A HIGH-DIMENSIONAL TWO-SAMPLE TEST FOR NON-GAUSSIAN DATA UNDER A STRONGLY SPIKED EIGENVALUE MODEL

Aki Ishii

In this paper, we discuss two-sample tests for high-dimension, non-Gaussian data. We suppose that two classes have a strongly spiked eigenvalue model. First, we investigate the noise space for high-dimension, non-Gaussian data. A two-sample test is proposed by using the cross-data-matrix (CDM) methodology and its power is derived under some regularity conditions when the dimension is very large. We discuss the validity of assumptions. We check the performance of the proposed two-sample test procedure by simulations. Finally, we demonstrate the proposed two-sample test in actual data analyses.

Key words and phrases: Cross-data-matrix methodology, eigenstructure, HDLSS, large $p$, small $n$.

1. Introduction

One of the features of modern data is that the data dimension is extremely high, however, the sample size is relatively low. We call such data "HDLSS" or "large $p$, small $n" data, where $p$ is the data dimension and $n$ is the sample size. In this paper, we consider the two-sample test in the HDLSS context. We aim to give a two-sample test procedure for high-dimensional non-Gaussian data such as in the area of gene-set testing.

Suppose we have two classes $\pi_i$, $i = 1, 2$, and define independent $p \times n_i$ data matrices, $X_i = [x_{i1}, \ldots, x_{in_i}]$, $i = 1, 2$, from $\pi_i$, $i = 1, 2$, where $x_{ij}$, $j = 1, \ldots, n_i$, are independent and identically distributed (i.i.d.) as a $p$-dimensional distribution with a mean vector $\mu_i$ and covariance matrix $\Sigma_i$ ($\geq O$). We assume $n_i \geq 4$, $i = 1, 2$. The eigen-decomposition of $\Sigma_i$ is given by $\Sigma_i = H_i \Lambda_i H_i^T$, where $\Lambda_i = \text{diag}(\lambda_{1(i)}, \ldots, \lambda_{p(i)})$ having $\lambda_{1(i)} \geq \cdots \geq \lambda_{p(i)}(\geq 0)$ and $H_i = [h_{1(i)}, \ldots, h_{p(i)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $X_i - [\mu_i, \ldots, \mu_i] = H_i \Lambda_i^{1/2} Z_i$ for $i = 1, 2$. Then, $Z_i$ is a $p \times n_i$ sphericed data matrix from a distribution with the zero mean and identity covariance matrix. Let $Z_i = [z_{i1}, \ldots, z_{ip}]^T$ and $z_{j(i)} = (z_{j1(i)}, \ldots, z_{jn_i(i)})^T$, $j = 1, \ldots, p$, for $i = 1, 2$. Note that $E(z_{jk(i)} z_{j'k(i)}) = 0$ ($j \neq j'$) and $\text{Var}(z_{j(i)}) = I_{n_i}$, where $I_{n_i}$ denotes the $n_i$-dimensional identity matrix. Also, note that if $X_i$ is Gaussian, $z_{jk(i)}$s are i.i.d. as the standard normal distribution, $N(0,1)$. We assume that the fourth moments of each variable in $Z_i$ are uniformly bounded for $i = 1, 2$. Let $z_{aj(i)} = z_{j(i)} - (\bar{z}_{j(i)}, \ldots, \bar{z}_{j(i)})^T$, $j = 1, \ldots, p; i = 1, 2$, where

Received July 18, 2017. Revised September 25, 2017. Accepted October 16, 2017.

*Department of Information Sciences, Tokyo University of Science, 2641 Yamazaki, Noda-shi, Chiba 278-8510, Japan.
\begin{equation}
\tilde{z}_{j(i)} = n_i^{-1} \sum_{k=1}^{n_i} \tilde{z}_{jk(i)}. \quad \text{We assume that } P(\lim_{p \to \infty} \|z_{o1(i)}\| \neq 0) = 1 \text{ for } i = 1, 2, \quad \text{where } \| \cdot \| \text{ denotes the Euclidean norm. As for the eigenvalues of } \Sigma_i, i = 1, 2, \text{ we consider the following assumption:}
\end{equation}

\begin{equation}
(A-i) \quad \frac{\text{tr}(\Sigma_i^2) - \lambda_{1(i)}^2}{\lambda_{1(i)}^2} = o(1), \quad p \to \infty.
\end{equation}

The above eigenvalue model is regarded as a strongly spiked eigenvalue model which was proposed by Aoshima and Yata (2017).

