension from the one-sample test and takes $\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ as the optimal projection direction, regardless of assumptions for distributions. The correspondingly test is exact under multivariate normal assumption while asymptotic otherwise. In both cases, we study the finite sample performance from simulation experiments to have an over-all assessment of the proposed algorithm. In multivariate normal case, we also provide the asymptotic power comparison with the major alternatives. We further extend the discussion to the multivariate normal populations with unequal covariance matrices. we show that the optimal projection direction is $\left(\Sigma_1 + \frac{N_1}{N_2}\Sigma_2\right)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ by first taking Bennett's transformation to obtain an one-sample sequence of size N_1 that is distributed as $N(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \Sigma_1 + \frac{N_1}{N_2}\Sigma_2)$, assuming $N_1 < N_2$. We provide comparison of the asymptotic power for the proposed test and the completing alternative which does not require the equivalence of the covariance matrices. Finite sample performance is also evaluated in the simulation study.

5.2 Future work

5.2.1 Improve efficiency of implementation

Current implementation scheme generates two exclusive random subsets of the original data, $\mathcal{S}_1$ and $\mathcal{S}_2$, with a splitting percentage. If $\mathcal{S}_1$ is used to estimate the direction, then $\mathcal{S}_2$ will be used to obtain the projected data and conduct the test. In our simulation studies and the real data applications, we have used 40% as the splitting percentage to partition the dataset, and the smaller set is used for estimating the direction. Despite being able to keep the projected data independent by separating the estimation of direction and the projection, the random splitting step reduces the degree of freedom of the t-test, and therefore affects the power of the test. For this reason, an algorithm that would more effectively utilize the two subsets could potentially increase the power of the test.

One possible solution is to evenly partition the data and conduct the test twice, with $\mathcal{S}_1$ and $\mathcal{S}_2$ switching the role. This procedure generates two p-values, p_1 and p_2. A rejection region should be carefully designed to control the Type I error when combining p_1 and p_2. We have experimented the following two reject regions (RRs).

(RR1) Reject if $(p_1 < (\alpha/2))$ or $(p_2 < (\alpha/2))$

(RR2) Reject if $(p_1 < (\alpha/2))$ or $(p_2 < (\alpha/2))$ or $(p_1 < \alpha$ and $p_2 < \alpha)$

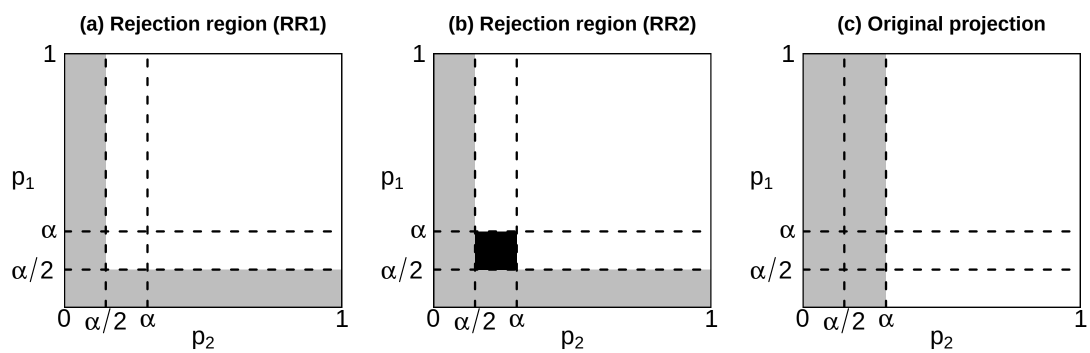


Figure 5.1: Illustration for the rejection regions.

We note that (RR1) is based on the Bonferroni correction, which is known to be conservative. Compared with (RR1), the reject region (RR2) adds one more rejection criterion. It is clear that if p_1 and p_2 are independent, the test based on (RR2) is exact. In our case, however, p_1 and p_2 are correlated due to the very nature that they come from the same original dataset. More specifically, p_1 and p_2 are positively correlated as they test the same pair of hypotheses. In this case, we expect (RR2) can still control the Type I error, but not give an exact test.

