Statistica Sinica

Least Favorable Direction Test for Multivariate Analysis of Variance in High Dimension

Rui Wang, Xingzhong Xu

Beijing Institute of Technology

Abstract: This paper considers the problem of multivariate analysis of variance for normal samples in the high dimension medium sample size setting. When the sample dimension is larger than the sample size, the classical likelihood ratio test is not defined since the likelihood function is unbounded. Based on the unboundedness of the likelihood function, we propose a new test called least favorable direction test. The asymptotic distributions of the test statistic are derived under both nonspiked and spiked covariances. The local asymptotic power function of the test is also given. The asymptotic power function and simulations show that the proposed test is particularly powerful under spiked covariance.

Key words and phrases: High dimensional data, least favorable direction test, multivariate analysis of variance, principal component analysis, spiked covariance.

1. Introduction

Suppose there are k ($k \geq 2$) independent samples of p-dimensional data. Within the ith sample ($1 \leq i \leq k$), the observations $\{X_{ij}\}_{j=1}^{n_i}$ are independent and identically distributed (iid) as $\mathcal{N}_p(\theta_i, \mathbf{\Sigma})$, the p-dimensional normal distribution with mean vector θ_i and common variance matrix $\mathbf{\Sigma}$. We would like to test the hypotheses

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_k \quad \text{v.s.} \quad H_1 : \theta_i \neq \theta_j \text{ for some } i \neq j. \tag{1.1}$$

This testing problem is known as one-way multivariate analysis of variance (MANOVA) and has been well studied when p is small compared with N, where $N = \sum_{i=1}^{k} n_i$ is the total sample size.

Let $\mathbf{H} = \sum_{i=1}^{k} n_i(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})(\bar{\mathbf{X}}_i - \bar{\mathbf{X}})^\top$ be the sum-of-squares between groups and $\mathbf{G} = \sum_{i=1}^{k} \sum_{j=1}^{n_i}(X_{ij} - \bar{\mathbf{X}}_i)(X_{ij} - \bar{\mathbf{X}}_i)^\top$ be the sum-of-squares within groups, where $\bar{\mathbf{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ is the sample mean of group i and $\bar{\mathbf{X}} = N^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}$ is the pooled sample mean. There are four classical test statistics for hypotheses (1.1), which are all based on the eigenvalues of $\mathbf{HG}^{-1}$.

Wilks' Lambda:	$\lvert\mathbf{G} + \mathbf{H}\rvert / \lvert\mathbf{G}\rvert$
Pillai trace:	$\mathrm{tr}[\mathbf{H}(\mathbf{G} + \mathbf{H})^{-1}]$
Hotelling-Lawley trace:	$\mathrm{tr}[\mathbf{HG}^{-1}]$
Roy's maximum root:	$\lambda_1(\mathbf{HG}^{-1})$

In some modern scientific applications, people would like to test hypotheses (1.1) in high dimensional setting, i.e., p is greater than N. See, e.g., Verstynen et al. (2005) and Tsai and Chen (2009). However, when $p \geq N$, the four classical test statistics are all not defined. Researchers have done extensive work to study the testing problem (1.1) in high dimensional setting. So far, numerous tests have been proposed for the case $k = 2$. See, e.g., Bai and Saranadasa (1996), Srivastava (2007), Chen and Qin (2010), Cai et al. (2014) and Feng et al. (2015). Some tests have also been introduced for the case of general $k \geq 2$. Schott (2007) modified Hotelling-Lawley trace and proposed the test statistic

$$T_{Sc} = \frac{1}{\sqrt{N-1}}\Big(\frac{1}{k-1}\operatorname{tr}\big(\mathbf{H}\big) - \frac{1}{N-k}\operatorname{tr}\big(\mathbf{G}\big)\Big).$$

Statistic T_{Sc} is a representative of the so-called sum-of-squares type statistics as it is based on an estimation of squared Euclidean norm $\sum_{i=1}^{k} n_i\|\theta_i - \bar{\theta}\|^2$, where $\bar{\theta} = N^{-1}\sum_{i=1}^{k} n_i\theta_i$. See Srivastava and Kubokawa (2013), Yamada and Himeno (2015), Hu et al. (2017), Zhang et al. (2017), Zhou et al. (2017) and Cao et al. (2019) for some other sum-of-squares type test statistics for general $k \geq 2$. It is known that the sum-of-squares type tests are particularly powerful against dense alternatives. In another work, Cai and

Xia (2014) proposed a test statistic

$$T_{CX} = \max_{1\le i\le p} \sum_{1\le j<l\le k} \frac{n_j n_l}{n_j + n_l} \frac{(\Omega(\bar{\mathbf{X}}_j - \bar{\mathbf{X}}_l))_i^2}{\omega_{ii}},$$

where $\Omega = (\omega)_{ij} = \mathbf{\Sigma}^{-1}$ is the precision matrix. When Ω is unknown, it is substituted by an estimator. Unlike T_{Sc}, the test statistic T_{CX} is an extreme value type one and is very powerful against sparse alternatives.

Most existing sum-of-squares type test procedures require the condition $\operatorname{tr}(\mathbf{\Sigma}^4)/\operatorname{tr}^2(\mathbf{\Sigma}^2) \to 0$, which is equivalent to

$$\frac{\boldsymbol{\lambda}_1}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} \to 0, \tag{1.2}$$

where $\boldsymbol{\lambda}_i$ is the ith largest eigenvalue of $\mathbf{\Sigma}$, $i = 1, \ldots, p$. In fact, the equivalence of these two conditions can be seen from the inequalities,

$$\frac{\boldsymbol{\lambda}_1^4}{\operatorname{tr}^2(\mathbf{\Sigma}^2)} \le \frac{\operatorname{tr}(\mathbf{\Sigma}^4)}{\operatorname{tr}^2(\mathbf{\Sigma}^2)} \le \frac{\boldsymbol{\lambda}_1^2 \operatorname{tr}(\mathbf{\Sigma}^2)}{\operatorname{tr}^2(\mathbf{\Sigma}^2)} = \frac{\boldsymbol{\lambda}_1^2}{\operatorname{tr}(\mathbf{\Sigma}^2)}.$$

Condition (1.2) is reasonable if $\mathbf{\Sigma}$ is nonspiked in the sense that it does not have significantly large eigenvalues. In some important situations, however, variables are heavily correlated with common factors, and the covariance matrix $\mathbf{\Sigma}$ is thus spiked in the sense that a few eigenvalues of $\mathbf{\Sigma}$ are significantly larger than the others (Fan et al., 2013; Cai et al., 2015; Wang and Fan, 2017). In such cases, condition (1.2) can be violated, and consequently, existing sum-of-squares type tests may not have correct level. Some

adjusted sum-of-squares type test procedures have been proposed to solve the problem. See, e.g., Katayama et al. (2013), Ma et al. (2015), Zhang et al. (2017) and Wang and Xu (2019). However, the power behavior of these corrected tests may not be satisfactory.

Recently, Aoshima and Yata (2018) and Wang and Xu (2018) considered two sample mean testing problem under the spiked covariance model. These tests have better power behavior compared with sum-of-squares type tests. However, both papers imposed strong conditions on the magnitude of p. For example, under the approximate factor model in Fan et al. (2013), the test in Aoshima and Yata (2018) requires $p/n \to 0$, while the test in Wang and Xu (2018) requires that $p/n^2 \to 0$ and the small eigenvalues of $\mathbf{\Sigma}$ are all equal.

The likelihood ratio test (LRT) method has been very successful in leading to satisfactory procedures in many specific problems. However, the LRT statistic for hypotheses (1.1), i.e. Wilks' Lambda statistic, is not defined for $p > N - k$. In high dimensional setting, both sum-of-squares type statistics and extreme value type statistics are not based on likelihood function. This motivates us to construct a likelihood-based test in high dimensional setting. In a recent work, Zhao and Xu (2016) proposed a generalized likelihood ratio test in the context of one sample mean vector test.

They used a least favorable argument to construct a generalized likelihood ratio test statistic. Their simulation results showed that their test has good power performance, especially when the variables are correlated. However, this phenomenon is not theoretically proved.

In this paper, we propose a generalized likelihood ratio test statistic for hypotheses (1.1) called least favorable direction (LFD) test statistic, which is a generalization of the test in Zhao and Xu (2016). We give the asymptotic distributions of the test statistic under both nonspiked and spiked covariances. An adaptive LFD test procedure is constructed by consistently detecting unknown covariance structure and estimating unknown parameters. The asymptotic local power function of the LFD test is also given. Our theoretical results show that the LFD test is particularly powerful under the spiked covariance. This explains the simulation results of Zhao and Xu (2016). Compared with the work of Zhao and Xu (2016), our main contribution is that we give a thorough theoretical analysis of the LFD test. Our theoretical analysis fall into the high dimension medium sample size setting, where both $n, p \to \infty$, but $p/n \to \infty$ (see Aoshima et al. (2018), Section 5). To prove our main results, we carefully study the high-order asymptotic behavior of the eigenvalues and eigenspaces of the sample covariance matrix. These results are also of independent interests. We further compare

the proposed test procedure with existing tests by simulations. It is shown that the LFD test has comparable behavior to existing sum-of-squares tests under the nonspiked covariance, while significantly outperforms competing tests under the spiked covariance.

