Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context

Aki Ishii\textsuperscript{a}, Kazuyoshi Yata\textsuperscript{b}, Makoto Aoshima\textsuperscript{b,1}

\textsuperscript{a}Graduate School of Pure and Applied Sciences, University of Tsukuba, Ibaraki, Japan
\textsuperscript{b}Institute of Mathematics, University of Tsukuba, Ibaraki, Japan

Abstract

A common feature of high-dimensional data is that the data dimension is high, however, the sample size is relatively low. We call such data HDLSS data. In this paper, we study asymptotic properties of the first principal component in the HDLSS context and apply them to equality tests of covariance matrices for high-dimensional data sets. We consider HDLSS asymptotic theories as the dimension grows for both the cases when the sample size is fixed and the sample size goes to infinity. We introduce an eigenvalue estimator by the noise-reduction methodology and provide asymptotic distributions of the largest eigenvalue in the HDLSS context. We construct a confidence interval of the first contribution ratio. We give asymptotic properties both for the first PC direction and PC score as well. We apply the findings to equality tests of two covariance matrices in the HDLSS context. We provide numerical results and discussions about the performances both on the estimates of the first PC and the equality tests of two covariance matrices.

Keywords: Contribution ratio, Equality test of covariance matrices, HDLSS, Noise-reduction methodology, PCA

2000 MSC: primary 34L20, secondary 62H25

Email address: aoshima@math.tsukuba.ac.jp (Makoto Aoshima)
\textsuperscript{1}Institute of Mathematics, University of Tsukuba, Ibaraki 305-8571, Japan; Fax: +81-298-53-6501

Preprint submitted to Elsevier
1. Introduction

One of the features of modern data is the data dimension $d$ is high and the sample size $n$ is relatively low. We call such data HDLSS data. In HDLSS situations such as $d/n \to \infty$, new theories and methodologies are required to develop for statistical inference based on the large sample theory. One of the approaches is to study geometric representations of HDLSS data and investigate the possibilities to make use of them in HDLSS statistical inference. Hall et al. (2005), Ahn et al. (2007), and Yata and Aoshima (2012) found several conspicuous geometric descriptions of HDLSS data when $d \to \infty$ while $n$ is fixed. The HDLSS asymptotic studies usually assume either the normality as the population distribution or a $\rho$-mixing condition as the dependency of random variables in a spherated data matrix. See Jung and Marron (2009) and Jung et al. (2012). However, Yata and Aoshima (2009) developed an HDLSS asymptotic theory without assuming those assumptions and showed that the conventional principal component analysis (PCA) cannot give consistent estimation in the HDLSS context. In order to overcome this inconvenience, Yata and Aoshima (2012) provided the noise-reduction (NR) methodology that can successfully give consistent estimators of both the eigenvalues and eigenvectors together with the principal component (PC) scores. Furthermore, Yata and Aoshima (2010, 2013) created the cross-data-matrix (CDM) methodology that is a nonparametric method to ensure consistent estimation of those quantities. Given this background, Aoshima and Yata (2011, 2013) developed a variety of inference for HDLSS data such as given-bandwidth confidence region, two-sample test, test of equality of two covariance matrices, classification, variable selection, regression, pathway analysis and so on along with the sample size determination to ensure prespecified accuracy for each inference.

In this paper, suppose we have a $d \times n$ data matrix, $X(d) = [x_{1d}, \ldots, x_{nd}]$, where $x_{jd} = (x_{j1d}, \ldots, x_{jd})^T$, $j = 1, \ldots, n$, are independent and identically distributed (i.i.d.) as a $d$-dimensional distribution with a mean vector $\mu_d$ and covariance matrix $\Sigma_d (\geq O)$. We assume $n \geq 3$. The eigen-decomposition of $\Sigma_d$ is given by $\Sigma_d = H_d \Lambda_d H_d^T$, where $\Lambda_d$ is a diagonal matrix of eigenvalues, $\lambda_1(d) \geq \cdots \geq \lambda_d(d) (\geq 0)$, and $H_d = [h_{1d}, \ldots, h_{dd}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $X(d) - [\mu_d, \ldots, \mu_d] = H_d \Lambda_d^{1/2} Z(d)$. Then, $Z(d)$ is a $d \times n$ spherated data matrix from a distribution with the zero mean and the identity covariance matrix. Here, we write $Z(d) = [z_{1d}, \ldots, z_{nd}]^T$ and $z_{jd} = (z_{j1d}, \ldots, z_{jnd})^T$, $j = 1, \ldots, d$. Note that $E(z_{ji(d)}z_{j'i(d)}) = 0$ ($j \neq j'$) and $\text{Var}(z_{jd}) = I_n$, where $I_n$ is the $n$-dimensional identity matrix. The $i$-th true PC
score of \( x_{j(d)} \) is given by \( h_{i(d)}^T(x_{j(d)} - \mu_d) = \lambda_{i(d)}^{1/2} z_{ij(d)} \) (hereafter called \( s_{ij(d)} \)). Note that \( \text{Var}(s_{ij(d)}) = \lambda_{i(d)} \) for all \( i, j \). Hereafter, the subscript \( d \) will be omitted for the sake of simplicity when it does not cause any confusion. We assume that \( \lambda_1 \) has multiplicity one in the sense that \( \lim_{d \to \infty} \lambda_1 / \lambda_2 > 1 \). Also, we assume that \( \lim_{d \to \infty} E(z_{ij}^2) < \infty \) for all \( i, j \) and \( P(\lim_{d \to \infty} ||z|| \neq 0) = 1 \). Note that if \( X \) is Gaussian, \( z_{ij} \)s are i.i.d. as the standard normal distribution, \( N(0, 1) \).

As necessary, we consider the following assumption for the normalized first PC scores, \( z_{1j} (= s_{1j} / \lambda_{1/2}) \), \( j = 1, \ldots, n \):

(A-i) \quad z_{1j}, \ j = 1, \ldots, n, \text{ are i.i.d. as } N(0, 1).

Note that \( P(\lim_{d \to \infty} ||z|| \neq 0) = 1 \) under (A-i). Let us write the sample covariance matrix as \( S = (n^{-1} - 1)(X - \bar{X})(X - \bar{X})^T = (n^{-1} - 1) \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})^T \), where \( \bar{X} = [\bar{x}, \ldots, \bar{x}] \) and \( \bar{x} = \sum_{j=1}^{n} x_j / n \). Then, we define the \( n \times n \) dual sample covariance matrix by \( S_D = (n^{-1} - 1)(X - \bar{X})(X - \bar{X})^T \). Let \( \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{n-1} \geq 0 \) be the eigenvalues of \( S_D \). Let us write the eigen-decomposition of \( S_D \) as \( S_D = \sum_{j=1}^{n-1} \hat{\lambda}_j \hat{u}_j \hat{u}_j^T \), where \( \hat{u}_j = (\hat{u}_{j1}, \ldots, \hat{u}_{jn})^T \) denotes a unit eigenvector corresponding to \( \hat{\lambda}_j \). Note that \( S \) and \( S_D \) share non-zero eigenvalues.

In this paper, we study asymptotic properties of the first principal component in the HDLSS context and apply them to equality tests of covariance matrices for high-dimensional data sets. We consider HDLSS asymptotic theories as \( d \to \infty \) for both the cases when \( n \) is fixed and \( n \to \infty \). In Section 2, we introduce an eigenvalue estimator by the NR methodology and provide asymptotic distributions of the largest eigenvalue in the HDLSS context. We construct a confidence interval of the first contribution ratio. In Section 3, we give asymptotic properties both for the first PC direction and PC score as well. In Section 4, we apply the findings to equality tests of two covariance matrices in the HDLSS context. Finally, in Section 5, we provide numerical results and discussions about the performances both on the estimates of the first PC and the equality tests of two covariance matrices.

### 2. Largest eigenvalue and its contribution rate

In this section, we give asymptotic distributions of the largest eigenvalue and construct a confidence interval of the first contribution rate.

#### 2.1. Asymptotic distributions of the largest eigenvalue

Let \( \delta_i = \text{tr}(\Sigma^2) - \sum_{s=1}^{i} \lambda_s^2 = \sum_{s=i+1}^{d} \lambda_s^2 \) for \( i = 1, \ldots, d - 1 \). We consider the following assumptions for the largest eigenvalue:
(A-ii) $\frac{\delta_i}{\lambda_i^2} = o(1)$ as $d \to \infty$ when $n$ is fixed; $\frac{\delta_i}{\lambda_i^2} = o(1)$ as $d \to \infty$ for some fixed $i_*$ $(<d)$ when $n \to \infty$.