We present a preliminary simulation results with the same settings as used in Section 3.3.3. The results are based on 10000 replicates and presented in percentage. Tables 5.1-5.3 show results of rejection region (RR1), (RR2), and the original projection test under 50% split, respectively. Recall that for a test with significant level 0.05 and 10000 replicates, the Monte Carlo error equals to

$1.96\sqrt{0.05 \times 0.95/10000} = 0.43\%$. Therefore, well-controlled Type I errors are expected to vary roughly between 4.57% and 5.43%.

The simulation results are expected for both Type I error and power. In terms of Type I error, the proposed method can manage to get it well-maintained. Both new rejection regions tend to be conservative. Rejection region (RR2) is slightly better by including one more rejection criterion. In terms of power, rejection region (RR2) gives higher power than (RR1) in all cases as comparing Table 5.1 and 5.2. We believe rejection region (RR2) is a better choice to apply. If we compare that the original method and rejection region (RR2), it is observed that (RR2) is a clear winner except the autocorrelation case with $(N, p) = (40, 1600)$. More generally, the power gain from (RR2) is larger for the compound symmetry case than the auto correlation case.

Table 5.1: Performance of new rejection region RR1 with 50% split

	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
N=40,p=400												
auto	4.08	3.99	4.26	4.31	46.19	33.15	18.86	14.79	100.00	99.76	92.61	80.77
combine	3.95	4.38	4.38	4.29	47.60	48.02	38.38	35.38	99.95	99.82	99.59	98.88
compound	4.45	4.32	4.26	4.42	60.31	85.34	99.29	100.00	99.96	100.00	100.00	100.00
N=40,p=1600												
auto	3.85	4.00	4.18	3.97	16.53	12.80	8.45	5.39	96.09	87.20	59.15	23.99
combine	4.62	4.40	4.64	4.23	15.36	16.55	14.71	9.09	86.36	86.29	82.55	55.30
compound	4.34	4.54	4.30	4.23	19.08	34.58	77.91	99.96	88.74	95.53	99.80	100.00
N=160,p=400												
auto	3.97	4.05	4.07	4.27	100.00	99.65	92.19	99.20	100.00	100.00	100.00	100.00
combine	3.79	3.98	3.82	4.16	100.00	100.00	99.98	100.00	100.00	100.00	100.00	100.00
compound	3.98	3.86	3.95	3.58	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
N=160,p=1600												
auto	3.35	3.70	3.70	4.64	97.13	88.32	60.10	44.46	100.00	100.00	100.00	99.96
combine	3.60	3.66	3.76	4.64	98.61	98.77	95.80	89.56	100.00	100.00	100.00	100.00
compound	3.86	3.85	3.78	3.79	99.66	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 5.2: Performance of new rejection region RR2 with 50% split

	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
N=40,p=400												
auto	4.42	4.31	4.63	4.46	48.27	34.83	19.85	15.40	100.00	99.82	93.32	81.68
combine	4.30	4.66	4.63	4.53	49.48	49.87	39.92	36.32	99.96	99.83	99.64	99.01
compound	4.62	4.65	4.55	4.74	61.94	86.01	99.33	100.00	99.97	100.00	100.00	100.00
N=40,p=1600												
auto	4.13	4.41	4.58	4.39	17.95	13.84	9.10	5.79	96.56	88.29	61.27	24.97
combine	4.70	4.51	4.87	4.57	16.07	17.05	15.26	9.65	86.75	86.65	83.26	56.59
compound	4.52	4.66	4.45	4.36	19.67	35.44	78.60	99.96	89.03	95.64	99.80	100.00
N=160,p=400												
auto	4.32	4.32	4.28	4.45	100.00	99.70	92.78	99.33	100.00	100.00	100.00	100.00
combine	4.23	4.29	4.10	4.45	100.00	100.00	99.98	100.00	100.00	100.00	100.00	100.00
compound	4.20	4.17	4.26	4.01	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
N=160,p=1600												
auto	3.93	4.24	4.04	4.79	97.60	89.64	62.09	45.80	100.00	100.00	100.00	99.97
combine	4.12	4.23	4.12	4.80	98.76	99.01	96.35	90.34	100.00	100.00	100.00	100.00
compound	4.42	4.52	4.34	4.36	99.69	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 5.3: Performance of the original projection test with 50% split