The rest of the paper is organized as follows. In Section 2, we propose the LFD test statistic and derive its explicit forms. The asymptotic distributions of the LFD test statistic under both nonspiked and spiked covariances are given in Section 3. Based on these theoretical results, an adaptive LFD test procedure is proposed. Section 4 complements our study with numerical simulations. In Section 5, we give a short discussion. Finally, the proofs are gathered in the supplementary material.

2. Least favorable direction test

We introduce some notations. Define the $p \times N$ pooled sample matrix $\mathbf{X}$ as

$$\mathbf{X} = (X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}, \ldots, X_{k1}, X_{k2}, \ldots, X_{kn_k}).$$

The sum-of-squares within groups $\mathbf{G}$ can be written as $\mathbf{G} = \mathbf{X}(\mathbf{I}_N - \mathbf{J}\mathbf{J}^\top)\mathbf{X}^\top$ where

$$\mathbf{J} = \begin{pmatrix} \frac{1}{\sqrt{n_1}}\mathbf{1}_{n_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{n_2}}\mathbf{1}_{n_2} & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sqrt{n_k}}\mathbf{1}_{n_k} \end{pmatrix}$$

is an $N \times k$ matrix and $\mathbf{1}_{n_i}$ is an n_i-dimensional vector with all elements equal to 1, $i = 1, \ldots, k$. Let $n = N - k$ be the degrees of freedom of $\mathbf{G}$. Construct an $N \times n$ matrix $\tilde{\mathbf{J}}$ as

$$\tilde{\mathbf{J}} = \begin{pmatrix} \tilde{\mathbf{J}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{J}}_2 & \mathbf{0} \\ \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{J}}_k \end{pmatrix},$$

where $\tilde{\mathbf{J}}_i$ is an $n_i \times (n_i - 1)$ matrix defined as

$$\tilde{\mathbf{J}}_i = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ 0 & -\frac{2}{\sqrt{6}} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & -\frac{n_i-2}{\sqrt{(n_i-2)(n_i-1)}} & \frac{1}{\sqrt{(n_i-1)n_i}} \\ 0 & 0 & \cdots & 0 & -\frac{n_i-1}{\sqrt{(n_i-1)n_i}} \end{pmatrix}.$$

The matrix $\tilde{\mathbf{J}}$ is a column orthogonal matrix satisfying $\tilde{\mathbf{J}}^\top\tilde{\mathbf{J}} = \mathbf{I}_n$ and $\tilde{\mathbf{J}}\tilde{\mathbf{J}}^\top = \mathbf{I}_N - \mathbf{J}\mathbf{J}^\top$. Define $\mathbf{Y} = \mathbf{X}\tilde{\mathbf{J}}$. Then $\mathbf{G}$ can be written as

$$\mathbf{G} = \mathbf{Y}\mathbf{Y}^\top.$$

The sum-of-squares between groups $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{X}(\mathbf{J}\mathbf{J}^\top - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top)\mathbf{X}^\top = \mathbf{X}\mathbf{J}(\mathbf{I}_k - \frac{1}{N}\mathbf{J}^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{J})\mathbf{J}^\top\mathbf{X}^\top.$$

By some matrix algebra, we have $\mathbf{I}_k - N^{-1}\mathbf{J}^\top\mathbf{1}_N\mathbf{1}_N^\top\mathbf{J} = \mathbf{C}\mathbf{C}^\top$ where $\mathbf{C}$ is a $k \times (k-1)$ matrix defined as $\mathbf{C} = \mathbf{C}_1\mathbf{C}_2$, and

$$\mathbf{C}_1 = \begin{pmatrix} \sqrt{n_1} & \sqrt{n_1} & \cdots & \sqrt{n_1} & \sqrt{n_1} \\ -\frac{n_1}{\sqrt{n_2}} & \sqrt{n_2} & \cdots & \sqrt{n_2} & \sqrt{n_2} \\ 0 & -\frac{n_1+n_2}{\sqrt{n_3}} & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & -\frac{\sum_{i=1}^{k-2} n_i}{\sqrt{n_{k-1}}} & \sqrt{n_{k-1}} \\ 0 & 0 & \cdots & 0 & -\frac{\sum_{i=1}^{k-1} n_i}{\sqrt{n_k}} \end{pmatrix},$$

$$\mathbf{C}_2 = \begin{pmatrix} \frac{n_1(n_1+n_2)}{n_2} & 0 & \cdots & 0 \\ 0 & \frac{(\sum_{i=1}^{2} n_i)(\sum_{i=1}^{3} n_i)}{n_3} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \frac{(\sum_{i=1}^{k-1} n_i)(\sum_{i=1}^{k} n_i)}{n_k} \end{pmatrix}^{-\frac{1}{2}}.$$

Then $\mathbf{H}$ can be written as

$$\mathbf{H} = \mathbf{X}\mathbf{J}\mathbf{C}\mathbf{C}^\top\mathbf{J}^\top\mathbf{X}^\top.$$

Define $\mathbf{\Theta} = (\sqrt{n_1}\theta_1, \ldots, \sqrt{n_k}\theta_k)$. Then the null hypothesis H_0 is equivalent to $\mathbf{\Theta C} = \mathbf{O}_{p\times(k-1)}$, where $\mathbf{O}_{p\times(k-1)}$ is a $p \times (k-1)$ matrix with 0 entries. Thus, the hypotheses (1.1) are equivalent to

$$H_0 : \mathbf{\Theta C} = \mathbf{O}_{p\times(k-1)} \quad \text{v.s.} \quad H_1 : \mathbf{\Theta C} \neq \mathbf{O}_{p\times(k-1)}.$$

In low dimensional setting, the testing problem (1.1) is well studied. A classical test statistic is Roy's maximum root which is constructed by Roy (1953) using his well-known union intersection principle. The key idea is to decompose data $\mathbf{X}$ into a set of univariate data $\{\mathbf{X}_a = a^\top \mathbf{X} : \ a \in \mathbb{R}^p, a^\top a = 1\}$. This induces a decomposition of the null hypothesis and the alternative hypothesis:

$$H_0 = \bigcap_{a\in\mathbb{R}^p, a^\top a=1} H_{0a} \quad \text{v.s.} \quad H_1 = \bigcup_{a\in\mathbb{R}^p, a^\top a=1} H_{1a},$$

where $H_{0a} : a^\top \mathbf{\Theta C} = \mathbf{O}_{1\times(k-1)}$ and $H_{1a} : a^\top \mathbf{\Theta C} \neq \mathbf{O}_{1\times(k-1)}$. Let $L_0(a)$ and $L_1(a)$ be the maximum likelihood of $\mathbf{X}_a$ under H_{0a} and H_{1a}, respectively. For each a satisfying $a^\top a = 1$, the component LRT statistic

$$\frac{L_1(a)}{L_0(a)} = \Big(\frac{a^\top(\mathbf{G}+\mathbf{H})a}{a^\top \mathbf{G} a}\Big)^{N/2}$$

can be used to test H_{0a} v.s. H_{1a}. Using union intersection principle, Roy proposed the test statistic $\max_{a^\top a=1} L_1(a)/L_0(a) = (1 + \lambda_1(\mathbf{H}\mathbf{G}^{-1}))^{N/2}$, where $\lambda_i(\cdot)$ means the ith largest eigenvalue. This statistic is an increasing function of Roy's maximum root.

From a likelihood point of view, log likelihood ratio is an estimator of the Kullback-Leibler divergence between the true distribution and the null distribution. Hence the component LRT statistic $L_1(a)/L_0(a)$ characterizes the discrepancy between the true distribution and the null distribution along the direction a. This motivates us to consider the direction

$$a^* = \arg\max_{a^\top a=1} \frac{L_1(a)}{L_0(a)} \tag{2.1}$$

which can hopefully achieve the largest discrepancy between the true distribution and the null distribution. Thus, H_{0a^*} is the component null hypothesis most likely to be not true. We shall call a^* the least favorable direction. Roy's maximum root is in fact the component LRT statistic along the least favorable direction.

