Two-Sample Testing of High-Dimensional Covariance Matrices

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Nan Sun
December 2021

Examining Committee Members:

Dr. Cheng Yong Tang, Advisory Chair, Department of Statistical Science
Dr. Yuexiao Dong, Department of Statistical Science
Dr. Xu Han, Department of Statistical Science
Dr. Cristian E. Gutiérrez, External Reader, Department of Mathematics, Temple University

ABSTRACT

Testing equality between two high dimensional covariance matrices is challenging. As the most efficient way to measure evidential discrepancies in observed data, the likelihood ratio test is expected to be powerful when the null hypothesis is violated. However, when the data dimensionality becomes large and potentially exceeds the sample size by a substantial margin, likelihood ratio based approaches face practical and theoretical challenges. To solve this problem, this study proposes a method by which we first randomly project the original high dimensional data into lower dimensional space, and then apply the corrected likelihood ratio tests developed with random matrix theory. We show that testing with a single random projection is consistent under the null hypothesis. Through evaluating the power function, which is challenging in this context, we provide evidence that the test with a single random projection based on a random projection matrix with reasonable column sizes is more powerful when the two covariance matrices are unequal but component-wise discrepancy could be small - a weak and dense signal setting. To more efficiently utilize data information, we propose combined tests from multiple random projections from the class of meta-analyses. We establish the foundation of the combined tests from our theoretical analysis that the p-values from multiple random projections are asymptotically independent in the high-dimensional covariance matrices testing problem. Then, we show that combined tests from multiple random projections are consistent under the null hypothesis. In addition, our theory presents the merit of certain meta-analysis approaches over testing with a single random projection. Numerical evaluation of the power function of the combined tests from multiple random projections is also provided based on numerical evaluation of power function of testing with a single random projection. Extensive simulations and two real genetic data analyses confirm the merits and potential applications of our test.

ACKNOWLEDGEMENTS

This research includes calculations carried out on High Performance Computing resources supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Comparing dependence among components of random vectors under two different treatments has strong motivation. Copy number measurement (Sebat et al., 2007) and gene expression measurement (Khan et al., 2001) are examples of such genetic studies. In statistics, interest in this topic relies on testing the equality between two covariance matrices. In classical multivariate analysis, such a problem is typically solved using the golden rule – a likelihood ratio approach – by assuming that the two-sample data independently follow some multivariate normal distributions; see the monograph in Anderson (1962). It is well recognized that the likelihood ratio approach is powerful in the context of the conventional setting when the number of the observations goes to infinity and the dimensionality of the random vectors is fixed. However, when data dimensionality diverges, both the practical applicability and statistical properties of the conventional likelihood ratio approach faces problems.

Recently, methods have been developed to solve the testing problem using high dimensional data; we refer to Hu & Bai (2016) and T. T. Cai (2017) as overviews on this topic. In these attempts to address the challenges from large dimensional covariance matrices in testing problems, random matrix theory (Bai et al., 2009) has been influential. In the context of likelihood ratio approaches, Bai et al. (2009) proposed corrections to the likelihood ratio statistic and established the limiting distribution of the corrected test statistic. Along this line, Zheng et al. (2015)

established the central limit theorem for linear spectral statistics with high-dimensional covariance matrices; see also Zheng et al. (2017).

In addition to likelihood ratio based approaches, other discrepancy measures are also utilized in testing statistical hypotheses for high-dimensional covariance matrices. Among them, the maximum element-wise discrepancy is a distinguished candidate. T. Cai et al. (2013) considered testing for sparse covariance matrices with the maximum element-wise discrepancy and established the limiting distribution of the test statistic. Accompanying this class of testing methods, multiplier bootstrap approaches supported by the high-dimensional Gaussian approximation (Chernozhukov et al., 2013) have demonstrated their usefulness when approximating the distribution of the maximum discrepancy measures. Chang et al. (2017) tested the equivalence of two high-dimensional covariance matrices with the Gaussian approximation method and theory; Chang et al. (2018) constructed a class of simultaneous confidence regions for a subset of the entries of a large precision matrix based on multiplier bootstrap procedure. Intuitively, the maximum discrepancy is suitable for detecting sparse and relatively stronger signals when the null hypothesis is violated, but other discrepancy measures by designated aggregations have also been investigated in the literature for testing problems; see, among others, Chen et al. (2010), Qiu et al. (2012), Li et al. (2012), Jiang et al. (2012), T. T. Cai et al. (2013), and He & Chen (2016).

Despite the merits of using corrected likelihood ratio testing approaches supported by the random matrix theory, there remain practical limitations and challenges when solving high-dimensional testing problems. In particular, the likelihood ratio test becomes unapplicable, theoretically and practically, when data dimensionality exceeds the sample size; see Zhang et al. (2019) for more on this recent development. This constitutes a major challenge for addressing high-dimensional testing problems. As for the approaches based on the maximum element-wise discrepancy, on the other hand, a natural concern is that they may be less powerful in case many but weaker element-wise violations occur when the null hypothesis is violated.

This dissertation develops new testing procedures based on the corrected likelihood ratio approach to accommodate high-dimensional testing problems. The foundational building block of this new method is the random matrix theory supported corrected likelihood ratio tests developed by, for example, Bai et al. (2009) and Zheng et al. (2015). With the likelihood ratio based discrepancy measure, our procedure captures the evidence from violating the null hypothesis most efficiently. To address challenges introduced from testing with high-dimensional random vectors, our method proposes to reduce the dimensionality of the problem with random projections – projecting the high-dimensional vectors onto lower-dimensional space. Practically, random projections form a class of computationally efficient dimension reduction approaches, and they have appealing properties when maintaining the geometric properties of the high-dimensional data; see Bingham & Mannila (2001) and Vempala (2005).

The random projection technique has been applied in the context of testing high-dimensional mean vectors: Lopes et al. (2011) proposed a two-sample mean test based on a single random projection. R. Srivastava et al. (2016) pointed out that a single randomly projected Hotelling's test could lead to different conclusions when applying different projections. To address this issue, R. Srivastava et al. (2016) averaged the p-values of tests from multiple random projections. In the context of testing high-dimensional covariance matrices, Wu & Li (2020) proposed randomly projecting the vectors to a one-dimensional subspace and then applying the conventional method for testing the equality of the two variances. They also pointed out the same aforementioned issue of using a single projection; as a remedy, they suggested the minimum p-value derived from multiple random projections.

Concretely, the approach offered in this dissertation applies random projections to high-dimensional random vectors, and then applies the corrected likelihood ratio approach to test the equality of the covariance matrices. We demonstrate that under the null hypothesis, the proposed test with projected high-dimensional random vectors follows the standard normal distribution with appropriate centering and normalizing. Simulations provide evidence that our approach is indeed more powerful for testing high-dimensional covariance matrices than maximum discrepancy based approaches, especially when component-wise discrepancy is small.

Based on testing with a single random projection, this dissertation also offers a combined test from multiple random projections. We consider the problem in the setting of meta-analysis using the unified framework of Xie et al. (2011) in the context of the confidence distributions. Meta-analysis has recently gained renewed interests in practical problems from genomics and functional magnetic resonance imaging; see Hedges & Olkin (2014) for an overview. Meta-analysis attempts to combine multiple p-values to utilize available information as much as possible; see Marden et al. (1991) for the four most widely-used methods: the Fisher's method, the Stouffer's method, the minimum method and the maximum method.

We provide a new methodological contribution through this investigation: Our problem connects to but differs from the conventional meta-analysis approaches for combining multiple tests. Specifically, we consider a designated problem that is developed from a huge original problem that is not tractable. Then, we attempt to scale down its size to be manageable by applying random projection multiple times. That is, our approach is constructive in building the feasible multiple tests, whereas the meta-analysis setting typically starts from multiple conclusions without access to the original data sets.

The aforementioned methodological feature introduces new technical challenges that differ from those in meta-analysis. Foremost, the dependence structure between the multiple tests from the same data set requires dedicated attention. Furthermore, such an issue needs to be considered in a challenging setting when testing high dimensional covariance matrices of very large size. In our analysis, high dimensionality of the covariance matrices plays a central role. Our technical analysis of multiple random projection based tests shows that independence between those tests are achievable upon appropriately designing the dimension of the random projections and the number of replications. As a result, we establish the consistency of the meta-analysis approaches in combining those multiple p-values.

This dissertation is organized as follows. Chapter 2 outlines potential applications that motivate this study. Chapter 3 provides brief methodological background on developments in the likelihood ratio approach to tackling two-sample covariance matrices testing when for low or moderate dimensions. Chapter 4 introduces the proposed testing procedure with a single random projection and a combined test from multiple random projections. In Chapter 5, we establish the consistency of both testing with a single random projection and combined tests from multiple projections; theories and numerical approximations of power function of both testing with a single random projection and combined tests from multiple random projections are also provided. Illustrations using simulated two-sample random vectors are presented in Chapter 6, and the proposed method is applied to real genetic data in Chapter 7.

CHAPTER 2

MOTIVATING STUDY

Copy number measurements are produced by testing DNA copy number variation in patient's sample. Analyzing copy number measurements helps identify chromosomal copy number variation that may increase risks of some disease, see Sebat et al. (2007). Gene expression measurements are quantification of level at which a particular gene is expressed within a cell, tissue or organism. In particular, oncogene expression provides valuable information in determining an individual's susceptibility to cancer, see Khan et al. (2001). Genetic data of aforementioned kinds are high-dimensional, but with relatively small sample sizes, due to limited amount of study objects.

In a genetic study, it is often important to check if two treatments in a study share the same distribution, or at least, have the same mean or covariance. Due to the high-dimensional nature, it is more tractable and interpretable to test the equality of mean or covariance of genetic data from two treatments, instead of testing if two high-dimensional distributions are equal.

Our study focuses on comparing dependence among components of random vectors under two different treatments. In copy number analysis, the dependence among the copy number measurements may be different for patients under different biological conditions. In genetic expression analysis, some genes may be more correlated in patients with disease A and may be decoupled in patients with disease B.

The characteristic for depicting such dependence is covariance matrix, and it is of great importance to consider testing the equality of covariance matrices constructed from components of random vectors under two different treatments.

CHAPTER 3

METHODOLOGICAL BACKGROUND, $p < n_i, i = 1, 2$

To motivate the proposed approach to comparing dependence among components of random vectors under two different treatments, we first provide a brief review of the likelihood ratio test used in a low-dimensional case. That is, when the amount of covariates p is fixed, and $n_i, i = 1, 2$ diverges.

3.1 Likelihood Ratio Test

Let $\boldsymbol{X} = (X_1, \cdots, X_p)^T$ and $\boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$ be two p-dimensional random vectors with means $\boldsymbol{\mu}_1 = (\mu_{11}, \cdots, \mu_{1p})^T$ and $\boldsymbol{\mu}_2 = (\mu_{21}, \cdots, \mu_{2p})^T$, and covariance matrices $\boldsymbol{\Sigma}_1 = (\sigma_{1,kl})_{1 \le k,l \le p}$ and $\boldsymbol{\Sigma}_2 = (\sigma_{2,kl})_{1 \le k,l \le p}$, respectively. We are interested in testing the equality of the two covariance matrices:

$$\mathrm{H}_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \text{ vs } \mathrm{H}_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2. \tag{3.1-1}$$

For testing (3.1-1), we observe two random samples, denoted by $\{\boldsymbol{X}_i\}_{i=1}^{n_1}$ and $\{\boldsymbol{Y}_i\}_{i=1}^{n_2}$ that are mutually independent and identically follow the respective distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

Working with the multivariate normal distribution, the likelihood ratio test applies. In particular, with regular cases, the test statistic is:

$$\lambda = \frac{|\boldsymbol{A}|^{n_1/2}\,|\boldsymbol{B}|^{n_2/2}}{\left|\frac{n_1}{n_1+n_2}\boldsymbol{A} + \frac{n_2}{n_1+n_2}\boldsymbol{B}\right|^{(n_1+n_2)/2}}, \tag{3.1-2}$$

where

$$\boldsymbol{A} = \frac{1}{n_1}\sum_{i=1}^{n_1}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)^T,$$

and

$$\boldsymbol{B} = \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)^T,$$

where $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ are the respective sample means, and $|\cdot|$ is the determinant of a matrix. Standard theory says that when p is fixed, as $n_1, n_2 \to \infty$, $-2\log\lambda$ converges to $\chi^2_{p(p+1)/2}$ distribution asymptotically under H_0; see Anderson (1962).

However, the likelihood ratio test is biased when the sample sizes are not equal; that is, the probability of rejecting H_0 when H_0 is false can be smaller than the probability of rejecting H_0 when H_0 is true, see Muirhead (2009). Instead, the modified likelihood ratio test proposed by Bartlett (1937) is unbiased, see Muirhead (2009): The modified likelihood ratio statistic λ^* is defined by replacing n_i, $i = 1,\ 2$, in (3.1-2) with $N_i = n_i - 1$. That is,

$$\lambda^* = \frac{|\boldsymbol{S}_1|^{N_1/2}\,|\boldsymbol{S}_2|^{N_2/2}}{|c_1\boldsymbol{S}_1 + c_2\boldsymbol{S}_2|^{N/2}}, \tag{3.1-3}$$

where

$$\boldsymbol{S}_1 = \frac{1}{N_1}\sum_{i=1}^{n_1}\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)^T,$$

and

$$\boldsymbol{S}_2 = \frac{1}{N_2}\sum_{i=1}^{n_2}\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)^T,$$

where $N = N_1 + N_2$, $c_i = N_i/N$, $i = 1,\ 2$. When p is fixed, as $n_1, n_2 \to \infty$, $-2\log\lambda^*$ also converges to $\chi^2_{p(p+1)/2}$ distribution asymptotically under H_0, see Rao et al. (1973).

3.2 Corrected Likelihood Ratio Test

Clearly, the chi-square limiting distribution results from asymptotic quadratic form. When p is also diverging with n_1, n_2, the chi-square limiting distribution may poorly approximate the test statistic; see Bai et al. (2009). Further, the likelihood ratio statistic (3.1-2) and the modified likelihood ratio statistic (3.1-3) fail to define when p is larger than n_i, $i = 1,\ 2$, because the determinants therein are seen as zeros.

Further adjustments have thus been developed for large sample covariance matrices, supported by random matrix theory (Bai et al., 2009). A prominent result for the properties of the corrected likelihood ratio statistic from Zheng et al. (2015) is presented as follows. Instead of the chi-square distribution, a standard normal distribution for the modified likelihood ratio statistic is established under the null hypothesis, with appropriate centering and normalization. Zheng et al. (2015) derived central limit theorems for linear spectral statistics from the model $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{u}$, where $\boldsymbol{\mu}$ is unknown and $\boldsymbol{u}$ consists of independent and identically distributed random variables. For general $\boldsymbol{u}$, the central limit theorems involve unknown parameters such as the fourth moment. For ease of analysis and presentation without compromising the spirit of our approach, we consider normal distributions: $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Assumption 3.1. $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, *and* $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$.

Assumption 3.2. *The ratio of dimension-to-sample size* $\hat{y}_1 = p/N_1 \to y_1 > 0$ *as* $n_1, p \to \infty$, *and* $\hat{y}_2 = p/N_2 \to y_2 \in (0, 1)$ *as* $n_2, p \to \infty$.

Assumption 3.3. *The sequences* $\left\{\boldsymbol{\Sigma}_1 = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^T\right\}_{p\geq 1}$ *and* $\left\{\boldsymbol{\Sigma}_2 = \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2^T\right\}_{p\geq 1}$ *are bounded in spectral norm, and empirical spectral distributions of* $H_{1,p}(t) = \sum_{j=1}^{p} I\left(\gamma_{j,1} \leq t\right)/p$,

$H_{2,p}(t) = \sum_{j=1}^{p} I\left(\gamma_{j,2} \leq t\right)/p$ *converge weakly to limiting spectral distributions* $H_1(t)$ *and* $H_2(t)$ *respectively, where* $\{\gamma_{j,1}\}_{j=1}^{p}$ *and* $\{\gamma_{j,2}\}_{j=1}^{p}$ *are eigenvalues of* $\mathbf{\Sigma}_1$ *and* $\mathbf{\Sigma}_2$ *respectively, and* $I(\cdot)$ *is an indicator function.*

Proposition 3.4. *Assume that* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *satisfy assumptions 3.1-3.2-3.3. Then under the null hypothesis* $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$,

$$v^{-1/2}\left(-\frac{2\log\lambda^*}{N_1+N_2} - pF - m\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\begin{aligned} v =& v(\hat{y}_1, \hat{y}_2, y_1, y_2) \\ =& -\frac{2\hat{y}_2^2}{\left(\hat{y}_1+\hat{y}_2\right)^2}\log\left(1-\hat{y}_1\right) \\ & -\frac{2\hat{y}_1^2}{\left(\hat{y}_1+\hat{y}_2\right)^2}\log\left(1-\hat{y}_2\right) \\ & +2\log\left(\frac{y_1+y_2-y_1y_2}{\hat{y}_1+\hat{y}_2}\right), \end{aligned}$$

$$\begin{aligned} m =& m(\hat{y}_1, \hat{y}_2, y_1, y_2) \\ =& \frac{1}{2}\log\left(\frac{y_1+y_2-y_1y_2}{\hat{y}_1+\hat{y}_2}\right) \\ & -\frac{1}{2}\frac{\hat{y}_1}{\hat{y}_1+\hat{y}_2}\log\left(1-\hat{y}_2\right) \\ & -\frac{1}{2}\frac{\hat{y}_2}{\hat{y}_1+\hat{y}_2}\log\left(1-\hat{y}_1\right), \end{aligned}$$

$$
\begin{aligned}
F =& F(\hat{y}_1, \hat{y}_2, y_1, y_2) \\
=& \frac{y_1 + y_2 - y_1 y_2}{\hat{y}_1 \hat{y}_2} \log\left(\frac{\hat{y}_1 + \hat{y}_2}{y_1 + y_2 - y_1 y_2}\right) \\
& + \frac{\hat{y}_1 \left(1 - \hat{y}_2\right)}{\hat{y}_2 \left(\hat{y}_1 + \hat{y}_2\right)} \log\left(1 - \hat{y}_2\right) \\
& + \frac{\hat{y}_2 \left(1 - \hat{y}_1\right)}{\hat{y}_1 \left(\hat{y}_1 + \hat{y}_2\right)} \log\left(1 - \hat{y}_1\right).
\end{aligned}
$$

CHAPTER 4

PROPOSED METHODOLOGY, $p > n_i, i = 1, 2$

In the previous chapter, relevant likelihood ratio statistics are provided relating to two-sample covariance matrices testing. When $p < n_i, i = 1, 2$, corrected likelihood ratio statistic developed with the random matrix theory has been shown to be powerful. However, the requirement of $p < n_1$ or $p < n_2$ must be satisfied to make the corrected likelihood ratio statistic applicable. Therefore, to utilize the powerful corrected likelihood ratio statistic when $p > n_i, i = 1, 2$, a new testing procedure is needed that is applicable in a high-dimensional setting. We now propose a testing procedure to compare dependence among components of random vectors from two treatments. To address the challenges from testing with high-dimensional random vectors, we propose to reduce the dimensionality of the problem with random projections – projecting the high-dimensional vectors to some lower-dimensional space. Concretely, our approach first applies random projections to high-dimensional random vectors, and then applies the corrected likelihood ratio approach for testing the equality of two covariance matrices. Then, we propose to combine the tests from multiple random projections to apply to the class of approaches of meta-analysis.

4.1 Testing with a Single Random Projection

As seen from (3.1-3) and Assumption 3.2 in Proposition 3.4, though dimensionality of the random vector is allowed to diverge, it is required to be smaller than the sample size. To address this problem, we propose to project the p-dimensional random vectors to a k-dimensional space. Let $\boldsymbol{R} \in \mathbb{R}^{p\times k}$, $k \in \{1, \cdots, \min(n_1, n_2, p)\}$, be a linear transformation matrix with orthogonal columns such that $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_k$, where $\boldsymbol{I}_k$ is the identity matrix of size k. Orthogonality is not a strong requirement, because the Gram-Schmidt process can ensure that otherwise. Clearly, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ implies $\boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R} = \boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R}$. Hence, we consider testing

$$\mathrm{H}_{0,\mathrm{proj}} : \boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R} = \boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R} \text{ vs } \mathrm{H}_{1,\mathrm{proj}} : \boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R} \neq \boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R}. \qquad (4.1\text{-}1)$$

This test (4.1-1) serves the same purpose as the original test (3.1-1) by validating that it is a necessary condition.

We propose to apply random projection in generating $\boldsymbol{R}$. That is, to randomly generate the entries of $\boldsymbol{R}$ from some distributions, for example, the standard normal distribution, and other sparse random matrices; see, for example, Achlioptas (2001). For high dimensional problems, though randomly generated $\boldsymbol{R}$ may not have exactly orthogonal columns, $\boldsymbol{R}^T\boldsymbol{R}$ is actually close to $\boldsymbol{I}_k$; see Bingham & Mannila (2001). In our implementation, we generate entries of $\boldsymbol{R}$ independently from standard normal distribution, and then apply the Gram-Schmidt process.