(A-iii) $\frac{\sum_{r,s \geq 2} \lambda_r \lambda_s}{n \lambda_1^2} E \{ (z_{rk}^2 - 1)(z_{sk}^2 - 1) \} = o(1)$ as $d \to \infty$ either when $n$ is fixed or $n \to \infty$.

Note that (A-iii) holds when $X$ is Gaussian and (A-ii) is met. Let $z_{oj} = z_j - (\bar{z}_j, \ldots, \bar{z}_j)^T$, $j = 1, \ldots, p$, where $\bar{z}_j = n^{-1} \sum_{k=1}^n z_{jk}$. Let $\kappa = \text{tr}(\Sigma) / \lambda_1 = \sum_{s=2}^d \lambda_s$. Then, we have the following result.

**Proposition 2.1.** Under (A-ii) and (A-iii), it holds that

$$\frac{\hat{\lambda}_1}{\lambda_1} - \frac{||z_{o1}/\sqrt{n-1}||^2 - \frac{\kappa}{\lambda_1(n-1)}}{\lambda_1(n-1)} = o_p(1)$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$.

**Remark 2.1.** Jung et al. (2012) gave a result similar to Proposition 2.1 when $X$ is Gaussian, $\mu = 0$ and $n$ is fixed.

It holds that $E(||z_{o1}/\sqrt{n-1}||^2) = 1$ and $||z_{o1}/\sqrt{n-1}||^2 = 1 + o_p(1)$ as $n \to \infty$. If $\kappa/(n\lambda_1) = o(1)$ as $d \to \infty$ and $n \to \infty$, $\hat{\lambda}_1$ is a consistent estimator of $\lambda_1$. When $n$ is fixed, the condition $\kappa/\lambda_1 = o(1)$ is equivalent to $\lambda_1/\text{tr}(\Sigma) = 1 + o(1)$, in which the contribution ratio of the first principal component is asymptotically 1. In that sense, $\kappa/\lambda_1 = o(1)$ is quite strict condition in real high-dimensional data analyses. Hereafter, we assume $\lim \inf_{d \to \infty} \kappa/\lambda_1 > 0$.

Yata and Aoshima (2012) proposed a method for eigenvalue estimation called the noise-reduction (NR) methodology that was brought by a geometric representation of $S_D$. If one applies the NR methodology to the present case, $\lambda_i$s are estimated by

$$\tilde{\lambda}_i = \hat{\lambda}_i - \frac{\text{tr}(S_D) - \sum_{j=1}^i \hat{\lambda}_j}{n - 1 - i} \quad (i = 1, \ldots, n-2). \quad (2.1)$$

Note that $\tilde{\lambda}_i \geq 0$ w.p.1 for $i = 1, \ldots, n-2$. Also, note that the second term in (2.1) with $i = 1$ is an estimator of $\kappa/(n-1)$. See Lemma 2.1 in Section 2.2 for the details. Yata and Aoshima (2012, 2013) showed that $\tilde{\lambda}_i$ has several consistency properties when $d \to \infty$ and $n \to \infty$. On the other hand, Ishii et al. (2014) gave asymptotic properties of $\hat{\lambda}_1$ when $d \to \infty$ while $n$ is fixed. The following theorem summarizes their findings.
**Theorem 2.1.** Under (A-ii) and (A-iii), it holds that as \( d \to \infty \)
\[
\frac{\hat{\lambda}_1}{\lambda_1} = \begin{cases} 
||z_{o1}/\sqrt{n-1}||^2 + o_p(1) & \text{when } n \text{ is fixed}, \\
1 + o_p(1) & \text{when } n \to \infty.
\end{cases}
\]

Under (A-i) to (A-iii), it holds that as \( d \to \infty \)
\[
(n-1)\frac{\hat{\lambda}_1}{\lambda_1} \Rightarrow \chi^2_{n-1} \quad \text{when } n \text{ is fixed},
\]
\[
\sqrt{\frac{n-1}{2}} \left(\frac{\lambda_1}{\hat{\lambda}_1} - 1\right) \Rightarrow N(0, 1) \quad \text{when } n \to \infty.
\]

Here, \( \Rightarrow \) denotes the convergence in distribution and \( \chi^2_{n-1} \) denotes a random variable distributed as \( \chi^2 \) distribution with \( n-1 \) degrees of freedom.

2.2. Confidence interval of the first contribution ratio

We consider a confidence interval for the contribution ratio of the first principal component. Let \( a \) and \( b \) be constants satisfying \( P(a \leq \chi^2_{n-1} \leq b) = 1 - \alpha \), where \( \alpha \in (0, 1) \). Then, from Theorem 2.1, under (A-i) to (A-iii), it holds that
\[
P\left(\frac{\lambda_1}{\text{tr}(\Sigma)} \in \left[\frac{(n-1)\hat{\lambda}_1}{b\kappa + (n-1)\hat{\lambda}_1}, \frac{(n-1)\hat{\lambda}_1}{a\kappa + (n-1)\hat{\lambda}_1}\right]\right)
\]
\[
= P\left(a \leq (n-1)\frac{\hat{\lambda}_1}{\lambda_1} \leq b\right) = 1 - \alpha + o(1) \quad (2.2)
\]
as \( d \to \infty \) when \( n \) is fixed. We need to estimate \( \kappa \) in (2.2). Here, we give a consistent estimator of \( \kappa \) by \( \tilde{\kappa} = (n-1)(\text{tr}(S_D) - \hat{\lambda}_1)/(n-2) = \text{tr}(S_D) - \hat{\lambda}_1 \).

Then, we have the following results.

**Lemma 2.1.** Under (A-ii) and (A-iii), it holds that
\[
\frac{\tilde{\kappa}}{\kappa} = 1 + o_p(1) \quad \text{and} \quad \frac{\tilde{\kappa}}{\lambda_1} = \frac{\kappa}{\lambda_1} + o_p(1)
\]
as \( d \to \infty \) either when \( n \) is fixed or \( n \to \infty \).

**Theorem 2.2.** Under (A-i) to (A-iii), it holds that
\[
P\left(\frac{\lambda_1}{\text{tr}(\Sigma)} \in \left[\frac{(n-1)\hat{\lambda}_1}{b\tilde{\kappa} + (n-1)\hat{\lambda}_1}, \frac{(n-1)\hat{\lambda}_1}{a\tilde{\kappa} + (n-1)\hat{\lambda}_1}\right]\right) = 1 - \alpha + o(1) \quad (2.3)
\]
as \( d \to \infty \) when \( n \) is fixed.
Remark 2.2. From Theorem 2.1 and Lemma 2.1, under (A-ii) and (A-iii), it holds that \( \frac{\text{tr}(S_D)}{\text{tr}(\Sigma)} = (\tilde{\kappa} + \tilde{\lambda}_1)/\text{tr}(\Sigma) = 1 + o_p(1) \) as \( d \to \infty \) and \( n \to \infty \). We have that
\[
\frac{\tilde{\lambda}_1}{\text{tr}(S_D)} = \frac{\lambda_1}{\text{tr}(\Sigma)} \{1 + o_p(1)\}.
\]

Remark 2.3. The constants \((a, b)\) should be chosen for (2.3) to have the minimum length. If \( \lambda_1/\kappa = o(1) \), the length of the confidence interval becomes close to \( \{(n - 1)\tilde{\lambda}_1/\tilde{\kappa}\}\{1/a - 1/b\} \) under (A-ii) and (A-iii) when \( d \to \infty \) and \( n \) is fixed. Thus, we recommend to choose constants \((a, b)\) such that
\[
\arg\min_{a, b} (1/a - 1/b) \quad \text{subject to} \quad G_{n-1}(b) - G_{n-1}(a) = 1 - \alpha,
\]
where \( G_{n-1}(\cdot) \) denotes the c.d.f. of \( \chi_n^2 \).

Let us construct a confidence interval for the contribution ratio of the first principal component. We used gene expression data by Armstrong et al. (2002) in which the data set consists of 12582 \((= d)\) genes. The data set has three leukemia subtypes: 24 samples from acute lymphoblastic leukemia (ALL), 20 samples from mixed-lineage leukemia (MLL), and 28 samples from acute myeloid leukemia (AML). We standardized each sample so as to have the unit variance. Then, it holds \( \text{tr}(S) (= \text{tr}(S_D)) = d \), so that \( \tilde{\lambda}_1 + \tilde{\kappa} = d \). From Theorem 2.2, we constructed a 95% confidence interval of the first contribution rate for each data set by choosing \((a, b)\) as in Remark 2.3. The results are summarized in Table 1.