Unfortunately, Roy's maximum root can only be defined when $n \geq p$, hence can not be used in high dimensional setting. In what follows, we assume $p > n$. In this case, the set

$$\mathcal{A} \stackrel{def}{=} \{a : L_1(a) = +\infty,\, a^\top a = 1\} = \{a : a^\top \mathbf{G} a = 0,\, a^\top a = 1\}$$

is not empty since $\mathbf{G}$ is singular. Consequently, the right hand side of (2.1) is not well defined since the ratio involves infinity. Hence we need a new definition for LFD in high dimensional setting. Define

$$\mathcal{B} = \{a : L_0(a) = +\infty,\, a^\top a = 1\} = \{a : a^\top (\mathbf{G} + \mathbf{H}) a = 0,\, a^\top a = 1\}.$$

It can be seen that $\mathcal{B} \subset \mathcal{A}$. Moreover, by the independence of $\mathbf{G}$ and $\mathbf{H}$, with probability 1, we have $\mathcal{A} \cap \mathcal{B}^c \neq \emptyset$. Then for any direction a, there are three possible scenarios: $L_1(a) < +\infty$ and $L_0(a) < +\infty$; $L_1(a) = +\infty$ and $L_0(a) < +\infty$; $L_1(a) = +\infty$ and $L_0(a) = +\infty$. To maximizes the discrepancy between $L_1(a)$ and $L_0(a)$, one may consider the direction a such that $L_1(a) = +\infty$ and $L_0(a) < +\infty$. This suggests that the least favorable direction a^*, which hopefully maximizes the discrepancy between $L_1(a)$ and $L_0(a)$, should be defined as $a^* = \arg\min_{a \in \mathcal{A} \cap \mathcal{B}^c} L_0(a)$. Equivalently,

$$a^* = \underset{a \in \mathcal{A} \cap \mathcal{B}^c}{\arg\min}\, L_0(a) = \underset{a^\top a = 1, a^\top G a = 0}{\arg\max}\; a^\top \mathbf{H} a.$$

Based on a^* and the likelihood $L_0(a)$, we propose a new test statistic

$$T(\mathbf{X}) = a^{*T} \mathbf{H} a^* = \max_{a^\top a = 1, a^\top \mathbf{G} a = 0} a^\top \mathbf{H} a.$$

The null hypothesis is rejected when $T(\mathbf{X})$ is large enough. We shall call $T(\mathbf{X})$ the LFD test statistic. Since the least favorable direction a^* is obtained from the component likelihood function, the statistic $T(\mathbf{X})$ is also a generalized likelihood ratio test statistic.

Now we derive the explicit forms of the LFD test statistic. Let $\mathbf{Y} = \mathbf{U_Y} \mathbf{D_Y} \mathbf{V_Y}^\top$ be the singular value decomposition of $\mathbf{Y}$, where $\mathbf{U_Y}$ and $\mathbf{V_Y}$ are $p \times \min(n, p)$ and $n \times \min(n, p)$ column orthogonal matrices, respectively, and $\mathbf{D_Y}$ is a $\min(n, p) \times \min(n, p)$ diagonal matrix whose diagonal elements

are the non-increasingly ordered singular values of $\mathbf{Y}$. If $p > n$, let $\mathbf{P_Y} = \mathbf{U_Y}\mathbf{U_Y}^\top$ be the projection matrix onto the column space of $\mathbf{Y}$. Then Lemma 1 in supplementary material implies that for $p > n$,

$$T(\mathbf{X}) = \lambda_1\big(\mathbf{C}^\top\mathbf{J}^\top\mathbf{X}^\top(\mathbf{I}_p - \mathbf{P_Y})\mathbf{X}\mathbf{J}\mathbf{C}\big). \tag{2.2}$$

While (2.2) is convenient for theoretical analysis, it is not convenient for computation. When $p > N$, another simple form of $T(\mathbf{X})$ can be used for computation. If $p > N$, then $\mathbf{X}^\top\mathbf{X}$ is invertible. By the relationship

$$\begin{aligned}\begin{pmatrix} \mathbf{J}^\top\mathbf{X}^\top\mathbf{X}\mathbf{J} & \mathbf{J}^\top\mathbf{X}^\top\mathbf{X}\tilde{\mathbf{J}} \\ \tilde{\mathbf{J}}^\top\mathbf{X}^\top\mathbf{X}\mathbf{J} & \tilde{\mathbf{J}}^\top\mathbf{X}^\top\mathbf{X}\tilde{\mathbf{J}} \end{pmatrix}^{-1} =& \Big(\begin{pmatrix} \mathbf{J}^\top \\ \tilde{\mathbf{J}}^\top \end{pmatrix} \mathbf{X}^\top\mathbf{X} \begin{pmatrix} \mathbf{J} & \tilde{\mathbf{J}} \end{pmatrix} \Big)^{-1} \\ =& \begin{pmatrix} \mathbf{J}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{J} & \mathbf{J}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\tilde{\mathbf{J}} \\ \tilde{\mathbf{J}}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{J} & \tilde{\mathbf{J}}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\tilde{\mathbf{J}} \end{pmatrix}\end{aligned}$$

and matrix inverse formula, we have that

$$\begin{aligned}\big(\mathbf{J}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{J}\big)^{-1} =& \mathbf{J}^\top\mathbf{X}^\top\mathbf{X}\mathbf{J} - \mathbf{J}^\top\mathbf{X}^\top\mathbf{X}\tilde{\mathbf{J}}(\tilde{\mathbf{J}}^\top\mathbf{X}^\top\mathbf{X}\tilde{\mathbf{J}})^{-1}\tilde{\mathbf{J}}^\top\mathbf{X}^\top\mathbf{X}\mathbf{J} \\ =& \mathbf{J}^\top\mathbf{X}^\top(\mathbf{I}_p - \mathbf{P_Y})\mathbf{X}\mathbf{J}.\end{aligned}$$

Thus,

$$T(\mathbf{X}) = \lambda_1\Big(\mathbf{C}^\top\big(\mathbf{J}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{J}\big)^{-1}\mathbf{C}\Big). \tag{2.3}$$

Compared with (2.2), the expression (2.3) doesn't involve $\mathbf{P_Y}$ and is more convenient for computation.

In the case of $k = 2$, it can be seen that the least favorable direction is proportional to $(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$ and the LFD test statistic has expression

$$T(\mathbf{X}) = \frac{n_1 n_2}{n_1 + n_2} \|(\mathbf{I}_p - \mathbf{P}_\mathbf{Y})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)\|^2.$$

In this case, the least favorable direction coincides with the maximal data piling direction proposed by Ahn and Marron (2010).

3. Theoretical analysis

We now turn to the analysis of the asymptotic distributions of the LFD test statistic. The normality of the observations is an important assumption for our results and will be assumed throughout this section. We shall give theoretical results under both nonspiked and spiked covariances. Based on these results, an adaptive test with asymptotically correct level can be constructed. Also, these results allow us to derive the local asymptotic power function of LFD test.

3.1 Nonspiked covariance

In this subsection, we establish the asymptotic distribution of $T(\mathbf{X})$ under the nonspiked covariance. Let $\mathbf{W}_{k-1}$ be a $(k-1) \times (k-1)$ symmetric random matrix whose entries above the main diagonal are iid $\mathcal{N}(0, 1)$ random variables and the entries on the diagonal are iid $\mathcal{N}(0, 2)$ random variables.

The following theorem establishes the asymptotic distribution of the LFD test statistic.

Theorem 1. *Suppose as* $n, p \to \infty$*, condition* (1.2) *holds. Furthermore, suppose that* $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ *and* $\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$*. Then under the local alternative hypothesis* $\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\| = O(\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$*,*

$$\frac{T(\mathbf{X}) - (\operatorname{tr}(\boldsymbol{\Sigma}) - n\operatorname{tr}(\boldsymbol{\Sigma}^2)/\operatorname{tr}(\boldsymbol{\Sigma}))}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}} \sim \lambda_1\left(\mathbf{W}_{k-1} + \frac{\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\right) + o_P(1),$$

where $\sim$ *means having the same distribution.*

Remark 1. The condition $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ implies $p/n \to \infty$. Hence $T(\mathbf{X})$ is well defined for large n. The condition $\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)})$ requires that the range of the eigenvalues of $\boldsymbol{\Sigma}$ is not too large.

To centralize $T(\mathbf{X})$ under the conditions of Theorem 1, the parameters $\operatorname{tr}(\boldsymbol{\Sigma})$ and $\operatorname{tr}(\boldsymbol{\Sigma}^2)$ should be estimated. Let $\hat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{G} = n^{-1}\mathbf{Y}\mathbf{Y}^\top$ be the sample covariance matrix. We use the following simple estimators,

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}), \quad \widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)} = \operatorname{tr}(\hat{\boldsymbol{\Sigma}}^2) - n^{-1}\operatorname{tr}^2(\hat{\boldsymbol{\Sigma}}).$$

Define

$$Q_1 = \frac{T(\mathbf{X}) - \left(\widehat{\operatorname{tr}(\boldsymbol{\Sigma})} - n\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Sigma})}\right)}{\sqrt{\widehat{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}}.$$

Let $F_1(x)$ be the cumulative distribution function of $\lambda_1(\mathbf{W}_{k-1})$. Then we reject the null hypothesis if $Q_1 > F_1^{-1}(1-\alpha)$. The following corollary

gives the asymptotic local power function of the proposed test under the nonspiked covariance.

Corollary 1. *Under the conditions of Theorem 1,*

$$\begin{aligned}&\Pr\left(Q_1 > F_1^{-1}(1-\alpha)\right)\\ =&\Pr\left(\lambda_1\left(\mathbf{W}_{k-1}+\frac{\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\boldsymbol{\Sigma}^2)}}\right) > F_1^{-1}(1-\alpha)\right)+o(1).\end{aligned}$$

Corollary 1 shows that under the nonspiked covariance, the LFD test has similar power behavior to existing sum-of-squares type tests. In fact, if $k = 2$, the asymptotic local power function given by Corollary 1 is equal to the asymptotic local power function of the tests in Bai and Saranadasa (1996) and Chen and Qin (2010).