	$c = 0$				$c = 0.5$				$c = 1$			
	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95	0.25	0.5	0.75	0.95
N=40,p=400												
auto	4.63	4.73	5.08	5.15	47.37	34.43	19.82	14.19	100.00	99.64	90.68	73.03
combine	4.78	4.91	4.75	4.90	45.90	45.79	37.33	32.71	99.59	99.17	98.66	96.41
compound	4.92	4.76	4.81	4.87	55.73	78.18	97.41	99.99	99.49	99.89	99.99	100.00
N=40,p=1600												
auto	4.69	4.80	5.12	4.73	18.05	14.13	9.58	6.30	95.32	86.31	59.04	23.44
combine	4.56	4.65	4.67	4.77	13.96	14.74	13.44	9.06	76.85	75.11	73.68	51.13
compound	4.43	4.50	4.59	4.57	16.46	28.05	62.95	99.37	78.49	88.61	98.38	100.00
N=160,p=400												
auto	5.12	5.01	4.72	4.75	99.99	99.22	88.30	97.98	100.00	100.00	100.00	100.00
combine	4.85	5.07	4.51	4.95	100.00	100.00	99.84	100.00	100.00	100.00	100.00	100.00
compound	4.72	4.75	4.68	4.62	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
N=160,p=1600												
auto	4.96	4.91	4.72	4.85	97.50	89.11	60.95	40.33	100.00	100.00	100.00	99.83
combine	4.95	4.89	4.56	4.84	98.26	98.59	95.34	85.05	100.00	100.00	100.00	100.00
compound	5.14	5.37	5.38	5.24	99.50	100.00	100.00	100.00	100.00	100.00	100.00	100.00

5.2.2 Extension to other setups

Our current method applies to one-sample and two-sample mean problems, which allow testing for one treatment effect. A test capable of accommodating multiple groups is appreciated due to the massive amount of experiments which collect data naturally in multiple groups. An search in the Gene Expression Omnibus dataset (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser/) can easily pinpoint a study with multiple treatments or being collected from subjects in different disease stages. For example, the dataset GDS4094 contains MMTV-Myc tumors data collected from subjects in an E2F wild-type, E2F1 null, E2F2 null or E2F3 heterozygous background (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS4094).

In the setup of the one-sample and two-sample test, it is required that the samples within each group are independent and identically distributed. It is likely that the samples are heterogeneous. One way to resolve the problem is by taking into account the characteristics of the group members provided in the dataset. For example, the dataset GDS5074 (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS5037) describes bronchial epithelial cells data collected from subjects of different stages of asthma. For each stage group, it also provides the gender information. Gender may have effects over the performance of cells data and the test could be more powerful with gender being considered in the analysis. The independence assumption would fail in the longitudinal study, in which each treatment group may contain observations collected at different length of time after taking the treatment. We are interested to extend our methods to accommodate the inter-correlation too.

Bibliography

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica sinica*, 6:311–329.

Bennett, B. (1950). Note on a solutions to the generalized Behrans-Fisher problem. *Annals of the Institute of Statistical Mathematics*, 2(1):87–90.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6):2350–2383.

Cai, J., Fan, J., Li, R., and Zhou, H. (2005). Variable selection for multivariate failure time data. *Biometrika*, 92(2):303–316.

Candes, E. and Tao, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, 35(6):2313–2351.

Chatterjee, A. and Lahiri, S. N. (2011). Bootstrapping lasso estimators. *Journal of the American Statistical Association*, 106(494):608–625.