We propose to apply the corrected likelihood ratio test of Zheng et al. (2015) on the projected random vectors. In particular, we investigate the modified likelihood ratio statistic with projected data using $\boldsymbol{R}$:

$$\lambda^{(\boldsymbol{R})} = \frac{\left|\boldsymbol{R}^T\boldsymbol{S}_1\boldsymbol{R}\right|^{N_1/2}\left|\boldsymbol{R}^T\boldsymbol{S}_2\boldsymbol{R}\right|^{N_2/2}}{\left|c_1\boldsymbol{R}^T\boldsymbol{S}_1\boldsymbol{R} + c_2\boldsymbol{R}^T\boldsymbol{S}_2\boldsymbol{R}\right|^{N/2}}. \tag{4.1-2}$$

Here we note that $\boldsymbol{R}^T\boldsymbol{S}_i\boldsymbol{R}$, $i = 1, 2$, are invertible with probability 1, as implied from Lemma 1 in R. Srivastava et al. (2016), provided that $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_k$ and $\boldsymbol{\Sigma}_i$, $i = 1, 2$, are positive definite. Based on (4.1-2) and following Zheng et al. (2015), we define a corrected likelihood ratio statistic as

$$Z^{(\boldsymbol{R})} = v^{*-1/2}\left(-\frac{2\log\lambda^{(\boldsymbol{R})}}{N_1 + N_2} - kF^* - m^*\right), \tag{4.1-3}$$

where $\hat{y}_i^* = k/N_i$, and y_i^* are limits of $\hat{y}_i^*$ $(i = 1, 2)$,

$$v^* = v\left(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*\right),$$

$$m^* = m\left(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*\right),$$

$$F^* = F(\hat{y}_1^*, \hat{y}_2^*, y_1^*, y_2^*).$$

The standard normal distribution of $Z^{(\boldsymbol{R})}$ is guaranteed by Theorem 5.3 in Section 5.1.1 of Chapter 5.

Concretely, we summarize the procedure with a single random projection as follows.

1. Randomly generate matrix $\boldsymbol{R} \in \mathbb{R}^{p\times k}$ with entries from the standard normal distribution; upon applying the Gram-Schmidt process, $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_k$;

2. Compute the modified likelihood ratio statistic $\lambda^{(\boldsymbol{R})}$ by (4.1-2);

3. Compute the corrected likelihood ratio statistic $Z^{(\boldsymbol{R})}$ by (4.1-3), with $\hat{y}_i^*$ replaced by y_i^*, $i = 1, 2$;

4. For Bartlett's modified likelihood ratio test, reject H_0 if $\lambda^{(\boldsymbol{R})}$ exceeds the $(1-\alpha)$-level quantile of $\chi^2_{k(k+1)/2}$ distribution; for the corrected likelihood ratio test, reject H_0 if $\left|Z^{(\boldsymbol{R})}\right|$ exceeds the $(1-\alpha/2)$-level quantile of the standard normal distribution.

4.2 Tests with Multiple Random Projections

A single randomly projected corrected likelihood ratio test could lead to different conclusions in the high-dimensional two-sample covariance matrices testing if one applies different projections. To remedy this issue, we propose to apply multiple random projections denoted by $\boldsymbol{R}_j$, $j = 1, \cdots, r$, and then combine the resulting p-values.

Clearly, multiple factors contribute to the joint distribution of the tests with multiple random projections – namely, the data dimensionality p, the projected dimension k, and the number of random projections r. Those components are quite involved in determining the dependence structure of the test statistics, which is the key in developing an effective combination of them. As a foundation of our approach, we first establish a result on the asymptotic independence between the p-values of the tests, in the context of testing high-dimensional covariance matrices.

The establishment of asymptotic independence between p-values relies on a result in Guo (2020). With $\boldsymbol{R}_j$, $j = 1, \cdots, r$, and under the null hypothesis $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, Guo (2020) considers a working sequence of $\boldsymbol{R}_j^*$ by

$$\boldsymbol{R}_1^* = \boldsymbol{R}_1,$$

and

$$\boldsymbol{R}_j^* = \boldsymbol{R}_j - \sum_{i=1}^{j-1} \boldsymbol{R}_i^* \left(\boldsymbol{R}_i^{*T} \boldsymbol{\Sigma} \boldsymbol{R}_i^*\right)^{-1} \boldsymbol{R}_i^{*T} \boldsymbol{\Sigma} \boldsymbol{R}_j,\ j = 2, \cdots, r.$$

As a result, $\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}_j^*$, $j = 1, \cdots, r$, are orthogonal. That is, $(\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}_i^*)^T \boldsymbol{\Sigma}^{1/2}\boldsymbol{R}_j^* = \boldsymbol{R}_i^{*T}\boldsymbol{\Sigma}\boldsymbol{R}_j^* = \boldsymbol{0}_k$ for $i \neq j$. Therefore, for multivariate normal data, $\left\{\boldsymbol{R}_j^{*T}\boldsymbol{S}_1\boldsymbol{R}_j^*\right\}_{j=1}^r$ are mutually independent, and so are $\left\{\boldsymbol{R}_j^{*T}\boldsymbol{S}_2\boldsymbol{R}_j^*\right\}_{j=1}^r$.

Guo (2020) then defines the corresponding working modified likelihood ratio statistic as

$$\lambda^{\left(\boldsymbol{R}_j^*\right)} = \frac{\left|\boldsymbol{R}_j^{*T}\boldsymbol{S}_1\boldsymbol{R}_j^*\right|^{N_1/2} \left|\boldsymbol{R}_j^{*T}\boldsymbol{S}_2\boldsymbol{R}_j^*\right|^{N_2/2}}{\left|c_1\boldsymbol{R}_j^{*T}\boldsymbol{S}_1\boldsymbol{R}_j^* + c_2\boldsymbol{R}_j^{*T}\boldsymbol{S}_2\boldsymbol{R}_j^*\right|^{N/2}}.$$

Hence, $\lambda^{\left(\boldsymbol{R}_j^*\right)}$, $j = 1, \cdots, r$, are mutually independent according to the definition of $\boldsymbol{R}_j^*$.

Despite that the working sequence $\left\{\boldsymbol{R}_j^*\right\}_{j=1}^r$ is not realistic in practice because $\boldsymbol{\Sigma}$ is unknown, based on theoretical result of Guo (2020) that the difference between $\log\left\{\lambda^{\left(\boldsymbol{R}_j^*\right)}\right\}/N$ and $\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N$ is negligible asymptotically, then $\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N$, $j = 1, \cdots, r$, themselves can be shown to be asymptotically independent. Such a property is ensured by the theorem in Guo (2020), and we summarize it as the following proposition.

Proposition 4.1. *If* $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, $0 < c \leq \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) \leq C < \infty$, *where* $\lambda_{\min}(\boldsymbol{\Sigma})$ *and* $\lambda_{\max}(\boldsymbol{\Sigma})$ *denote the smallest and the largest eigenvalues of* $\boldsymbol{\Sigma}$. *If*

Assumption 5.1 is satisfied, then there exists $r = O(N^{c_0})$ for some $c_0 > 0$, such that

$$\max_{1 \le j \le r} \left| \log \left\{ \lambda^{(\boldsymbol{R}_j)} \right\} / N - \log \left\{ \lambda^{\left(\boldsymbol{R}_j^*\right)} \right\} / N \right| = O_{\mathbb{P}} \left(r^{1/2} k^{3/2} p^{-1/2} \right).$$

According to (4.1-3), we rewrite $Z^{(\boldsymbol{R}_j)} = C_1 \log \left\{ \lambda^{(\boldsymbol{R}_j)} \right\} / N + C_2$, where both C_1 and C_2 are functions of $\hat{y}_i^*$ and its limit y_i^*, $i = 1,\ 2$. Both C_1 and C_2 tend to be constants as $k, N_1, N_2 \to \infty$. In particular,

$$C_1 = -2 v^{* -1/2},$$

$$C_2 = -v^{* -1/2} \left(k F^* + m^* \right).$$

Based on $\left\{ \lambda^{\left(\boldsymbol{R}_j^*\right)} \right\}_{j=1}^{r}$, we define the working sequence of

$$Z^{\left(\boldsymbol{R}_j^*\right)} = C_1 \log \left\{ \lambda^{\left(\boldsymbol{R}_j^*\right)} \right\} / N + C_2,$$

and the resulting working sequence of p-values $p_j^* = 2\Phi \left\{ - \left| Z^{\left(\boldsymbol{R}_j^*\right)} \right| \right\}$, $j = 1, \cdots, r$, which are mutually independent.

Define $p_j = 2\Phi \left\{ - \left| Z^{(\boldsymbol{R}_j)} \right| \right\}$, $j = 1, \ldots, r$, and as an implication from Proposition 4.1, the condition under which $\{p_j\}_{j=1}^r$ are asymptotically independent is established in the following theorem.

Theorem 4.2. *Assuming conditions of Proposition 4.1,*

$$\max_{1 \le j \le r} \left| p_j - p_j^* \right| = O_{\mathbb{P}} \left(r^{1/2} k^{3/2} p^{-1/2} \right).$$

The proof of Theorem 4.2 is given in Appendix A.

Theorem 4.2 implies that the p-values $\left[p_j = 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\}\right]_{j=1}^{r}$ are asymptotically mutually independent with probability tending to one if $r^{1/2}k^{3/2}p^{-1/2} \to 0$. The high dimensionality $p \to \infty$ is a key ensuring the independence, which is exactly the context of the testing problem we are considering. $r^{1/2}k^{3/2}p^{-1/2} \to 0$ indicates some condition on r, implying that one cannot arbitrarily apply many random projections. If r is too large say $rk = 2p$, then we can construct a set of tests $\left[p_j = 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\}\right]_{j=1}^{r}$ such that the independence does not hold.

4.3 Combining Tests from Multiple Random Projections

We now propose a new approach combining the covariance tests from multiple random projections by considering the problem in a setting of meta-analysis. Our presentation uses the unified framework of Xie et al. (2011) in the context of the confidence distributions and is based on Theorem 4.2.

The ingredients we are working with are those p-values, $p_j = 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\}$, $j = 1, \ldots, r$. $p_j = 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\}$ has the standard uniform distribution when H_0 is true given $\boldsymbol{R}_j$, see Singh et al. (2007).

Let $\boldsymbol{\Theta} = \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2$, then the null hypothesis is $\mathrm{H}_0 : \boldsymbol{\Theta} = \mathbf{0}$ and the alternative hypothesis is $\mathrm{H}_1 : \boldsymbol{\Theta} \neq \mathbf{0}$. According to Singh et al. (2007), we define the following confidence distribution function for $\boldsymbol{\Theta}$ as

$$c_j(\boldsymbol{s}) = c_j\left\{Z^{(\boldsymbol{R}_j)}, \boldsymbol{s}\right\} = 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\} \mid \boldsymbol{\Theta} = \boldsymbol{s}, j = 1, \cdots, r.$$

$c_j(\boldsymbol{s})$ is a continuous cumulative distribution function given $Z^{(\boldsymbol{R}_j)}$ and $c_j(\mathbf{0})$ has the standard uniform distribution as a function of $Z^{(\boldsymbol{R}_j)}$, which means $c_j(\boldsymbol{s})$ is a well-defined confidence distribution function.

Let $g_c(u_1, \cdots, u_r)$ be a continuous function on $[0,1]^r \to \mathbb{R}$ which is increasing in each coordinate. Singh et al. (2005) suggested to combine r independent confidence distribution functions by

$$c^c(\boldsymbol{s}) = G_c\left[g_c\left\{c_1(\boldsymbol{s}), \cdots, c_r(\boldsymbol{s})\right\}\right], \tag{4.3-4}$$

where $G_c(t) = \mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq t\right\}$ with $U_1, \cdots, U_r$ are independent $\mathrm{U}[0,1]$ random variables.

$c^c(\boldsymbol{s})$ in (4.3-4) is also a confidence distribution function since $G_c(\cdot)$ is defined to be a continuous cumulative distribution function given $\left\{Z^{(\boldsymbol{R}_j)}\right\}_{j=1}^r$; and

$$c^c(\mathbf{0}) = \mathbb{P}\left[g_c(U_1, \cdots, U_r) \leq g_c\left\{c_1(\mathbf{0}), \cdots, c_r(\mathbf{0})\right\}\right]$$

has the standard uniform distribution as a function of $\left\{Z^{(\boldsymbol{R}_j)}\right\}_{j=1}^r$ if $\left\{c_j(\mathbf{0})\right\}_{j=1}^r$ are independent standard uniform random variables, which can be fulfilled with the help of Theorem 4.2.

We consider the equal-weight combination as in Xie et al. (2011) with

$$g_c(u_1, \cdots, u_r) = F_0^{-1}(u_1) + \cdots + F_0^{-1}(u_r), \tag{4.3-5}$$

where $F_0(\cdot)$ is any continuous cumulative distribution function and $F_0^{-1}(\cdot)$ is its inverse function. In this case, $G_c = F_0 \otimes \cdots \otimes F_0$, where $\otimes$ stands for convolution. Consequently, $c^c(\mathbf{0})$ is the combined p-value.

If one chooses $F_0(t) = \exp(t)$, for $t \leq 0$, then

$$c_{\mathrm{F}}^c(\boldsymbol{s}) = \mathbb{P}\left[-\chi_{2r}^2/2 \leq \sum_{j=1}^r \log\left\{c_j(\boldsymbol{s})\right\}\right],$$

where χ^2_{2r} is a chi-square random variable with degrees of freedom $2r$. $c^c_{\mathrm{F}}(\mathbf{0})$ is the combined p-value for the Fisher's method, denoted by T_{F} hereinafter.

If one chooses $F_0(t) = \Phi(t)$, for $t \in \mathbb{R}$, then

$$c^c_{\mathrm{S}}(\boldsymbol{s}) = \mathbb{P}\left[r^{1/2}Z \leq \sum_{j=1}^{r} \Phi^{-1}\left\{c_j(\boldsymbol{s})\right\}\right],$$

where Z is a standard normal random variable. $c^c_{\mathrm{S}}(\mathbf{0})$ is the combined p-value for the Stouffer's method, denoted by T_{S}.

The g_c choice for the minimum method is $u_{(1)} = \min(u_1, \cdots, u_r)$, then

$$c^c_{\min}(\boldsymbol{s}) = \mathbb{P}\left[U_{(1)} \leq \min_{1\leq j\leq r}\left\{c_j(\boldsymbol{s})\right\}\right],$$

where $U_{(1)}$ is the minimum of r independent standard uniform random variables. $c^c_{\min}(\mathbf{0})$ is the combined p-value for the minimum method, denoted by $T_{\min}$.

The g_c choice for the maximum method is $u_{(r)} = \max(u_1, \cdots, u_r)$, then

$$c^c_{\max}(\boldsymbol{s}) = \mathbb{P}\left[U_{(r)} \leq \max_{1\leq j\leq r}\left\{c_j(\boldsymbol{s})\right\}\right],$$

where $U_{(r)}$ is the maximum of r independent standard uniform random variables. $c^c_{\max}(\mathbf{0})$ is the combined p-value for the maximum method, denoted by $T_{\max}$.

Minimum method and Maximum method are not an equal-weight combination, but they are also considered in our framework of combining p-values.

Concretely, we summarize our procedure as in the following unified algorithm.

1. Input: $\{\boldsymbol{X}_i\}_{i=1}^{n_1}$,$\{\boldsymbol{Y}_i\}_{i=1}^{n_2}$, α, and $\{\boldsymbol{R}_j\}_{j=1}^{r}$, such that $\boldsymbol{R}_j^T\boldsymbol{R}_j = \boldsymbol{I}_k$;

2. $\bar{\boldsymbol{X}} \leftarrow \sum_{i=1}^{n_1}\boldsymbol{X}_i/n_1$, $\bar{\boldsymbol{Y}} \leftarrow \sum_{i=1}^{n_2}\boldsymbol{Y}_i/n_2$;

3. $\boldsymbol{S}_1 \leftarrow \sum_{i=1}^{n_1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T/N_1$, $\boldsymbol{S}_2 \leftarrow \sum_{i=1}^{n_2}(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}})^T/N_2$, where $N_i = n_i - 1$;

4. $\lambda^{(\boldsymbol{R}_j)} \leftarrow \frac{\left|\boldsymbol{R}_j^T\boldsymbol{S}_1\boldsymbol{R}_j\right|^{N_1/2}\left|\boldsymbol{R}_j^T\boldsymbol{S}_2\boldsymbol{R}_j\right|^{N_2/2}}{\left|\boldsymbol{R}_j^T(c_1\boldsymbol{S}_1+c_2\boldsymbol{S}_2)\boldsymbol{R}_j\right|^{N/2}}$, where $c_i = N_i/N,\ N = N_1 + N_2$;

5. $Z^{(\boldsymbol{R}_j)} \leftarrow v^{*-1/2}\left[-2\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N - kF^* - m^*\right]$ according to (4.1-3), with $y_i^* \leftarrow \hat{y}_i^*$, where y_i^* are limits of $\hat{y}_i^*$;

6. $c_j\left(\boldsymbol{0}\right) \leftarrow 2\Phi\left\{-|Z^{(\boldsymbol{R}_j)}|\right\}$;

7. $T_{\mathrm{F}} \leftarrow \mathbb{P}\left[\chi^2_{2r} \geq -2\sum_{j=1}^{r}\log\left\{c_j(\boldsymbol{0})\right\}\right]$, $T_{\mathrm{S}} \leftarrow \mathbb{P}\left[Z \leq r^{-1/2}\sum_{j=1}^{r}\Phi^{-1}\left\{c_j(\boldsymbol{0})\right\}\right]$, $T_{\min} \leftarrow \mathbb{P}\left[U_{(1)} \leq \min_{1\leq j\leq r}\left\{c_j(\boldsymbol{0})\right\}\right]$, and $T_{\max} \leftarrow \mathbb{P}\left[U_{(r)} \leq \max_{1\leq j\leq r}\left\{c_j(\boldsymbol{0})\right\}\right]$;

8. Output: reject H_0 if $T_* < \alpha$, and $* \in \{\mathrm{F}, \mathrm{S}, \min, \max\}$ denotes the Fisher's method, the Stouffer's method, the minimum method, and the maximum method respectively.

The validity of decision rules stated in the aforementioned algorithm is guaranteed in Theorem 5.7 in Section 5.2.1 of Chapter 5 based on Theorem 4.2.

CHAPTER 5

THEORY

In this chapter, we establish the consistency of testing with a single random projection and the consistency of combined tests with multiple random projections under the null. Under the alternative hypothesis, when two high-dimensional covariance matrices are unequal, it is a known challenging problem to depict the power function explicitly. To address this problem, we devised numerical approximation to power functions of interest.

5.1 Theory of Testing with a Single Random Projection

5.1.1 Consistency of Testing with a Single Random Projection

Assumption 5.1. *The ratios of dimension-to-sample size* $\hat{y}_1^* = k/N_1 \to y_1^* > 0$ *as* $n_1, k \to \infty$, *and* $\hat{y}_2^* = k/N_2 \to y_2^* \in (0,1)$ *as* $n_2, k \to \infty$.

Assumption 5.2. *The sequences of* $\left\{\boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R} = \boldsymbol{\Gamma}_1^*\boldsymbol{\Gamma}_1^{*T}\right\}_{k\geq 1}$, $\left\{\boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R} = \boldsymbol{\Gamma}_2^*\boldsymbol{\Gamma}_2^{*T}\right\}_{k\geq 1}$ *are bounded in spectral norm, and empirical spectral distributions of both* $H_{1,k}^*(t) = \sum_{j=1}^k I\left(\gamma_{j,1}^* \leq t\right)/k$ *and* $H_{2,k}^*(t) = \sum_{j=1}^k I\left(\gamma_{j,2}^* \leq t\right)/k$ *converge weakly to limiting spectral distributions* $H_1^*(t)$ *and* $H_2^*(t)$ *respectively, where* $\left\{\gamma_{j,1}^*\right\}_{j=1}^k$ *and* $\left\{\gamma_{j,2}^*\right\}_{j=1}^k$ *are eigenvalues of* $\boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R}$ *and* $\boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R}$ *respectively.*

Theorem 5.3. *Assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy assumptions 3.1-5.1-5.2. Conditioning on $\boldsymbol{R}$ and under the null hypothesis $H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$,*

$$\mathbb{P}\left(Z^{(\boldsymbol{R})} \leq x \mid \boldsymbol{R}\right) \to \Phi(x),$$

where $x \in \mathbb{R}$.

The proof of Theorem 5.3 is given in Appendix B.

Theorem 5.3 shows that the random projection based test is consistent for any random projection matrix. Clearly, the random projection matrix $\boldsymbol{R}$ together with the specific form of the alternative hypothesis jointly determine the power of the test; see our result in Section 5.1.2. We also will show that testing with a single random projection is advantageous when testing against alternatives with weak and dense signals. We do so both approximately using theory and empirically using simulations in Section 6.1 of Chapter 6. See our evaluations of the power functions in Section 5.1.3.

5.1.2 Power of Testing with a Single Random Projection

Clearly, the power of the test using $Z^{(\boldsymbol{R})}$ is determined by its distribution when the null hypothesis is violated. We have the following result for the distribution of $Z^{(\boldsymbol{R})}$ under a more general data generating process.