3.2 Spiked covariance

Now we derive the asymptotic results under the spiked covariance, which are much more involved than the nonspiked case. Let $\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$ denote the eigenvalue decomposition of $\boldsymbol{\Sigma}$, where $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_p)$ and $\mathbf{U}$ is an orthogonal matrix. Suppose that $\boldsymbol{\Sigma}$ has r spiked eigenvalues, where $1 \leq r \leq p$ can also vary as $n, p \to \infty$. We shall first assume the spiked number r is known. Adaptation to unknown r will be considered latter. Denote $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r)$ and $\boldsymbol{\Lambda}_2 = \operatorname{diag}(\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p)$. Correspondingly, we

denote $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$ where $\mathbf{U}_1$ and $\mathbf{U}_2$ are the first r columns and the last $p - r$ columns of $\mathbf{U}$. Then $\boldsymbol{\Sigma} = \mathbf{U}_1 \boldsymbol{\Lambda}_1 \mathbf{U}_1^\top + \mathbf{U}_2 \boldsymbol{\Lambda}_2 \mathbf{U}_2^\top$.

First we shall derive the asymptotic properties of the eigenvalues and eigenspaces of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$ since they play a key role in our latter analysis. The following proposition gives the asymptotic behavior of $\lambda_1(\hat{\boldsymbol{\Sigma}}), \ldots, \lambda_r(\hat{\boldsymbol{\Sigma}})$ and $\sum_{i=r+1}^{n} \lambda_i(\hat{\boldsymbol{\Sigma}})$.

Proposition 1. *Suppose that $r \leq n$. Then uniformly for $i = 1, \ldots, r$,*

$$\lambda_i(\hat{\boldsymbol{\Sigma}}) = \boldsymbol{\lambda}_i + n^{-1} \operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(\boldsymbol{\lambda}_i \sqrt{\frac{r}{n}} + \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right);$$

and

$$\sum_{i=r+1}^{n} \lambda_i(\hat{\boldsymbol{\Sigma}}) = \left(1 - \frac{r}{n}\right) \operatorname{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r \sqrt{\frac{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right).$$

Remark 2. Recently, the asymptotic behavior of the spiked eigenvalues of the sample covariance matrix is actively studied. See, e.g., Yata and Aoshima (2013); Shen et al. (2016); Wang and Fan (2017); Cai et al. (2019). An important improvement of Proposition 1 over existing results is that Proposition 1 does not impose any condition for the structure of $\boldsymbol{\Sigma}$ while still gives the correct convergence rate.

Based on Proposition 1, we propose the following estimators of $\operatorname{tr}(\boldsymbol{\Lambda}_2)$

and $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$,

$$\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)} = \left(1 - \frac{r}{n}\right)^{-1} \sum_{i=r+1}^{n} \lambda_i(\hat{\boldsymbol{\Sigma}}), \quad \hat{\boldsymbol{\lambda}}_i = \lambda_i(\hat{\boldsymbol{\Sigma}}) - n^{-1}\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}, \quad i = 1, \ldots, r.$$

Moreover, our latter analysis requires an estimator of $\mathrm{tr}(\boldsymbol{\Lambda}_2^2)$. We propose the following estimator of $\mathrm{tr}(\boldsymbol{\Lambda}_2^2)$,

$$\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)} = \sum_{i=r+1}^{n} \left(\lambda_i(\hat{\boldsymbol{\Sigma}}) - n^{-1}\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)}\right)^2.$$

The following proposition gives the convergence rate of these estimators.

Proposition 2. *Suppose that $r = o(n)$. Then uniformly for $i = 1, \ldots, r$,*

$$\hat{\boldsymbol{\lambda}}_i = \boldsymbol{\lambda}_i + O_P\left(\boldsymbol{\lambda}_i\sqrt{\frac{r}{n}} + \sqrt{\frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + \boldsymbol{\lambda}_{r+1}\right);$$

and

$$\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2)} = \mathrm{tr}(\boldsymbol{\Lambda}_2) + O_P\left(r\sqrt{\frac{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{n}} + r\boldsymbol{\lambda}_{r+1}\right),$$
$$\widehat{\mathrm{tr}(\boldsymbol{\Lambda}_2^2)} = \mathrm{tr}(\boldsymbol{\Lambda}_2^2) + O_P\left(\frac{r\,\mathrm{tr}(\boldsymbol{\Lambda}_2^2)}{n} + r\boldsymbol{\lambda}_{r+1}^2\right).$$

Remark 3. Our estimators of $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$ and $\mathrm{tr}(\boldsymbol{\Lambda}_2)$ are similar to some existing estimators, e.g., the noise-reduction estimators in Yata and Aoshima (2012) and the estimators in Wang and Fan (2017). However, their theoretical results require that r is fixed, p is not large and $\boldsymbol{\Sigma}$ satisfies certain spiked covariance models.

Remark 4. The estimation of $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ is relatively unexplored. Recently, Aoshima and Yata (2018) proposed an estimator of $\operatorname{tr}(\mathbf{\Lambda}_2^2)$ by using the cross-data-matrix methodology. They also proved the consistency of their estimator. Their method relies, however, on an arbitrary split of the data into two samples of equal size.

Next we consider the asymptotic behavior of the eigenspaces of $\hat{\mathbf{\Sigma}}$. Let $\mathbf{U}_{\mathbf{Y},1}$ denote the first r columns of $\mathbf{U}_{\mathbf{Y}}$. Then the columns of $\mathbf{U}_{\mathbf{Y},1}$ are the principal eigenvectors of $\hat{\mathbf{\Sigma}}$, and $\mathbf{P}_{\mathbf{Y},1} = \mathbf{U}_{\mathbf{Y},1}\mathbf{U}_{\mathbf{Y},1}^\top$ is the projection matrix onto the rank r principal subspace of $\hat{\mathbf{\Sigma}}$. The properties of $\mathbf{P}_{\mathbf{Y},1}$ and individual principal eigenvectors have been extensively studied. See Cai et al. (2015), Shen et al. (2016), Wang and Fan (2017) and the references therein. The existing results include the consistency of the principal subspace and the high-order asymptotic behavior of the individual principal eigenvectors. However, these results are not enough for our latter analysis. The following proposition gives the high-order asymptotic behavior of $\mathbf{P}_{\mathbf{Y},1}$. To the best of our knowledge, such result has never appeared in the literature before.

Write $\mathbf{Y} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Z}$, where $\mathbf{Z}$ is a $p \times n$ random matrix with iid $\mathcal{N}(0,1)$ entries. Then $\mathbf{Y} = \mathbf{U}_1\mathbf{\Lambda}_1^{1/2}\mathbf{Z}_1 + \mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2$, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the first r rows and last $p-r$ rows of $\mathbf{Z}$.

Proposition 3. *Suppose that* $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ *and* $r\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to$

0. *Then*

$$\left\|\mathbf{P}_{\mathbf{Y},1} - \mathbf{P}^{\dagger}_{\mathbf{Y},1}\right\| = O_P\left(\frac{\mathrm{tr}(\boldsymbol{\Lambda}_2)}{n\boldsymbol{\lambda}_r} + \frac{\boldsymbol{\lambda}_{r+1}}{\boldsymbol{\lambda}_r}\right),$$

where $\|\cdot\|$ *is the spectral norm,* $\mathbf{P}^{\dagger}_{\mathbf{Y},1} = \mathbf{U}_1\mathbf{U}_1^{\top} + \mathbf{U}_1\mathbf{Q}^{\top}\mathbf{U}_2^{\top} + \mathbf{U}_2\mathbf{Q}\mathbf{U}_1^{\top}$ *and* $\mathbf{Q} = \boldsymbol{\Lambda}_2^{1/2}\mathbf{Z}_2\mathbf{Z}_1^{\top}(\mathbf{Z}_1\mathbf{Z}_1^{\top})^{-1}\boldsymbol{\Lambda}_1^{-1/2}$.

Remark 5. The condition $\mathrm{tr}(\boldsymbol{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ is commonly adopted in the study of the principal subspaces. In fact, when this condition is violated, the principal subspace will lose its relation to the rank r eigenspace of $\boldsymbol{\Sigma}$. See, e.g., Nadler (2008).

Remark 6. Recently, some high-order Davis-Kahan theorems are established, e.g., Lemma 2 in Koltchinskii and Lounici (2016) and Lemma 2 in Fan et al. (2019). These general results explicitly characterizes the linear term and high-order error on rank r eigenspace due to matrix perturbation. By applying these results to $\hat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}$, one can obtain similar results to Proposition 3. Compared with Proposition 3, however, the results so obtained are slightly weaker and requires stronger conditions.

If $p > n$, let $\mathbf{U}_{\mathbf{Y},2}$ be the $r+1$ to nth columns of $\mathbf{U}_{\mathbf{Y}}$. Then $\mathbf{P}_{\mathbf{Y},2} = \mathbf{U}_{\mathbf{Y},2}\mathbf{U}_{\mathbf{Y},2}^{\top}$ is the projection matrix onto the eigenspace spanned by the $r+1$ to nth eigenvectors of $\hat{\boldsymbol{\Sigma}}$. Our latter analysis also requires the asymptotic properties of $\mathbf{P}_{\mathbf{Y},2}$, which has not been considered in the literature. Let

$\mathbf{V}_{\mathbf{Z}_1} = \mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1/2}$. Then $\mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top = \mathbf{Z}_1^\top(\mathbf{Z}_1\mathbf{Z}_1^\top)^{-1}\mathbf{Z}_1$ is the projection matrix onto the row space of $\mathbf{Z}_1$. Let $\tilde{\mathbf{V}}_{\mathbf{Z}_1}$ be a $n\times(n-r)$ column orthogonal matrix which satisfies $\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top = \mathbf{I}_n - \mathbf{V}_{\mathbf{Z}_1}\mathbf{V}_{\mathbf{Z}_1}^\top$. The following proposition gives the asymptotic behavior of $\mathbf{P}_{\mathbf{Y},2}$.