In this paper, we consider the two-sample test:

\begin{equation}
H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2.
\end{equation}

We define \( \overline{x}_{in_i} = \sum_{j=1}^{n_i} x_{ij}/n_i \) and \( S_{in_i} = \sum_{j=1}^{n_i}(x_{ij} - \overline{x}_{in_i})(x_{ij} - \overline{x}_{in_i})^T/(n_i - 1) \) for \( i = 1, 2 \). Note that \( S_{in_i}^{-1} \) does not exist in the HDLSS context such as \( n_i/p \to 0 \). Under the assumption that \( \pi_1 \) and \( \pi_2 \) are Gaussian, there are a lot of literatures about the two-sample problem in the HDLSS context. When \( \pi_1 \) and \( \pi_2 \) are non-Gaussian, Bai and Saranadasa (1996) and Cai et al. (2014) considered the two-sample test under homoscedasticity, \( \Sigma_1 = \Sigma_2 \). Chen and Qin (2010) and Aoshima and Yata (2011, 2015) considered the two-sample test under heteroscedasticity, \( \Sigma_1 \neq \Sigma_2 \). Particularly, Aoshima and Yata (2011) proposed a two-sample test procedure to ensure prespecified accuracies regarding both the size and power. We note that the above literatures considered constructing two-sample test procedures under the eigenvalue condition as follows:

\begin{equation}
\frac{\lambda_{1(i)}^2}{\text{tr}(\Sigma_i^2)} \to 0 \quad \text{as} \quad p \to \infty \quad \text{for} \quad i = 1, 2.
\end{equation}

However, (1.2) sometimes fails in actual high-dimensional analyses. See Aoshima and Yata (2017) for the details. Aoshima and Yata (2017) proposed to develop high-dimensional inference not only for the “non-strongly spiked eigenvalue (NSSE) model” defined by (1.2) but also for the “strongly spiked eigenvalue (SSE) model” defined by

\begin{equation}
\liminf_{p \to \infty} \left\{ \frac{\lambda_{1(i)}^2}{\text{tr}(\Sigma_i^2)} \right\} > 0 \quad \text{for} \quad i = 1 \text{ or } 2.
\end{equation}

In this paper, we assume (A-i) and focus on the SSE model. We aim to develop a two-sample test procedure in the HDLSS context when \( \pi_1 \) and \( \pi_2 \) are non-Gaussian.

The rest of the paper is organized as follows. In Section 2, we investigate the noise space for high-dimensional non-Gaussian data. In Section 3, we consider the cross-data-matrix (CDM) methodology to estimate the first eigenvalue and give its bias-correction. In Section 4, we propose a two-sample test procedure by the CDM method and derive its asymptotic power under some regularity conditions. In Section 5, we discuss the validity of assumptions. In Section 6, we check the performance of the proposed two-sample test procedure by simulations. In Section 7, we give actual data analyses by using the proposed two-sample test. Finally, in Section 8, we give a concluding remark.
2. Noise space for high-dimensional non-Gaussian data

Throughout this section, we omit the population index to avoid confusion. In this section, we investigate the noise space for high-dimensional data. We have the sample covariance matrix, 

\[ S = (n - 1)^{-1}(X - \bar{X})(X - \bar{X})^T, \]

where \( \bar{X} = (\bar{x}, \ldots, \bar{x}) \) having \( \bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j / n \). We define the dual sample covariance matrix by 

\[ S_D = (n - 1)^{-1}(X - \bar{X})^T(X - \bar{X}). \]

Note that \( S \) and \( S_D \) share the non-zero eigenvalues. Now we write that

\[ (n - 1)S_D = \lambda_1 z_{o1} z_{o1}^T + \sum_{s=2}^{p} \lambda_s z_{os} z_{os}^T. \]  

Under (A-i) and the condition that

\[ \sum_{s=2}^{p} \lambda_s \sum_{r,s \geq 2} \lambda_r \lambda_s E \left\{ (z_{rk}^2 - 1) (z_{sk}^2 - 1) \right\} = o(1), \]

the second term in (2.1) can be evaluated as follows.

\[ \sum_{s=2}^{p} \frac{\lambda_s}{\lambda_1} z_{os} z_{os}^T \rightarrow \sum_{s=2}^{p} \frac{\lambda_s}{\lambda_1} (I_n - n^{-1} 1_n 1_n^T) \rightarrow O \]