Chatterjee, A. and Lahiri, S. N. (2013). Rates of convergence of the adaptive lasso estimators to the oracle distribution and higher order refinements by the bootstrap. *The Annals of Statistics*, 41(3):1232–1259.

Chen, L., Paul, D., Prentice, R. L., and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *Journal of the American Statistical Association*, 106(496):1345–1360.

Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Chipuk, J. E., Kuwana, T., Bouchier-Hayes, L., Droin, N. M., Newmeyer, D. D., Schuler, M., and Green, D. (2004). Direct activation of bax by p53 mediates mitochondrial membrane permeabilization and apoptosis. *Science*, 303(5660):1010–1014.

Cui, H., Li, R., and Zhong, W. (2014). Model-free feature screening for ultrahigh dimensional discriminant analysis. *Journal of the American Statistical Association*, (Accepted).

Dempster, A. (1958). A high dimensional two sample significance test. *The Annals of Mathematical Statistics*, 29(4):995–1010.

Dempster, A. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics*, 16(1):41–50.

Dezeure, R., Bühlmann, P., Meier, L., and Meinshausen, N. (2014). High-dimensional inference: Confidence intervals, p-values and R-Software hdi. *arXiv preprint arXiv:1408.4026*.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. *AMS Math Challenges Lecture*, pages 1–32.

Donoho, D. L. (2004). Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical Report, Department of Statistics, Stanford University.

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829.

Donoho, D. L. and Elad, M. (2003). Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202.

Dudoit, S., Shaffer, J. P., and Boldrick, J. C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477):93–103.

Efron, B. (2014). Estimation and accuracy after model selection (with discussions). *Journal of the American Statistical Association*, 109(57):991–1007.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2):407–499.

Fan, J., Feng, Y., and Song, R. (2011). Nonparametric independence screening in sparse ultra-high dimensional additive models. *Journal of the American Statistical Association*, 106:544–557.

Fan, J., Han, F., and Liu, H. (2014a). Challenges of big data analysis. *National Science Review*, (1):293–314.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *The Annals of Statistics*, 30(1):74–99.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710–723.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *Proceeding of the International Congress of Mathematicians*, pages 595–622.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:849–911.

Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, 20(1):101–148.

Fan, J. and Lv, J. (2011). Nonconcave penalized likelihood with NP-dimensionality. *Information Theory, IEEE Transactions on*, 57(8):5467–5484.

Fan, J., Ma, Y., and Dai, W. (2014b). Nonparametric independence screening in sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association*, 109(507):1270–1284.

Fan, J. and Ren, Y. (2006). Statistical analysis of DNA microarray data in cancer research. *Clinical Cancer Research*, 12(15):4469–4473.

Fan, J., Samworth, R., and Wu, Y. (2009). Ultrahigh dimensional feature selection: Beyond the linear model. *Journal of Machine Learning Research*, 10:2013–2038.

Fan, J. and Song, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics*, 38(6):3567–3604.

Frick, H. (1996). On the power behavior of Lauter's exact multivariate one-sided tests. *Biometrical Journal*, 38(4):405–414.

Ghosh, B. (1973). Some monotonicity theorems for χ^2, F and t distribution with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 35(3):480–492.

Hall, P. and Miller, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics*, 18(3):533–550.

He, X., Wang, L., and Hong, H. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *The Annals of Statistics*, 41(1):342–369.

Huang, J., Horowitz, J. L., and Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, 36(2):587–613.

Hunter, D. and Lange, K. (2000). Rejoinder to discussion of "optimization transfer using surrogate objective functions". *Journal of Computational and Graphical Statistics*, 9:52–59.

Hunter, D. R. and Li, R. (2005). Variable selection using MM algorithms. *The Annals of statistics*, 33(4):1617–1642.

Javanmard, A. and Montanari, A. (2013a). Confidence intervals and hypothesis testing for high-dimensional regression. *arXiv preprint arXiv:1306.3171*.