Proposition 4. *Suppose that* $r = o(n)$, $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$ *and* $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$. *Then*

$$\left\|\mathbf{P}_{\mathbf{Y},2} - \mathbf{P}_{\mathbf{Y},2}^\dagger\right\| = O_P\left(\sqrt{\frac{\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1}{n\boldsymbol{\lambda}_r^2}} + \sqrt{\frac{n\boldsymbol{\lambda}_{r+1}}{\operatorname{tr}(\mathbf{\Lambda}_2)}}\right),$$

where $\mathbf{P}_{\mathbf{Y},2}^\dagger = (\operatorname{tr}(\mathbf{\Lambda}_2))^{-1}\,\mathbf{U}_2\mathbf{\Lambda}_2^{1/2}\mathbf{Z}_2\tilde{\mathbf{V}}_{\mathbf{Z}_1}\tilde{\mathbf{V}}_{\mathbf{Z}_1}^\top\mathbf{Z}_2^\top\mathbf{\Lambda}_2^{1/2}\mathbf{U}_2^\top$.

Remark 7. The condition $\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$ is stronger than the condition $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\boldsymbol{\lambda}_r) \to 0$ in Proposition 3. These two conditions are equivalent if $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_r$ are of the same order.

Now we are ready to derive the asymptotic properties of $T(\mathbf{X})$ under the spiked covariance. Let $\mathbf{W}_{k-1}^*$ be a $(k-1)\times(k-1)$ symmetric random matrix distributed as Wishart$(r, \mathbf{I}_{k-1})$ and is independent of $\mathbf{W}_{k-1}$, where Wishart$(m, \mathbf{\Psi})$ is the Wishart distribution with parameter $\mathbf{\Psi}$ and m degrees of freedom. The following theorem gives the asymptotic distribution of $T(\mathbf{X})$ under the null and the local alternative hypothesis.

Theorem 2. *Suppose that* $r = o(\sqrt{n})$, $r\operatorname{tr}(\mathbf{\Lambda}_2)\boldsymbol{\lambda}_1/(n\boldsymbol{\lambda}_r^2) \to 0$, $rn\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\mathbf{\Lambda}_2) \to 0$, $r\boldsymbol{\lambda}_{r+1}/\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)} \to 0$ *and* $\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p = O(n^{-1}\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)})$. *Then*

(i) under the null hypothesis $\boldsymbol{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$,

$$\begin{aligned}&\frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\\ \sim&\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1})\right.\\ &\left.+\frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\mathbf{W}_{k-1}\right) + o_P(1);\end{aligned}$$

(ii) if $r \to \infty$ *or* $\operatorname{tr}(\boldsymbol{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}) \to 0$, *then under the local alternative hypothesis* $\|\mathbf{C}^\top\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\mathbf{C}\| = O(\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)})$,

$$\begin{aligned}&\frac{T(\mathbf{X}) - ((1+r/n)\operatorname{tr}(\boldsymbol{\Lambda}_2) - n\operatorname{tr}(\boldsymbol{\Lambda}_2^2)/\operatorname{tr}(\boldsymbol{\Lambda}_2))}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\\ \sim&\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1})\right.\\ &+\frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\mathbf{W}_{k-1}\\ &\left.+\frac{\mathbf{C}^\top\boldsymbol{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\boldsymbol{\Theta}\mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}\right) + o_P(1).\end{aligned}$$

Remark 8. Suppose the approximate factor model in Fan et al. (2013) holds. That is, r is fixed, $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_r$ diverge at rate $O(p)$ and $\boldsymbol{\lambda}_{r+1}, \ldots, \boldsymbol{\lambda}_p$ are bounded. Then the conditions of Theorem 2 become $p/n \to \infty$ and $\boldsymbol{\lambda}_{r+1} - \boldsymbol{\lambda}_p = O(\sqrt{p}/n)$. Hence Theorem 2 holds for ultra-high dimensional data. In contrast, recently proposed tests under the spiked covariance model can only be used for lower dimensional data. In fact, under the approximate factor model in Fan et al. (2013), Aoshima and Yata (2018) requires $p/n \to$

0, while Wang and Xu (2018) requires $p/n^2 \to 0$ and $\boldsymbol{\lambda}_{r+1} = \cdots = \boldsymbol{\lambda}_p$. We note that if $k = 2$ and $p/n^2 \to 0$, then the coefficient of $\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}$ is negligible, and consequently, $T(\mathbf{X})$ is asymptotically normal distributed. Thus, Theorem 2 gives the high-order behavior of $T(\mathbf{X})$.

Now we formulate a test procedure with asymptotically correct level. Define the standardized statistic as

$$Q_2 = \frac{T(\mathbf{X}) - \left((1 + r/n)\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)} - n\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}/\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}\right)}{\sqrt{rn^{-2}(\widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)})^2 + \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}}.$$

Let $F_2(x; \operatorname{tr}(\boldsymbol{\Lambda}_2), \operatorname{tr}(\boldsymbol{\Lambda}_2^2))$ be the cumulative distribution function of

$$\lambda_1 \left(\frac{n^{-1}\operatorname{tr}(\boldsymbol{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} (\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1}) + \frac{\sqrt{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\boldsymbol{\Lambda}_2) + \operatorname{tr}(\boldsymbol{\Lambda}_2^2)}} \mathbf{W}_{k-1} \right).$$

Then we reject the null hypothesis if

$$Q_2 > F_2^{-1}\left(1 - \alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right).$$

The following corollary shows that this test procedure has asymptotically correct level, and also gives the asymptotic local power function.

Corollary 2. *Suppose the conditions of Theorem 2 hold. Then*

(i) under the null hypothesis $\boldsymbol{\Theta}\mathbf{C} = \mathbf{O}_{p\times(k-1)}$,

$$\Pr\left(Q_2 > F_2^{-1}\left(1 - \alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)}\right)\right) = \alpha + o(1);$$

(ii) if $r \to \infty$ *or* $\operatorname{tr}(\mathbf{\Lambda}_2)/(n\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}) \to 0$*, then under the local alternative hypothesis* $\|\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}\| = O(\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)})$,

$$
\begin{aligned}
&\Pr\left(Q_2 > F_2^{-1}\left(1-\alpha; \widehat{\operatorname{tr}(\mathbf{\Lambda}_2)}, \widehat{\operatorname{tr}(\mathbf{\Lambda}_2^2)}\right)\right) \\
=&\Pr\left(\lambda_1\left(\frac{n^{-1}\operatorname{tr}(\mathbf{\Lambda}_2)}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}(\mathbf{W}_{k-1}^* - r\mathbf{I}_{k-1})\right.\right. \\
&\qquad + \frac{\sqrt{\operatorname{tr}(\mathbf{\Lambda}_2^2)}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}\mathbf{W}_{k-1} \\
&\qquad \left.\left. + \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}}\right)\right. \\
&\qquad \left.\left. > F_2^{-1}\left(1-\alpha; \operatorname{tr}(\mathbf{\Lambda}_2), \operatorname{tr}(\mathbf{\Lambda}_2^2)\right)\right)\right) + o(1).
\end{aligned}
$$

To gain some insight into the asymptotic behavior of $T(\mathbf{X})$, we consider $k = 2$ and compare the LFD test with the tests in Bai and Saranadasa (1996) and Chen and Qin (2010). Corollary 2 implies that if

$$
\liminf_{n\to\infty} \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2) + \operatorname{tr}(\mathbf{\Lambda}_2^2)}} > 0,
$$

then the LFD test has nontrivial power asymptotically. In contrast, if

$$
\limsup_{n\to\infty} \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}} = 0,
$$

then the tests in Bai and Saranadasa (1996) and Chen and Qin (2010) has trivial power asymptotically. To compare $\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}$ and $\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}$, we temporarily place a prior on $\mathbf{\Theta}$. Suppose that $\sqrt{n_i}\theta_i$ has prior distribution $\mathcal{N}_p(\mathbf{0}_p, \psi \mathbf{I}_p)$, $i = 1, 2$. Then $\psi^{-1}\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}$ is distributed as χ^2 distribution with p degrees of freedom. On the other hand, $\psi^{-1}\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{U}_2 \mathbf{U}_2^\top \mathbf{\Theta} \mathbf{C}$

is distributed as χ^2 distribution with $p-r$ degrees of freedom. Then we have

$$\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}}{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}} \xrightarrow{P} 1.$$

So in average, the signal contained in $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{U}_2\mathbf{U}_2^\top\mathbf{\Theta}\mathbf{C}$ is roughly the same as that in $\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}$. Now we compare the asymptotic variance. It is not hard to see that

$$\frac{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}{\operatorname{tr}(\mathbf{\Sigma}^2)} \to 0.$$

That is, the asymptotic variance of $T(\mathbf{X})$ is much smaller than the tests in Bai and Saranadasa (1996) and Chen and Qin (2010). To appreciate this phenomenon, we note that in the expression (2.2), $(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{X}\mathbf{J}\mathbf{C}|\mathbf{P}_\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}_p,(\mathbf{I}_p-\mathbf{P}_\mathbf{Y})\mathbf{\Sigma}(\mathbf{I}_p-\mathbf{P}_\mathbf{Y}))$. But $\mathbf{I}_p-\mathbf{P}_\mathbf{Y}$ tends to be orthogonal to $\mathbf{U}_1\mathbf{U}_1^\top$ which is the projection matrix onto the eigenspace corresponding to the leading eigenvalues of $\mathbf{\Sigma}$. Hence the projection by $\mathbf{I}_p-\mathbf{P}_\mathbf{Y}$ helps reduce the variance of $\mathbf{X}\mathbf{J}\mathbf{C}$.