**Table 1.** The 95\% confidence interval (CI) of the first contribution ratio, together with \( \tilde{\lambda}_1 \) and \( \tilde{\kappa} \), for Armstrong et al. (2002)'s data sets having \( d = 12582 \).

|        | CI            | \( \tilde{\lambda}_1 \) | \( \tilde{\kappa} \) |
|--------|---------------|--------------------------|-----------------------|
| ALL \((n = 24)\) | [0.0557, 0.1663] | 1256 | 11326 |
| MLL \((n = 20)\) | [0.1201, 0.3458] | 2717 | 9865 |
| AML \((n = 28)\) | [0.0706, 0.1884] | 1501 | 11081 |

3. First PC direction and PC score

In this section, we give asymptotic properties of the first PC direction and PC score in the HDLSS context.
3.1. Asymptotic properties of the first PC direction

Let $\hat{H} = [\hat{h}_1, ..., \hat{h}_d]$, where $\hat{H}$ is a $d \times d$ orthogonal matrix of the sample eigenvectors such that $\hat{H}^T S \hat{H} = \hat{\Lambda}$ having $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_d)$. We assume $\hat{h}_i^T \hat{h}_i \geq 0$ w.p.1 for all $i$ without loss of generality. Note that $\hat{h}_i$ can be calculated by $\hat{h}_i = \{(n - 1)\hat{\lambda}_i\}^{-1/2}(X - \bar{X})\hat{u}_i$. First, we have the following result.

Lemma 3.1. Under (A-ii) and (A-iii), it holds that

$$\hat{h}_1^T h_1 - \left(1 + \frac{\kappa}{\hat{\lambda}_1||z_{o1}||^2}\right)^{-1/2} = o_p(1)$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$.

If $\kappa / (n\hat{\lambda}_1) = o(1)$ as $d \to \infty$ and $n \to \infty$, $\hat{h}_1$ is a consistent estimator of $h_1$ in the sense that $\hat{h}_1^T h_1 = 1 + o_p(1)$. When $n$ is fixed, $\hat{h}_1$ is not a consistent estimator because $\lim_{d \to \infty} \kappa / \hat{\lambda}_1 > 0$. In order to overcome this inconvenience, we consider applying the NR methodology to the PC direction vector. Let $\tilde{h}_i = \{(n - 1)\hat{\lambda}_i\}^{-1/2}(X - \bar{X})\hat{u}_i$. From Lemma 3.1, we have the following result.

Theorem 3.1. Under (A-ii) and (A-iii), it holds that

$$\tilde{h}_1^T h_1 = 1 + o_p(1)$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$.

Note that $||\tilde{h}_1||^2 = \hat{\lambda}_1 / \hat{\lambda}_1 \geq 1$ w.p.1. We emphasize that $\tilde{h}_1$ is a consistent estimator of $h_1$ in the sense of the inner product even when $n$ is fixed though $\tilde{h}_1$ is not a unit vector. We give an application of $\tilde{h}_1$ in Section 4.

3.2. Asymptotic properties of the first PC score

Let $z_{oij} = z_{ij} - \bar{z}_i$ for all $i, j$. First, we have the following result.

Lemma 3.2. Under (A-ii) and (A-iii), it holds that

$$\hat{u}_{1j} = z_{o1j} / ||z_{o1}|| + o_p(1) \quad \text{for } j = 1, ..., n$$

as $d \to \infty$ when $n$ is fixed.

Remark 3.1. By using Lemma 3.2 and the test of normality such as Jarque-Bera test, one can check whether (A-i) holds or not.
By applying the NR methodology to the first PC score, we obtain an estimate by 
\[ \tilde{s}_{1j} = \sqrt{(n-1)\lambda_1} \hat{u}_{1j}, \quad j = 1, ..., n. \] A sample mean squared error of the first PC score is given by 
\[ \text{MSE}(\tilde{s}_1) = n^{-1} \sum_{j=1}^{n} (\tilde{s}_{1j} - s_{1j})^2. \] Then, from Theorem 2.1 and Lemma 3.2, we have the following result.

**Theorem 3.2.** Under (A-ii) and (A-iii), it holds that
\[ \frac{1}{\sqrt{\lambda_1}} (\tilde{s}_{1j} - s_{1j}) = -\tilde{z}_1 + o_p(1) \quad \text{for } j = 1, ..., n \]
as \( d \to \infty \) when \( n \) is fixed. Under (A-i) to (A-iii), it holds that
\[ \sqrt{n} \lambda_1 (\tilde{s}_{1j} - s_{1j}) \Rightarrow N(0, 1) \quad \text{for } j = 1, ..., n; \quad \text{and} \quad n \frac{\text{MSE}(\tilde{s}_1)}{\lambda_1} \Rightarrow \chi_1^2 \]
as \( d \to \infty \) when \( n \) is fixed.

**Remark 3.2.** The conventional estimator of the first PC score is given by 
\[ \hat{s}_{1j} = \sqrt{(n-1)\lambda_1} \hat{u}_{1j}, \quad j = 1, ..., n. \] From Theorems 8.1 and 8.2 in Yata and Aoshima (2013), under (A-ii) and (A-iii), it holds that as \( d \to \infty \) and \( n \to \infty \)
\[ \frac{\text{MSE}(\hat{s}_1)}{\lambda_1} = o_p(1) \quad \text{if } \kappa/(n\lambda_1) = o(1), \quad \text{and} \quad \frac{\text{MSE}(\tilde{s}_1)}{\lambda_1} = o_p(1). \]

4. Equality tests of two covariance matrices

In this section, we consider the test of equality of two covariance matrices in 
the HDLSS context. Even though there are a variety of tests to deal with covariance matrices when \( d \to \infty \) and \( n \to \infty \), there seem to be no tests available in the HDLSS context such as \( d \to \infty \) while \( n \) is fixed. Suppose we have two independent \( d \times n_i \) data matrices, \( \mathbf{X}_i = [\mathbf{x}_{1(i)}, ..., \mathbf{x}_{n_i(i)}], \quad i = 1, 2, \) where \( \mathbf{x}_{j(i)}, \quad j = 1, ..., n_i, \) are i.i.d. as a \( d \)-dimensional distribution, \( \pi_i, \) having a mean vector \( \mu_i \) and covariance matrix \( \Sigma_i (\geq 0) \). We assume \( n_i \geq 3, \quad i = 1, 2. \) The eigen-decomposition of \( \Sigma_i \) is given by \( \Sigma_i = \mathbf{H}_i \Lambda_i \mathbf{H}_i^T, \) where \( \Lambda_i = \text{diag}(\lambda_{1(i)}, ..., \lambda_{d(i)}) \) having \( \lambda_{1(i)} \geq \cdots \geq \lambda_{d(i)}(\geq 0) \) and \( \mathbf{H}_i = [\mathbf{h}_{1(i)}, ..., \mathbf{h}_{d(i)}] \) is an orthogonal matrix of the corresponding eigenvectors.

8
4.1. Equality test using the largest eigenvalues

We consider the following test for the largest eigenvalues:

\[ H_0 : \lambda_1(1) = \lambda_1(2) \quad \text{vs.} \quad H_a : \lambda_1(1) \neq \lambda_1(2) \quad \text{(or} \quad H_b : \lambda_1(1) < \lambda_1(2)) \quad (4.1) \]

Let \( \tilde{\lambda}_{1(i)} \) be the estimate of \( \lambda_{1(i)} \) by the NR methodology as in (2.1) for \( \pi_i \). Let \( \nu_1 = n_1 - 1 \) and \( \nu_2 = n_2 - 1 \). From Theorem 2.1, we have the following result.

**Corollary 4.1.** Under (A-i) to (A-iii) for each \( \pi_i \), it holds that

\[ \frac{\tilde{\lambda}_{1(1)}}{\lambda_{1(1)}} \frac{\tilde{\lambda}_{1(2)}}{\lambda_{1(2)}} \Rightarrow F_{\nu_{1},\nu_{2}} \]

as \( d \to \infty \) when \( n_i \)s are fixed, where \( F_{\nu_{1},\nu_{2}} \) denotes a random variable distributed as \( F \) distribution with degrees of freedom, \( \nu_1 \) and \( \nu_2 \).