Javanmard, A. and Montanari, A. (2013b). Hypothesis testing in high-dimensional regression under the Gaussian random design model: Asymptotic theory. *arXiv preprint arXiv:1301.4240*.

Kai, B., Li, R., and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *The Annals of statistics*, 39(1):305–332.

Kallenberg, W. C. (1990). Inequalities for noncentral chi-square distributions. *Statistics & Probability Letters*, 9(3):273–278.

Kropf, S., Hothorn, L., and Lauter, J. (1997). Multivariate many-to-one procedures with applications to preclinical trials. *Drug Information Journal*, 31(2):433–447.

Lange, K., Hunter, D., and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). *Journal of Computational and Graphical Statistics*, 9(1):1–20.

Lauter, J. (1996). Exact t and F tests for analyzing studies with multiple endpoints. *Biometrics*, 52(3):964–970.

Lauter, J., GLimm, E., and Kropf, S. (1998). Multivariate tests based on left-spherically distributed linear scores. *The Annals of Statistics*, 26(5):1972–2988.

Lee, S. H., Lim, J., Li, E., Vannucci, M., and Petkova, E. (2012). Order test for high-dimensional two-sample means. *Journal of Statistical Planning and Inference*, 142(9):2719–2725.

Li, R. and Liang, H. (2008). Variable selection in semiparametric regression modeling. *The Annals of Statistics*, 36(1):261–286.

Li, R. and Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian kriging models. *Technometrics*, 47(2):111–120.

Li, R., Zhong, W., and Zhu, L. (2012). Feature selection via distance correlation learning. *Journal of American Statistical Association*, 107:1129–1139.

Liang, H. and Li, R. (2009). Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association*, 104(485):234–248.

Liang, H., Liu, X., Li, R., and Tsai, C. (2010). Estimation and testing for partially linear single-index models. *The Annals of statistics*, 38(6):3811–3836.

Liang, J. and Tang, M. (2009). Generalized F-tests for the multivariate normal mean. *Computational Statistics and Data Analysis*, 53(4):1177–1190.

Liu, J., Li, R., and Wu, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association*, 109(505):266–274.

Lockhart, R., Taylor, J., Tibshirani, R. J., and Tibshirani, R. (2014). A significance test for the lasso. *The Annals of Statistics*, 42(2):413–468.

Logan, B. and Tamhane, A. (2004). On O'Brien's OLS and GLS tests for multiple endpoints. *IMS Lecture Notes-Monograph Series*, 47:76–88. Recent Developments in MultipleComparison Procedures.

Lopes, M., Jacob, L., and Wainwright, M. (2011a). A more powerful two-sample test in high dimensions using random projection. *Advances in Neural Information Processing Systems 24*, pages 1206–1214. Curran Associates, Inc.

Lopes, M., Jacob, L., and Wainwright, M. (2011b). A more powerful two-sample test in high dimensions using random projection. *Manuscript*. arXiv:1108.2401v2.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34:1436–1462.

Meinshausen, N., Meier, L., and Bühlmann, P. (2009). P-values for high-dimensional regression. *Journal of the American Statistical Association*, 104(488):1671–1681.

Pan, G. M. and Zhou, W. (2011). Central limit theorem for Hotelling's T^2 statistic under large dimension. *The Annals of Applied Probability*, 21(5):1860–1910.

Park, M. Y. and Hastie, T. (2007). L_1 regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):659–677.

Percival, C. J., Huang, Y., Jabs, E. W., Li, R., and Richtsmeier, J. T. (2014). Embryonic craniofacial bone volume and bone mineral density in $Fgfr2^{+/P253R}$ and nonmutant mice. *Developmental Dynamics*, 243(4):541–551.

Ravikumar, P., Lafferty, J., Liu, H., and Wasserman, L. (2009). Sparse additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):1009–1030.

Rosset, S. and Zhu, J. (2007). Piecewise linear regularized solution paths. *The Annals of Statistics,*, 35:1012–1030.