Thus, if $\mathbf{\Theta}$ satisfies

$$\liminf_{n\to\infty}\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{rn^{-2}\operatorname{tr}^2(\mathbf{\Lambda}_2)+\operatorname{tr}(\mathbf{\Lambda}_2^2)}}>0, \quad \limsup_{n\to\infty}\frac{\mathbf{C}^\top\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{C}}{\sqrt{\operatorname{tr}(\mathbf{\Sigma}^2)}}=0,$$

the LFD test has nontrivial power while the tests in Bai and Saranadasa (1996) and Chen and Qin (2010) has trivial power. Hence the LFD test tends to be more powerful than the tests in Bai and Saranadasa (1996)

and Chen and Qin (2010).

In practice, one may not know weather the covariance matrix is spiked. Even if it is known that the covariance matrix is spiked, the spike number r may be unknown. So we would like to propose an adaptive test procedure. Note that Theorem 1 requires $n\boldsymbol{\lambda}_1/\operatorname{tr}(\boldsymbol{\Sigma}) \to 0$ while Theorem 2 requires $\operatorname{tr}(\boldsymbol{\Lambda}_2)/n\boldsymbol{\lambda}_r \to 0$ and $n\boldsymbol{\lambda}_{r+1}/\operatorname{tr}(\boldsymbol{\Lambda}_2) \to 0$. This motivates us to consider the following adaptive test procedure. Let $\tau > 1$ be a hyperparameter. If

$$\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau,$$

then we reject the null hypothesis if $Q_1 > F^{-1}(1-\alpha)$. Otherwise, we reject the null hypothesis if $Q_2 > F_2^{-1}(1-\alpha; \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2,)}, \widehat{\operatorname{tr}(\boldsymbol{\Lambda}_2^2)})$ where the unknown r is substituted by the estimator

$$\hat{r} = \min\left\{1 \le i < n : \frac{n\lambda_{i+1}(\hat{\boldsymbol{\Sigma}})}{\sum_{j=i+1}^{n}\lambda_j(\hat{\boldsymbol{\Sigma}})} < \tau\right\}.$$

We have the following proposition.

Proposition 5. *Let $\tau > 1$ be a constant.*

(i) Under the conditions of Theorem 1,

$$\Pr\left(\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau\right) \to 1;$$

(ii) Under the conditions of Theorem 2,

$$\Pr\left(\frac{n\lambda_1(\hat{\boldsymbol{\Sigma}})}{\operatorname{tr}(\hat{\boldsymbol{\Sigma}})} < \tau\right) \to 0, \quad \Pr(\hat{r} = r) \to 1.$$

Proposition 5 implies that the spiked covariance structure can be consistently detected. So the proposed adaptive LFD test procedure can indeed adapt to the unknown covariance structure.

4. Numerical study

In this section, we compare the numerical performance of the adaptive LFD test procedure with some existing tests, including the MANOVA tests in Schott (2007), Cai and Xia (2014), Hu et al. (2017) and Zhang et al. (2017). These competing tests are denoted by Sc, CX, HBWW and ZGZ, respectively. Throughout the simulations, we take the nominal test level $\alpha = 0.05$ and the group number $k = 3$. For the adaptive LFD test, we take $\tau = 5$. For CX, we use their oracle procedure. All the simulation results are based on 5000 replications.

First, we simulate the empirical level and power under various models of $\mathbf{\Sigma}$ and $\mathbf{\Theta}$. To characterize the signal strength, we define signal-to-noise ratio (SNR) as

$$\text{SNR} = \frac{\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C}}{\sqrt{\text{tr}(\mathbf{\Sigma}^2)}}.$$

We consider four models for $\mathbf{\Sigma}$ where the first two of them are nonspiked and the last two of them are spiked.

- Model I: $\mathbf{\Sigma} = \mathbf{I}_p$.

- Model II: $\mathbf{\Sigma} = (\sigma_{ij})$ where $\sigma_{ij} = 0.6^{|i-j|}$.

- Model III: $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ where $\mathbf{U}$ is a $p \times p$ orthogonal matrix generated from Haar distribution and $\mathbf{\Lambda} = \text{diag}(3p, 2p, p, 1, \ldots, 1)$.

- Model IV: $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top + \mathbf{A}\mathbf{A}^\top$ where $\mathbf{U}$ is a $p \times p$ orthogonal matrix generated from Haar distribution, $\mathbf{\Lambda} = \text{diag}(p, p, 1, \ldots, 1)$ and $\mathbf{A}$ is a $p \times p$ matrix whose elements are independently generated from Bernoulli distribution with success probability 0.01.

Under the null hypothesis, we shall always take $\theta_1 = \cdots = \theta_k = \mathbf{0}_p$. We consider two different structures of alternative hypotheses: the non-sparse alternative and the sparse alternative. In the non-sparse case, we take $\theta_1 = \kappa\mathbf{1}_p$, $\theta_2 = -\kappa\mathbf{1}_p$ and $\theta_3 = \mathbf{0}_p$, where κ is selected to make SNR equal to specific values. In the sparse case, we take $\theta_1 = \kappa(\mathbf{1}_{p/5}^\top, \mathbf{0}_{4p/5}^\top)^\top$, $\theta_2 = \kappa(\mathbf{0}_{p/5}^\top, \mathbf{1}_{p/5}^\top, \mathbf{0}_{3p/5}^\top)^\top$ and $\theta_3 = \mathbf{0}_p$. Again, κ is selected to make SNR equal to specific values. The simulation results are summarized in Figures 1-4. It can be seen that in all scenarios, the empirical sizes of the LFD test are reasonably close to the nominal level 0.05. Under model I and model II, where the covariance matrices are nonspiked, the empirical power of the LFD test is slightly lower than the sum-of-squares type tests, but is higher than the CX test. Under model III and model IV, where the covariance

matrices are spiked, the empirical power of the LFD test is significantly higher than the sum-of-squares type tests. Also, the LFD test offers higher empirical power than the CX test in most cases, except for model IV with sparse means. These simulation results verify our theoretical results that the LFD test is particularly powerful under the spiked covariance.

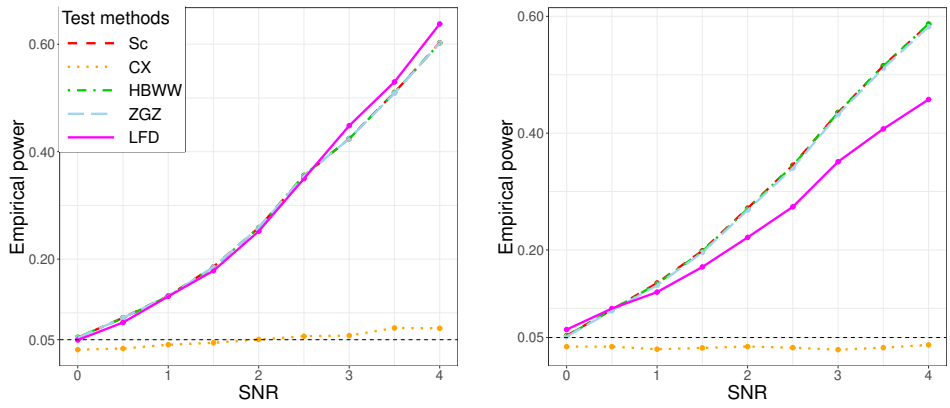
In our second simulation study, we would like to investigate the effect of correlations between variables. We consider the compound symmetry structure, that is, the diagonal elements of $\mathbf{\Sigma}$ are 1 and the off-diagonal elements are ρ with $0 \leq \rho < 1$. The parameter ρ characterizes the correlations between variables. We take $\theta_1 = \kappa(\mathbf{1}_{p/5}^\top, \mathbf{0}_{4p/5}^\top)^\top$, $\theta_2 = \kappa(\mathbf{0}_{p/5}^\top, \mathbf{1}_{p/5}^\top, \mathbf{0}_{3p/5}^\top)^\top$ and $\theta_3 = \mathbf{0}_p$, where κ is selected such that $\mathbf{C}^\top \mathbf{\Theta}^\top \mathbf{\Theta} \mathbf{C} / (\sum_{i=2}^p \boldsymbol{\lambda}_i^2)^{1/2} = 5$. Figure 5 plots the empirical powers of various tests versus ρ. We can see that the empirical power of the LFD test holds nearly constant as ρ varies while the empirical powers of competing sum-of-squares type tests decrease rapidly as ρ increases. When ρ is non-zero, the LFD test outperforms competing tests significantly.