Let \( F_1 = \frac{\tilde{\lambda}_{1(1)}}{\tilde{\lambda}_{1(2)}} \). From Corollary 4.1, we test (4.1) for given \( \alpha \in (0, 1/2) \) by

\[
\begin{align*}
\text{accepting } H_a & \iff F_1 \notin \left[ F_{\nu_{1},\nu_{2}}(\alpha/2) \right]^{-1}, \left[ F_{\nu_{1},\nu_{2}}(\alpha/2) \right]^{-1} \\
\text{or } \text{accepting } H_b & \iff F_1 < \left[ F_{\nu_{1},\nu_{2}}(\alpha) \right]^{-1},
\end{align*}
\]

(4.2) (4.3)

where \( F_{\nu,\nu}(\alpha) \) denotes the upper \( \alpha \% \) point of \( F \) distribution with degrees of freedom, \( \nu_1 \) and \( \nu_2 \). Then, under (A-i) to (A-iii) for each \( \pi_i \), it holds that

\[ \text{size} = \alpha + o(1) \]

as \( d \to \infty \) when \( n_i \)s are fixed.

Now, we check the performance of the test by (4.2) or (4.3). We also consider a test by the conventional estimator, \( \hat{\lambda}_{1(i)} \). Let \( \kappa_i = \text{tr}(\Sigma_i) - \lambda_{1(i)} = \sum_{s=2}^{d} \lambda_s(i) \) for \( i = 1, 2 \). From Proposition 2.1, if \( \kappa_{i}/\lambda_{1(i)} = o(1), i = 1, 2 \), under (A-i) to (A-iii) for each \( \pi_i \) it holds that

\[ \frac{\hat{\lambda}_{1(1)}}{\lambda_{1(1)}} \frac{\hat{\lambda}_{1(2)}}{\lambda_{1(2)}} \Rightarrow F_{\nu_{1},\nu_{2}} \]

as \( d \to \infty \) when \( n_i \)s are fixed. As mentioned in Section 2, the condition \( \kappa_{i}/\lambda_{1(i)} = o(1) \) for \( i = 1, 2 \) is quite strict in real high-dimensional data analyses. Hereafter, we assume \( \liminf_{d \to \infty} \kappa_{i}/\lambda_{1(i)} > 0 \) for \( i = 1, 2 \). We analyzed the same gene expression data as in Table 1. We set \( \alpha = 0.05 \). We considered two cases: (I)
\( \pi_1 \): ALL \((n_1 = 24)\) and \( \pi_2 \): MLL \((n_2 = 20)\), and (II) \( \pi_1 \): AML \((n_1 = 28)\) and \( \pi_2 \): MLL \((n_2 = 20)\). As for \( F'_1 = \hat{\lambda}_{1(1)}/\hat{\lambda}_{1(2)} \), we considered (4.2) and (4.3) by replacing \( F_1 \) with \( F'_1 \). The results are summarized in Table 2. We observed from Table 2 that only \( H_0 \) for (I) was accepted by \( F_1 \), namely, only \( F_1 \) for (I) rejected \( H_0 \) vs. \( H_b \). One should note that the condition ‘\( \kappa_i/\lambda_{1(i)} = o(1) \) for \( i = 1, 2 \)’ does not hold both for (I) and (II) as observed in Table 1.

**Table 2.** Tests of \( H_0 : \lambda_{1(1)} = \lambda_{1(2)} \) vs. \( H_a : \lambda_{1(1)} \neq \lambda_{1(2)} \) or \( H_b : \lambda_{1(1)} < \lambda_{1(2)} \) with size 0.05 for Armstrong et al. (2002)’s data sets having \( d = 12582 \).

|                | \( H_a \) by \( F_1 \) | \( H_a \) by \( F'_1 \) | \( H_b \) by \( F_1 \) | \( H_b \) by \( F'_1 \) |
|----------------|-----------------|-----------------|-----------------|-----------------|
| (I) \( \pi_1 \): ALL, \( \pi_2 \): MLL | Reject          | Reject          | Accept          | Reject          |
| (II) \( \pi_1 \): AML, \( \pi_2 \): MLL | Reject          | Reject          | Reject          | Reject          |

4.2. Equality test using the largest eigenvalues and their PC directions

We consider the following test using the largest eigenvalues and their PC directions:

\[
H_0 : (\lambda_{1(1)}, h_{1(1)}) = (\lambda_{1(2)}, h_{1(2)}) \quad \text{vs.} \quad H_a : (\lambda_{1(1)}, h_{1(1)}) \neq (\lambda_{1(2)}, h_{1(2)}).
\]

(4.4)

Let \( \tilde{h}_{1(i)} \) be the estimator of the first PC direction for \( \pi_i \) by the NR methodology given in Section 3.1. We assume \( h_{1(i)}^T \tilde{h}_{1(i)} \geq 0 \) w.p.1 for \( i = 1, 2 \), without loss of generality. Here, we have the following result.

**Lemma 4.1.** Under (A-ii) and (A-iii) for each \( \pi_i \), it holds that

\[
\tilde{h}_{1(1)}^T \tilde{h}_{1(2)} = h_{1(1)}^T h_{1(2)} + o_p(1)
\]

as \( d \to \infty \) either when \( n_i \) is fixed or \( n_i \to \infty \) for \( i = 1, 2 \).

Let \( \tilde{h} = |\tilde{h}_{1(1)}^T \tilde{h}_{1(2)}|/2 + |\tilde{h}_{1(1)}^T \tilde{h}_{1(2)}|^{-1}/2 \). Note that \( \tilde{h} \geq 1 \). Then, from Lemma 4.1, we give a test statistic for (4.4) as follows:

\[
F_2 = \frac{\lambda_{1(1)}}{\lambda_{1(2)}} \tilde{h}_*,
\]

where

\[
\tilde{h}_* = \begin{cases} 
\tilde{h} & \text{if } \lambda_{1(1)} \geq \lambda_{1(2)}, \\
\tilde{h}^{-1} & \text{otherwise}.
\end{cases}
\]

From Lemma 4.1, we have the following result.
Theorem 4.1. Under (A-i) to (A-iii) for each $\pi_i$, it holds that

$$F_2 \Rightarrow F_{\nu_1,\nu_2} \text{ under } H_0$$

as $d \to \infty$ when $n_i$s are fixed.

From Theorem 4.1, we consider testing (4.4) by (4.2) with $F_2$ instead of $F_1$. Then, the size becomes close to $\alpha$ as $d$ increases. For the same gene expression data sets as in Section 4.1, we tested (4.4) with $\alpha = 0.05$ for the cases of (I) and (II). We observed that only $H_a$ for (II) was accepted by $F_2$, namely, only $F_2$ for (II) rejected $H_0$ vs. $H_a$ in (4.4).

4.3. Equality test of the covariance matrices

We consider the following test for the covariance matrices:

$$H_0 : \Sigma_1 = \Sigma_2 \text{ vs. } H_a : \Sigma_1 \neq \Sigma_2. \quad (4.5)$$

When $d \to \infty$ and $n_i$s are fixed, one cannot estimate $\lambda_{j(i)}$s and $h_{j(i)}$s for $j = 2, \ldots, d$. Instead, we consider estimating $\kappa_i$s. Let $S_D(i)$ be the dual sample covariance matrix for $\pi_i$. We estimate $\kappa_i$ by $\tilde{\kappa}_i = \text{tr}(S_{D(i)}) - \tilde{\lambda}_{1(i)}$ for $i = 1, 2$. From Lemma 2.1, under (A-ii) and (A-iii) for each $\pi_i$, $\tilde{\kappa}_i$s are consistent estimators of $\kappa_i$s in the sense that $\tilde{\kappa}_i/\kappa_i = 1 + o_p(1)$ as $d \to \infty$ when $n_i$s are fixed. Let $\tilde{\gamma} = \max\{\tilde{\kappa}_1/\tilde{\kappa}_2, \tilde{\kappa}_2/\tilde{\kappa}_1\}$. Now, we give a test statistic for (4.5) as follows:

$$F_3 = \frac{\tilde{\lambda}_{1(1)}}{\tilde{\lambda}_{1(2)}} \tilde{h}_* \tilde{\gamma}_*,$$

where

$$\tilde{\gamma}_* = \begin{cases} \tilde{\gamma} & \text{if } \tilde{\lambda}_{1(1)} \geq \tilde{\lambda}_{1(2)}; \\ \tilde{\gamma}^{-1} & \text{otherwise}. \end{cases}$$

Then, we have the following result.