Assumption 5.4. *Define*

$$T^{(\boldsymbol{R})} = \left(\boldsymbol{R}^T \boldsymbol{\Sigma}_2 \boldsymbol{R}\right)^{-1/2} \left(\boldsymbol{R}^T \boldsymbol{\Sigma}_1 \boldsymbol{R}\right) \left(\boldsymbol{R}^T \boldsymbol{\Sigma}_2 \boldsymbol{R}\right)^{-1/2}.$$

Sequence of $\left\{T^{(\boldsymbol{R})} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^T\right\}_{k\geq 1}$ *is bounded in spectral norm, as well as the empirical spectral distribution of* $H_k(t) = \sum_{j=1}^{k} I\left(\lambda_j \leq t\right)/k$ *converges weakly to limiting spectral distribution* $H(t)$*, where* $\{\lambda_j\}_{j=1}^{k}$ *are eigenvalues of* $T^{(\boldsymbol{R})}$*.*

Theorem 5.5. *Assume that* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *satisfy assumptions 3.1-5.1-5.4. Conditioning on* $\boldsymbol{R}$ *and under the alternative hypothesis* $H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$*,* $v^{*1/2}Z^{(\boldsymbol{R})} + kF^* + m^*$ *converges in distribution to a normal distribution with mean* $kF_1^* + m_1^*$ *and variance* v_1^**. That is,*

$$\begin{aligned}&\mathbb{P}\left(v^{*1/2}Z^{(\boldsymbol{R})} + kF^* + m^* \leq x \mid \boldsymbol{R}\right)\\&\to \Phi\left(\frac{x - kF_1^* - m_1^*}{v_1^*}\right),\end{aligned}$$

where $x \in \mathbb{R}$*.*

Define

$$f(x) = \log\left(\hat{y}_1^* + \hat{y}_2^* x\right) - \frac{\hat{y}_2^*}{\hat{y}_1^* + \hat{y}_2^*}\log x - \log\left(\hat{y}_1^* + \hat{y}_2^*\right),$$

then

$$\begin{aligned}m_1^* =& -\frac{1}{4\pi i}\oint_{\mathcal{C}} f'(z)\log\left\{\frac{y_1^* + y_2^* - y_1^* y_2^*}{y_2^*} - \frac{y_1^*}{y_2^*}\frac{\left(1 - y_2^*\int \frac{m_0(z)}{t+m_0(z)}dH(t)\right)^2}{1 - y_2^*\int \frac{m_0^2(z)}{(t+m_0(z))^2}dH(t)}\right\}dz \\ &- \frac{1}{4\pi i}\oint_{\mathcal{C}} f'(z)\log\left\{1 - y_2^*\int \frac{m_0^2(z)}{\left(t+m_0(z)\right)^2}dH(t)\right\}dz,\end{aligned} \tag{5.1-1}$$

where

$$m_0(z) = z\left\{\frac{y_1^* + y_2^* - y_1^* y_2^*}{y_2^*\left(-1 + y_2^*\int \frac{m_0(z)}{t+m_0(z)}dH(t)\right)} + \frac{y_1^*}{y_2^*}\right\}^{-1}, \tag{5.1-2}$$

and $H(t)$ *is the limiting spectral distribution of* $T^{(\boldsymbol{R})}$*.*

Besides, by Theorem B.10 of Bai & Silverstein (2010), define

$$\underline{m}(z) = \frac{1}{m_0(z)} - y_2^* \int \frac{1}{t + m_0(z)} dH(t), \tag{5.1-3}$$

$$u^*(x) = \frac{1}{\pi y_1^*} \lim_{\epsilon \to 0+} \Im \left\{ \underline{m}(x + \epsilon i) \right\}, \tag{5.1-4}$$

$$\underline{m}(x) = \lim_{\epsilon \to 0+} \underline{m}(x + \epsilon i), \tag{5.1-5}$$

then

$$F_1^* = \int_{c_1}^{c_2} f(x) u^*(x) dx, \tag{5.1-6}$$

where (c_1, c_2) is the support of the limiting spectral density u^* of the general Fisher matrix $\left(\boldsymbol{R}^T \boldsymbol{S}_1 \boldsymbol{R}\right) \left(\boldsymbol{R}^T \boldsymbol{S}_2 \boldsymbol{R}\right)^{-1}$ conditional on $\boldsymbol{R}$, see the definition of general Fisher matrix in Zheng et al. (2017). $\mathcal{C}$ in the formula of m_1^* is the contour enclosing the support set (c_1, c_2). $\Im(\cdot)$ denotes the imaginary part of a complex number.

Finally,

$$v_1^* = \frac{1}{\pi^2} \int \int f'(x) f'(y) \log \left| \frac{\underline{m}(x) - \overline{\underline{m}(y)}}{\underline{m}(x) - \underline{m}(y)} \right| dx dy, \tag{5.1-7}$$

where $\overline{z}$ denotes conjugation of a complex number z.

The proof of Theorem 5.5 is given in Appendix B.

From Theorem 5.5, we have the following result.

Corollary 5.6. *Under the alternative hypothesis $H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$,*

$$
\begin{aligned}
&\mathbb{P}\left\{\Phi^{-1}\left(\frac{\alpha}{2}\right) < Z^{(\boldsymbol{R})} < \Phi^{-1}\left(1-\frac{\alpha}{2}\right) \mid \boldsymbol{R}\right\} = \\
&\Phi\left[v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}\right] \\
&-\Phi\left[v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}\right].
\end{aligned}
$$

The corollary follows by observing that

$$
\begin{aligned}
&v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* + m^* \\
&< v^{*1/2}Z^{(\boldsymbol{R})} + kF^* + m^* \\
&< v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* + m^*,
\end{aligned}
$$

which is equivalent to

$$
\begin{aligned}
&v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\} \\
&< v_1^{*-1/2}\left\{v^{*1/2}Z^{(\boldsymbol{R})} + kF^* - kF_1^* + m^* - m_1^*\right\} \\
&< v_1^{*-1/2}\left\{v^{*1/2}\Phi^{-1}\left(1-\frac{\alpha}{2}\right) + kF^* - kF_1^* + m^* - m_1^*\right\}.
\end{aligned}
$$

Then applying the asymptotic normality result from Theorem 5.5, the corollary follows.

From Theorem 5.5, we note that the power of the test is decided by the spectrum properties of $T^{(\boldsymbol{R})} = \left(\boldsymbol{R}^T\mathbf{\Sigma}_2\boldsymbol{R}\right)^{-1/2}\left(\boldsymbol{R}^T\mathbf{\Sigma}_1\boldsymbol{R}\right)\left(\boldsymbol{R}^T\mathbf{\Sigma}_2\boldsymbol{R}\right)^{-1/2}$. However, it is known that it is generally difficult to evaluate the quantities in Theorem 5.5; see Bai &

Silverstein (2010). In Section 5.1.3, we develop a framework to numerically approximate the power function.

5.1.3 Evaluating the Power Function of Testing with a Single Random Projection

To evaluate the power function given in Corollary 5.6, we need to approximate the limiting mean and covariance functions in the central limit theorem for the general Fisher matrix; see those items in Theorem 5.5. The general Fisher matrix is defined as the product of one sample covariance matrix and the inverse of another sample covariance matrix, where the population covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ can be arbitrary. When the population covariance matrices are equal, the general Fisher matrix is referred to by a simplified name as the Fisher matrix.

Evaluating the limiting mean and covariance functions for the Fisher matrix has been studied in Dobriban (2015) and Lin et al. (2016), which is a known challenging problem. Wang (2014) derived asymptotic power for corrected likelihood ratio test based on random matrix theory concerning one-sample covariance test, however, an explicit power expression is much harder for two-sample covariance test. Evaluating the limiting mean and covariance functions for general Fisher matrix is nevertheless much harder, because those functions are complex contour integrals though their values are real. Furthermore, since those functions have a very complex structure depending on non-trivial contour integrals, it is more challenging to approximate them accurately, see Zheng et al. (2017).

To approximate the power function in Corollary 5.6, we develop a numerical approximation procedure inspired by the approach of Zheng et al. (2017). The starting point is to estimate the limiting spectral distribution H of $T^{(\boldsymbol{R})}$. We use the empirical spectral distribution $H_k(t) = \sum_{j=1}^{k} \mathrm{I}\left(\lambda_j \leq t\right)/k$ to approximate the limiting spectral distribution $H(t)$. The next and an important step is to detect support (c_1, c_2) where $H(t)$ is defined. For such a purpose, we compute the Stieltjes transform $\underline{m}(z)$ by combining fixed point algorithm on $m_0(z)$, given by (5.1-2) and (5.1-3), on a grid of points $z = x + i\epsilon$, for a range of x, and a small ϵ, e.g., 0.001. Then the support is decided by examining the imaginary part of the $\underline{m}(z)$. That is, when the imaginary part is small, e.g., less than ϵ, we declare that the corresponding value of x is outside of the support.

On the detected support set (c_1, c_2), we specify

$$\{z_q = x_q + i\epsilon,\ x_q = c_1 + (c_2 - c_1)\, q/Q\}_{q=0}^{Q},$$

where ϵ is a small step size, e.g., 0.001, and Q is a large number, e.g., 1000. Then, we compute the Stieltjes transform $\underline{m}\left(z_q\right)$ by combining fixed point algorithm on $m_0(z_q)$ with (5.1-2) and (5.1-3). $u^*\left(x_q\right)$ is approximated by $\Im\left(\underline{m}\left(z_q\right)\right)/(\pi y_1^*)$ according to (5.1-4). F_1^* is approximated by $(c_2 - c_1)\sum_{q=0}^{Q} f\left(x_q\right) u^*\left(x_q\right)/(Q+1)$ according to equation (5.1-6).

To approximate m_1^* and v_1^*, we denote two grid sets as follows:

$$
\begin{aligned}
\mathcal{A}_1 = \Big\{ & z_k = c_1 - \epsilon + \left(\zeta - \frac{2\zeta k}{K_1}\right) i, \\
& k = 0, \cdots, K_1, \\
& z_{K_1+j} = c_1 - \epsilon + \frac{(c_2 - c_1 + 2\epsilon)\, j}{K_2} - \zeta i, \\
& j = 1, \cdots, K_2 - 1, \\
& z_{K_1+K_2+k} = c_2 + \epsilon + \left(-\zeta + \frac{2\zeta k}{K_1}\right) i, \\
& k = 0, \cdots, K_1, \\
& z_{2K_1+K_2+j} = c_2 + \epsilon - \frac{(c_2 - c_1 + 2\epsilon)\, j}{K_2} + \zeta i, \\
& j = 1, \cdots, K_2 \Big\}, \\
\mathcal{A}_2 = \Big\{ & z_k = c_1 - \frac{\epsilon}{2} + \left(\frac{\zeta}{2} - \frac{\zeta k}{K_1}\right) i, \\
& k = 0, \cdots, K_1, \\
& z_{K_1+j} = c_1 - \frac{\epsilon}{2} + \frac{(c_2 - c_1 + \epsilon)\, j}{K_2} - \frac{\zeta}{2} i, \\
& j = 1, \cdots, K_2 - 1, \\
& z_{K_1+K_2+k} = c_2 + \frac{\epsilon}{2} + \left(-\frac{\zeta}{2} + \frac{\zeta k}{K_1}\right) i, \\
& k = 0, \cdots, K_1 \\
& z_{2K_1+K_2+j} = c_2 + \frac{\epsilon}{2} - \frac{(c_2 - c_1 + \epsilon)\, j}{K_2} + \frac{\zeta}{2} i, \\
& j = 1, \cdots, K_2 \Big\},
\end{aligned}
$$

where K_1 and K_2 are large integers, e.g., $1,000$, and ϵ and ζ are small numbers, e.g., 0.001.

Then approximation to m_1^* is as follows according to equation (5.1-1):

$$
\begin{aligned}
m_1^* \approx \\
& -\frac{1}{4\pi}\sum_{j=0}^{2K_1+2K_2-1}\Im\left[f'(z_j)(z_{j+1}-z_j)\times\log\left\{\frac{h^{*2}}{y_2^*}-\frac{y_1^*}{y_2^*}\frac{\left(1-y_2^*\int\frac{m_0(z_j)}{t+m_0(z_j)}dH(t)\right)^2}{1-y_2^*\int\frac{m_0^2(z_j)}{(t+m_0(z_j))^2}dH(t)}\right\}\right] \\
& -\frac{1}{4\pi}\sum_{j=0}^{2K_1+2K_2-1}\Im\left[f'(z_j)(z_{j+1}-z_j)\times\log\left\{1-y_2^*\int\frac{m_0^2(z_j)}{(t+m_0(z_j))^2}dH(t)\right\}\right],
\end{aligned}
$$

where $h^{*2}=y_1^*+y_2^*-y_1^*y_2^*$. And approximation to m_1^* can be computed based on either $\mathcal{A}_1$ or $\mathcal{A}_2$.

Approximation to v_1^* is as follows by equation (5.1-5) and (5.1-7):

$$
v_1^* \approx \frac{(c_2-c_1)^2}{\pi^2K_1^2}\sum_{\substack{j,k=0\\ j\neq k}}^{K_2}\Re\left(f'(z_{K_1+j})\,f'(z_{K_1+k})\times\log\left|\frac{\underline{m}(z_j)-\overline{\underline{m}(z_k)}}{\underline{m}(z_j)-\underline{m}(z_k)}\right|\right),
$$

where $z_{K_1+j},\ z_j\in\mathcal{A}_1$ and $z_{K_1+k},\ z_k\in\mathcal{A}_2$, and $\Re(\cdot)$ is the real part of a complex number.

With the above procedure, we numerically evaluate the theoretical power function which is one minus the result in Corollary 5.6 in Section 5.1.2.

5.2 Theory of Combined Tests from Multiple Random Projections

5.2.1 Consistency of Combined Tests from Multiple Random Projections

Theorem 5.7. *Assuming conditions of Theorem 4.2, with $r^{1/2}k^{3/2}p^{-1/2}\to 0$, all p-value combining methods subsumed under our framework are consistent with probability* *tending to* 1. *Namely,*

$$\mathbb{P}\left[\mathbb{P}\left\{c_*^c(\mathbf{0}) < \alpha\right\} = \alpha\right] \to 1, * \in \{F, S, \min, \max\}.$$

The proof of Theorem 5.7 is given in Appendix C.

Theorem 5.7 indicates that combined tests from multiple random projections are consistent with high probability when $r^{1/2}k^{3/2}p^{-1/2}$ is small, which guarantees the validity of combined tests from multiple random projections when testing the equality of two high-dimensional covariance matrices.

5.2.2 Power Function of Combined Tests from Multiple Random Projections

Given $\boldsymbol{R}$, we have devised a numerical approximation of the distribution of $Z^{(\boldsymbol{R})} \sim \mathcal{N}(\mu, \sigma^2)$ when $\boldsymbol{s} \neq \mathbf{0}$ in Section 5.1.3, where

$$\mu = v^{*-1/2}\left\{k\left(F_1^* - F^*\right) + m_1^* - m^*\right\},$$

and

$$\sigma^2 = v_1^*/v^*.$$

Consequently, both μ and σ^2 depend on the spectrum property of

$$T^{(\boldsymbol{R})} = \left(\boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R}\right)^{-1/2}\left(\boldsymbol{R}^T\boldsymbol{\Sigma}_1\boldsymbol{R}\right)\left(\boldsymbol{R}^T\boldsymbol{\Sigma}_2\boldsymbol{R}\right)^{-1/2}.$$

And the result of Corollary 5.6 can be re-written as,

$$\mathbb{P}\left\{c(\boldsymbol{s}) < \alpha\right\} = 1 - \Phi\left\{-\frac{\mu}{\sigma} - \frac{\Phi^{-1}(\alpha/2)}{\sigma}\right\} + \Phi\left\{-\frac{\mu}{\sigma} + \frac{\Phi^{-1}(\alpha/2)}{\sigma}\right\}.$$

Then, given $\{\boldsymbol{R}_j\}_{j=1}^r$, the minimum power of testing with a single random projection among r random projections is

$$\min_{1\le j\le r} \mathbb{P}\left\{c_j(\boldsymbol{s}) < \alpha\right\} = \min_{1\le j\le r}\left[1 - \Phi\left\{-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\} + \Phi\left\{-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\}\right].$$

Conventional meta-analysis for p-value combining methods considers the situation when does not have access to test statistics. For practical purpose, Owen (2009) computed the power of the Fisher's method considering the case that independent test statistics following $\mathcal{N}(0,1)$ under H_0 and $\mathcal{N}(\beta_j,1)$ under H_1, which can be adapted to our situation. To depict powers of the Fisher's method, the Stouffer's method, the minimum method, and the maximum method, we have the following theorem.

Theorem 5.8. *Under the alternative hypothesis $H_1 : \boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2 = \boldsymbol{s}$. Given $\{\boldsymbol{R}_j\}_{j=1}^r$, assume $Z^{(\boldsymbol{R}_j)} \sim \mathcal{N}\left(\mu_j, \sigma_j^2\right)$, $j = 1, \cdots, r$, independently.*

- $\mathbb{P}\left\{c_F^c(\boldsymbol{s}) < \alpha\right\} = \mathbb{P}\left\{\sum_{j=1}^r X_{F,j} > \left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\}$, *where $\left(\chi^2_{2r}\right)^{-1}(t)$ is the inverse function of the cumulative distribution function of χ^2_{2r} random variable, and $X_{F,j} > 0$,*

$$\mathbb{P}\left(X_{F,j} \le t\right) = \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right] - \Phi\left[-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right].$$

- $\mathbb{P}\left\{c_S^c(\boldsymbol{s}) < \alpha\right\} = \mathbb{P}\left\{\sum_{j=1}^r X_{S,j} < r^{1/2}\Phi^{-1}(\alpha)\right\}$, *and $X_{S,j} \in \mathbb{R}$,*

$$\mathbb{P}\left(X_{S,j} \ge t\right) = \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\right] - \Phi\left[-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\right].$$

- $\mathbb{P}\left\{c^c_{\min}(\boldsymbol{s}) < \alpha\right\} =$

$$1 - \prod_{j=1}^{r}\left(\Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\}\right]\right.$$
$$\left.-\Phi\left[-\frac{\mu_j}{\sigma_j} + \frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\}\right]\right).$$

- $\mathbb{P}\left\{c^c_{\max}(\boldsymbol{s}) < \alpha\right\} =$

$$\prod_{j=1}^{r}\left[1 - \Phi\left\{-\frac{\mu_j}{\sigma_j} - \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right)\right\} + \Phi\left\{-\frac{\mu_j}{\sigma_j} + \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right)\right\}\right].$$

To show merit of the Fisher's method and the Stouffer's method for combining p-values given $\{\boldsymbol{R}_j\}_{j=1}^r$ analytically, we compare power of the Fisher's method and the Stouffer's method with the minimum power of testing with a single random projection given $\{\boldsymbol{R}_j\}_{j=1}^r$ in the following Theorem 5.9.

Theorem 5.9. *Assuming conditions of Theorem 5.8, both the Fisher's method and the Stouffer's method have better power than the minimum power of testing with a single random projection given* $\{\boldsymbol{R}_j\}_{j=1}^r$. *That is,*

$$\mathbb{P}\left\{c^c_*(\boldsymbol{s}) < \alpha\right\} > \min_{1\le j\le r}\mathbb{P}\left\{c_j(\boldsymbol{s}) < \alpha\right\}, * \in \{F, S\}.$$

Theorem 5.9 shows that it is beneficial to consider the combined test using multiple random projections, since the power of combined tests using either the Fisher's method or the Stouffer's method is greater than the possible low power from testing with a single random projection among r random projections.

5.2.3 Evaluating the Power Function of Combined Tests from Multiple Random Projections

It is clear that the power functions of both the minimum method and the maximum method in Theorem 5.8 are straightforward to calculate once μ_j and σ_j, $j = 1, \cdots, r$, have been computed using the approximation introduced in Section 5.1.3. To calculate the power function of the Fisher's method, one needs to convolve cumulative distribution functions of $X_{\mathrm{F},j}$, $j = 1, \cdots, r$, which has been studied in Owen (2009). However, it is challenging to calculate the power function of the Stouffer's method since the support of $X_{\mathrm{S},j}$ is infinite; see Owen (2009).

To compute the power of Fisher's method, it is crucial to get the distribution of

$$X_{\mathrm{Q}} := \sum_{j=1}^{r} X_{\mathrm{F},j},$$

where

$$F_{\mathrm{F},j} = \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right] - \Phi\left[-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right], X_{\mathrm{F},j} > 0.$$

$F_{\mathrm{F},j}$ is the cumulative distribution function of $X_{\mathrm{F},j}$.

Assuming $X_{\mathrm{F},j}$, $j = 1, \cdots, r$, are independent, Owen (2009) proposed to convolve cumulative distribution functions to fulfill the goal to approximate

$$F_{\mathrm{Q}} = \bigotimes_{j=1}^{r} F_{\mathrm{F},j},$$

where F_{Q} is the cumulative distribution function of X_{Q}, by bounding each $F_{\mathrm{F},j}$ between a cumulative mass function $F^{-}_{\mathrm{F},j}$ associated with a stochastically smaller discrete random variable and $F^{+}_{\mathrm{F},j}$ associated with a stochastically larger discrete random variable. Denote the convolutions of the aforementioned discrete cumulative mass functions as

$$F_{\mathrm{Q}}^{r*} = \bigotimes_{j=1}^{r} F_{\mathrm{F},j}^{*},\ * \in \{+,-\}.$$

We use the first convolution to determine the critical levels corresponding to the cumulative mass functions F_{Q}^{r-} and F_{Q}^{r+} under H_0, denoted by $\left(\chi^2_{2r}\right)^{-1}(1-\alpha)^-$ and $\left(\chi^2_{2r}\right)^{-1}(1-\alpha)^+$. Then a second convolution is used under H_1 to compute power of the Fisher's method. Specifically, the upper limit of power of Fisher's method comes from $1-F_{\mathrm{Q}}^{r+}\left\{\left(\chi^2_{2r}\right)^{-1}(1-\alpha)^-\right\}$, and the lower limit of the power of Fisher's method comes from $1-F_{\mathrm{Q}}^{r-}\left\{\left(\chi^2_{2r}\right)^{-1}(1-\alpha)^+\right\}$.

To bound some cumulative distribution function F with non-negative support, we consider the probability mass function associated with a stochastically larger discrete random variable defined as follows,

$$\begin{aligned}
\lceil f\rceil_N(0) &= 0,\\
\lceil f\rceil_N(i\eta) &= F(i\eta) - F\{(i-1)\eta\}, i = 1,\cdots,N-1,\\
\lceil f\rceil_N(\infty) &= 1 - \sum_{i=1}^{N-1}\lceil f\rceil_N(i\eta).
\end{aligned}$$

And one associated with a stochastically smaller discrete random variable, defined as follows,

$$\begin{aligned}
\lfloor f\rfloor_N(0) &= 0,\\
\lfloor f\rfloor_N(i\eta) &= F(i\eta) - F\{(i-1)\eta\}, i = 1,\cdots,N-2,\\
\lfloor f\rfloor_N\{i(N-1)\} &= 1 - \sum_{i=1}^{N-2}\lfloor f\rfloor_N(i\eta).
\end{aligned}$$

N is some large integer, e.g., 100000, and η is some small number, e.g., 0.001.