5. Concluding remarks

In this paper, using the idea of least favorable direction, we proposed the LFD test for MANOVA in high dimensional setting. We derived the asymp-

(a) Model I, non-sparse case

(b) Model II, non-sparse case



(c) Model I, sparse case

(d) Model II, sparse case

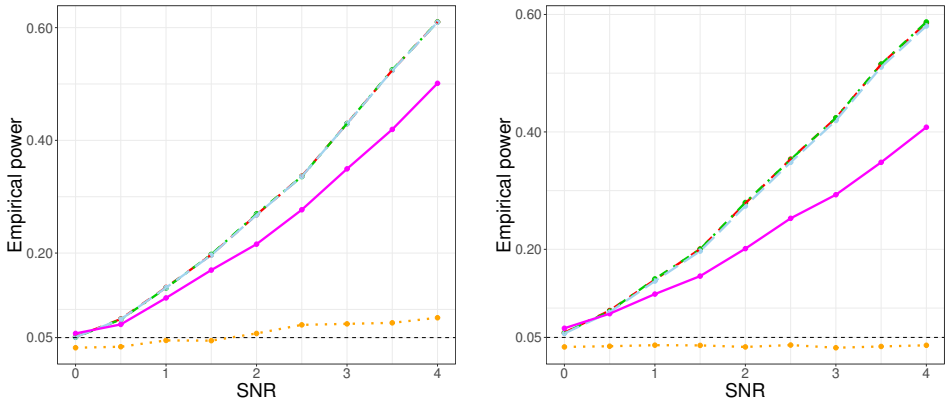
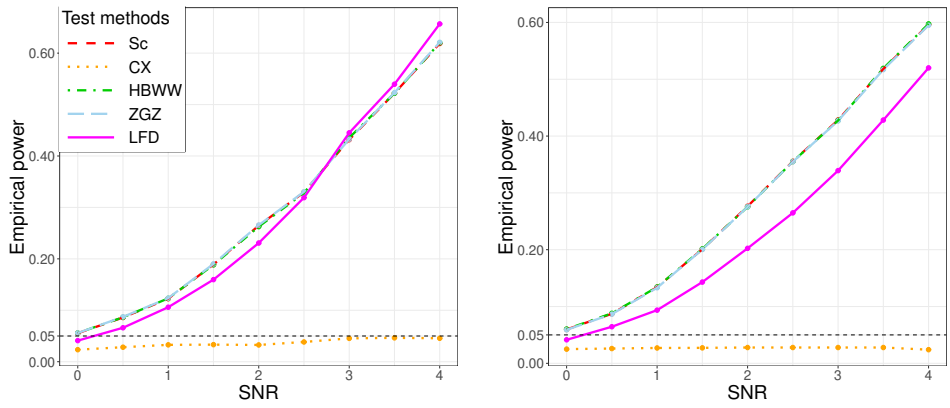


Figure 1: Empirical sizes and powers of tests under model I and model II.

$n_1 = n_2 = n_3 = 20$, $p = 300$.

(a) Model I, non-sparse case

(b) Model II, non-sparse case



(c) Model I, sparse case

(d) Model II, sparse case

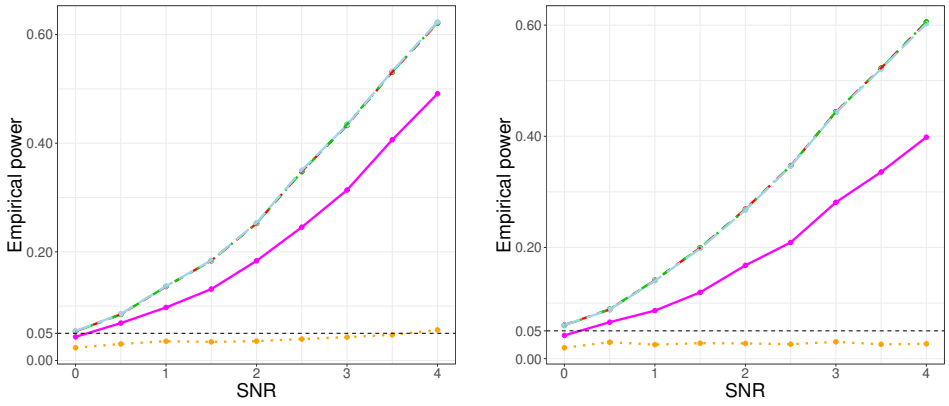
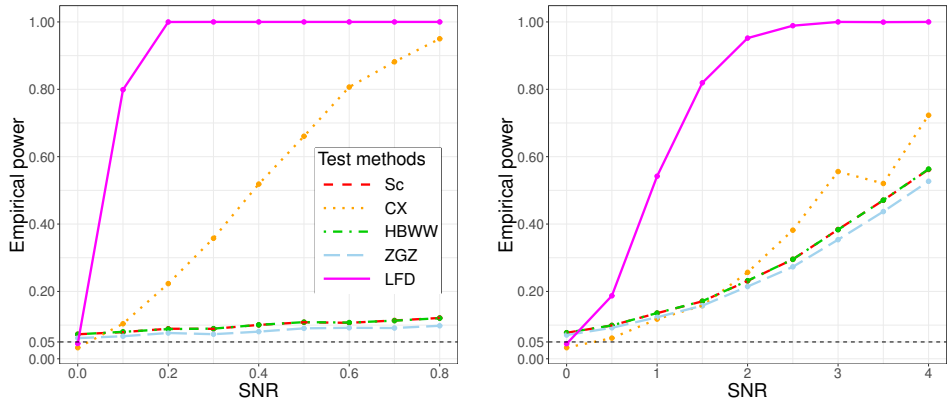


Figure 2: Empirical sizes and powers of tests under model I and model II.

$n_1 = n_2 = n_3 = 25$, $p = 800$.

(a) Model III, non-sparse case

(b) Model IV, non-sparse case



(c) Model III, sparse case

(d) Model IV, sparse case

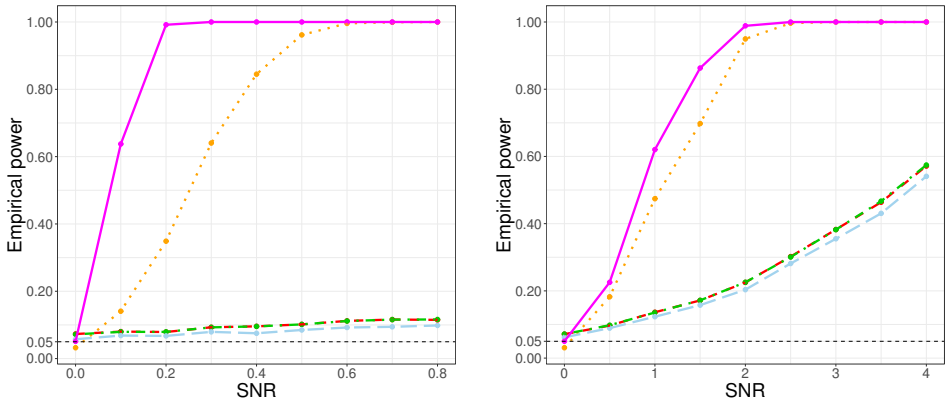
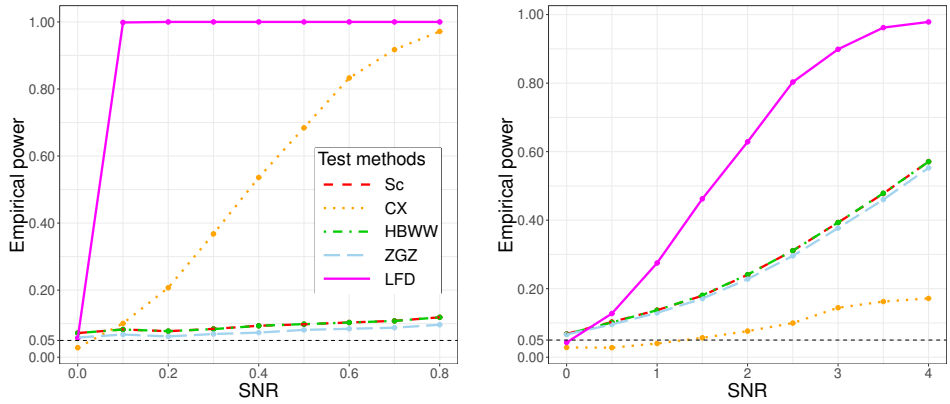


Figure 3: Empirical sizes and powers of tests under model III and model IV. $n_1 = n_2 = n_3 = 20$, $p = 300$.