Theorem 4.2. Under (A-i) to (A-iii) for each $\pi_i$, it holds that

$$F_3 \Rightarrow F_{\nu_1,\nu_2} \text{ under } H_0$$

as $d \to \infty$ when $n_i$s are fixed.
From Theorem 4.2, we consider testing (4.5) by (4.2) with $F_3$ instead of $F_1$. Then, the size becomes close to $\alpha$ as $d$ increases. For the same gene expression data sets as in Section 4.1, we tested (4.5) with $\alpha = 0.05$ for the cases of (I) and (II). We compared the performance of $F_3$ with two other test statistics: $Q^2$ and $T^2$ by Srivastava and Yanagihara (2010). The results are summarized in Table 3. We observed that $H_a$ was accepted by $F_3$ both for (I) and (II), namely, $F_3$ rejected $H_0$ vs. $H_a$ in (4.5) for both the cases. On the other hand, $Q^2$ and $T^2$ did not work for these data sets. It should be noted that $Q^2$ and $T^2$ require to meet the conditions that $0 < \lim_{d \to \infty} \text{tr}(\Sigma) / d < \infty$ ($i = 1, \ldots, 4$) and $d^{1/2} / n = o(1)$. As observed in Table 1, the conditions seem not to hold for these data sets with $d = 12582$ and $n \leq 28$. Hence, there is no theoretical guarantee for the results by $Q^2$ and $T^2$.

Table 3. Tests of $H_0 : \Sigma_1 = \Sigma_2$ vs. $H_a : \Sigma_1 \neq \Sigma_2$ with size 0.05 for Armstrong et al. (2002)'s data sets having $d = 12582$.

| Case | $H_a$ by $F_3$ | $H_a$ by $Q^2$ | $H_a$ by $T^2$ |
|------|----------------|----------------|----------------|
| (I) $\pi_1$: ALL, $\pi_2$: MLL | Accept | Reject | Reject |
| (II) $\pi_1$: AML, $\pi_2$: MLL | Accept | Reject | Reject |

5. Numerical results and discussions

5.1. Comparisons of the estimates on the first PC

In this section, we compared the performance of $\hat{\lambda}_1$, $\hat{h}_1$ and $\tilde{z}_{ij}$ with their conventional counterparts by Monte Carlo simulations. We set $d = 2^k$, $k = 3, \ldots, 11$ and $n = 10$. We considered two cases for $\lambda_i$: (a) $\lambda_i = d^{1/i}, i = 1, \ldots, d$ and (b) $\lambda_i = d^{3/(2+2i)}, i = 1, \ldots, d$. Note that $\lambda_1 = d$ for (a) and $\lambda_1 = d^{3/4}$ for (b). Also, note that (A-ii) holds both for (a) and (b). Let $d_* = \lceil d^{1/2} \rceil$, where $[x]$ denotes the smallest integer $\geq x$. We considered a non-Gaussian distribution as follows: $(z_{1j}, \ldots, z_{d-d,j})^T, j = 1, \ldots, n$, are i.i.d. as $N_{d-d_*}(0, I_{d-d_*})$ and $(z_{d-d_*+1j}, \ldots, z_{dj})^T, j = 1, \ldots, n$, are i.i.d. as the $d_*$-variate t-distribution, $t_{d_*}(0, I_{d_*}, 10)$ with mean zero, covariance matrix $I_{d_*}$ and degrees of freedom 10, where $(z_{1j}, \ldots, z_{d-d_*})^T$ and $(z_{d-d_*+1j}, \ldots, z_{dj})^T$ are independent for each $j$. Note that (A-i) and (A-iii) hold both for (a) and (b) from the fact that $\sum_{r,s \geq 2} \lambda_r \lambda_s E \{ (z_{rk}^2 - 1) (z_{sk}^2 - 1) \} = 2 \sum_{s=2}^{d-d_*} \lambda_s^2 + O(\sum_{r,s \geq 2}^{d-d_*+1} \lambda_r \lambda_s) = o(\lambda_1^2)$.

The findings were obtained by averaging the outcomes from 2000 ($= R$, say) replications. Under a fixed scenario, suppose that the $r$-th replication ends with
The values of $A: \hat{\lambda}/\lambda_1$ and $B: \tilde{\lambda}/\lambda_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The asymptotic variance of $\tilde{\lambda}/\lambda_1$ was given by $\text{Var}\left\{\chi^2_{n-1}/(n-1)\right\} = 0.222$.

Figure 1. The values of $A: \hat{\lambda}/\lambda_1$ and $B: \tilde{\lambda}/\lambda_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of $A: \text{var}(\hat{\lambda}/\lambda_1)$ and $B: \text{var}(\tilde{\lambda}/\lambda_1)$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The asymptotic variance of $\tilde{\lambda}/\lambda_1$ was given by $\text{Var}\left\{\chi^2_{n-1}/(n-1)\right\} = 0.222$ and denoted by the solid line in the left panel. We observed that the sample mean and variance of $\tilde{\lambda}/\lambda_1$ become close to those asymptotic values as $d$ increases.

Similarly, we plotted $(\hat{\lambda}_r, \hat{h}_r, \text{MSE}(\hat{s}_1)_r)$ and $(\tilde{\lambda}_r, \tilde{h}_r, \text{MSE}(\tilde{s}_1)_r)$ $(r = 1, \ldots, R)$. Let us simply write $\hat{\lambda}_1 = R^{-1}\sum_{r=1}^{R} \hat{\lambda}_r$ and $\tilde{\lambda}_1 = R^{-1}\sum_{r=1}^{R} \tilde{\lambda}_r$. We also considered the Monte Carlo variability by $\text{var}(\hat{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R} (\hat{\lambda}_r - \hat{\lambda}_1)^2/\lambda_1^2$ and $\text{var}(\tilde{\lambda}_1/\lambda_1) = (R-1)^{-1}\sum_{r=1}^{R} (\tilde{\lambda}_r - \tilde{\lambda}_1)^2/\lambda_1^2$. Figure 1 shows the behaviors of $(\hat{\lambda}_1/\lambda_1, \tilde{\lambda}_1/\lambda_1)$ in the left panel and $(\text{var}(\hat{\lambda}_1/\lambda_1), \text{var}(\tilde{\lambda}_1/\lambda_1))$ in the right panel for (a) and (b). We gave the asymptotic variance of $\hat{\lambda}_1/\lambda_1$ by $\text{Var}\left\{\chi^2_{n-1}/(n-1)\right\} = 0.222$ from Theorem 2.1 and showed it by the solid line in the right panel.

We also gave the asymptotic variance of $\text{MSE}(\tilde{s}_1)/\lambda_1$ by $\text{Var}(\chi^2_{n-1}/n) = 0.02$ in the right panel of Figure 3. Throughout, the estimators by the NR method gave good performances both for (a) and (b) when $d$ is large. However, the conventional estimators gave poor performances especially for (b). This is probably because the bias of the conventional estimators, $\kappa/2(n\lambda_1)$, is large for (b) compared to (a). See Proposition 2.1 for the details.
Figure 2. The values of A: $\hat{h}_1^T h_1$ and B: $\tilde{h}_1^T h_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of A: var($\tilde{h}_1^T h_1$) and B: var($\hat{h}_1^T h_1$) are denoted by the dashed lines for (a) and by the dotted lines for (b) in the right panel.