Scheffe, H. (1943). On solutions of the Behrans-Fisher problem based on the t-distribution. *The Annals of Mathematics Statistics*, 14(1):35–44.

Srivastava, M. (2007). Multivariate theory for analyzing high dimensional data. *Journal of the Japan Statistical Society*, 37(1):53–86.

Srivastava, M. (2009). A test of the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis*, 100(3):518–532.

Srivastava, M. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis*, 99(3):386–402.

Srivastava, M. S., Katayama, S., and Kano, Y. (2013). A two sample test in high dimensional data. *Journal of Multivariate Analysis*, 114:349–358.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550.

Thulin, M. (2014). A high-dimensional two-sample test for the mean using random subspaces. *Computational Statistics & Data Analysis*, 74:26–38.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267–288.

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in medicine*, 16(4):385–395.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42:1166–1202.

Wang, H., Li, R., and Tsai, C. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94(3):553–568.

Wang, L., Kim, Y., and Li, R. (2013). Calibrating nonconvex penalized regression in ultra-high dimension. *The Annals of Statistics*, 41(5):2505–2536.

Wang, L., Peng, B., and Li, R. (2014a). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*. To appear.

Wang, L., Sherwood, B., and Li, R. (2014b). Discussion of "estimation and accuracy after model selection" by Brad Efron. *Journal of the American Statistical Association*, 109(57):1007–1010.

Wasserman, L. and Roeder, K. (2009). High dimensional variable selection. *The Annals of statistics*, 37:2178–2201.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zhang, C. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942.

Zhang, C. and Zhang, S. (2013). Confidence intervals for low-dimensional parameters in high-dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. Early view DOI: 10.1111/rssb.12026.

Zhang, C. H. and Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, 38:1567–1594.

Zhang, Y., Li, R., and Tsai, C. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, 105(489):312–323.

Zhu, L., Li, L., Li, R., and Zhu, L. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, 106:1464–1475.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *The Annals of Statistics*, 36(4):1509–1533.

Vita

Yuan Huang

326 Thomas Building, University Park, PA 16802
Email: yuanhuang.stat@psu.edu

Education

Ph.D. in Statistics, Penn State University (expected 2015)

M.S. in Statistics, Renmin University of China (2010)

B.S. in Statistics, Renmin University of China (2008)

Publications

Li, R., Huang, Y., Wang, L., and Xu, C. (2015). Projection Test for High-Dimensional Mean Vectors with Optimal Direction. Submitted to JASA.

Huang, Y., Zhang, Q., Zhang, S., Huang, J., and Ma, S. (2015). Promoting Similarity of Sparsity Structures in Integrative Analysis. Submitted to JASA.

Huang, Y., Liu, J., Yi, H., and Ma, S. (2014). Promoting Similarity of Model Sparsity Structures in Integrative Analysis of Cancer Genetic Data. Submitted to Statistics in Medicine.

Percival, C. J., Huang, Y., Jabs, E. W., Li, R., and Richtsmeier, J. T. (2014). Embryonic craniofacial bone volume and bone mineral density in Fgfr2 +/P 253R and nonmutant mice. Developmental Dynamics, 243(4):541551.

Ma, S., Zhang, Y., Huang, J., Huang, Y., Lan, Q., Rothman, N., and Zhang, T. (2012). Integrative analysis of cancer prognosis data with multiple subtypes using regularized gradient descent. Genetic Epidemiology, 36(8):829-38.

Huang, Y., Huang, J., Shia, BC., and Ma, S. (2012). Identification of cancer genomic markers via integrative sparse boosting. Biostatistics, 13(3): 509-522.

Ma, S., Huang, Y., Huang, J., and Fang, K. (2012). Gene network-based cancer prognosis analysis with sparse boosting. Genetic Research, 94(4): 205-221.

Honors

2010 Pennsylvania State University Distinguished Graduate Fellowship

2008 Distinguished Graduate Student of Beijing, China

2006 Award for Outstanding Leadership, Renmin University of China