To convolve F with another cumulative distribution function G, we use Fourier transform and inverse Fourier transform to calculate convolution of probability mass functions fast as follows, by using the "fastest Fourier transform in the west" algorithm wrapped in the r package "fftw".

$$\lceil f \otimes g \rceil_N = \mathcal{F}^{-1}\left\{\mathcal{F}\left(\lceil f \rceil_N\right) \odot \mathcal{F}\left(\lceil g \rceil_N\right)\right\},$$
$$\lceil f \otimes g \rceil_N(\infty) = \lceil f \rceil_N(\infty) + \lceil g \rceil_N(\infty) - \lceil f \rceil_N(\infty)\lceil g \rceil_N(\infty),$$
$$\lfloor f \otimes g \rfloor_N = \mathcal{F}^{-1}\left\{\mathcal{F}\left(\lfloor f \rfloor_N\right) \odot \mathcal{F}\left(\lfloor g \rfloor_N\right)\right\},$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote Fourier and inverse Fourier transforms, respectively, $\odot$ denotes the element-wise product, and

$$\lceil f \rceil_N = \left[\lceil f \rceil_N(0), \lceil f \rceil_N(\eta), \cdots, \lceil f \rceil_N\{(N-1)\eta\}\right],$$

$$\lfloor f \rfloor_N = \left[\lfloor f \rfloor_N(0), \lfloor f \rfloor_N(\eta), \cdots, \lfloor f \rfloor_N\{(N-1)\eta\}\right],$$

$\lceil g \rceil_N$ and $\lfloor g \rfloor_N$ are defined similarly. Hence, $\lceil F \otimes G \rceil_N$ and $\lfloor F \otimes G \rfloor_N$ are computed as follows.

$$\lceil F \otimes G \rceil_N(i\zeta) = \sum_{k=0}^{i} \lceil f \otimes g \rceil_N(k\zeta), i = 0, 1, \cdots, N-1, \infty,$$
$$\lfloor F \otimes G \rfloor_N(i\zeta) = \sum_{k=0}^{i} \lfloor f \otimes g \rfloor_N(k\zeta), i = 0, 1, \cdots, N-1.$$

Consequently, F_{Q}^{r+} and F_{Q}^{r-} are computed as follows.

$$F_{\mathrm{Q}}^{j+} = \lceil F_{\mathrm{Q}}^{(j-1)+} \otimes F_{\mathrm{F},j}^{+} \rceil_N,\ j = 1, \cdots, r,$$

$$F_{\mathrm{Q}}^{j-} = \lfloor F_{\mathrm{Q}}^{(j-1)-} \otimes F_{\mathrm{F},j}^{-} \rfloor_N,\ j = 1, \cdots, r,$$

where $F_{\mathrm{Q}}^{0\pm}$ is the cumulative mass function of a point mass at 0.

CHAPTER 6

NUMERICAL EXAMPLES

To better illustrate the proposed methodology, we conduct simulation studies to validate the performance of the proposed tests with single random projection and multiple random projections. In our numerical examples, we generate independent samples from normal distributions with a zero mean, and set $n_1 = n_2 = n$. For all settings, the simulations are repeated $1,000$ times.

Let $\boldsymbol{D} = \mathrm{diag}(d_{11}, \ldots, d_{pp})$ be a diagonal matrix with d_{ii}, $i = 1, \cdots, p$, randomly generated from $\mathrm{U}(0.5, 2.5)$ distribution. We consider the following four models to generate simulated data with $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}^{(i)}$, $i = 1, \cdots, 4$:

1. $\boldsymbol{\Sigma}^{*(1)} = \left(\sigma_{ij}^{*(1)}\right)$, where $\sigma_{ii}^{*(1)} = 1$, $\sigma_{ij}^{*(1)} = 0.5$ for $5(k-1)+1 \leq i \neq j \leq 5k$, where $k = 1, \ldots, p/5$, and $\sigma_{ij}^{*(1)} = 0$ otherwise. $\boldsymbol{\Sigma}^{(1)} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^{*(1)}\boldsymbol{D}^{1/2}$.

2. $\boldsymbol{\Sigma}^{*(2)} = \left(\sigma_{ij}^{*(2)}\right)$, where $\sigma_{ij}^{*(2)} = 0.5^{|i-j|}$ for $1 \leq i, j \leq p$. $\boldsymbol{\Sigma}^{(2)} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^{*(2)}\boldsymbol{D}^{1/2}$.

3. $\boldsymbol{\Sigma}^{*(3)} = \left(\sigma_{ij}^{*(3)}\right)$, where $\sigma_{ii}^{*(3)} = 1$, $\sigma_{ij}^{*(3)} = 0.5 * \mathrm{Bernoulli}(1, 0.05)$ for $i < j$ and $\sigma_{ji}^{*(3)} = \sigma_{ij}^{*(3)}$. $\boldsymbol{\Sigma}^{(3)} = \boldsymbol{D}^{1/2}\left(\boldsymbol{\Sigma}^{*(3)} + \delta \boldsymbol{I}\right)/(1+\delta)\boldsymbol{D}^{1/2}$ with $\delta = \left|\lambda_{\min}\left\{\boldsymbol{\Sigma}^{*(3)}\right\}\right| + 0.05$.

4. $\boldsymbol{\Sigma}^{(4)} = \boldsymbol{O}\boldsymbol{\Delta}\boldsymbol{O}$, where $\boldsymbol{O} = \mathrm{diag}(\omega_1, \cdots, \omega_p)$ and $\omega_1, \cdots, \omega_p \sim \mathrm{U}(1, 5)$ independently and $\boldsymbol{\Delta} = (a_{ij})$ and $a_{ij} = (-1)^{i+j} 0.4^{|i-j|^{1/10}}$.

With $\boldsymbol{\Sigma}_1$ taken from the four models, respectively, we then set $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1 + \boldsymbol{U}$ in an additive manner with dense and weak signals. Here $\boldsymbol{U}$ is a symmetric matrix with upper triangular and diagonal entries independently generated from $a \times V \times \max_{1 \le j \le p} \sigma_{jj}$, where $\{\sigma_{jj}\}_{j=1}^{p}$ are diagonal entries of $\mathbf{\Sigma}_1$ and V follows the uniform distribution over the interval $(0, 4)$. When $a = 0$, the null hypothesis is true such that we are evaluating the consistency of the proposed test; and when $a = 0.1$, the power performance of the proposed test under many small entry-wise departures from H_0 is investigated.

6.1 Simulation for Testing with a Single Random Projection

The test with a single random projection is evaluated from different combinations of n and p. Our method is implemented as described in Section 4.1. We implement the random projections with different k's: $k = [qn^{1/3}]$ with $q = 5, 6, 7$, where $[\cdot]$ denotes the operation of taking integer. For given matrix $\boldsymbol{R}$, we implement the Bartlett's modified likelihood ratio statistic (4.1-2) and compare with the chi-square distribution, and denote it by "RP-BLRT". We implement the corrected likelihood ratio statistic (4.1-3) with random matrix theory, and denote it by "RP-CLRT". For comparisons, we also implement the test of T. Cai et al. (2013), denoted by "CLX", and the bootstrap-assisted test of Chang et al. (2017), denoted by "CZZW". Both tests are based on $\max_{1 \le k,l \le p} |\sigma_{1,kl} - \sigma_{2,kl}|$.

6.1.1 Simulation for Consistency of Testing with a Single Random Projection

The results are reported in Table 6.1. From Table 6.1, we observe that the corrected test adjusted with the random matrix theory works satisfactorily with the randomly projected vectors. From Table 6.1, we also see that the adjustment with the random matrix theory is necessary, because we observe that the sizes with the Bartlett's modification are way off from the nominal level. This is due to the fact that the chi-square distribution poorly approximates that of the likelihood ratio even when the data dimensionality is moderate.

For all models, we also see that the empirical sizes of the tests of T. Cai et al. (2013) and Chang et al. (2017) are satisfactory.

6.1.2 Simulation for Power Comparisons with Additive, Dense and Weak Signals

To evaluate the empirical powers of the tests, we generate data from scenarios with H_0 violated by setting $a = 0.1$, under which there are many small entry-wise departures from H_0. Results of the simulations are reported in Table 6.2. From Table 6.2, we can see that the corrected test adjusted with the random matrix theory is either more powerful than or comparable to tests of T. Cai et al. (2013) and Chang et al. (2017) under all different combinations of n and p, and all four different models, when the signals are additive, dense, and weak. In particular, when $n = 300$, $p = 400$, the power superiority of the corrected test adjusted with the random matrix theory over competing tests are over 50% for $k = [7n^{1/3}]$, over 45% for $k = [6n^{1/3}]$, over 35% for $k = [5n^{1/3}]$, under all four different models with competing tests having power around 35%. This reflects the merit of the likelihood ratio approach in aggregating the evidence from violating the null hypothesis.

6.1.3 Simulation with Rotational Transformations

In this study, we consider a different way of generating the alternative $\mathbf{\Sigma}_2$ from sequentially rotating $\mathbf{\Sigma}_1$ in multiple subspaces with a small angle θ accompanied with a small extension factor $(1 + d)$. By doing so, $\mathbf{\Sigma}_2$ also deviate from $\mathbf{\Sigma}_1$ with small component-wise difference. Specifically, for a given small θ, we apply the Givens rotation matrix for such a purpose:

Table 6.1: Empirical percentage of rejecting H_0 for testing with a single random projection: H_0 is true ($a = 0$), $\alpha = 0.05$.

		Model 1					Model 2					Model 3					Model 4				
n	p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
Empirical percentage of rejecting H_0; $k = [7n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	97.3	97.4	97.7	97.6	97.4	96.9	97.3	97.6	97.2	97.6	96.4	96.8	96.5	96.8	97.5	97.1	96.7	98.1	97.1	97.4
	RP-CLRT	5.8	6.0	5.5	6.0	5.8	6.5	6.2	5.6	6.7	5.1	6.2	6.6	4.8	4.7	5.2	4.9	5.1	6.3	6.1	5.4
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	85.4	86.7	86.3	87.8	85.1	86.8	86.7	85.8	85.8	84.1	85.2	86.3	85.7	86.8	86.1	84.4	85.9	84.9	85.7	85.5
	RP-CLRT	5.5	6.5	5.4	5.1	5.2	5.2	5.4	4.2	6.2	5.3	5.5	6.0	5.8	4.5	4.5	4.9	5.1	4.3	5.5	5.5
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	52.6	55.2	53.1	51.2	54.2	52.7	54.2	50.4	53.0	51.9	51.8	55.0	51.0	51.9	53.8	50.4	53.7	50.8	54.1	52.4
	RP-CLRT	5.3	4.8	5.9	3.9	3.8	6.1	4.9	4.5	4.8	5.3	6.1	5.9	4.9	4.5	4.9	4.6	5.6	5.2	5.7	5.0
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6
Empirical percentage of rejecting H_0; $k = [6n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	79.8	78.5	80.1	80.1	80.2	79.4	78.3	82.8	78.8	75.8	80.0	79.9	78.9	78.5	81.2	77.2	79.0	80.0	80.9	79.6
	RP-CLRT	6.9	7.2	5.2	5.4	4.3	5.9	4.3	5.9	5.6	6.3	4.9	6.7	5.1	5.0	5.5	4.7	6.5	6.3	5.7	3.4
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	64.4	64.6	63.2	65.0	66.5	64.7	64.2	64.9	63.4	63.7	62.7	65.5	63.3	64.5	65.1	65.3	65.0	61.2	64.5	64.1
	RP-CLRT	5.2	4.5	6.5	6.4	5.6	5.6	5.6	5.8	4.8	5.1	5.0	4.9	5.0	5.9	5.9	6.6	5.0	5.9	5.5	4.7
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	33.7	32.0	29.7	30.9	28.5	31.9	34.6	33.5	30.5	30.0	34.2	30.3	29.5	28.5	31.1	34.2	30.8	32.5	29.8	32.0
	RP-CLRT	6.1	5.0	6.0	5.4	5.5	6.4	5.1	5.1	4.8	6.0	6.1	6.4	4.5	6.7	5.3	6.5	3.7	6.2	4.3	6.8
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6
Empirical percentage of rejecting H_0; $k = [5n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	55.3	55.1	56.0	54.0	53.2	53.5	57.4	51.7	56.2	55.9	54.5	55.3	56.1	58.6	55.9	54.3	56.1	57.1	53.5	51.7
	RP-CLRT	4.4	5.6	5.2	6.0	6.1	4.7	5.2	7.1	6.6	3.8	6.2	4.9	5.9	6.0	7.2	6.1	5.9	5.0	5.5	5.9
	CLX	3.8	4.6	5.1	4.6	3.5	3.9	5.9	5.4	3.2	5.6	6.0	5.5	5.5	5.6	3.4	4.9	4.0	4.5	4.1	4.0
	CZZW	4.9	5.7	6.3	6.3	5.8	4.9	7.0	7.3	5.3	7.8	6.3	6.3	6.6	7.6	4.8	6.6	6.5	6.9	9.1	9.1
100	RP-BLRT	34.0	33.3	35.8	34.1	33.2	34.5	34.8	33.2	35.1	34.1	37.2	34.5	35.3	33.4	31.9	35.4	35.4	36.1	35.1	34.7
	RP-CLRT	5.9	5.1	6.5	4.9	5.1	4.6	4.6	5.1	4.9	5.3	5.7	6.2	4.6	5.8	6.3	7.0	5.3	5.4	6.7	6.1
	CLX	3.6	4.4	4.2	4.9	3.1	4.6	4.5	4.4	5.1	5.3	4.1	3.8	4.7	6.3	4.9	3.9	2.6	5.1	4.7	4.2
	CZZW	4.2	6.2	5.1	5.5	4.4	5.0	5.2	5.1	5.2	5.4	4.6	4.8	5.2	6.6	6.7	5.4	4.4	7.1	5.8	5.3
300	RP-BLRT	15.5	15.4	14.8	17.6	18.0	16.0	17.2	17.4	15.8	15.1	15.9	16.8	18.2	16.8	16.3	16.7	17.0	15.8	17.0	17.7
	RP-CLRT	4.8	4.9	4.3	5.7	5.8	6.2	4.2	5.1	5.5	4.9	5.4	5.3	6.0	6.0	6.2	4.9	5.6	5.4	4.3	4.2
	CLX	3.5	4.5	4.2	3.4	4.6	4.0	4.2	5.2	2.4	7.6	3.9	3.4	5.3	4.3	3.8	4.2	3.3	3.3	4.2	6.6
	CZZW	3.8	5.1	4.5	3.4	4.8	4.5	4.8	5.6	2.4	9.0	4.2	4.0	5.7	4.6	3.7	5.6	5.2	4.5	5.0	7.6

Table 6.2: Empirical percentage of rejecting H_0 for testing with a single random projection: H_0 is not true ($a = 0.1$), $\alpha = 0.05$.

		Model 1					Model 2					Model 3					Model 4				
n	p	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800	50	100	200	400	800
Empirical percentage of rejecting H_0; $k = [7n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	100	99.6	99.4	98.9	98.4	100	99.9	99.7	99.2	99.2	100	100	99.4	99.0	99.5	100	99.8	99.8	98.6	98.2
	RP-CLRT	65.3	23.6	13.9	8.2	6.3	71.6	29.7	10.5	6.8	6.1	69.8	30.3	14.9	8.3	5.8	77.2	26.2	11.1	7.6	7.3
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	100	100	99.6	96.6	92.8	100	100	99.3	97.8	95.7	100	100	99.2	96.7	93.2	100	100	99.3	97.5	94.9
	RP-CLRT	99.7	68.0	39.6	15.4	9.0	99.3	77.0	32.2	18.8	8.6	99.6	76.5	34.5	18.8	9.3	100	76.9	43.8	15.4	9.6
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	100	99.8	97.4	100	100	100	99.9	98.1	100	100	100	100	97.9	100	100	100	100	97.5
	RP-CLRT	100	100	100	95.1	61.4	100	100	100	84.3	50.7	100	100	100	95.5	45.0	100	100	100	93.8	42.8
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9
Empirical percentage of rejecting H_0; $k = [6n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	99.7	97.0	92.4	88.9	85.1	99.7	97.4	92.3	89.6	84.7	99.7	98.2	92.8	87.1	85.8	99.8	98.0	92.2	89.2	85.3
	RP-CLRT	61.6	25.5	8.4	7.8	6.6	59.7	18.6	12.1	8.2	6.4	75.7	31.7	17.2	6.3	6.5	67.8	28.0	17.9	6.7	6.7
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	100	99.9	94.4	86.5	79.2	100	99.1	94.1	85.5	79.0	100	99.7	95.5	89.1	78.8	100	99.8	94.5	86.7	77.2
	RP-CLRT	98.8	63.4	22.3	17.2	8.7	98.0	66.7	41.8	12.8	12.1	99.7	72.4	26.2	12.5	7.8	99.6	67.8	53.5	11.6	6.1
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	100	99.4	88.7	100	100	100	99.5	88.9	100	100	100	98.9	88.9	100	100	100	99.6	90.3
	RP-CLRT	100	100	99.7	85.4	36.6	100	100	99.8	96.3	37.6	100	100	99.9	91.2	48.1	100	100	99.9	93.7	65.7
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9
Empirical percentage of rejecting H_0; $k = [5n^{1/3}]$ for $\boldsymbol{R}$																					
60	RP-BLRT	96.7	89.5	78.1	70.4	63.2	98.0	88.7	75.5	71.0	64.2	97.9	88.1	78.1	69.7	63.8	98.8	87.4	77.9	67.6	65.2
	RP-CLRT	48.5	18.9	14.3	9.6	5.4	49.3	28.0	11.0	5.9	8.7	53.8	29.9	15.6	7.7	8.0	72.4	31.0	10.3	8.5	7.1
	CLX	29.6	19.6	13.5	7.6	6.9	27.9	20.9	11.8	9.3	11.1	35.4	20.8	12.2	7.7	4.6	35.3	19.2	12.4	7.7	7.8
	CZZW	33.7	22.0	16.5	9.8	9.2	31.6	23.2	13.4	11.5	14.3	38.2	24.0	14.4	9.6	6.2	37.6	23.2	15.4	11.1	9.5
100	RP-BLRT	99.7	94.4	77.8	61.2	50.2	100	93.4	78.2	62.7	51.3	100	95.0	78.4	61.4	51.1	100	96.2	78.6	61.7	47.6
	RP-CLRT	97.1	65.6	17.7	16.2	9.0	95.9	49.3	20.4	10.6	5.1	98.8	62.4	39.7	9.7	6.5	97.6	68.2	43.1	11.5	7.1
	CLX	59.9	35.5	19.1	10.4	5.4	63.3	36.5	19.1	16.9	3.9	69.0	41.3	19.8	11.4	4.6	68.1	35.5	17.9	10.1	4.4
	CZZW	62.4	39.4	21.1	11.6	6.9	66.0	39.6	21.5	18.6	5.4	70.9	43.8	21.9	12.9	5.0	70.2	38.7	19.6	10.7	5.7
300	RP-BLRT	100	100	100	94.5	72.6	100	100	100	94.1	72.5	100	100	99.9	93.9	75.9	100	100	100	96.8	75.9
	RP-CLRT	100	100	99.8	90.5	32.2	100	100	99.5	65.6	63.5	100	100	99.9	92.8	44.3	100	100	99.6	88.1	46.6
	CLX	100	98.4	74.1	33.3	23.9	100	99.0	78.9	34.7	22.1	100	98.7	78.2	44.0	14.5	100	99.4	78.1	37.3	21.3
	CZZW	100	98.5	76.5	34.1	24.6	100	98.9	80.7	36.1	23.1	100	98.9	80.2	44.8	17.0	100	99.7	80.2	37.6	22.9

$$
\boldsymbol{G}(i,j,\theta,d) = \begin{array}{cc}
\begin{array}{ccccccc} & & j & & i & & \end{array} & \\
\begin{pmatrix}
1 & \dots & 0 & \dots & 0 & \dots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots \\
0 & \dots & \cos(\theta)\times(1+d)^{1/2} & \dots & -\sin(\theta)\times(1+d)^{1/2} & \dots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \dots & \sin(\theta)\times(1+d)^{1/2} & \dots & \cos(\theta)\times(1+d)^{1/2} & \dots & 0 \\
\vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \dots & 0 & \dots & 0 & \dots & 1
\end{pmatrix} &
\begin{array}{c} \\ \\ j \\ \\ i \\ \\ \\ \end{array}
\end{array}.
$$

We set $\boldsymbol{\Sigma}_2 = \{\prod_{g=1}^{p/2} \boldsymbol{G}(i_g, j_g, \theta, d)\} \times \boldsymbol{\Sigma}_1 \times \{\prod_{g=1}^{p/2} \boldsymbol{G}(i_g, j_g, \theta, d)\}^T$. We consider two cases for $\boldsymbol{\Sigma}_1$ respectively being the identity matrix of appropriate size, and $\boldsymbol{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$.

We set $n_1 = n_2 = n = 200$, $p = 400$, $k = [7n^{1/3}]$, and compare the powers with different settings of θ. The results are reported in Figure 6.1 and Figure 6.2. Since it is evident that Bartlett's modified likelihood ratio statistic using (4.1-2) and chi-square distribution as the reference distribution does not work well, we do not include the corresponding results in our figures.

Both Figures 6.1 and 6.2 show promising performance of our test. From Figure 6.1, we can see that the corrected likelihood ratio test adjusted with the random matrix theory is more powerful than the tests designed for detecting sparse and strong signals of T. Cai et al. (2013) and Chang et al. (2017) under all combinations of angle θ and extension $e = (1+d)^{1/2}$, under the setting $\boldsymbol{\Sigma}_1 = \boldsymbol{I}$. Under the setting $\boldsymbol{\Sigma}_1 = \boldsymbol{I}$, sequentially rotating $\boldsymbol{\Sigma}_1$ in multiple subspaces with a small angle θ accompanied by small extension e will result in the signals differing $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are dense and weak.