Figure 3. The values of A: $\text{MSE}((\hat{s}_1)/\lambda_1$ and B: $\text{MSE}((\tilde{s}_1)/\lambda_1$ are denoted by the dashed lines for (a) and by the dotted lines for (b) in the left panel. The values of A: var(MSE($\hat{s}_1)/\lambda_1$) and B: var(MSE($\tilde{s}_1)/\lambda_1$) are denoted by the dashed lines for (a) and by the dotted lines for (b) in the right panel. The asymptotic mean and variance of $\text{MSE}((\tilde{s}_1)/\lambda_1$ were given by $E(\chi_1^2/n) = 0.1$ and Var($\chi_1^2/n) = 0.02$ and denoted by the solid lines in both the panels.
5.2. Equality tests of two covariance matrices

We used computer simulations to study the performance of the test procedures by \( F_1 \) for (4.1), \( F_2 \) for (4.4) and \( F_3 \) for (4.5). We set \( \alpha = 0.05 \). Independent pseudo-random normal observations were generated from \( \pi_i : N_d(0, \Sigma_i) \), \( i = 1, 2 \). We set \((n_1, n_2) = (10, 20)\). We considered the cases: \( d = 2^k, k = 3, \ldots, 11, \) and

\[
\Sigma_i = \begin{pmatrix}
\Sigma_{i(1)} & O_{d-2,2} \\
O_{d-2,2} & \Sigma_{i(2)}
\end{pmatrix}, \ i = 1, 2,
\]  

(5.1)

where \( O_{k,l} \) is the \( k \times l \) zero matrix, \( \Sigma_{1(1)} = \text{diag}(d^{3/4}, d^{1/2}) \) and \( \Sigma_{1(2)} = (0.3^{l-t}) \). When considered the alternative hypotheses, we set

\[
\Sigma_{2(1)} = \left( \begin{array}{cc}
1/3 & \sqrt{8}/3 \\
\sqrt{8}/3 & -1/3
\end{array} \right) \text{diag}(3d^{3/4}, 1.5d^{1/2}) \left( \begin{array}{cc}
1/3 & \sqrt{8}/3 \\
\sqrt{8}/3 & -1/3
\end{array} \right)
\]  

(5.2)

and \( \Sigma_{2(2)} = 1.5(0.3^{l-t}) \). Note that \( \lambda_{1(2)}/\lambda_{1(1)} = 3, \ \kappa_2/\kappa_1 = 1.5, \ \mathbf{h}_{1(1)} = (1, 0, \ldots, 0)^T \) and \( \mathbf{h}_{1(2)} = (1/3, \sqrt{8}/3, 0, \ldots, 0)^T \), so that \( \mathbf{h}^T_{1(1)} \mathbf{h}_{1(2)} = 1/3 \). Also, note that (A-i) to (A-iii) hold for each \( \pi_i \). Let \( h = (|\mathbf{h}^T_{1(1)} \mathbf{h}_{1(2)}| + 1/|\mathbf{h}^T_{1(1)} \mathbf{h}_{1(2)}|)/2 \) and \( \gamma = \max\{\kappa_1/\kappa_2, \kappa_2/\kappa_1\} \). From Lemmas 2.1 and 4.1, it holds that \( \hat{h} = h + o_p(1) \) and \( \hat{\gamma} = \gamma + o_p(1) \). Thus, from Corollary 4.1, Theorems 4.1 and 4.2, we obtained the asymptotic powers of \( F_1, F_2 \) and \( F_3 \) with \( (\hat{h}_*, \hat{\gamma}_*) = (h^{-1}, \gamma^{-1}) \) as follows:

- \[
\text{Power}(F_1) = P\{ (\lambda_{1(1)}/\lambda_{1(2)}) f \notin [F_{\nu_2, \nu_1}(\alpha/2)]^{-1}, F_{\nu_1, \nu_2}(\alpha/2)] \} = 0.39,
\]

- \[
\text{Power}(F_2) = P\{ h^{-1}(\lambda_{1(1)}/\lambda_{1(2)}) f \notin [F_{\nu_2, \nu_1}(\alpha/2)]^{-1}, F_{\nu_1, \nu_2}(\alpha/2)] \} = 0.726
\]

and \[
\text{Power}(F_3) = P\{ \gamma^{-1}h^{-1}(\lambda_{1(1)}/\lambda_{1(2)}) f \notin [F_{\nu_2, \nu_1}(\alpha/2)]^{-1}, F_{\nu_1, \nu_2}(\alpha/2)] \}
\]

\[
= 0.908,
\]

where \( f \) denotes a random variable distributed as \( F \) distribution with degrees of freedom, \( \nu_1 \) and \( \nu_2 \). Note that \( \text{Power}(F_2) \) and \( \text{Power}(F_3) \) give lower bounds of the asymptotic powers when \( \hat{h}_* = h^{-1} \) and \( \hat{\gamma}_* = \gamma^{-1} \).

In Figure 4, we summarized the findings obtained by averaging the outcomes from 4000 (= \( R \), say) replications. Here, the first 2000 replications were generated by setting \( \Sigma_2 = \Sigma_1 \) as in (5.1) and the last 2000 replications were generated by setting \( \Sigma_2 \) as in (5.2). Let \( F_{ir} \ (i = 1, 2, 3) \) be the \( r \)th observation of \( F_i \) for \( r = 1, \ldots, 4000 \). We defined \( P_r = 1 \) (or 0) when \( H_0 \) was falsely rejected (or not) for \( r = 1, \ldots, 2000 \), and \( H_a \) was falsely rejected (or not) for \( r = 2001, \ldots, 4000 \). We defined \( \overline{P} = (R/2)^{-1} \sum_{r=1}^{R/2} P_r \) to estimate the size and
Sizes of $F_1$, $F_2$ and $F_3$

Powers of $F_1$, $F_2$ and $F_3$

Figure 4. The values of $\alpha$ are denoted by the dashed lines in the left panel and the values of $1 - \beta$ are denoted by the dashed lines in the right panel for $F_1$, $F_2$ and $F_3$. The asymptotic powers were given by $\text{Power}(F_1) = 0.39$, $\text{Power}(F_1) = 0.726$ and $\text{Power}(F_3) = 0.908$ which were denoted by the solid lines in the right panel.

$1 - \beta = 1 - (R/2)^{-1}\sum_{r=R/2+1} P_r$ to estimate the power. Their standard deviations are less than 0.011. Throughout, the tests gave adequate performances for the high-dimensional cases.

Appendix A.

Throughout, let $P_n = I_n - 1_n1_n^T/n$, where $1_n = (1, \ldots, 1)^T$. Let $e_n = (e_1, \ldots, e_n)^T$ be an arbitrary (random) $n$-vector such that $||e_n|| = 1$ and $e_n^T1_n = 0$.

Proof of Proposition 2.1. We assume $\mu = 0$ without loss of generality. We write that $X^T X = \sum_{s=1}^{i_*} \lambda_s z_s z_s^T + \sum_{s=i_*+1}^{d} \lambda_s z_s z_s^T$ for $i_* = 1$ when $n$ is fixed, and for some fixed $i_* (\geq 1)$ when $n \to \infty$. Here, by using Markov’s inequality, for any $\tau > 0$, under (A-ii) and (A-iii), we have that

$$P\left\{ \sum_{j=1}^{n} \left( \sum_{s=i_*+1}^{d} \frac{\lambda_s (z_{sj}^2 - 1)}{n\lambda_1} \right)^2 > \tau \right\} \leq \sum_{r,s \geq 2} \lambda_r \lambda_s E\left\{ (z_{rk}^2 - 1)(z_{sk}^2 - 1) \right\} \frac{\tau n\lambda_1^2}{\lambda_1^2} \to 0$$

and

$$P\left\{ \sum_{j \neq j'}^{n} \left( \sum_{s=i_*+1}^{d} \frac{\lambda_s z_{sj} z_{sj'}}{n\lambda_1} \right)^2 > \tau \right\} \leq \frac{\delta_{i_*}}{\tau \lambda_1^2} \to 0$$
as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. Note that $\sum_{j=1}^{n} e_j^4 \leq 1$ and $\sum_{j \neq j'} e_j^2 e_{j'}^2 \leq 1$. Then, under (A-ii) and (A-iii), we have that

$$
\left| \sum_{j=1}^{n} e_j^2 \sum_{s=i+1}^{d} \frac{\lambda_s (z_{s,j}^2 - 1)}{n \lambda_1} \right| \leq \left\{ \sum_{j=1}^{n} e_j^4 \right\}^{1/2} \left\{ \sum_{j=1}^{n} \left( \sum_{s=i+1}^{d} \frac{\lambda_s (z_{s,j}^2 - 1)}{n \lambda_1} \right)^2 \right\}^{1/2}
= o_p(1)
$$

and

$$
\left| \sum_{j \neq j'} e_j e_{j'} \sum_{s=i+1}^{d} \frac{\lambda_s z_{s,j} z_{s,j'}}{n \lambda_1} \right| \leq \left\{ \sum_{j \neq j'} e_j e_{j'}^2 \right\}^{1/2} \left\{ \sum_{j \neq j'} \left( \sum_{s=i+1}^{d} \frac{\lambda_s z_{s,j} z_{s,j'}}{n \lambda_1} \right)^2 \right\}^{1/2}
= o_p(1)
$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. Thus, we claim that

$$
e_n^T X^T X e_n = e_n^T \sum_{s=1}^{i_s} \lambda_s z_s z_s^T e_n + \frac{\kappa}{(n - 1) \lambda_1} + o_p(1)
$$