in probability as \( p \to \infty \),

where \( 1_n = (1, \ldots, 1)^T \). Now, let us observe (2.2) by using computer simulations. We took \( n = 3 \) samples from (a) \( N_p(0, \Sigma) \) and (b) \( p \)-variate \( t \)-distribution with 5 degrees of freedom, \( t_p(0, \Sigma, 5) \), where for (a) and (b) the mean is \( 0 \) and the covariance matrix is \( \Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \) having \( \lambda_1 = \frac{p^2}{3} \) and \( \lambda_2 = \cdots = \lambda_p = 1 \). Note that (A-i) holds in (a) and (b). Also, note that (a) satisfies (C-i). We considered four cases: \( p = 4, 32, 256, 2048 \). For each case we calculated \( \lambda_1^{-1} \sum_{s=2}^{p} \lambda_s z_{os} z_{os}^T \) and its non-zero eigenvalues, \( \delta_1 \geq \delta_2 (> 0) \), together with the corresponding unit eigenvectors, \( v_1 \) and \( v_2 \). Let \( w_i = \delta_i v_i, \ i = 1, 2 \). We repeated this procedure twenty times and plotted \( \pm w_i \)s in Fig. 1. Both for (a) and (b) one can observe that the noise space is growing as \( p \) becomes large. One of the approaches to HDLSS data is studying its geometric representations. Hall et al. (2005), Ahn et al. (2007) and Yata and Aoshima (2012) gave several noticeable geometric representations of HDLSS data. Also, Jung and Marron (2009) and Jung et al. (2012) extended the result of Hall et al. (2005). As for (a), the noise space has a geometric representation when \( p \) is large. Yata and Aoshima (2012) evaluated the amount of the noise by using the geometric representation and gave the noise-reduction (NR) methodology. Ishii et al. (2014, 2016) derived several asymptotic properties of the first eigenspace and constructed an equality test of covariance matrices by using the NR method when \( p \to \infty \) while \( n \) is fixed. Also, Ishii (2017) considered a two-sample test procedure based on the NR method for high-dimension, Gaussian data satisfying (C-i). As for (b), one cannot observe any geometric characteristic. Actually, one cannot handle the noise space by \( S_D \) for high-dimension, non-Gaussian data. In order to handle the two-sample problem for high-dimensional non-Gaussian data, we shall consider applying the CDM method given by Yata and Aoshima (2010).
Figure 1. We took $n = 3$ samples from (a) $N_p(0, \Sigma)$ and (b) $t_p(0, \Sigma_5)$. We set the covariance matrix as $\Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ with $\lambda_1 = p^{2/3}$ and $\lambda_2 = \cdots = \lambda_p = 1$. We plotted $\pm w_1$ (black circle) and $\pm w_2$ (white cuboid) when $p = 4, 32, 256, 2048$. 

(a) When $X$ is distributed as $N_p(0, \Sigma)$.

(b) When $X$ is distributed as $t_p(0, \Sigma_5)$. 
3. Bias-corrected CDM estimator

The cross-data-matrix (CDM) methodology is high-dimensional Principal Component Analysis developed by Yata and Aoshima (2010). Again, throughout this section, we omit the population index to avoid confusion. Let $n(1) = \lfloor n/2 \rfloor$ and $n(2) = n - n(1)$, where $\lfloor x \rfloor$ denotes the smallest integer $\geq x$. We divide the data matrix $X$ into $X(1) = [x_{(1)1}, \ldots, x_{(1)n(1)}]$ and $X(2) = [x_{(2)1}, \ldots, x_{(2)n(2)}]$ at random. We define a cross data matrix with having $p$-vector $\bar{x}(i) = n(i)^{-1}\sum_{j=1}^{n(i)} x_{(i)j}$ ($i = 1, 2$). Let $r = n(2) - 1$. We calculate the singular value decomposition of $S_{D(1)}$ by $S_{D(1)} = \sum_{j=1}^{r} \lambda_j \hat{u}_{(1)j} \hat{u}_{(2)j}^T$, where $\lambda_1 \geq \cdots \geq \lambda_r \geq 0$ denote singular values of $S_{D(1)}$, and $\hat{u}_{(1)j}$ (or $\hat{u}_{(2)j}$) denotes a unit left- (or right-) singular vector corresponding to $\lambda_j$ ($j = 1, \ldots, r$). Yata and Aoshima (2010, 2013) showed that $\hat{\lambda}_j$ enjoys several consistency properties to estimate $\lambda_j$ without any assumptions about the population distribution when $p \to \infty$ and $n \to \infty$ even in the HDLSS context.

Let us write $X(i) - [\mu_1, \ldots, \mu_p] = H\Lambda^{1/2}Z(i)$, where $Z(i) = [z_{(i)1}, \ldots, z_{(i)p}]^T$ and $z_{(i)j} = (z_{(i)1j}, \ldots, z_{(i)pj})^T$, $i = 1, 2$; $j = 1, \ldots, p$. Let $z_{o(i)j} = z_{(i)j} - (\bar{z}_{(i)1}, \ldots, \bar{z}_{(i)p})^T$, $j = 1, \ldots, p$, where $\bar{z}_{(i)j} = n(i)^{-1}\sum_{k=1}^{n(i)} z_{(i)jk}$ ($i = 1, 2$; $j = 1, \ldots, p$). We assume $P(\lim_{p \to \infty} \|z_{o(i)1}\| \neq 0) = 1$, $i = 1, 2$. We have that

$$\sqrt{(n(1) - 1)(n(2) - 1)}S_{D(1)} = \lambda_1 z_{o(1)1} z_{o(2)1}^T + \sum_{j=2}^{p} \lambda_j z_{o(1)j} z_{o(2)j}^T. \quad (3.1)$$