From Figure 6.2, we see same phenomenon when $\theta = 0.01\pi$, 0.03π, accompanied with all pre-specified values of small extension e under the setting $\mathbf{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$. Since sequential rotation by not so small angle θ accompanied with quite small extension e on matrix with all non-zero off-diagonal entries tend to generate sparse signals, for the case $\theta = 0.05\pi$, 0.07π, the tests of T. Cai et al. (2013) and Chang et al. (2017) display its sensitivity to sparse signals, but still show unsatisfactory power. On the other hand, when $\theta = 0.05\pi$, 0.07π, accompanied with a not so small extension e, the corrected test adjusted by random matrix theory better captures those dense and weak signals generated by getting $\mathbf{\Sigma}_2$ from sequential rotation at not so small an angle θ accompanied by not so small an extension e on $\mathbf{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$.

This setting is also informative in demonstrating the merits of our approach compared with others, including those targeting at aggregating the component-wise discrepancies, for example, M. S. Srivastava & Yanagihara (2010), which is based on $\operatorname{tr}\left(\mathbf{\Sigma}_1^2\right)/\operatorname{tr}^2\left(\mathbf{\Sigma}_1\right) - \operatorname{tr}\left(\mathbf{\Sigma}_2^2\right)/\operatorname{tr}^2\left(\mathbf{\Sigma}_2\right)$, where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix, and we denote it by "SY". We consider the case for $\mathbf{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$, and then $\mathbf{\Sigma}_2 = \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\} \times \mathbf{\Sigma}_1 \times \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\}^T$ as well as $\mathbf{\Sigma}_2 = \left\{\prod_{g=1}^{p} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\} \times \mathbf{\Sigma}_1 \times \left\{\prod_{g=1}^{p} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\}^T$, of which two settings of $\theta = 0.01\pi$ and 0.07π are considered.

We set (n, p) as $(100, 200)$ and $k = n/2$. We compare the powers in two settings: 1) with d ranging in $0.03, 0.06, \cdots, 0.57, 0.6$ and $\mathbf{\Sigma}_2 = \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\} \times \mathbf{\Sigma}_1 \times \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\}^T$; 2) with d ranging in $0.07, 0.14, \cdots, 1.26, 1.40$ and $\mathbf{\Sigma}_2 = \left\{\prod_{g=1}^{p} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\} \times \mathbf{\Sigma}_1 \times \left\{\prod_{g=1}^{p} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\}^T$. Here a larger d means relatively more discrepancy between the two covariance matrices; and $\mathbf{\Sigma}_2$ in the 2nd setting differs more substantially from $\mathbf{\Sigma}_1$ than that in the 1st setting. The results are reported in Figure 6.3 and Figure 6.4, respectively. From Figure 6.3, we can see that the corrected likelihood ratio test with a single random projection is becoming more

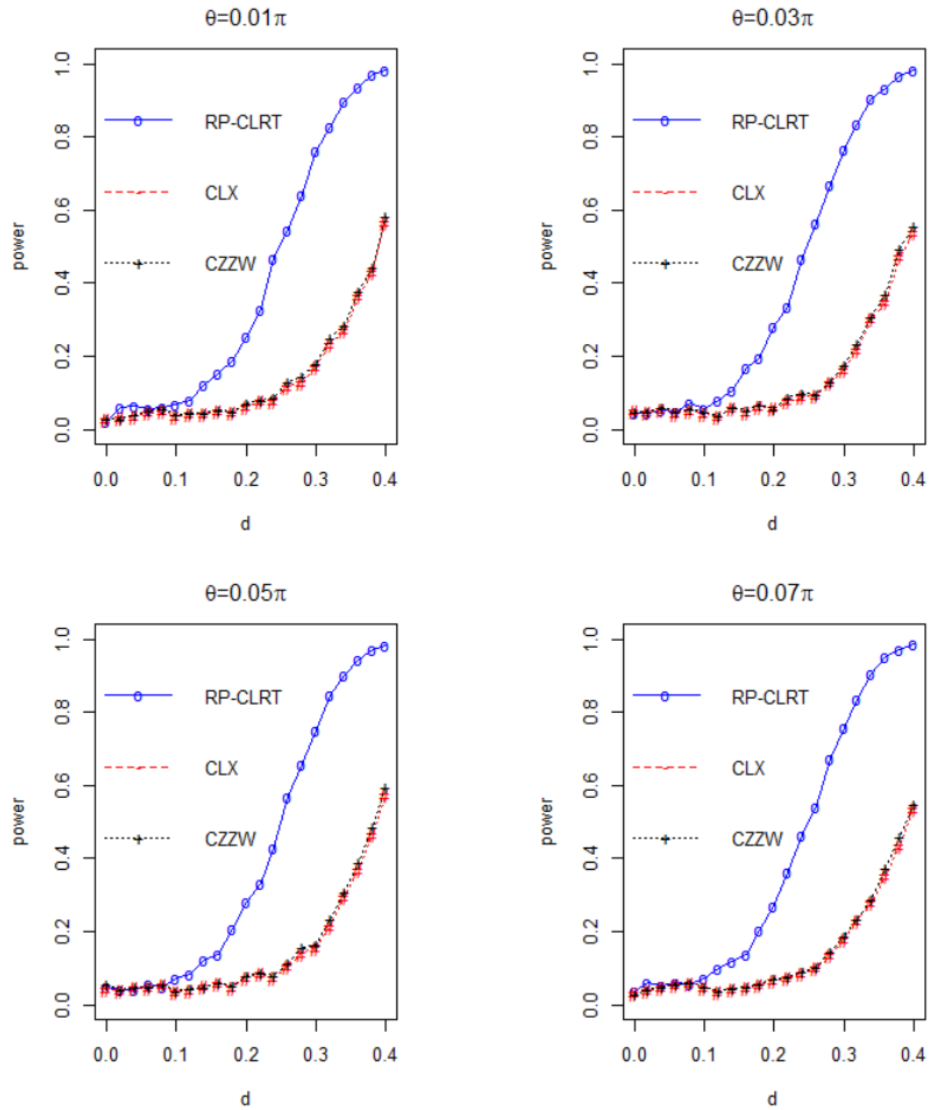

Figure 6.1: Graphs of divergence between powers of corrected likelihood ratio test from a single random projection and "CLX" as well as "CZZW" when $\Sigma_1 = I$ and $\Sigma_2 = \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\} \times \Sigma_1 \times \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\}^T$: $(n,p) = (200, 400)$, $k = [7n^{1/3}]$.

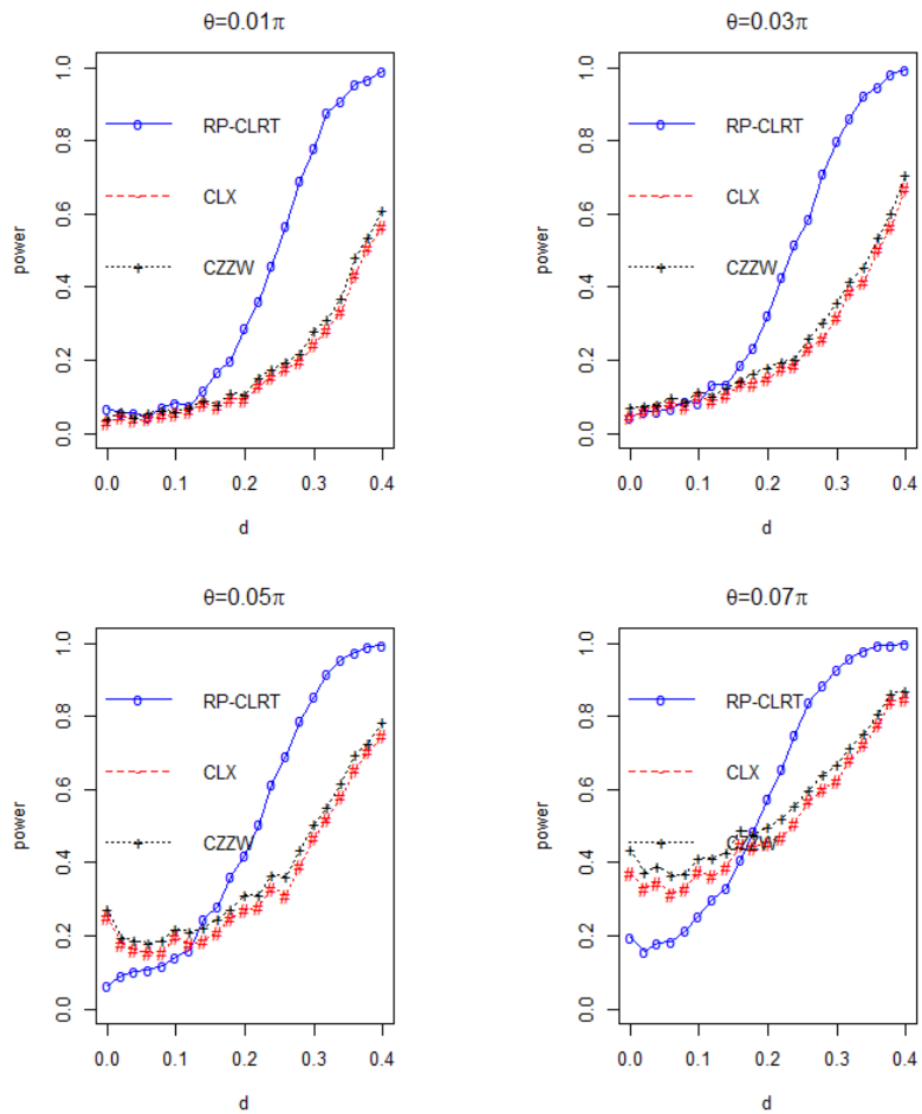

Figure 6.2: Graphs of divergence between powers of corrected likelihood ratio test from a single random projection and "CLX" as well as "CZZW" when $\Sigma_1 = 0.8I + 0.2 \times 11^T$ and $\Sigma_2 = \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\} \times \Sigma_1 \times \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\}^T$: $(n, p) = (200, 400)$, $k = [7n^{1/3}]$.

and more powerful when d is increasing, whereas "SY" remains not powerful in this setting. In Figure 6.4, though "SY" can gain its power with a relatively larger d in this setting with stronger signal, the corrected likelihood ratio test with a single random projection reaches 100% very quickly. This demonstrates the advantage of the proposed approach for detecting weak and dense signal, thanks to the powerful nature of the corrected likelihood ratio statistics.

6.1.4 Impact from Different k's for Testing with a Single Random Projection

In this section, we examine the impact from different values of k on the power of the corrected likelihood ratio testing with a single random projection. For each k, we repeat the simulation 1,000 times, and report the averaged powers and their standard deviations.

We conduct simulations with different k's being $0.05n$, $0.1n$, $\cdots$, $0.45n$ and $0.5n$ for different combinations of (n, p) being (100,200), (200,400), and (300,600). We use the rotational alternative setting of $\boldsymbol{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$ and $\boldsymbol{\Sigma}_2 = \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\} \times \boldsymbol{\Sigma}_1 \times \left\{\prod_{g=1}^{p/2} \boldsymbol{G}\left(i_g, j_g, \theta, d\right)\right\}^T$ with $\theta = 0.07\pi$ as well as $d = 0.25$, and 0.3.

The results are reported in Table 6.3. Since all standard deviations in Table 6.3 are quite small, it can be concluded that the powers are stable for a given k.

We also examine the powers with different k's with Table 6.3. Collectively, we observe that the powers of the corrected likelihood ratio testing with a single random projection is stable for k in a reasonable range. Nonetheless, we observe that when the sample size is not large enough, and k becomes excessively large, there is some power loss, compared with using smaller k. As demonstrated in previous examples, the corrected likelihood ratio testing with a single random projection takes advantage of the likelihood ratio approach in detecting signals that the null hypothesis is violated.

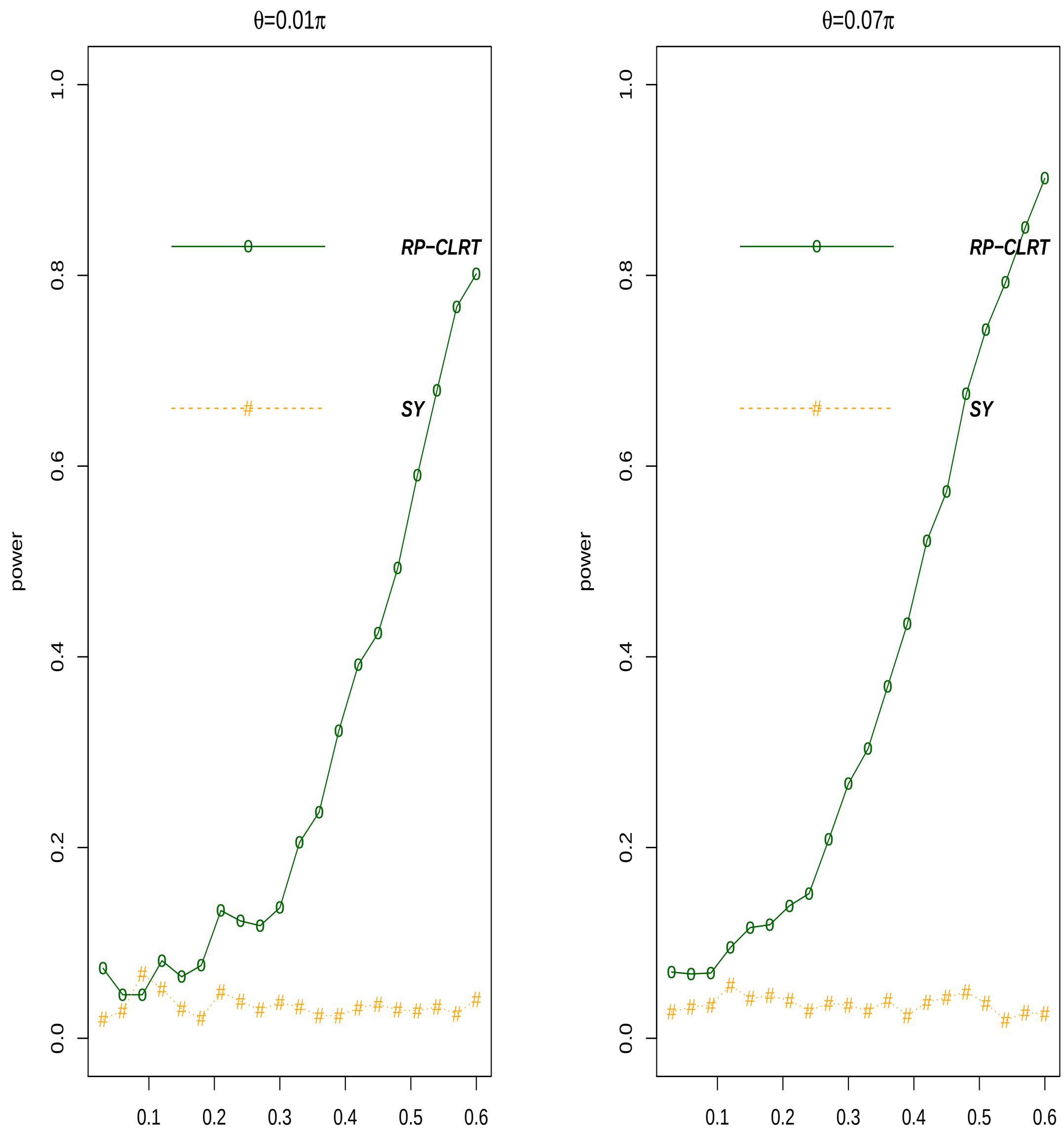

Figure 6.3: Graphs of divergence between powers of corrected likelihood ratio test from a single random projection and "SY" when $\Sigma_1 = 0.8I + 0.2 \times 11^T$ and $\Sigma_2 = \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\} \times \Sigma_1 \times \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\}^T$: $(n, p) = (100, 200)$, $k = n/2$.

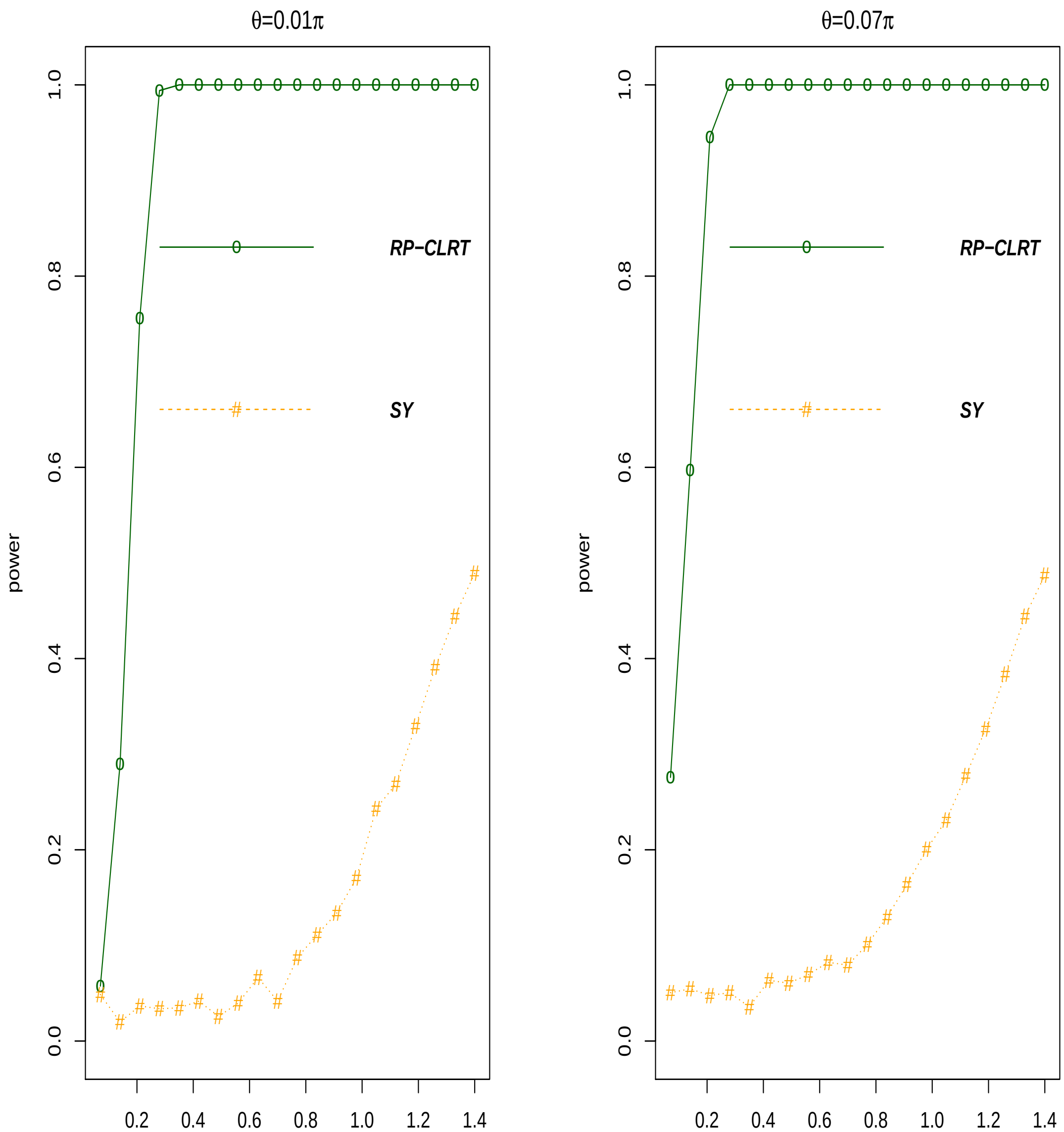

Figure 6.4: Graphs of divergence between powers of corrected likelihood ratio test from a single random projection and "SY" when $\Sigma_1 = 0.8I + 0.2 \times 11^T$ and $\Sigma_2 = \{\prod_{g=1}^{p} G(i_g, j_g, \theta, d)\} \times \Sigma_1 \times \{\prod_{g=1}^{p} G(i_g, j_g, \theta, d)\}^T$: $(n, p) = (100, 200)$, $k = n/2$.

When k becomes sufficiently large with a given n, it may exceed the asymptotic powerful zone of the likelihood ratio test statistic, especially when n is not large enough. In our experience, as long as k not exceeding $n/2$, the tests work reasonably well; we recommend checking multiple choices in the range between $n/4$ and $n/2$ for its stability.

6.1.5 Numerical Approximation to Power Function of Testing with a Single Random Projection

In Theorem 5.5, we have shown that the power of the corrected likelihood ratio testing with a single random projection depends on the matrix $T^{(\boldsymbol{R})}$. Given the order of k, we could utilize the numerical evaluation, which is introduced in Section 5.1.3, of the power of the corrected likelihood ratio testing with any single random projection matrix $\boldsymbol{R} \in \mathbb{R}^{p\times k}$. Fixing the order of k being $[7n^{1/3}]$, we now demonstrate that the corrected likelihood ratio testing with a single random projection is able to asymptotically distinguish H_1 from H_0 with random projection matrix $\boldsymbol{R}$ of column size $k = [7n^{1/3}]$.

Denote $\boldsymbol{U} = \boldsymbol{U}^T = \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1 = (u_{kl})_{1\le k,l\le p}$, and $\{\sigma_{jj}\}_{j=1}^p$ being diagonal entries of $\boldsymbol{\Sigma}_1$, we consider testing the following class

$$\mathcal{W}(s,\lambda) = \left\{ u_{kl} \sim s \times W \times \max_{1\le j\le p} \sigma_{jj}, \lambda = \frac{p}{n}, W \sim \mathrm{Unif}(0,1) \right\},$$

where s denotes the order of the signal and λ denotes the dimension-to-sample size ratio when both n and p tend to be large.

In particular, we numerically approximate the power functions of the corrected likelihood ratio testing with five random projection matrices of column size k being determined as $[7n^{1/3}]$ for the testing class $\mathcal{W}(0.04, 2)$, where the order of the signal

Table 6.3: Averages (*avg.*) and standard deviations (*std.dev.*) of the empirical percentage of rejecting H_0 under rotational alternative of $\Sigma_1 = 0.8I + 0.2 \times 11^T$, $\Sigma_2 = \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\} \times \Sigma_1 \times \left\{\prod_{g=1}^{p/2} G\left(i_g, j_g, \theta, d\right)\right\}^T$, $\theta = 0.07\pi$; $\alpha = 0.05$; $k = yn$.