(A.1)

from the fact that $\sum_{s=i+1}^{d} \lambda_s / (n \lambda_1) = \kappa / (n \lambda_1) + o(1)$ when $n \to \infty$. Note that $e_n^T P_n = e_n^T$ and $P_n z_s = z_{os}$ for all $s$. Also, note that $(z_{os}/n^{1/2})^T (z_{os}/n^{1/2}) = o_p(1)$ for $s \neq s'$ as $n \to \infty$ from the fact that $E \{(z_{os}/n)^2\} = o(1)$ as $n \to \infty$. Then, by noting that $P(\lim_{d \to \infty} ||z_{o1}|| \neq 0) = 1$, $\lim_{d \to \infty} \lambda_1/\lambda_2 > 1$ and $z_{o1}^T 1_n = 0$, it holds that

$$
\max_{e_n} \left\{ e_n^T \sum_{s=1}^{i_s} \lambda_s z_s z_s^T e_n \right\} = \max_{e_n} \left\{ e_n^T \sum_{s=1}^{i_s} \lambda_s z_{os} z_{os}^T e_n \right\} = ||z_{o1}/\sqrt{n-1}||^2 + o_p(1)
$$

(A.2)

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. Note that $\hat{u}_1^T 1_n = 0$ and $\hat{u}_1^T P_n = \hat{u}_1^T$ when $S_D \neq O$. Then, from (A.1), (A.2) and $P_n X^T X P_n/(n - 1) = S_D$, under (A-ii) and (A-iii), we have that

$$
\hat{u}_1^T S_D \hat{u}_1 = \hat{u}_1^T X^T X \hat{u}_1 = ||z_{o1}/\sqrt{n-1}||^2 + \frac{\kappa}{(n - 1) \lambda_1} + o_p(1)
$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. It concludes the result. □

Proof of Lemma 2.1. By using Markov’s inequality, for any $\tau > 0$, under (A-ii)
and (A-iii), we have that

\[
P\left\{ \frac{\sum_{s=2}^{d} \lambda_s \{ ||z_{os}||^2 - (n-1) \}}{(n-1)\lambda_1} > \tau \right\}
\]

\[
= P\left\{ \frac{\sum_{s=2}^{d} \lambda_s \{ (n-1) \sum_{k=1}^{n} \left( z_{sk}^2 - 1 \right)/n - \sum_{k\neq k'}^{n} z_{sk} z_{sk'}/n \}}{(n-1)\lambda_1} > \tau \right\}
\]

\[
= O\left\{ \frac{\sum_{r,s \geq 2} \lambda_r \lambda_s E \left\{ (z_{rk}^2 - 1)(z_{sk}^2 - 1) \right\}}{n\lambda_1^2} + O\left\{ \frac{\delta_1}{(n\lambda_1)^2} \right\} \right\} \rightarrow 0
\]
as \(d \rightarrow \infty\) either when \(n\) is fixed or \(n \rightarrow \infty\). Thus it holds that \(\text{tr}(S_D)/\lambda_1 = \kappa/\lambda_1 + ||z_{o1}/\sqrt{n-1}||^2 + o_p(1)\) from the fact that \(\text{tr}(S_D) = \lambda_1 ||z_{o1}||^2/(n-1) + \sum_{s=2}^{d} \lambda_s ||z_{os}||^2/(n-1)\). Then, from Proposition 2.1 and \(\lim \inf_{d \rightarrow \infty} \kappa/\lambda_1 > 0\), we can claim the results.

**Proof of Theorem 2.1.** When \(n \rightarrow \infty\), we can claim the results from Theorems 4.1, 4.2 and Corollary 4.1 in Yata and Aoshima (2013). When \(n\) is fixed, by combining Proposition 2.1 with Lemma 2.1, we can claim the results because \(||z_{o1}||^2 = \sum_{k=1}^{n} z_{ik}^2 - n\bar{z}_1^2\) is distributed as \(\chi_{n-1}^2\) if \(z_{1j}, j = 1, ..., k,\) are i.i.d. as \(N(0, 1)\).

**Proof of Theorem 2.2.** From Theorem 2.1 and Lemma 2.1, under (A-i) to (A-iii), it holds that

\[
P\left( \frac{\lambda_1}{\text{tr}(\Sigma)} \in \left[ \frac{(n-1)\bar{\lambda}_1}{b\kappa + (n-1)\lambda_1}, \frac{(n-1)\bar{\lambda}_1}{a\kappa + (n-1)\lambda_1} \right] \right)
\]

\[
= P\left( \frac{(n-1)\bar{\lambda}_1}{b\kappa + (n-1)\lambda_1} \leq \frac{\lambda_1}{\text{tr}(\Sigma)} \leq \frac{(n-1)\bar{\lambda}_1}{a\kappa + (n-1)\lambda_1} \right)
\]

\[
= P\left( \frac{a\kappa}{(n-1)\lambda_1} \leq \frac{\kappa}{\lambda_1} \leq \frac{b\kappa}{(n-1)\lambda_1} \right) = P\left( a \leq (n-1)\frac{\bar{\lambda}_1\kappa}{\lambda_1^2} \leq b \right)
\]

\[= 1 - \alpha + o(1)\]
as \(d \rightarrow \infty\) when \(n\) is fixed. It concludes the result.

**Proofs of Lemmas 3.1 and 3.2.** We note that \(||z_{o1}||^2/n = 1 + o_p(1)\) as \(n \rightarrow \infty\). From (A.3), under (A-ii) and (A-iii), we have that

\[
\hat{u}_1^T z_{o1}/||z_{o1}|| = 1 + o_p(1) \quad \text{(A.4)}
\]
as $d \to \infty$ either when $n$ is fixed or $n \to \infty$, so that $\hat{u}_1^T z_{o1} = ||z_{o1}|| + o_p(n^{1/2})$. Thus, we can claim the result of Lemma 3.2. On the other hand, with the help of Proposition 2.1, under (A-ii) and (A-iii), it holds that from (A.4)

$$h_1^T \hat{h}_1 = \frac{h_1^T (X - \bar{X}) \hat{u}_1}{\{(n - 1)\lambda_1\}^{1/2}} = \frac{\lambda_1^{1/2} z_{o1}^T \hat{u}_1}{\{(n - 1)\lambda_1\}^{1/2}} = \frac{||z_{o1}|| + o_p(n^{1/2})}{||z_{o1}||^2 + \kappa/\lambda_1 + o_p(n)^{1/2}} \frac{1}{\{(1 + \kappa/(\lambda_1 ||z_{o1}||^2))^{1/2} + o_p(1)^{1/2}}$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. It concludes the result of Lemma 3.1.

Proof of Theorem 3.1. With the help of Theorem 2.1, under (A-ii) and (A-iii), we have that from (A.4)

$$h_1^T \hat{h}_1 = \frac{h_1^T (X - \bar{X}) \hat{u}_1}{\{(n - 1)\lambda_1\}^{1/2}} = \frac{||z_{o1}|| + o_p(n^{1/2})}{||z_{o1}|| + \kappa/\lambda_1 + o_p(n)} = 1 + o_p(1)$$

as $d \to \infty$ either when $n$ is fixed or $n \to \infty$. It concludes the result.

Proof of Theorem 3.2. By combining Theorem 2.1 with Lemma 3.2, under (A-ii) and (A-iii), we have that

$$\tilde{s}_{1j} / \sqrt{\lambda_1} = \hat{u}_{1j} \sqrt{(n - 1)\lambda_1} / \lambda_1 = \hat{u}_{1j} ||z_{o1}|| + o_p(1) = z_{o1j} + o_p(1)$$

as $d \to \infty$ when $n$ is fixed. By noting that $z_{o1j} = z_{1j} - \bar{z}_1$ and $\bar{z}_1$ is distributed as $N(0, 1/n)$ under (A-i), we have the results.

Proof of Corollary 4.1. From Theorem 2.1, the result is obtained straightforwardly.