Here, for any $(i, j)$ element of $\sum_{s=2}^{p} \lambda_s z_{o(1)s} z_{o(2)s}^T$, it holds that as $p \to \infty$

$$\frac{\text{Var}\{\sum_{s=2}^{p} \lambda_s (z_{(1)s} - \bar{z}_{(1)s})(z_{(2)s} - \bar{z}_{(2)s})\}}{\lambda_1^2} = \frac{(n(1) - 1)(n(2) - 1) \text{tr}(\Sigma^2) - \lambda_1^2}{n(1)n(2)} \to 0$$

under (A-i). Hence, we can claim under (A-i) that as $p \to \infty$

$$\sum_{j=2}^{p} \frac{\lambda_j z_{o(1)j} z_{o(2)j}^T}{\lambda_1} \to O \quad \text{in probability.} \quad (3.2)$$

Let us observe (3.2) by using computer simulations. We took $n = 6$ samples from $p$-variate $t$-distribution, $t_p(\mathbf{0}, \Sigma, 5)$, with mean $\mathbf{0}$, covariance matrix $\Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ having $\lambda_1 = p^{2/3}$ and $\lambda_2 = \cdots = \lambda_p = 1$, and 5 degrees of freedom. We considered four cases: $p = 6, 60, 600, 6000$. For each case we calculated $\lambda_1^{-1}\{\sqrt{(n(1) - 1)(n(2) - 1)}S_{D(1)} - \lambda_1 z_{o(1)1} z_{o(2)1}^T\} = (w_1, w_2, w_3)^T$, where $w_j$ is a 3 x 1 vector. Note that $w_j$ are orthogonal to $\mathbf{1}_3$ in view of (3.1). We plotted $w_1$ (white triangle), $w_2$ (black circle) and $w_3$ (cross mark) twenty times on the compliment space of $\mathbf{1}_3$ in Fig. 2. One can observe $w_j$ converge to zero when $p$ is large as expected theoretically in (3.2).
Figure 2. The behaviors of $\lambda^{-1}\{\sqrt{(n(1)-1)(n(2)-1)}S_{D(1)} - \lambda_1 z_{o(1)} z_{o(2)}^T\} = (w_1, w_2, w_3)^T$ on the compliment space of 13. We plotted $w_1$ (white triangle), $w_2$ (black circle) and $w_3$ (cross mark) when $n = 6$ samples are taken from $t_p(0, \Sigma, 5)$ with $\Sigma = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ having $\lambda_1 = p^{2/3}$ and $\lambda_2 = \cdots = \lambda_p = 1$.

From (3.2) we have under (A-i) that as $p \to \infty$

(3.3) \[
\frac{\hat{\lambda}_1}{\lambda_1} = \hat{u}_{1(1)}^T S_{D(1)} \hat{u}_{1(2)} = \hat{u}_{1(1)}^T \frac{z_{o(1)} z_{o(2)}^T}{\sqrt{(n(1) - 1)(n(2) - 1)}} \hat{u}_{1(2)} + o_p(1).
\]

Then, we have the following result.

**Theorem 3.1.** It holds under (A-i) that as $p \to \infty$

\[
\frac{\hat{\lambda}_1}{\lambda_1} = \begin{cases} 
\|z_{o(1)}1/\sqrt{n(1)} - I\| \|z_{o(2)}1/\sqrt{n(2)} - I\| + o_p(1) & \text{when } n \text{ is fixed}, \\
1 + o_p(1) & \text{when } n \to \infty.
\end{cases}
\]
Now, we consider the following assumption for the first normalized PC scores, $z_{(i)j}$, $j = 1, \ldots, n_{(i)}$ ($i = 1, 2$):

(A-ii) $z_{(i)j}$, $j = 1, \ldots, n_{(i)}$, are i.i.d. as $N(0, 1)$ for $i = 1, 2$.

**Corollary 3.1.** It holds under (A-i) and (A-ii) that as $p \to \infty$

$$\lambda_1 \Rightarrow \sqrt{\frac{\chi^2_{(1)n_{(1)}-1}}{n_{(1)}-1} \frac{\chi^2_{(2)n_{(2)}-1}}{n_{(2)}-1}},$$

where “$\Rightarrow$” denotes the convergence in distribution, and $\chi^2_{(i)n_{(i)}-1}$, $i = 1, 2$, are mutually independent random variables distributed as the chi-squared distribution with $n_{(i)} - 1$, degrees of freedom.