(n,p)		(100,200)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	34.5	32.1	31.6	28.8	27.5	27.1	25.3	22.8	20.4	18.9
	std. dev.	2.3	2.4	2.2	2.0	2.1	2.2	2.3	2.0	2.2	2.1
0.3	avg.	45.1	44.4	43.9	42.5	39.4	38.0	35.7	33.1	31.8	29.4
	std. dev.	2.1	2.4	2.3	2.3	2.2	2.4	2.0	2.3	2.4	2.2
(n,p)		(200,400)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	75.3	81.2	79.5	79.0	78.3	74.6	72.8	69.2	64.1	59.6
	std. dev.	2.1	2.0	1.9	2.3	2.1	2.0	2.3	2.1	2.0	1.9
0.3	avg.	91.0	92.9	95.1	93.7	92.1	89.3	88.6	83.5	80.0	76.4
	std. dev.	1.2	1.2	1.1	1.3	1.2	1.2	1.4	1.4	1.9	2.0
(n,p)		(300,600)									
d	y	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
0.25	avg.	97.9	98.3	98.1	98.4	98.3	97.0	95.6	92.8	91.1	90.0
	std. dev.	0.2	0.3	0.1	0.2	0.3	0.2	0.4	0.4	0.3	0.2
0.3	avg.	99.5	99.9	99.9	99.9	99.7	99.8	99.5	99.3	98.3	98.1
	std. dev.	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

is set as 0.04, and $p = 2n$, a high-dimensional testing scenario, where a set of (n, p) combinations is $(100, 200)$, $(150, 300)$, $(200, 400)$, $(300, 600)$, $(400, 800)$, $(600, 1200)$, $(800, 1600)$, $(1200, 2400)$ and $(1600, 3200)$.

The results for Model 1 are reported in Figure 6.5. Since the results for all four models are quite similar, we do not include corresponding results in our figures. From Figure 6.5, it can be seen that as long as the column size of the random projection matrix $\boldsymbol{R}$ is set as $k = [7n^{1/3}]$, the approximated power functions for a given combination of (n, p) and given random projection matrix $\boldsymbol{R}_i$ of column size $[7n^{1/3}]$, $i = 1, 2, 3, 4, 5$, of the testing class $\mathcal{W}(0.04, 2)$ are quite close to each other. The similarities among the approximated theoretical power functions along with Table 6.3 together deliver the message that for reasonable number of k's, the powers are stable among different random projects.

We observe that as both n and p increase in the testing class $\mathcal{W}(0.04, 2)$, the new test has its power approaching 1. We also report the maximum element-wise difference in Figure 6.5, which is seen relatively stable even as both n and p getting larger. This means that our test can still achieve promising power with high-dimensional data in a setting with weak maximum discrepancy between $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, thanks to the aggregation of evidence from high-dimensional data.

6.2 Simulation for Combined Tests with Multiple Random Projections

In this section, we conduct simulation for combined tests with multiple random projections using the four models introduced at the beginning of Chapter 6 to investigate additive, dense, and weak signals as well as rotational transformation to further check

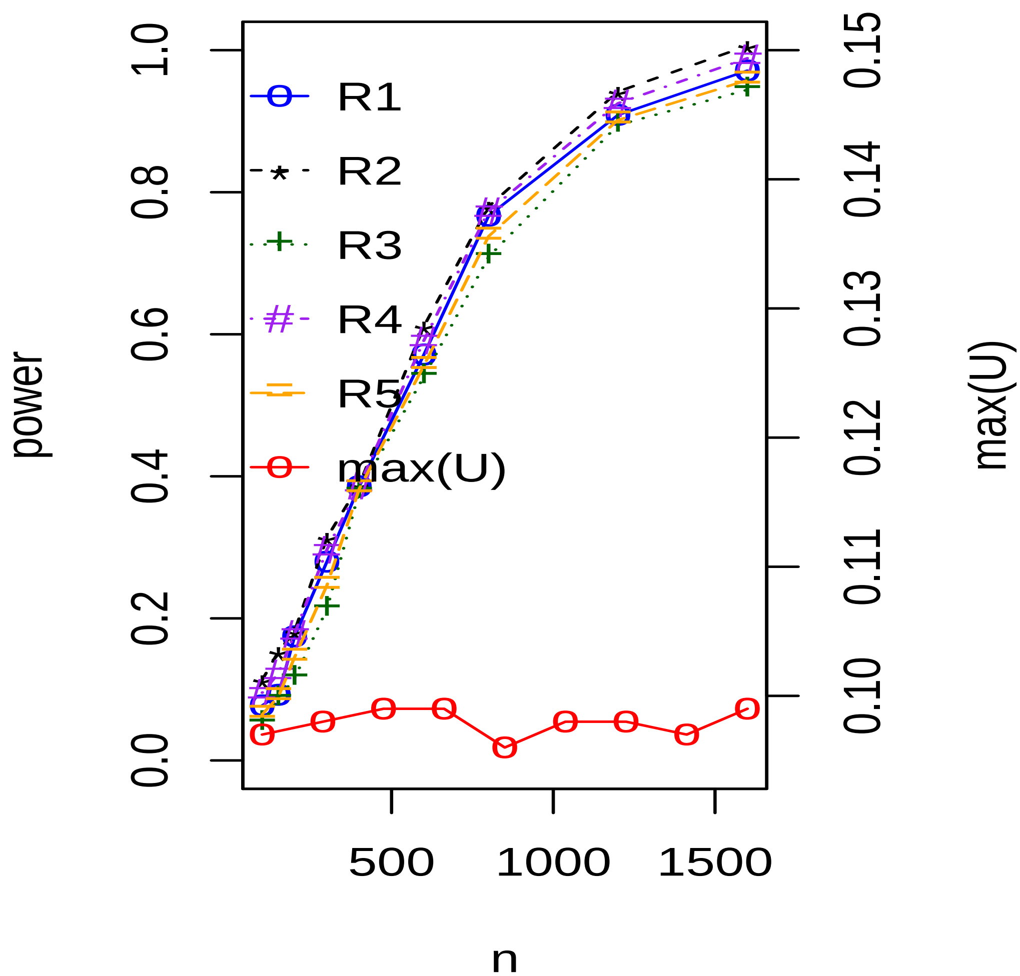

Figure 6.5: Numerical evaluation of the theoretical power functions when $k = [7n^{1/3}]$ under Model 1 in $\mathcal{W}(0.04, 2)$ for testing with a single random projection.

power performance. A numerical approximation example of the Fisher's method is also provided in Section 6.2.3.

6.2.1 Examples with Additive, Dense and Weak Signals

For corrected likelihood ratio testing with multiple random projections, simulation needs to be evaluated for different combinations of n, p, k, and r, as guided by the theory in Theorem 5.7; we choose $k = [7n^{1/3}]$.

Our method is implemented as described in Section 4.3. For a given number of multiple random projections r, we compute r corrected likelihood ratio statistics in (4.1-3), then we combine p-values using the methods of meta-analysis; they are respectively denoted by T_{F} for the Fisher's method, T_{S} for the Stouffer's method, T_{min} for the minimum method, T_{max} for the maximum method. Specifically, when $r = 1$, T_{F}, T_{S}, T_{min} and T_{max} are the same, which is the test with a single random projection, denoted by T_1 later in the simulation of rotational transformations. We also implement the test of T. Cai et al. (2013), denoted by T_{clx} for comparison purpose.

Results of this simulation study are reported in Table 6.4. The four segments of Table 6.4 contain simulation results corresponding to the four models introduced at the beginning of Chapter 6. First, we observe that proposed methods for combining tests from multiple random projections can maintain their size at a satisfactory level, indicating the consistency of the combined test. Second, we find that the combined tests have better power than those based on a single random projection. For example, when $(n, p) = (200, 1500)$, simulations demonstrate that combining tests with $r = 12$ random projections improves the power from 12.6% when $r = 1$ to 35.9% with the Fisher's method. Third, the Fisher's method has the best overall power. As a designated approach for detecting sparse signals, the method using T_{clx} can only achieve the power at around 10%, demonstrating its inadequacy in the setting of

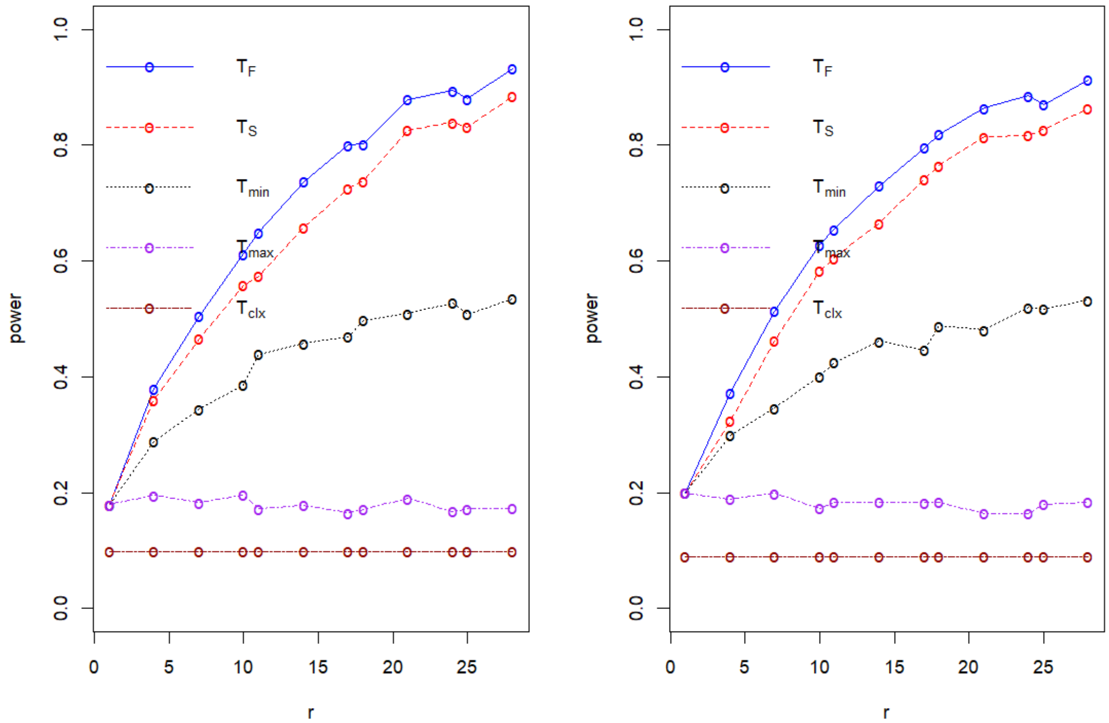

Figure 6.6: Graphs of divergence between powers of combined test from multiple random projections and that of T_{clx} for Model 1 to the left, and for Model 2 to the right, $(n, p) = (400, 3000)$, $k = [7n^{1/3}]$.

additive, dense, and weak signals. The left part of Figure 6.6 is the accompanying figure for the scenario of $(n, p) = (400, 3000)$ in the first part of Table 6.4, which depicts the divergence of our proposed testing procedure with T_{clx} in terms of power comparison. Model 2, which corresponds to the scenario of $(n, p) = (400, 3000)$ is plotted in the right part of Figure 6.6. Figure 6.7 visualizes the same message for Model 3 and Model 4 corresponding to the scenario of $(n, p) = (400, 3000)$.

6.2.2 Simulations with Rotational Transformations

In this study, we consider two cases for $\mathbf{\Sigma}_1$ respectively being $\mathbf{\Sigma}_1 = 0.9\boldsymbol{I} + 0.1 \times \mathbf{1}\mathbf{1}^T$, and $\mathbf{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$. We set $n_1 = n_2 = n = 200$, $p = 1500$, $k =$

Table 6.4: Empirical percentage of rejecting H_0 when $a = 0$ (empirical size) and when $a = 0.1$ (empirical power) for combined test from multiple random projections: $k = [7n^{1/3}]$, $\alpha = 0.05$.

(n,p)	(200,1500)						(300,2250)						(400,3000)					
r	1		6		12		1		10		20		1		14		28	
a	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1
Empirical sizes and powers in Model 1																		
T_{F}	5.8	12.6	5.7	30.0	4.4	35.9	5.3	15.4	4.7	45.6	4.9	70.3	4.8	17.7	6.2	75.1	5.8	91.1
T_{S}	5.8	12.6	4.9	25.6	4.6	30.7	5.3	15.4	5.3	39.2	5.5	62.1	4.8	17.7	5.8	68.4	5.5	86.8
$T_{\min}$	5.8	12.6	6.5	22.0	5.2	26.5	5.3	15.4	5.4	32.2	5.4	36.9	4.8	17.7	5.7	45.4	6.6	53.1
$T_{\max}$	5.8	12.6	5.2	13.0	5.0	11.5	5.3	15.4	5.7	13.4	4.2	13.8	4.8	17.7	4.9	18.8	6.2	17.0
T_{clx}	4.2	7.0	4.2	7.0	4.2	7.0	4.1	9.0	4.1	9.0	4.1	9.0	5.1	9.9	5.1	9.9	5.1	9.9
Empirical sizes and powers in Model 2																		
T_{F}	4.4	14.7	6.1	26.8	6.1	40.0	5.8	15.7	5.9	52.8	5.7	69.9	6.4	21.8	5.3	50.9	5.5	93.6
T_{S}	4.4	14.7	5.6	23.9	5.3	33.5	5.8	15.7	5.9	45.1	6.3	62.3	6.4	21.8	5.0	47.0	4.9	88.0
$T_{\min}$	4.4	14.7	5.6	21.7	5.9	25.6	5.8	15.7	6.3	35.0	5.6	40.9	6.4	21.8	5.7	37.7	4.8	52.2
$T_{\max}$	4.4	14.7	5.4	10.8	4.8	9.1	5.8	15.7	4.1	15.3	6.1	13.6	6.4	21.8	5.7	18.1	5.3	18.9
T_{clx}	5.1	9.7	5.1	9.7	5.1	9.7	3.1	9.0	3.1	9.0	3.1	9.0	3.7	9.7	3.7	9.7	3.7	9.7
Empirical sizes and powers in Model 3																		
T_{F}	5.0	12.4	5.7	28.2	6.1	40.9	4.0	16.8	4.7	48.6	5.2	71.5	5.0	19.0	6.9	73.0	5.5	91.4
T_{S}	5.0	12.4	5.5	24.4	5.6	33.2	4.0	16.8	4.8	43.9	5.5	65.6	5.0	19.0	6.9	68.2	5.9	87.2
$T_{\min}$	5.0	12.4	5.6	21.8	5.8	25.7	4.0	16.8	5.3	32.8	6.2	40.8	5.0	19.0	6.6	46.2	6.2	54.8
$T_{\max}$	5.0	12.4	5.0	10.9	5.8	11.4	4.0	16.8	4.8	12.7	5.1	13.0	5.0	19.0	5.7	21.3	3.9	16.9
T_{clx}	3.2	7.6	3.2	7.6	3.2	7.6	3.8	8.3	3.8	8.3	3.8	8.3	5.4	8.4	5.4	8.4	5.4	8.4
Empirical sizes and powers in Model 4																		
T_{F}	4.7	13.5	6.3	27.4	5.9	40.3	5.4	18.7	5.6	51.3	5.3	75.4	5.6	18.8	6.6	74.6	6.5	91.5
T_{S}	4.7	13.5	6.2	24.6	5.2	34.1	5.4	18.7	5.6	44.6	5.0	66.8	5.6	18.8	5.7	67.3	5.5	86.8
$T_{\min}$	4.7	13.5	6.1	19.6	5.4	25.9	5.4	18.7	4.9	32.5	5.1	36.1	5.6	18.8	6.5	45.7	5.8	53.1
$T_{\max}$	4.7	13.5	6.2	11.2	3.3	10.5	5.4	18.7	5.0	14.0	4.4	15.6	5.6	18.8	4.7	18.0	5.9	16.7
T_{clx}	4.7	7.7	4.7	7.7	4.7	7.7	3.9	8.3	3.9	8.3	3.9	8.3	4.5	8.4	4.5	8.4	4.5	8.4

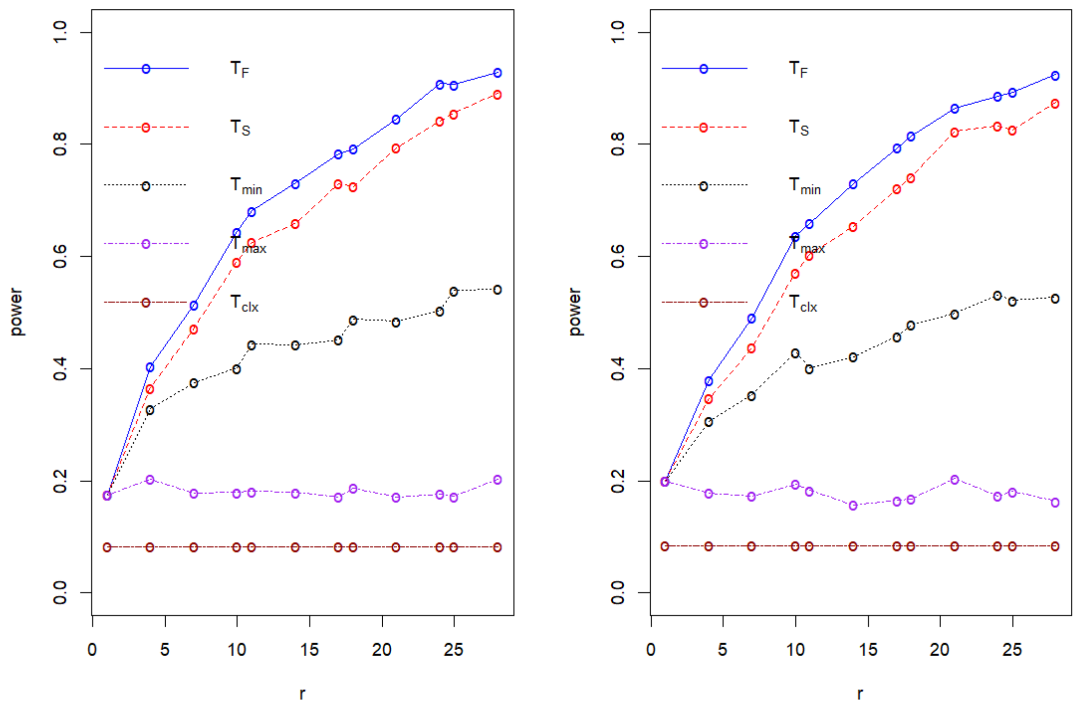

Figure 6.7: Graphs of divergence between powers of combined test from multiple random projections and that of T_{clx} for Model 3 to the left, and for Model 4 to the right, $(n, p) = (400, 3000)$, $k = [7n^{1/3}]$.

$[7n^{1/3}]$, $r = 12$, and compare the powers with two settings of $\theta = 0.01\pi$ and $\theta = 0.03\pi$. Given the value of θ, we investigate the power performance with varying $d = 0, 0.03, \ldots, 0.21, 0.24$. The same sets of tests as those in Section 4.3 are implemented in this study. Meanwhile, as simulations also showed the superiority of the Fisher's method, we only plot the results of the Fisher's method T_{F}, the single projected test T_1, and T_{clx}.

Results of this study are reported in Figure 6.8. In the upper part of Figure 6.8, we plot the case that rotational transformation based on $\boldsymbol{\Sigma}_1 = 0.9\boldsymbol{I} + 0.1 \times \mathbf{1}\mathbf{1}^T$. When $\theta = 0.01\pi$, our proposed testing procedure with the Fisher's method can reach the power of 95.4% when d increases to 0.21 compared with single projected test statistic, namely, T_1, having power of 30.3%. When $\theta = 0.03\pi$, our proposed testing procedure with the Fisher's method can reach the power of 96.1% when d increases to 0.21 compared with single projected test statistic, namely, T_1, having a power of 34.6%. For the test developed by T. Cai et al. (2013), when $\theta = 0.01\pi$, it can reach the power of 6% when $d = 0.24$; when $\theta = 0.03\pi$, it can reach the power of 15% when $d = 0.24$. In the lower part of Figure 6.8, we plot the case wherein rotational transformation is based on $\boldsymbol{\Sigma}_1 = 0.8\boldsymbol{I} + 0.2 \times \mathbf{1}\mathbf{1}^T$. When $\theta = 0.01\pi$, our proposed testing procedure with the Fisher's method can reach the power of 94.3% when d increases to 0.21. In contrast, the single projected test statistic, namely, T_1, only has a power of 33.4%. When $\theta = 0.03\pi$, our testing procedure with the Fisher's method can reach powers as high as 98.5%, while the single projected test statistic, namely, T_1, has a power of about 37.9%. The test from T. Cai et al. (2013), when $\theta = 0.01\pi$, it can reach a power of 16% when $d = 0.24$; when $\theta = 0.03\pi$, it can reach the power of 24% when $d = 0.24$.

From the simulations, we conclude that the proposed method of combined test from multiple random projections has promising performance when testing alternatives with weak but dense signals.

6.2.3 Numerical Approximation to Power Function of Combined Tests from Multiple Random Projections

In this section, we utilize the numerical approximation introduced in Section 5.1.3 as well as Section 5.2.3 to depict the power function of combined test from multiple random projections choosing the Fisher's method as an example.

In particular, we consider Model 1 introduced at the beginning of Chapter 6 when getting $\boldsymbol{\Sigma}_1$, and let $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_1 + \boldsymbol{U}$. We consider two testing classes. In the first targeted testing class, we consider $\mathcal{W}(0.4, 7.5)$. In particular, we set $(n, p) = (200, 1500)$, $(228, 1710)$, $(256, 1920)$, $(284, 2130)$, $(312, 2340)$, $(340, 2550)$, $(368, 2760)$, $(394, 2970)$, $(424, 3180)$, and then numerically evaluate power of the Fisher's method when $r = 1, 2, 4$ using methods described in Section 5.1.3 and Section 5.2.3, where $r = 1$ is power of a single-projected test. In the second targeted testing class, we consider another testing class

$$\mathcal{V}(a) = \left\{ u_{kl} \sim a \times V \times \max_{1 \le j \le p} \sigma_{jj}, p = 424, n = 3180, V \sim \text{Unif}(0, 4) \right\},$$

where $\{\sigma_{jj}\}_{j=1}^{p}$ are diagonal entries of $\boldsymbol{\Sigma}_1$. We set $a = 0.102$, 0.090, 0.078, 0.066, 0.054, 0.042, 0.030, 0.018, 0.006, and then numerically evaluate power of the Fisher's method when $r = 1, 2, 4$. In both cases, we set $k = [7n^{1/3}]$.