(a) Model III, non-sparse case

(b) Model IV, non-sparse case



(c) Model III, sparse case

(d) Model IV, sparse case

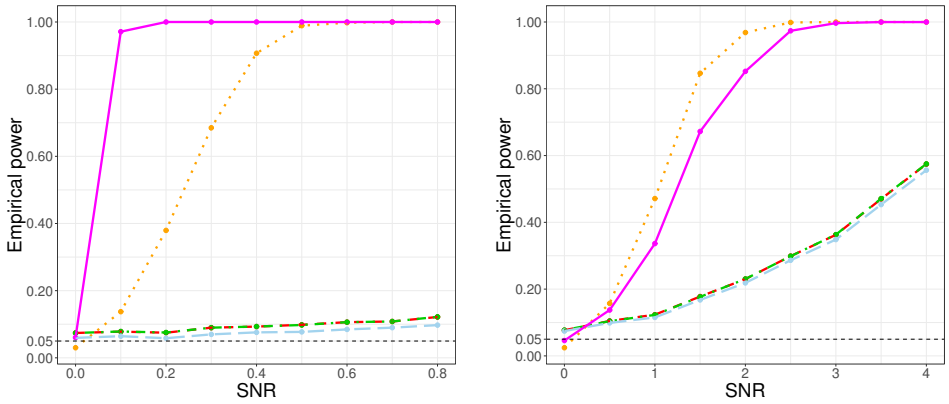


Figure 4: Empirical sizes and powers of tests under model III and model IV. $n_1 = n_2 = n_3 = 25$, $p = 800$.

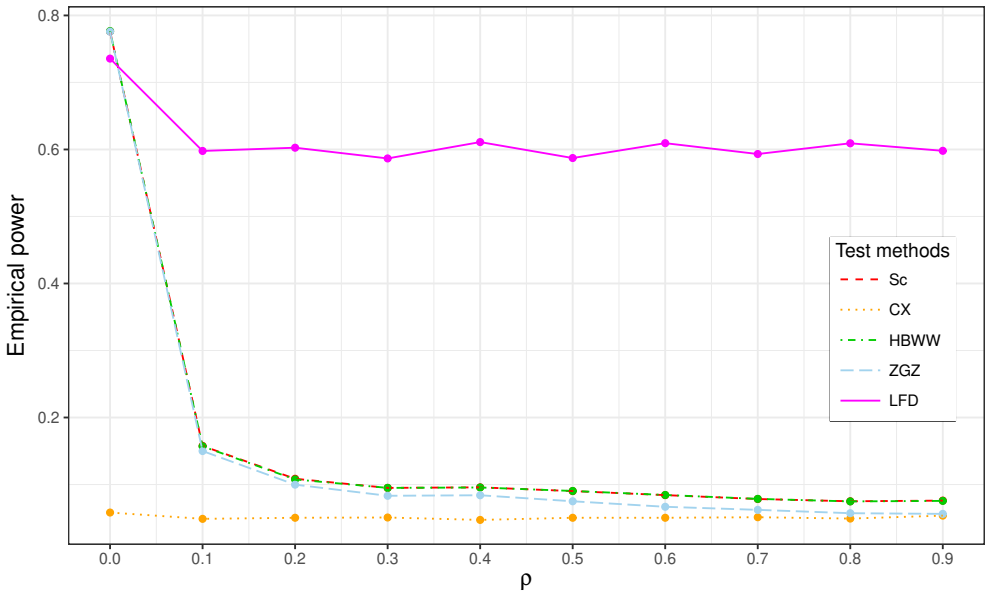


Figure 5: Empirical powers of tests. $n_1 = n_2 = n_3 = 35$, $p = 1000$.

totic distribution of the LFD test statistic under both nonspiked and spiked covariances. The asymptotic local power functions are also given. From our theoretic results and simulation studies, it is seen that the LFD test has comparable power behavior to existing tests when the covariance matrix is nonspiked, while tends to be much more powerful than existing tests when the covariance matrix is spiked.

There are several interesting but challenging problems yet to be solved. First, for the case where the covariance structure is unknown, we proposed an adaptive LFD test procedure by consistently detecting unknown covariance structure and estimating the unknown r. However, this procedure

relies on a hyperparameter τ. How to choose an optimal τ remains an interesting problem. Second, our theoretical results rely on the normality of the observations. In fact, our proofs utilize the independence of $\mathbf{XJC}$ and $\mathbf{Y}$. Note that $\mathbf{XJC}$ and $\mathbf{Y} = \mathbf{X}\tilde{\mathbf{J}}$ are both the linear combinations of indepedent random vectors X_{ij}. It is known that the independence of linear combinations of independent random variables essentially characterizes the normality of the variables (see, e.g., Kagan et al. (1973), Section 3.1). Hence our strategy is not feasible without the normality assumption. It is unclear whether the conclutions of our theorems hold without normal assumption. Third, our theoretical results require $p/n \to \infty$. In fact, the asymptotic behavior of $T(\mathbf{X})$ will be different in the regime where $p/n \to$ constant. Random matrix theory may be useful to investigate the asymptotic behavior of $T(\mathbf{X})$ in this regime. We leave these topics for future research.

Supplementary Materials

The online supplementary material presents proofs of the propositions and theorems.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant Nos. 11471035, 11471030.

References

Ahn, J. and J. S. Marron (2010). The maximal data piling direction for discrimination. *Biometrika 97*(1), 254–259.

Aoshima, M., D. Shen, H. Shen, K. Yata, Y.-H. Zhou, and J. S. Marron (2018). A survey of high dimension low sample size asymptotics. *Australian & New Zealand Journal of Statistics 60*(1), 4–19.

Aoshima, M. and K. Yata (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica 28*(1), 43–62.

Bai, Z. and H. Saranadasa (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica 6*(2), 311–329.

Cai, T., X. Han, and G. Pan (2019). Limiting Laws for Divergent Spiked Eigenvalues and Largest Non-spiked Eigenvalue of Sample Covariance Matrices. *The Annals of Statistics In press*.

Cai, T., Z. Ma, and Y. Wu (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability Theory & Related Fields 161*(3-4), 781–815.

Cai, T. T., W. Liu, and Y. Xia (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(2), 349–372.

Cai, T. T. and Y. Xia (2014). High-dimensional sparse MANOVA. *Journal of Multivariate*

Analysis 131, 174–196.

Cao, M.-X., J. Park, and D.-J. He (2019). A test for the k sample behrensfisher problem in high dimensional data. *Journal of Statistical Planning and Inference 201*, 86–102.

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

Fan, J., D. Wang, K. Wang, and Z. Zhu (2019). Distributed Estimation of Principal Eigenspaces. *The Annals of Statistics In press*.

Fan, J., L. Yuan, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Feng, L., C. Zou, Z. Wang, and L. Zhu (2015). Two-sample behrens-fisher problem for high-dimensional data. *Statistica Sinica 25*(4), 1297–1312.

Hu, J., Z. Bai, C. Wang, and W. Wang (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics 69*(2), 365–387.

Kagan, A., Y. Linnik, and C. Rao (1973). *Characterization Problems in Mathematical Statistics* (1st ed.). New York: Wiley.

Katayama, S., Y. Kano, and M. S. Srivastava (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Mul-*

tivariate Analysis 116, 410–421.

Koltchinskii, V. and K. Lounici (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 52*(4), 1976–2013.

Ma, Y., W. Lan, and H. Wang (2015). A high dimensional two-sample test under a low dimensional factor structure. *Journal of Multivariate Analysis 140*, 162–170.

Nadler, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics 36*(6), 2791–2817.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics 24*(2), 220–238.

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *Journal of Multivariate Analysis 98*(9), 1825–1839.

Shen, D., H. Shen, and J. S. Marron (2016). A general framework for consistency of principal component analysis. *Journal of Machine Learning Research 17*(150), 1–34.

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

Srivastava, M. S. and T. Kubokawa (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis 115*, 204–216.

Tsai, C.-A. and J. J. Chen (2009). Multivariate analysis of variance test for gene set analysis.

Bioinformatics 25(7), 897–903.

Verstynen, T., J. Diedrichsen, N. Albert, P. Aparicio, and R. Ivry (2005). Ipsilateral motor cortex activity during unimanual hand movements relates to task complexity. *Journal of Neurophysiology 93*(3), 1209–1222.

Wang, R. and X. Xu (2018). On two-sample mean tests under spiked covariances. *Journal of Multivariate Analysis 167*, 225 – 249.

Wang, R. and X. Xu (2019). A feasible high dimensional randomization test for the mean vector. *Journal of Statistical Planning and Inference 199*, 160–178.

Wang, W. and J. Fan (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *The Annals of Statistics 45*(3), 1342–1374.

Yamada, T. and T. Himeno (2015). Testing homogeneity of mean vectors under heteroscedasticity in high-dimension. *Journal of Multivariate Analysis 139*, 7–27.

Yata, K. and M. Aoshima (2012). Effective pca for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis 105*(1), 193 – 215.

Yata, K. and M. Aoshima (2013). Pca consistency for the power spiked model in high-dimensional settings. *Journal of Multivariate Analysis 122*, 334 – 354.

Zhang, J.-T., J. Guo, and B. Zhou (2017). Linear hypothesis testing in high-dimensional one-way manova. *Journal of Multivariate Analysis 155*, 200 – 216.

Zhao, J. and X. Xu (2016). A generalized likelihood ratio test for normal mean when p is greater than n. *Computational Statistics & Data Analysis 99*, 91–104.

Zhou, B., J. Guo, and J.-T. Zhang (2017). High-dimensional general linear hypothesis testing under heteroscedasticity. *Journal of Statistical Planning and Inference 188*, 36–54.

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081,China

E-mail: wangruiphd@bit.edu.cn

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, 100081,China

and Beijing Key Laboratory on MCAACI, Beijing Institute of Technology, Beijing 100081,China

E-mail: xuxz@bit.edu.cn