From Corollary 3.1 we have that

$$E \left( \sqrt{\frac{\chi^2_{(1)n_{(1)}-1}}{n_{(1)}-1} \frac{\chi^2_{(2)n_{(2)}-1}}{n_{(2)}-1}} \right) = \frac{c}{\sqrt{(n_{(1)}-1)(n_{(2)}-1)}}$$

with $c = 2\Gamma\left(\frac{n_{(1)}}{2}\right) \Gamma\left(\frac{n_{(2)}}{2}\right) \Gamma\left(\frac{n_{(1)}-1}{2}\right)^{-1} \Gamma\left(\frac{n_{(2)}-1}{2}\right)^{-1}$,

where $\Gamma(\cdot)$ denotes the gamma function. Thus we can give a bias-correction of the CDM estimator for the first eigenvalue by

$$\hat{\lambda}_1 = \frac{\sqrt{(n_{(1)}-1)(n_{(2)}-1)}}{c} \hat{\lambda}_1.$$

Then, we have the following result.

**Corollary 3.2.** It holds under (A-i) and (A-ii) that as $p \to \infty$

$$\frac{\hat{\lambda}_1}{\lambda_1} \Rightarrow \frac{1}{c} \sqrt{\chi^2_{(1)n_{(1)}-1} \chi^2_{(2)n_{(2)}-1}}$$

Now, we simply show numerical results of the first eigenvalue estimation. We compared the performances of $\hat{\lambda}_{1s}$ with $\hat{\lambda}_1$.

We set $n = 10$ and $p = 2^s$ for $s = 3, \ldots, 11$. We also set $\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_p)$, where $\lambda_1 = p^{2/3}$, $\lambda_2 = p^{1/3}$ and $\lambda_3 = \cdots = \lambda_p = 1$. We considered two cases: (a) $X$ is distributed as $N_p(0, \Sigma)$ and (b) $X$ is distributed as $p$-variate $t$-distribution with 5 degrees of freedom, $t_p(0, \Sigma, 5)$, where the mean is zero and the covariance matrix is $\Sigma$. Note that (A-i) holds both for (a) and (b), however, (C-i) holds only for (a).

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 estimates, $\hat{\lambda}_{1sr}$ and $\hat{\lambda}_{1r}$ ($r = 1, \ldots, R$). Let us simply write $\hat{\lambda}_{1s} = \hat{\lambda}_1}$
When $X$ is distributed as $N_p(0, \Sigma)$. 

When $X$ is distributed as $t_p(0, \Sigma, 5)$. 

Figure 3. The performances of $\hat{\lambda}_1^*$ and $\hat{\lambda}_1$. We set $n = 10$ and $p = 2^s$, $s = 3, \ldots, 11$. The values of $\hat{\lambda}_1^*/\lambda_1$ and $\hat{\lambda}_1/\lambda_1$ are denoted by the solid lines and dashed lines for (a) and (b) in the left panels. The values of $\text{var}(\hat{\lambda}_1^*/\lambda_1)$ and $\text{var}(\hat{\lambda}_1/\lambda_1)$ are denoted by the solid lines and dashed lines for (a) and (b) in the right panels. 

We also considered the Monte Carlo variability by $\text{var}(\hat{\lambda}_1^*/\lambda_1) = (R - 1)^{-1} \sum_{r=1}^{R} (\hat{\lambda}_1^* - \hat{\lambda}_1)^2/\lambda_1^2$ and $\text{var}(\hat{\lambda}_1/\lambda_1) = (R - 1)^{-1} \sum_{r=1}^{R} (\hat{\lambda}_1 - \lambda_1)^2/\lambda_1^2$. Fig. 3 shows the behaviors of $(\hat{\lambda}_1^*/\lambda_1, \hat{\lambda}_1/\lambda_1)$ in the left panels and $(\text{var}(\hat{\lambda}_1^*/\lambda_1), \text{var}(\hat{\lambda}_1/\lambda_1))$ in the right panels for (a) and (b). One can observe $\hat{\lambda}_1^*$ gave good performances both for (a) and (b). We recommend to use $\hat{\lambda}_1^*$ when the HDLSS data is non-Gaussian type and the sample-size is quite small.

4. Two-sample test procedure based on the CDM method

In this section, we construct a two-sample test procedure by using the CDM method. We start with the following test statistic:

$$T_n = \|\mathbf{x}_{1n_1} - \mathbf{x}_{2n_2}\|^2 - \sum_{i=1}^{2} \text{tr}(S_{in_i})/n_i.$$ 

Note that $T_n$ was discussed by Chen and Qin (2010) and Aoshima and Yata (2011, 2015) under the NSSE model. We consider $T_n$ under the SSE model, (A-i). We assume the following assumption:
\((A-\text{iii})\) \(\frac{\lambda_{1(1)}}{\lambda_{1(2)}} = 1 + o(1)\) and \(h_{1(1)}^T h_{1(2)} = 1 + o(1)\) as \(p \to \infty\).