The left panel of Figure 6.9 draws results from the first targeted testing class. As p increase from 1500 to 3180, we keep dimension-to-sample size ratio fixed at 7.5 and setting $k = [7n^{1/3}]$. We observe that a single projected test has power

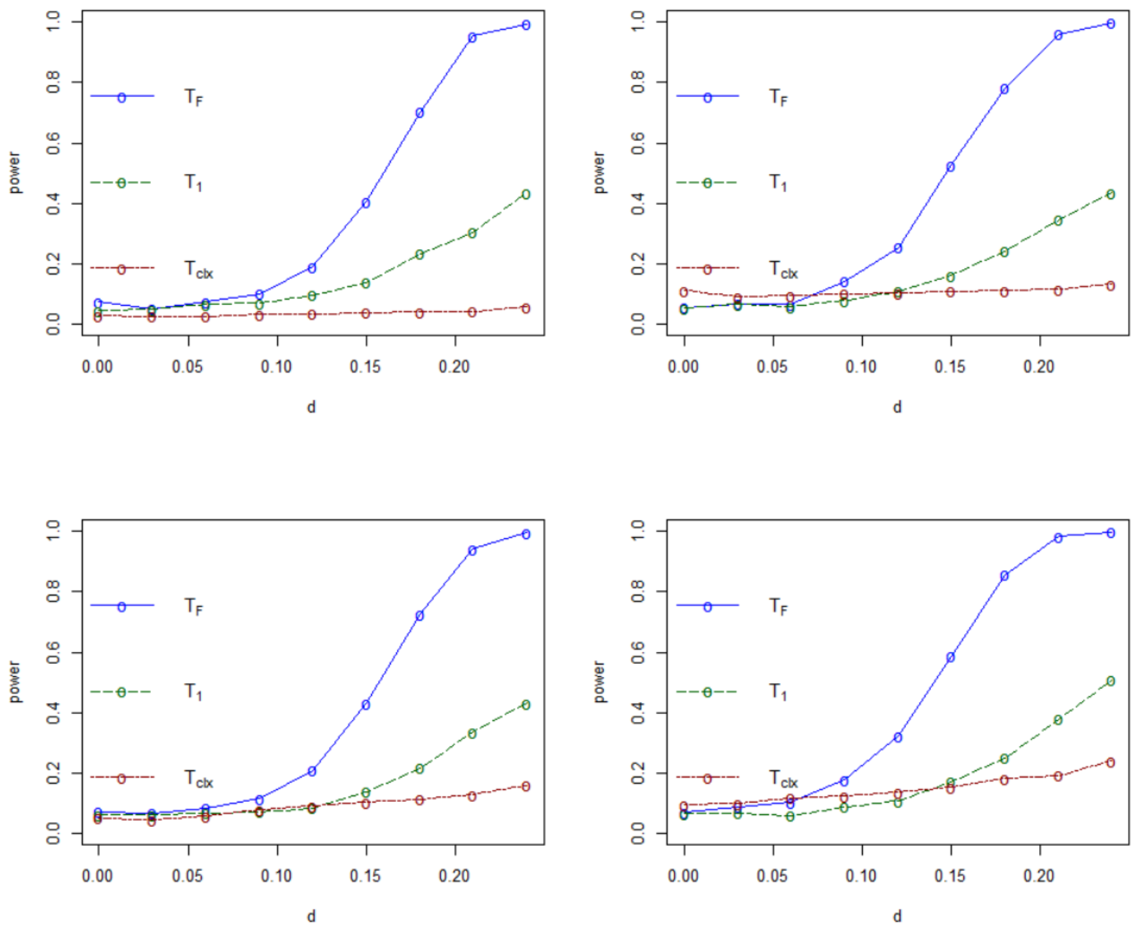

Figure 6.8: Graphs of divergence of powers between combined test from multiple random projections, corrected likelihood ratio test from a single random projection and T_{clx} under rotational transformations: $\Sigma_1 = 0.9I + 0.1 \times 11^T$ when $\theta = 0.01\pi$ to the upper left, and $\theta = 0.03\pi$ to the upper right; $\Sigma_1 = 0.8I + 0.2 \times 11^T$ when $\theta = 0.01\pi$ to the lower left, and $\theta = 0.03\pi$ to the lower right; $\Sigma_2 = \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\} \times \Sigma_1 \times \{\prod_{g=1}^{p/2} G(i_g, j_g, \theta, d)\}^T$; $(n, p) = (200, 1500)$, $k = [7n^{1/3}]$, $r = 12$.

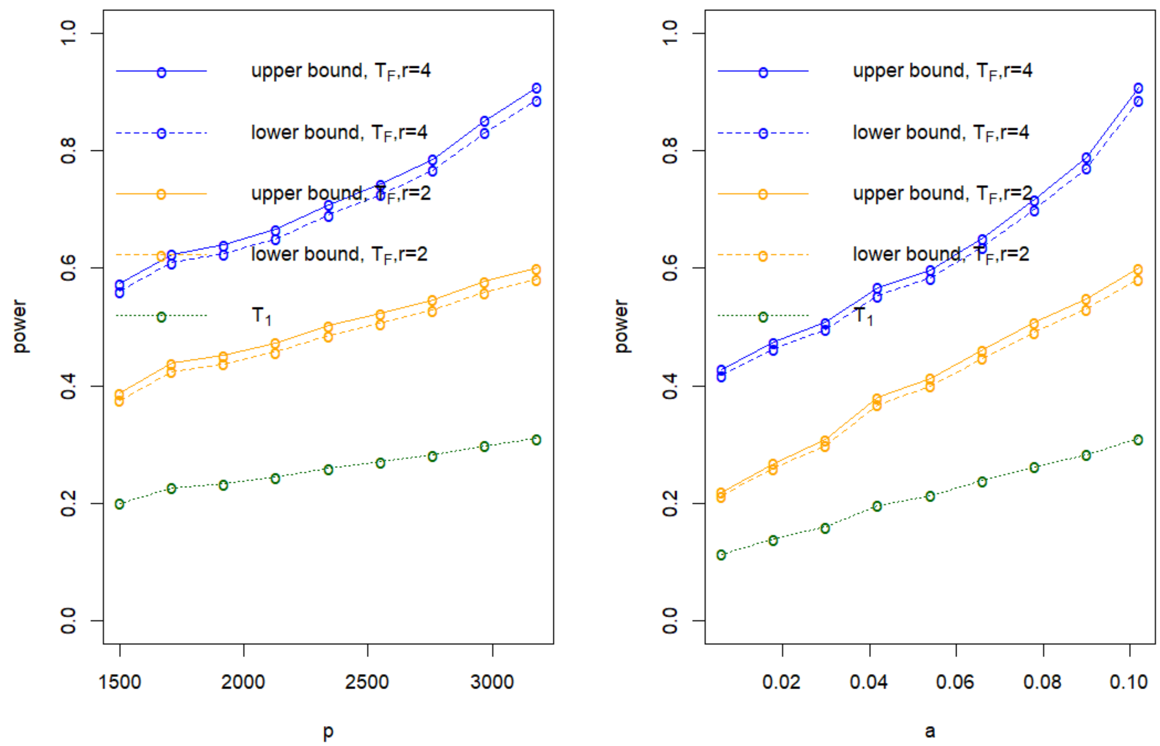

Figure 6.9: Graphs of divergence of the Fisher's method and single projected test for two classes: $\mathcal{W}(0.4, 7.5)$ to the left; $\mathcal{V}(a)$ to the right; $k = [7n^{1/3}]$; $r = 1$, $r = 2$ as well as $r = 4$.

from 20% to 31%. In contrast, the Fisher's method combined twice has power from 38% to 60%, and the Fisher's method combined four times has power from 56% to 90%. The right of Figure 6.9 shows result for the second targeted testing class, and the observations are qualitatively similar. From the comparisons, we see promising evidence of the merits from combining tests with multiple random projections - substantially improved powers.

CHAPTER 7

APPLICATION TO GENETIC DATA

We now present the analytical results of applying the proposed method to the motivating study described in Chapter 2, which consists of two sets of data: (1) copy number measurement data and (2) gene expression data.

7.1 Real Data Analysis for Testing with a Single Random Projection

We apply our testing with a single random projection on the "chr1qseg" data in the R package "highD2pop". The data contains $n_1 = 92$ by $p = 400$ matrix of copy number measurements for 92 long term survivors, and $n_2 = 138$ by $p = 400$ matrix of copy number measurements for 138 short term survivors. Around 3% of the values are missing. We replace the missing copy number measurements at a certain location with average copy number measurements at a certain location for long term and short term survivors respectively. We then apply the corrected test adjusted with the random matrix theory with $k = [7n^{1/3}]$ where $n = \min(n_1, n_2)$ to test the null hypothesis. The value of the test statistic is 8.1194 with a p-value being 4.4409e-16; thus, the hypothesis is rejected. The value of test statistic in T. Cai et al. (2013) is 15.124 with p-value being 0.9989, wihch does not reject the hypothesis. The value of test statistic in Chang et al. (2017) is 3.889 with p-value being 0.1773, which also does not reject the hypothesis. In this case, it is reasonable to reject the null hypothesis because the two covariance matrices are constructed from copy number measurements for short-term and long-term survivors respectively. The hypothesis testing oriented to sparse and relatively strong signals, namely, "CLX" and "CZZW", cannot detect the difference between two covariance matrices in this specific example. But our proposed corrected test adjusted method with the random matrix theory confirms the difference between short-term and long-term survivors, which shows the merit of this method with the random matrix theory in aggregating many less-strong signals.

7.2 Real Data Analysis for Combined Test from Multiple Random Projections

In this study, we consider a small round blue cell tumors (SRBCTs) dataset that is publicly available in R-package "plsgenomics". SRBCTs are four different childhood tumors, namely, Ewing's family of tumors (EWS), neuroblastoma (NB), Burkitt's lymphoma (BL), and rhabdomyosarcoma (RMS), details of which can be found in Khan et al. (2001). The data contains $p = 2308$ gene expression of EWS, NB, BL, and RMS with sample sizes of 29, 11, 18, and 25, respectively. We choose to test the equality between covariance matrices for BL ($n_1 = 18$) and RMS ($n_2 = 25$). To apply our proposed testing procedure from multiple random projections, there are two parameters to be selected. One is the column number of random projection matrix, namely, k: $\boldsymbol{R}_j \in \mathbb{R}^{p\times k}$; the other is the replication of random projections, namely, r: we attempt to combine p-values to improve power over single projected test. Our theoretical analysis in Section 4.2 suggests that $r^{1/2}k^{3/2}p^{-1/2} = o(1)$ could ensure the asymptotic independence to hold under H_0. In particular, we consider $k = [cn^{1/3}]$, where $n = \min(n_1, n_2)$. For this data set, we implement our testing procedure with $rc^3 \ll p/n = 128.22$. We select $c = 2$, namely, $k = [2n^{1/3}]$ and $r = 5$ with our proposed testing procedure with the Fisher's method to combine the likelihood ratio statistics from multiple random projections. Consequently, we get five projected corrected likelihood ratio statistics $|Z^{(\boldsymbol{R}_1)}| = 0.0672$, $|Z^{(\boldsymbol{R}_2)}| = 2.5436$, $|Z^{(\boldsymbol{R}_3)}| = 4.8885$, $|Z^{(\boldsymbol{R}_4)}| = 3.2916$, and $|Z^{(\boldsymbol{R}_5)}| = 1.5094$ with p-values being 0.9464, 0.011, 0.000001, 0.001, 0.1312. For the test of T. Cai et al. (2013), the test statistic is $T_{\mathrm{clx}} = 38.54$ with p-value being 0.002. For the combined test from multiple random projections with the Fisher's method as a test statistic, we obtain T_{F} as a combined p-value less than 0.00001, indicating a stronger evidential outcome from the combined test.

BIBLIOGRAPHY

Achlioptas, D. (2001). Database-friendly random projections. In *Proceedings of the twentieth acm sigmod-sigact-sigart symposium on principles of database systems* (pp. 274–281).

Anderson, T. W. (1962). *An introduction to multivariate statistical analysis* (Tech. Rep.). Wiley New York.

Bai, Z., Jiang, D., Yao, J.-F., Zheng, S., et al. (2009). Corrections to lrt on large-dimensional covariance matrix by rmt. *The Annals of Statistics*, *37*(6B), 3822–3840.

Bai, Z., & Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices* (Vol. 20). Springer.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, *160*(901), 268–282.

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh acm sigkdd international conference on knowledge discovery and data mining* (pp. 245–250).

Cai, T., Liu, W., & Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, *108*(501), 265–277.

Cai, T. T. (2017). Global testing and large-scale multiple testing for high-dimensional covariance structures. *Annual Review of Statistics and Its Application*, *4*, 423–446.

Cai, T. T., Ma, Z., et al. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli*, *19*(5B), 2359–2388.

Chang, J., Qiu, Y., Yao, Q., & Zou, T. (2018). Confidence regions for entries of a large precision matrix. *Journal of Econometrics*, *206*(1), 57–82.

Chang, J., Zhou, W., Zhou, W.-X., & Wang, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics*, *73*(1), 31–41.

Chen, S. X., Zhang, L.-X., & Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, *105*(490), 810–819.

Chernozhukov, V., Chetverikov, D., Kato, K., et al. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, *41*(6), 2786–2819.

Dobriban, E. (2015). Efficient computation of limit spectra of sample covariance matrices. *Random Matrices: Theory and Applications*, *4*(04), 1550019.

Guo, X. (2020). Difference between $\log\left\{\lambda^{(r_j^*)}\right\}/n$ and $\log\left\{\lambda^{(r_j)}\right\}/n$ is negligible asymptotically. *Technical Report*.

He, J., & Chen, S. X. (2016). Testing super-diagonal structure in high dimensional covariance matrices. *Journal of Econometrics*, *194*(2), 283–297.

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis.* Academic press.

Hu, J., & Bai, Z. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Science China Mathematics*, *59*(12), 2281–2300.

Jiang, D., Jiang, T., & Yang, F. (2012). Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *Journal of Statistical Planning and Inference*, *142*(8), 2241–2256.

Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., ... others (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature medicine*, *7*(6), 673–679.

Li, J., Chen, S. X., et al. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, *40*(2), 908–940.

Lin, R., Liu, Z., Zheng, S., & Yin, G. (2016). Power computation for hypothesis testing with high-dimensional covariance matrices. *Computational Statistics & Data Analysis*, *104*, 10–23.

Lopes, M. E., Jacob, L. J., & Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. *arXiv preprint arXiv:1108.2401*.

Marden, J. I., et al. (1991). Sensitive and sturdy p-values. *The Annals of Statistics*, *19*(2), 918–934.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory* (Vol. 197). John Wiley & Sons.

Owen, A. B. (2009). Karl pearson's meta-analysis revisited. *The annals of statistics*, *37*(6B), 3867–3892.

Qiu, Y., Chen, S. X., et al. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics*, *40*(3), 1285–1314.

Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., & Rao, C. R. (1973). *Linear statistical inference and its applications* (Vol. 2). Wiley New York.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., ... others (2007). Strong association of de novo copy number mutations with autism. *Science*, *316*(5823), 445–449.

Singh, K., Xie, M., Strawderman, W. E., et al. (2005). Combining information from independent sources through confidence distributions. *Annals of statistics*, *33*(1), 159–183.

Singh, K., Xie, M., Strawderman, W. E., et al. (2007). Confidence distribution (cd)–distribution estimator of a parameter. In *Complex datasets and inverse problems* (pp. 132–150). Institute of Mathematical Statistics.

Srivastava, M. S., & Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, *101*(6), 1319–1329.

Srivastava, R., Li, P., & Ruppert, D. (2016). Raptt: An exact two-sample test in high dimensions using random projections. *Journal of Computational and Graphical Statistics*, *25*(3), 954–970.

Vempala, S. S. (2005). *The random projection method* (Vol. 65). American Mathematical Soc.

Wang, C. (2014). Asymptotic power of likelihood ratio tests for high dimensional data. *Statistics & Probability Letters*, *88*, 184–189.

Wu, T.-L., & Li, P. (2020). Projected tests for high-dimensional covariance matrices. *Journal of Statistical Planning and Inference*, *207*, 73–85.

Xie, M., Singh, K., & Strawderman, W. E. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association*, *106*(493), 320–333.

Zhang, Q., Hu, J., Bai, Z., et al. (2019). Invariant test based on the modified correction to lrt for the equality of two high-dimensional covariance matrices. *Electronic Journal of Statistics*, *13*(1), 850–881.

Zheng, S., Bai, Z., Yao, J., et al. (2015). Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Annals of Statistics*, *43*(2), 546–591.

Zheng, S., Bai, Z., Yao, J., et al. (2017). Clt for eigenvalue statistics of large-dimensional general fisher matrices with applications. *Bernoulli*, *23*(2), 1130–1178.

APPENDIX A

PROOF FOR ASYMPTOTIC INDEPENDENCE OF P-VALUES

This section presents the proof related to the asymptotic independence of p-values in Section 4.2 of Chapter 4.

A.1 Proof of Theorem 4.2

Proof.

$$
\begin{aligned}
&\left|p_j - p_j^*\right| \\
=&\left|2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\} - 2\Phi\left\{-\left|Z^{(\boldsymbol{R}_j^*)}\right|\right\}\right| \\
=&2\left|\Phi\left\{-\left|Z^{(\boldsymbol{R}_j)}\right|\right\} - \Phi\left\{-\left|Z^{(\boldsymbol{R}_j^*)}\right|\right\}\right| \\
=&2\left|\Phi\left[-\left|C_1\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N + C_2\right|\right] - \Phi\left[-\left|C_1\log\left\{\lambda^{(\boldsymbol{R}_j^*)}\right\}/N + C_2\right|\right]\right| \\
\leq&2\phi(0)\left|-\left|C_1\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N + C_2\right| + \left|C_1\log\left\{\lambda^{(\boldsymbol{R}_j^*)}\right\}/N + C_2\right|\right| \\
\leq&2\phi(0)\left|C_1\log\left\{\lambda^{(\boldsymbol{R}_j^*)}\right\}/N - C_1\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N\right| \\
=&2\phi(0)\left|C_1\right|\left|\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N - \log\left\{\lambda^{(\boldsymbol{R}_j^*)}\right\}/N\right|.
\end{aligned}
$$

The first inequality is due to definition of cumulative distribution function of standard normal random variable; The second inequality is by triangle inequality.

$$
\begin{aligned}
&\max_{1\leq j\leq r}\left|p_j - p_j^*\right| \\
\leq&\max_{1\leq j\leq r}\left[2\phi(0)\left|C_1\right|\left|\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N - \log\left\{\lambda^{\left(\boldsymbol{R}_j^*\right)}\right\}/N\right|\right] \\
=&2\phi(0)\left|C_1\right|\max_{1\leq j\leq r}\left|\log\left\{\lambda^{(\boldsymbol{R}_j)}\right\}/N - \log\left\{\lambda^{\left(\boldsymbol{R}_j^*\right)}\right\}/N\right| \\
=&O_{\mathbb{P}}\left(r^{1/2}k^{3/2}p^{-1/2}\right).
\end{aligned}
$$

$\square$

APPENDIX B

PROOF OF THEORY OF TESTING WITH A SINGLE RANDOM PROJECTION

This section presents proofs for Section 5.1 in Chapter 5.

B.1 Proof of Theorem 5.5

Proof. Assuming that assumptions 3.1-5.1-5.2 hold, we express $\boldsymbol{X}_i$ as $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{u}_i$, and $\boldsymbol{Y}_i$ as $\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{v}_i$. Conditioning on $\boldsymbol{R}$, $\boldsymbol{R}^T\boldsymbol{u}_i,\ i = 1, \cdots, n_1$, and $\boldsymbol{R}^T\boldsymbol{v}_i,\ i = 1, \cdots, n_2$, both independently follows $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ subject to $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_k$, which validates assumption(a)-(a') in Zheng et al. (2015) with the case $\beta_x = \beta_y = 0$; Assumptions (b)-(c) and assumptions (b')-(c') have also been assumed explicitly here. Then Theorem 4.2 in Zheng et al. (2015) guarantees Theorem 5.5. □

B.2 Proof of Theorem 5.6

Proof. Assuming that assumptions 3.1-5.1-5.4 hold, we express $\boldsymbol{X}_i$ as $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{u}_i$, and $\boldsymbol{Y}_i$ as $\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_2^{1/2}\boldsymbol{v}_i$. Conditioning on $\boldsymbol{R}$, $\boldsymbol{R}^T\boldsymbol{u}_i,\ i = 1, \cdots, n_1$ and $\boldsymbol{R}^T\boldsymbol{v}_i,\ i = 1, \cdots, n_2$ both independently follow $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ subject to $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_k$, which validates assumption(a)-$(b1)$-$(b2)$ in Zheng et al. (2017) with the case $\beta_x = \beta_y = 0$; Assumption (c) in Zheng et al. (2017) has also been assumed explicitly here; Conditioning on $\boldsymbol{R}$, assumption (d) in Zheng et al. (2017) has also been assumed explicitly here. Then Theorem 3.1 in Zheng et al. (2017) guarantees Theorem 5.6. □

APPENDIX C

PROOF OF THEORY OF COMBINED TESTS FROM MULTIPLE RANDOM PROJECTIONS

This section presents proofs for Section 5.2 in Chapter 5.