Proof of Lemma 4.1. Let $Z_i = [z_{1(i)}, ..., z_{d(i)}]^T$ be a sphered data matrix of $\pi_i$ for $i = 1, 2$, where $z_{j(i)} = (z_{j1(i)}, ..., z_{jn(i)})^T$. We assume $\mu_1 = \mu_2 = 0$ without loss of generality. Let $\beta_{st} = (\lambda_{s(1)} \lambda_{t(2)})^{1/2} h_{s(1)}^T h_{t(2)}$ for all $s, t$. Let $\iota_s$ be a fixed constant such that $\sum_{s=i+1}^d \lambda_{s(j)}^2 / \lambda_{1(j)}^2 = o(1)$ as $d \to \infty$ for $j = 1, 2$. Note that
$i_*$ exists under (A-ii) for each $\pi_i$. We write that

$$X_1^T X_2 = \sum_{s, t \leq i_*} \beta_{st} z_{s(1)}^T z_{t(2)} + \sum_{s, t \geq i_* + 1} \beta_{st} z_{s(1)}^T z_{t(2)}$$

$$+ \sum_{s = i_* + 1}^{d} \sum_{t = 1}^{i_*} \beta_{st} z_{s(1)}^T z_{t(2)} + \sum_{s = 1}^{d} \sum_{t = i_* + 1}^{d} \beta_{st} z_{s(1)}^T z_{t(2)}.$$ 

Note that

$$E\left\{ \left( \sum_{s = i_* + 1}^{d} \sum_{t = 1}^{i_*} \beta_{st} z_{s(1)}^T z_{t(2)} \right)^2 \right\} = \text{tr}\left( \sum_{s = i_* + 1}^{d} \lambda_{s(1)} h_{s(1)} h_{s(1)}^T \sum_{t = 1}^{i_*} \lambda_{t(2)} h_{t(2)} h_{t(2)}^T \right) \leq i_* \lambda_{i_* + 1(1)} \lambda_{1(2)}$$

for all $j, j'$. Also, note that

$$E\left\{ \left( \sum_{s, t \geq i_* + 1} \beta_{st} z_{s(1)}^T z_{t(2)} \right)^2 \right\} = \text{tr}\left( \sum_{s = i_* + 1}^{d} \lambda_{s(1)} h_{s(1)} h_{s(1)}^T \sum_{t = i_* + 1}^{d} \lambda_{t(2)} h_{t(2)} h_{t(2)}^T \right)$$

$$\leq \left( \sum_{s = i_* + 1}^{d} \lambda_{s(1)}^2 \sum_{t = i_* + 1}^{d} \lambda_{t(2)}^2 \right)^{1/2}$$

for all $j, j'$. Then, by using Markov’s inequality, for any $\tau > 0$, under (A-ii) for each $\pi_i$, we have that

$$P\left\{ \sum_{j = 1}^{n_1} \sum_{j' = 1}^{n_2} \left( \sum_{s = i_* + 1}^{d} \sum_{t = 1}^{i_*} \frac{\beta_{st} z_{s(1)}^T z_{t(2)}}{(n_1 n_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} \right)^2 > \tau \right\} \rightarrow 0,$$

$$P\left\{ \sum_{j = 1}^{n_1} \sum_{j' = 1}^{n_2} \left( \sum_{s = i_* + 1}^{d} \sum_{t = i_* + 1}^{d} \frac{\beta_{st} z_{s(1)}^T z_{t(2)}}{(n_1 n_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} \right)^2 > \tau \right\} \rightarrow 0$$

and

$$P\left\{ \sum_{j = 1}^{n_1} \sum_{j' = 1}^{n_2} \left( \sum_{s = 1}^{d} \sum_{t \geq i_* + 1}^{d} \frac{\beta_{st} z_{s(1)}^T z_{t(2)}}{(n_1 n_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} \right)^2 > \tau \right\} \rightarrow 0$$

as $d \rightarrow \infty$ either when $n_i$ is fixed or $n_i \rightarrow \infty$ for $i = 1, 2$. Hence, similar to (A.1), it holds that

$$\frac{e_{n_1}^T X_1^T X_2 e_{n_2}}{(\nu_1 \nu_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} = \frac{e_{n_1}^T \sum_{s, t \leq i_*} \beta_{st} z_{s(1)}^T z_{t(2)} e_{n_2}}{(\nu_1 \nu_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} + o_p(1).$$
Note that $e^T_{n_i} P_{n_i} = e^T_{n_i}$ and $P_{n_i} z_{1(i)} = z_{o1(i)}$ for $i = 1, 2$, where $z_{o1(i)} = z_{1(i)} - (\bar{z}_{1(i)}, \ldots, \bar{z}_{1(i)})^T$ and $\bar{z}_{1(i)} = n_i^{-1} \sum_{k=1}^{n_i} z_{1k(i)}$. Also, note that $X_i P_{n_i} = (X_i - \bar{X}_i)$ for $i = 1, 2$, where $\bar{X}_i = [\bar{x}_1, \ldots, \bar{x}_i]$ and $\bar{x}_i = \sum_{j=1}^{n_i} x_{j(i)}/n_i$. Let $\hat{u}_{1(i)}$ be the first (unit) eigenvector of $(X_i - \bar{X}_i)^T(X_i - \bar{X}_i)$ for $i = 1, 2$. Note that $\hat{u}_{1(i)}^T P_{n_i} = \hat{u}_{1(i)}^T$ when $(X_i - \bar{X}_i)^T(X_i - \bar{X}_i) \neq O$ for $i = 1, 2$. Then, under (A-ii) for each $\pi_i$, we have that

$$\frac{\hat{u}_{1(1)}^T (X_1 - \bar{X}_1)^T (X_2 - \bar{X}_2) \hat{u}_{1(2)}}{(\nu_1 \nu_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} = \frac{\hat{u}_{1(1)}^T \sum_{s,t \leq i} \beta_{st} z_{os(1)} z_{ot(2)} \hat{u}_{1(2)}}{(\nu_1 \nu_2 \lambda_{1(1)} \lambda_{1(2)})^{1/2}} + o_p(1)$$

(A.5)

as $d \to \infty$ either when $n_i$ is fixed or $n_i \to \infty$ for $i = 1, 2$. Note that $\tilde{h}_{1(i)} = \{\nu_i \lambda_{1(i)}\}^{-1/2}(X_i - \bar{X}_i) \hat{u}_{1(i)}$ for $i = 1, 2$. Also, note that $z_{os(i)}^T z_{os'(i)}/n_i = o_p(1)$ ($s \neq s'$) when $n_i \to \infty$ for $i = 1, 2$. Then, by combining (A.5) with Theorem 2.1 and (A.4), we can claim the result.

Proofs of Theorems 4.1 and 4.2. By combining Theorem 2.1, Lemmas 2.1 and 4.1, we can claim the results.

Acknowledgements

Research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078. Research of the third author was partially supported by Grants-in-Aid for Scientific Research (B) and Challenging Exploratory Research, JSPS, under Contract Numbers 22300094 and 26540010.

References

Ahn, J., Marron, J.S., Muller, K.M., Chi, Y.-Y., 2007. The high-dimension, low-sample-size geometric representation holds under mild conditions. Biometrika 94, 760-766.

Aoshima, M., Yata, K., 2011. Two-stage procedures for high-dimensional data. Sequential Anal. (Editor’s special invited paper) 30, 356-399.

Aoshima, M., Yata, K., 2013. Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. Methodol. Comput. Appl. Probab., in press (DOI: 10.1007/s11009-013-9370-7).
Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R., Korsmeyer, S.J., 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature Genetics 30, 41-47.

Hall, P., Marron, J.S., Neeman, A., 2005. Geometric representation of high dimension, low sample size data. J. R. Statist. Soc. B 67, 427-444.

Ishii, A., Yata, K., Aoshima, M., 2014. Asymptotic distribution of the largest eigenvalue via geometric representations of high-dimension, low-sample-size data. Sri Lankan J. Appl. Statist., Special Issue: Modern Statistical Methodologies in the Cutting Edge of Science (ed. Mukhopadhyay, N.), 81-94.

Jung, S., Marron, J.S., 2009. PCA consistency in high dimension, low sample size context. Ann. Statist. 37, 4104-4130.

Jung, S., Sen, A., Marron, J.S., 2012. Boundary behavior in high dimension, low sample size asymptotics of PCA. J. Multivariate Anal. 109, 190-203.

Srivastava, M.S., Yanagihara, H., 2010. Testing the equality of several covariance matrices with fewer observations than the dimension. J. Multivariate Anal. 101, 1319-1329.

Yata, K., Aoshima, M., 2009. PCA consistency for non-Gaussian data in high dimension, low sample size context. Commun. Statist. Theory Methods, Special Issue Honoring Zacks, S. (ed. Mukhopadhyay, N.) 38, 2634-2652.

Yata, K., Aoshima, M., 2010. Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. J. Multivariate Anal. 101, 2060-2077.

Yata, K., Aoshima, M., 2012. Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. J. Multivariate Anal. 105, 193-215.

Yata, K., Aoshima, M., 2013. PCA consistency for the power spiked model in high-dimensional settings. J. Multivariate Anal. 122, 334-354.

22