\((A-\text{iii})\) means that the two classes share the first eigenspace when \(p\) is large. In the next section, we shall discuss the validity of \((A-\text{iii})\).

Let \(n_{\min} = \min\{n_1, n_2\}\). From Lemma 1 in Ishii (2017), under (A-i), (A-iii) and \(H_0\) in (1.1), it holds that as \(p \to \infty\)

\[
T_n = (\bar{z}_{1(1)} - \bar{z}_{1(2)})^2 - \sum_{i=1}^{2} \frac{\|z_{o1(i)}\|^2}{n_i(n_i - 1)} + o_p(n_{\min}^{-1}) \tag{4.1}
\]

either when \(n_i\)s are fixed or \(n_{\min} \to \infty\). Under (A-iii) we can estimate \(\lambda_{1(1)}\) by using the CDM method as follows: We regard \(X_1 = [x_{11}, \ldots, x_{n1}]\) and \(X_2 = [x_{21}, \ldots, x_{2n2}]\) as \(X_1\) and \(X_2\) in the CDM method, respectively. We define the cross data matrix by \(S_{Dn_1} = \frac{1}{2} (X_1 - X_1)^T (X_2 - X_2)\).

We define the singular value decomposition by \(S_{Dn_1} = \sum_{j=1}^{n_{\min}-1} \hat{\lambda}_j U_{jn1} U_{jn2}^T\).

Then, from Theorem 3.1 we have the following result.

**Lemma 4.1.** Under (A-i) and (A-iii), it holds that as \(p \to \infty\)

\[
\frac{\hat{\lambda}_{1n}}{\lambda_{1(1)}} = \begin{cases} \|z_{o1(1)}/\sqrt{n_1 - 1}\| \|z_{o1(2)}/\sqrt{n_2 - 1}\| + o_p(1) & \text{when } n_i\text{s are fixed,} \\ 1 + o_p(1) & \text{when } n_{\min} \to \infty. \end{cases} \tag{4.1}
\]

Let \(u_n = (1/n_1 + 1/n_2)^{-1}\). From (4.1) and Lemma 4.1 we consider the following test statistic:

\[
F = u_n T_n / \hat{\lambda}_{1n} + 1. \tag{4.2}
\]

By using the above lemma, we have the following result.

**Theorem 4.1.** Assume (A-ii) for each class. Under (A-i), (A-iii) and \(H_0\) in (1.1), it holds that as \(p \to \infty\)

\[
\nu_n \chi_1^2 - \{(n_1 + n_2)\nu_n\}^{-1} \sum_{i \neq i'} n_{i'}(n_{i'} - 1) \chi_{n_{i'} - 1}^2 + 1 \tag{4.3}
\]

\[
\sqrt{\chi_{n_1 - 1}^2 \chi_{n_2 - 1}^2} \tag{4.4}
\]

where \(\nu_n = \sqrt{(n_1 - 1)(n_2 - 1)}\) and \(\chi_1^2, \chi_{n_1 - 1}^2\) and \(\chi_{n_2 - 1}^2\) are mutually independent random variables distributed as the chi-squared distribution with degrees of freedom, 1, \(n_1 - 1\) and \(n_2 - 1\), respectively.

**Remark 4.1.** When \(p \to \infty\) and \(n_{\min} \to \infty\), the result in Theorem 4.1 holds without (A-ii).
For a given $\alpha \in (0, 1/2)$ let $f_{n_1, n_2}(\alpha)$ be the upper $\alpha$ point of (4.3). From Theorem 4.1 one can test (1.1) by

(4.5) \quad \text{rejecting } H_0 \text{ in (1.1)} \iff F \geq f_{n_1, n_2}(\alpha).

Then, it holds under (A-i) to (A-iii) that

$$\text{size} = \alpha + o(1).$$

**Remark 4.2.** When $n_i$s are fixed, one can calculate $f_{n_1, n_2}(\alpha)$ numerically. First, one generates pseudo-random variables from $\chi^2_\nu$ distribution with $\nu = 1, n_1 - 1, n_2 - 1$ degrees of freedom, independently. Then, one can calculate (4.3). After iterating these calculation, one can use the upper $\alpha$th value of (4.3) as $f_{n_1, n_2}(\alpha)$.

Next, we consider the power of the test by (4.5). Let $\mu_{12} = \mu_1 - \mu_2$. Under $H_1$ in (1.1), we consider the regularity condition:

(4.4) \quad \frac{n_{\min} \mu_{12}' \Sigma_i \mu_{12}}{\lambda_{1(1)}^2} \to 0 \text{ as } p \to \infty \text{ either when } n_{\min} \text{ is fixed or } n_{\min} \to \infty.