C.1 Proof of Theorem 5.7

Proof. We firstly show that, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$\mathbb{P}\left\{g_c\left(p_1, \cdots, p_r\right) - g_c\left(p_1^*, \cdots, p_r^*\right) = 0\right\} \to 1.$$

For the Fisher's method,

$$\begin{aligned}
& g_c\left(p_1, \cdots, p_r\right) - g_c\left(p_1^*, \cdots, p_r^*\right) \\
= & \sum_{j=1}^r \log p_j - \sum_{j=1}^r \log p_j^* \\
= & \sum_{j=1}^r \left(\log p_j - \log p_j^*\right) \\
= & \sum_{j=1}^r \log\left(\frac{p_j}{p_j^*}\right) \\
= & \log\left(\prod_{j=1}^r \frac{p_j}{p_j^*}\right).
\end{aligned}$$

From Theorem 4.2, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$p_j - p_j^* \to 0, \text{ for all } j = 1, \cdots, r,$$

with probability tending to 1. That is,

$$\frac{p_j}{p_j^*} \to 1, \text{ for all } j = 1, \cdots, r,$$

with probability tending to 1. Hence,

$$\prod_{j=1}^{r} \frac{p_j}{p_j^*} \to 1,$$

with probability tending to 1. Therefore, for the Fisher's method, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) = \log\left(\prod_{j=1}^{r} \frac{p_j}{p_j^*}\right) \to 0,$$

with probability tending to 1.

For the Stouffer's method,

$$\begin{aligned} & g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) \\ = & \sum_{j=1}^{r} \Phi^{-1}(p_j) - \sum_{j=1}^{r} \Phi^{-1}\left(p_j^*\right) \\ = & \sum_{j=1}^{r} \left\{\Phi^{-1}(p_j) - \Phi^{-1}\left(p_j^*\right)\right\}. \end{aligned}$$

Due to same reason,

$$\Phi^{-1}(p_j) - \Phi^{-1}\left(p_j^*\right) \to 0, \text{ for all } j = 1, \cdots, r,$$

with probability tending to 1. Therefore, for the Stouffer's method, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) = \sum_{j=1}^{r} \left\{\Phi^{-1}(p_j) - \Phi^{-1}\left(p_j^*\right)\right\} \to 0,$$

with probability tending to 1.

For the minimum method, due to same reason, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) = \min_{1\leq j\leq r} p_j - \min_{1\leq j\leq r} p_j^* \to 0,$$

with probability tending to 1.

For the maximum method, due to same reason, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) = \max_{1\leq j\leq r} p_j - \max_{1\leq j\leq r} p_j^* \to 0,$$

with probability tending to 1.

Next, we show that, with $r^{1/2}k^{3/2}p^{-1/2} \to 0$,

$$\mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1, \cdots, p_r)\right\} \to \mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1^*, \cdots, p_r^*)\right\},$$

with probability tending to 1.

Note that

$$\begin{aligned}
&\mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1, \cdots, p_r)\right\} \\
=&\mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1^*, \cdots, p_r^*) + g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*)\right\} \\
=&\mathbb{E}_X\left[\mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1^*, \cdots, p_r^*) + x\right\}\mathbb{P}\left\{g_c(p_1, \cdots, p_r) - g_c(p_1^*, \cdots, p_r^*) = x\right\}\right] \\
\to&\mathbb{P}\left\{g_c(U_1, \cdots, U_r) \leq g_c(p_1^*, \cdots, p_r^*)\right\}.
\end{aligned}$$

since

$$\mathbb{P}\left\{g_c\left(p_1, \cdots, p_r\right) - g_c\left(p_1^*, \cdots, p_r^*\right) = 0\right\} \to 1.$$

Finally, note that

$$\begin{aligned}
&\mathbb{P}\left\{c_*^c(\mathbf{0}) < \alpha\right\} \\
=&\mathbb{P}\left(G_c\left[g_c\left\{c_1(\mathbf{0}), \cdots, c_r(\mathbf{0})\right\}\right] < \alpha\right) \\
=&\mathbb{P}\left[G_c\left\{g_c\left(p_1, \cdots, p_r\right)\right\} < \alpha\right] \\
=&\mathbb{P}\left[\mathbb{P}\left\{g_c\left(U_1, \cdots, U_r\right) \le g_c\left(p_1, \cdots, p_r\right)\right\} < \alpha\right] \\
\to&\mathbb{P}\left[\mathbb{P}\left\{g_c\left(U_1, \cdots, U_r\right) \le g_c\left(p_1^*, \cdots, p_r^*\right)\right\} < \alpha\right] \\
=&\alpha,
\end{aligned}$$

with probability tending to 1. The last equality is because $p_1^*, \cdots, p_r^*$ are independent standard uniform random variables. $\square$

C.2 Proof of Theorem 5.8

Proof.

$$
\begin{aligned}
&\mathbb{P}\left\{c_{\mathrm{F}}^{c}(\boldsymbol{s})<\alpha\right\} \\
=&\mathbb{P}\left(\mathbb{P}\left[\chi_{2r}^{2}\geq -2\sum_{j=1}^{r}\log\left\{c_j(\boldsymbol{s})\right\}\right]<\alpha\right) \\
=&\mathbb{P}\left[-2\sum_{j=1}^{r}\log\left\{c_j(\boldsymbol{s})\right\}>\left(\chi_{2r}^{2}\right)^{-1}(1-\alpha)\right] \\
=&\mathbb{P}\left(\sum_{j=1}^{r}\left[-2\log\left\{c_j(\boldsymbol{s})\right\}\right]>\left(\chi_{2r}^{2}\right)^{-1}(1-\alpha)\right) \\
=&\mathbb{P}\left\{\sum_{j=1}^{r}X_{\mathrm{F},j}>\left(\chi_{2r}^{2}\right)^{-1}(1-\alpha)\right\},
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left(X_{\mathrm{F},j}\leq t\right) \\
=&\mathbb{P}\left[-2\log\left\{c_j(\boldsymbol{s})\right\}\leq t\right] \\
=&\mathbb{P}\left[-2\log\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\right\}\leq t\mid \boldsymbol{\Theta}=\boldsymbol{s}\right] \\
=&\mathbb{P}\left[\log\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\right\}\geq -t/2\mid \boldsymbol{\Theta}=\boldsymbol{s}\right] \\
=&\mathbb{P}\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\geq \exp(-t/2)\mid \boldsymbol{\Theta}=\boldsymbol{s}\right\} \\
=&\mathbb{P}\left\{\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\geq \exp(-t/2)/2\mid \boldsymbol{\Theta}=\boldsymbol{s}\right\} \\
=&\mathbb{P}\left[-\left|Z^{(\boldsymbol{R}_j)}\right|\geq \Phi^{-1}\left\{\exp(-t/2)/2\right\}\mid \boldsymbol{\Theta}=\boldsymbol{s}\right] \\
=&\mathbb{P}\left[\left|Z^{(\boldsymbol{R}_j)}\right|\leq -\Phi^{-1}\left\{\exp(-t/2)/2\right\}\mid \boldsymbol{\Theta}=\boldsymbol{s}\right] \\
=&\mathbb{P}\left[\Phi^{-1}\left\{\exp(-t/2)/2\right\}\leq Z^{(\boldsymbol{R}_j)}\leq -\Phi^{-1}\left\{\exp(-t/2)/2\right\}\mid \boldsymbol{\Theta}=\boldsymbol{s}\right] \\
=&\mathbb{P}\left[-\frac{\mu_j}{\sigma_j}+\frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\leq \frac{Z^{(\boldsymbol{R}_j)}-\mu_j}{\sigma_j}\leq -\frac{\mu_j}{\sigma_j}-\frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right] \\
=&\Phi\left[-\frac{\mu_j}{\sigma_j}-\frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right]-\Phi\left[-\frac{\mu_j}{\sigma_j}+\frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right].
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left\{c_{\mathrm{S}}^{c}(\boldsymbol{s})<\alpha\right\}\\
=&\mathbb{P}\left(\mathbb{P}\left[Z \leq r^{-1/2}\sum_{j=1}^{r}\Phi^{-1}\left\{c_j(\boldsymbol{s})\right\}\right]<\alpha\right)\\
=&\mathbb{P}\left[r^{-1/2}\sum_{j=1}^{r}\Phi^{-1}\left\{c_j(\boldsymbol{s})\right\}<\Phi^{-1}(\alpha)\right]\\
=&\mathbb{P}\left[\sum_{j=1}^{r}\Phi^{-1}\left\{c_j(\boldsymbol{s})\right\}<r^{1/2}\Phi^{-1}(\alpha)\right]\\
=&\mathbb{P}\left[\sum_{j=1}^{r}X_{\mathrm{S},j}<r^{1/2}\Phi^{-1}(\alpha)\right],
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left(X_{\mathrm{S},j}\geq t\right)\\
=&\mathbb{P}\left[\Phi^{-1}\left\{c_j(\boldsymbol{s})\right\}\geq t\right]\\
=&\mathbb{P}\left[\Phi^{-1}\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\right\}\geq t \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&\mathbb{P}\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\geq \Phi(t) \mid \boldsymbol{\Theta}=\boldsymbol{s}\right\}\\
=&\mathbb{P}\left\{\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\geq \Phi(t)/2 \mid \boldsymbol{\Theta}=\boldsymbol{s}\right\}\\
=&\mathbb{P}\left[-\left|Z^{(\boldsymbol{R}_j)}\right|\geq \Phi^{-1}\left\{\Phi(t)/2\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&\mathbb{P}\left[\left|Z^{(\boldsymbol{R}_j)}\right|\leq -\Phi^{-1}\left\{\Phi(t)/2\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&\mathbb{P}\left[\Phi^{-1}\left\{\Phi(t)/2\right\}\leq Z^{(\boldsymbol{R}_j)}\leq -\Phi^{-1}\left\{\Phi(t)/2\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&\mathbb{P}\left[-\frac{\mu_j}{\sigma_j}+\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\leq \frac{Z^{(\boldsymbol{R}_j)}-\mu_j}{\sigma_j}\leq -\frac{\mu_j}{\sigma_j}-\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\right]\\
=&\Phi\left[-\frac{\mu_j}{\sigma_j}-\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\right]-\Phi\left[-\frac{\mu_j}{\sigma_j}+\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma_j}\right].
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left\{c^c_{\min}(\boldsymbol{s}) < \alpha\right\}\\
=&\mathbb{P}\left(\mathbb{P}\left[U_{(1)} \le \min_{1\le j\le r}\{c_j(\boldsymbol{s})\}\right] < \alpha\right)\\
=&\mathbb{P}\left[\min_{1\le j\le r}\{c_j(\boldsymbol{s})\} < F^{-1}_{U_{(1)}}(\alpha)\right]\\
=&1-\mathbb{P}\left[\max_{1\le j\le r}\{c_j(\boldsymbol{s})\} \ge F^{-1}_{U_{(1)}}(\alpha)\right]\\
=&1-\mathbb{P}\left[\max_{1\le j\le r}\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\right\} \ge F^{-1}_{U_{(1)}}(\alpha) \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&1-\mathbb{P}\left\{2\max_{1\le j\le r}\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) \ge 1-(1-\alpha)^{1/r} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right\}\\
=&1-\mathbb{P}\left\{\max_{1\le j\le r}\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) \ge \frac{1-(1-\alpha)^{1/r}}{2} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right\}\\
=&1-\mathbb{P}\left[\max_{1\le j\le r}\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) \ge \Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&1-\mathbb{P}\left[\min_{1\le j\le r}\left|Z^{(\boldsymbol{R}_j)}\right| \le -\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&1-\prod_{j=1}^{r}\mathbb{P}\left[\left|Z^{(\boldsymbol{R}_j)}\right| \le -\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&1-\prod_{j=1}^{r}\mathbb{P}\left[\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \le Z^{(\boldsymbol{R}_j)} \le -\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \mid \boldsymbol{\Theta}=\boldsymbol{s}\right]\\
=&1-\prod_{j=1}^{r}\mathbb{P}\left[-\frac{\mu_j}{\sigma_j}+\frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\} \le \frac{Z^{(\boldsymbol{R}_j)}-\mu_j}{\sigma_j} \le -\frac{\mu_j}{\sigma_j}-\frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\}\right]\\
=&1-\prod_{j=1}^{r}\left(\Phi\left[-\frac{\mu_j}{\sigma_j}-\frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\}\right]-\Phi\left[-\frac{\mu_j}{\sigma_j}+\frac{1}{\sigma_j}\Phi^{-1}\left\{\frac{1-(1-\alpha)^{1/r}}{2}\right\}\right]\right).
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{P}\left\{c^c_{\max}(\boldsymbol{s}) < \alpha\right\}\\
=&\mathbb{P}\left(\mathbb{P}\left[U_{(r)} \leq \max_{1\leq j\leq r}\left\{c_j(\boldsymbol{s})\right\}\right] < \alpha\right)\\
=&\mathbb{P}\left[\max_{1\leq j\leq r}\left\{c_j(\boldsymbol{s})\right\} < F^{-1}_{U_{(r)}}(\alpha)\right]\\
=&\mathbb{P}\left[\max_{1\leq j\leq r}\left\{2\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right)\right\} < F^{-1}_{U_{(r)}}(\alpha) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right]\\
=&\mathbb{P}\left\{2\max_{1\leq j\leq r}\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) < \alpha^{1/r} \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\\
=&\mathbb{P}\left\{\max_{1\leq j\leq r}\Phi\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) < \frac{\alpha^{1/r}}{2} \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\\
=&\mathbb{P}\left\{\max_{1\leq j\leq r}\left(-\left|Z^{(\boldsymbol{R}_j)}\right|\right) < \Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\\
=&\mathbb{P}\left\{\min_{1\leq j\leq r}\left|Z^{(\boldsymbol{R}_j)}\right| > -\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\\
=&\prod_{j=1}^{r}\mathbb{P}\left\{\left|Z^{(\boldsymbol{R}_j)}\right| > -\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\\
=&\prod_{j=1}^{r}\left[1 - \mathbb{P}\left\{\left|Z^{(\boldsymbol{R}_j)}\right| \leq -\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\right]\\
=&\prod_{j=1}^{r}\left[1 - \mathbb{P}\left\{\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \leq Z^{(\boldsymbol{R}_j)} \leq -\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \mid \boldsymbol{\Theta} = \boldsymbol{s}\right\}\right]\\
=&\prod_{j=1}^{r}\left[1 - \mathbb{P}\left\{-\frac{\mu_j}{\sigma_j} + \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right) \leq \frac{Z^{(\boldsymbol{R}_j)} - \mu_j}{\sigma_j} \leq -\frac{\mu_j}{\sigma_j} - \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right)\right\}\right]\\
=&\prod_{j=1}^{r}\left[1 - \Phi\left\{-\frac{\mu_j}{\sigma_j} - \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right)\right\} + \Phi\left\{-\frac{\mu_j}{\sigma_j} + \frac{1}{\sigma_j}\Phi^{-1}\left(\frac{\alpha^{1/r}}{2}\right)\right\}\right].
\end{aligned}
$$

$\square$

Proof of Theorem 5.9

Proof. From Theorem 5.8,

$$\mathbb{P}\left(X_{\mathrm{F},j} > t\right) = 1 - \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right] + \Phi\left[-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right].$$

Re-ordering $X_{\mathrm{F},j}$, $j = 1, \cdots, r$, stochastically as

$$X_{\mathrm{F},(1)} \preceq \cdots \preceq X_{\mathrm{F},(r)},$$

where $\prec$ denotes the stocahstic order, with the following associative "one minus cumulative distribution functions",

$$\begin{aligned}
&1 - \Phi\left[-\frac{\mu^{(1)}}{\sigma^{(1)}} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma^{(1)}}\right] + \Phi\left[-\frac{\mu^{(1)}}{\sigma^{(1)}} + \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma^{(1)}}\right] \\
\leq &\cdots \\
\leq &1 - \Phi\left[-\frac{\mu^{(r)}}{\sigma^{(r)}} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma^{(r)}}\right] + \Phi\left[-\frac{\mu^{(r)}}{\sigma^{(r)}} + \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma^{(r)}}\right].
\end{aligned}$$

From Theorem 5.8,

$$
\begin{aligned}
&\mathbb{P}\left\{c_{\mathrm{F}}^{c}(\boldsymbol{s}) < \alpha\right\} \\
=&\mathbb{P}\left\{\sum_{j=1}^{r} X_{\mathrm{F},j} > \left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\} \\
=&\mathbb{P}\left\{\frac{1}{r}X_{\mathrm{F},1} + \cdots + \frac{1}{r}X_{\mathrm{F},r} > r^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\} \\
\geq&\mathbb{P}\left\{\frac{1}{r}X_{\mathrm{F},(1)} + \cdots + \frac{1}{r}X_{\mathrm{F},(1)} > r^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\} \\
=&\mathbb{P}\left\{X_{\mathrm{F},(1)} > r^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\} \\
=&1-\Phi\left(-\frac{\mu^{(1)}}{\sigma^{(1)}} - \frac{\Phi^{-1}\left[\exp\left\{-(2r)^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\}/2\right]}{\sigma^{(1)}}\right) \\
&+\Phi\left(-\frac{\mu^{(1)}}{\sigma^{(1)}} + \frac{\Phi^{-1}\left[\exp\left\{-(2r)^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\}/2\right]}{\sigma^{(1)}}\right) \\
>&1-\Phi\left\{-\frac{\mu^{(1)}}{\sigma^{(1)}} - \frac{\Phi^{-1}(\alpha/2)}{\sigma^{(1)}}\right\} + \Phi\left\{-\frac{\mu^{(1)}}{\sigma^{(1)}} + \frac{\Phi^{-1}(\alpha/2)}{\sigma^{(1)}}\right\} \\
=&\min_{1\leq j\leq r}\left[1-\Phi\left\{-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\} + \Phi\left\{-\frac{\mu_j}{\sigma_j} + \frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\}\right] \\
=&\min_{1\leq j\leq r}\mathbb{P}\left\{c_j(\boldsymbol{s}) < \alpha\right\}.
\end{aligned}
$$

The first inequality is due to stochastic order between independent random variables; The second inequality is due to the fact that $\exp\left\{-(2r)^{-1}\left(\chi^2_{2r}\right)^{-1}(1-\alpha)\right\} > \alpha$.

From Theorem 5.8,

$$
\mathbb{P}\left(X_{\mathrm{S},j} \geq t\right) = \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right] + \Phi\left[-\frac{\mu_j}{\sigma_j} - \frac{\Phi^{-1}\left\{\exp(-t/2)/2\right\}}{\sigma_j}\right].
$$

Re-ordering $X_{\mathrm{S},j}$, $j = 1, \cdots, r$, stochastically as

$$
X_{\mathrm{S},(1)} \preceq \cdots \preceq X_{\mathrm{S},(r)},
$$

with the following associative “one minus cumulative distribution functions’,

$$
\begin{aligned}
&\Phi\left[-\frac{\mu^{(r)}}{\sigma^{(r)}}-\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma^{(r)}}\right]-\Phi\left[-\frac{\mu^{(r)}}{\sigma^{(r)}}+\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma^{(r)}}\right]\\
\leq&\cdots\\
\leq&\Phi\left[-\frac{\mu^{(1)}}{\sigma^{(1)}}-\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma^{(1)}}\right]-\Phi\left[-\frac{\mu^{(1)}}{\sigma^{(1)}}+\frac{\Phi^{-1}\left\{\Phi(t)/2\right\}}{\sigma^{(1)}}\right].
\end{aligned}
$$

From Theorem 5.8,

$$
\begin{aligned}
&\mathbb{P}\left\{c_{\mathrm{S}}^{c}(\boldsymbol{s})<\alpha\right\}\\
=&\mathbb{P}\left\{\sum_{j=1}^{r}X_{\mathrm{S},j}<r^{1/2}\Phi^{-1}(\alpha)\right\}\\
=&\mathbb{P}\left\{\frac{1}{r}X_{\mathrm{S},1}+\cdots+\frac{1}{r}X_{\mathrm{S},r}<r^{-1/2}\Phi^{-1}(\alpha)\right\}\\
\geq&\mathbb{P}\left\{\frac{1}{r}X_{\mathrm{S},(r)}+\cdots+\frac{1}{r}X_{\mathrm{S},(r)}<r^{-1/2}\Phi^{-1}(\alpha)\right\}\\
=&\mathbb{P}\left\{X_{\mathrm{S},(r)}<r^{-1/2}\Phi^{-1}(\alpha)\right\}\\
=&1-\Phi\left(-\frac{\mu^{(1)}}{\sigma^{(1)}}-\frac{\Phi^{-1}\left[\Phi\left\{r^{-1/2}\Phi^{-1}(\alpha)\right\}/2\right]}{\sigma^{(1)}}\right)\\
&+\Phi\left(-\frac{\mu^{(1)}}{\sigma^{(1)}}+\frac{\Phi^{-1}\left[\Phi\left\{r^{-1/2}\Phi^{-1}(\alpha)\right\}/2\right]}{\sigma^{(1)}}\right)\\
>&1-\Phi\left\{-\frac{\mu^{(1)}}{\sigma^{(1)}}-\frac{\Phi^{-1}(\alpha/2)}{\sigma^{(1)}}\right\}+\Phi\left\{-\frac{\mu^{(1)}}{\sigma^{(1)}}+\frac{\Phi^{-1}(\alpha/2)}{\sigma^{(1)}}\right\}\\
=&\min_{1\leq j\leq r}\left[1-\Phi\left\{-\frac{\mu_j}{\sigma_j}-\frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\}+\Phi\left\{-\frac{\mu_j}{\sigma_j}+\frac{\Phi^{-1}(\alpha/2)}{\sigma_j}\right\}\right]\\
=&\min_{1\leq j\leq r}\mathbb{P}\left\{c_j(\boldsymbol{s})<\alpha\right\}.
\end{aligned}
$$

The first inequality is due to stochastic order between independent random variables; The second inequality is due to the fact that $\Phi\left\{r^{-1/2}\Phi^{-1}(\alpha)\right\}>\alpha$. □