We have the following result.

**Lemma 4.2.** Under (A-i), (A-iii) and (A-iv), it holds that as $p \to \infty$

$$T_n \left( \frac{\lambda_{1(1)}}{\lambda_{1(1)}} \right) = (\bar{z}_{1(1)} - \bar{z}_{1(2)})^2 - 2 \sum_{i=1}^2 \frac{\|z_{o1(i)}\|^2}{n_i(n_i - 1)} + \frac{\|\mu_{12}\|^2}{\lambda_{1(1)}} + o_p(n_{\min}^{-1})$$

either when $n_{\min}$ is fixed or $n_{\min} \to \infty$.

By using the above lemma, we have the following result.

**Theorem 4.2.** Under (A-i), (A-iii) and (A-iv), the test by (4.5) holds that as $p \to \infty$ and $n_{\min} \to \infty$

$$\text{Power} = 1 - F_{\chi^2_1} \left( \frac{2}{\lambda_{1(1)}}(\alpha) - \frac{u_n \|\mu_{12}\|^2}{\lambda_{1(1)}} \right) + o(1),$$

where $F_{\chi^2_1}(\cdot)$ denotes the cumulative distribution function of the chi-squared distribution with 1 degree of freedom.

**Remark 4.3.** If $u_n \|\mu_{12}\|^2/\lambda_{1(1)} \to \infty$ as $p \to \infty$, the test by (4.5) holds under (A-i) to (A-iv) that Power$= 1 + o(1)$ as $p \to \infty$ either when $n_i$s are fixed or $n_{\min} \to \infty$.

5. **How to check the assumptions**

In this section, we provide methods to check the validity of the assumptions in actual data analyses. Aoshima and Yata (2017) gave a method to check (A-i). So, we focus on (A-ii) and (A-iii).
5.1. (A-ii): The Gaussianity of the first PC score

We estimate the first PC score by using the CDM method as follows: Let \( n_{(1)i} = [n_i/2] \) and \( n_{(2)i} = n_i - n_{(1)i} \) for \( i = 1, 2 \). For each class we divide the data matrix \( X_i \) into \( X_{(1)i} : p \times n_{(1)i} \) and \( X_{(2)i} : p \times n_{(2)i} \) at random. Similar to Section 3, we construct the cross data matrix by using \( X_{(1)i} \) and \( X_{(2)i} \), and calculate the first singular value \( \hat{\lambda}_{1(i)} \) and the corresponding unit left- (or right-) singular vector \( \hat{u}_{(1)i} \) (or \( \hat{u}_{(2)i} \)) for each class. Similarly, let \( z_{o(j)i} \) be the centered first PC vector for the \( j \)th division of class \( i \). We assume \( P(\lim_{p \to \infty} \| z_{o(j)i} \| \neq 0) = 1 \) for \( i = 1, 2; \ j = 1, 2 \). According to Yata and Aoshima (2010), we also calculate \( \hat{h}_{(j)i} = \{(n_{(j)i} - 1)\hat{\lambda}_{1(i)}\}^{-1/2}(X_{(j)i} - \bar{X}_{(j)i})\hat{u}_{(j)i} \) for \( i = 1, 2; \ j = 1, 2 \). Then, we have the following result.

**Lemma 5.1.** Under (A-i), it holds that as \( p \to \infty \)

\[ \hat{u}_{(j)i} \to z_{a(j)i}/\|z_{a(j)i}\| \quad \text{in probability for } \ i = 1, 2; \ j = 1, 2 \]
when \( n_i \)'s are fixed.

From Lemma 5.1 one can check the validity of (A-ii) by applying the test of the normality such as the Jarque-Bera test to \( \hat{u}_{(j)i} \).

5.2. (A-iii): The equality of the first eigenspace

Next, we consider the equality test of the first eigenspace in order to check (A-iii).

**Lemma 5.2.** Under (A-i), it holds that as \( p \to \infty \)

\[ \hat{h}_{(j)T}^T \hat{h}_{(j)12} = \left\{ \frac{(n_{(j')1} - 1)(n_{(j')2} - 1)}{(n_{(j)1} - 1)(n_{(j)2} - 1)} \|z_{o(j)11}\|^2\|z_{o(j)12}\|^2 \right\}^{1/4} \times h_{1(1)}^T h_{1(2)} + o_p(1) \]
for \( j = 1, 2; \ j \neq j' \), when \( n_i \)'s are fixed.

**Lemma 5.3.** Under (A-i), it holds that as \( p \to \infty \)

\[ (\hat{h}_{(j)11}^T \hat{h}_{(k)12})(\hat{h}_{(j')11}^T \hat{h}_{(k')12}) = \{h_{1(1)}^T h_{1(2)}\}^2 + o_p(1) \]
for \( j, k = 1, 2; \ (j, k) \neq (j', k') \), when \( n_i \)'s are fixed.

By using the above lemmas, we consider the equality test of the first eigenspace:

\[(5.1) \ H_0 : (\lambda_{1(1)}, h_{1(1)}) = (\lambda_{1(2)}, h_{1(2)}) \quad \text{vs.} \quad H_1 : (\lambda_{1(1)}, h_{1(1)}) \neq (\lambda_{1(2)}, h_{1(2)}). \]

Let \( h = \{h_{1(1)}^T h_{1(11)}^T h_{1(2)}^T h_{1(2)12}\}^{1/2} \) and \( h_* = \max\{h, h^{-1}\} \). From Corollary 3.1 and Lemma 5.3, we consider the test statistic:

\[ F_1 = \frac{\hat{\lambda}_{1(1)} \hat{h}_*}{\hat{\lambda}_{1(2)}} \]
where
\[ h_* = \begin{cases} 
    h_* & \text{when } \hat{\lambda}_{1(1)} \geq \hat{\lambda}_{1(2)}, \\
    1/h_* & \text{otherwise}.
\end{cases} \]

Then, we have the following result.

**Theorem 5.1.** Assume (A-ii) for each class. Under (A-i), it holds that
\[ F_1 \Rightarrow \{F_{n(1)1-1,n(1)2-1} \times F_{n(2)1-1,n(2)2-1}\}^{1/2} \quad \text{under } H_0 \text{ in (5.1)} \]
as \( p \to \infty \) when \( n_i \)'s are fixed, where \( F_{\nu_1,\nu_2} \) denotes a random variable distributed as \( F \)-distribution with \( (\nu_1,\nu_2) \) degrees of freedom and \( F_{n(1)1-1,n(1)2-1} \) and \( F_{n(2)1-1,n(2)2-1} \) are mutually independent.

For a given \( \alpha \in (0, 1/2) \), let \( g(\alpha) \) be the upper \( \alpha \) point of \( \{F_{n(1)1-1,n(1)2-1} \times F_{n(2)1-1,n(2)2-1}\}^{1/2} \). One can test (5.1) by
\[ (5.2) \quad \text{rejecting } H_0 \text{ in (5.1) } \iff F_1 \not\in \{g(1-\alpha/2), g(\alpha/2)\}. \]

Then, under (A-i) and (A-ii), it holds that as \( p \to \infty \)
\[ \text{size}(F_1) = \alpha + o(1) \]
when \( n_i \)'s are fixed.

**6. Performances of the two-sample test procedure**

In this section, we summarize simulation studies of the findings. We used computer simulations to study the performance of the test procedure by (4.5). We set \( \alpha = 0.05 \), \( \mu_1 = 0 \) and
\[ (6.1) \quad \Sigma_i = \begin{pmatrix} \Sigma_{(1)} & O_{2,p-2} \\ O_{p-2,2} & c_i \Sigma_{(2)} \end{pmatrix}, \quad i = 1, 2, \]
where \( O_{k,l} \) is the \( k \times l \) zero matrix, \( \Sigma_{(1)} = \text{diag}(p^{3/4}, p^{1/2}) \), \( \Sigma_{(2)} = (0.3|i-j|^{1/2}) \) and \( (c_1, c_2) = (1, 1.5) \). Note that (A-i) is met.

Independent pseudo-random observations were generated from \( N_p-p^* \cdot (0, I_{p-p^*}) \) for \( (z_{1j(i)}, \ldots, z_{p-p^*j(i)})^T, j = 1, \ldots, n_i \) and from \( p^* \)-variate \( t \)-distribution, \( t_{p^*}(0, I_{p^*}, \nu) \) for \( (z_{p-p^*+1j(i)}, \ldots, z_{pj(i)})^T, j = 1, \ldots, n_i, i = 1, 2. \) We considered three cases:

(a) \( p = 2^s \) for \( s = 3, \ldots, 11, p^* = p - 1, \nu = 5 \) and \( (n_1, n_2) = (10, 10), \)
(b) \( p = 2^s \) for \( s = 3, \ldots, 11, p^* = [p^{1/2}], \nu = 10 \) and \( (n_1, n_2) = (12, 7), \) and
(c) \( p = 1000, p^* = [p^{1/2}], \nu = 10 \) and \( n_1 = n_2 = 3 + 6s \) for \( s = 1, \ldots, 9. \)

Note that (C-i) is not satisfied for (a). We considered \( \mu_2 = 0 \) for \( H_0 \) and \( \mu_2 = (0, \ldots, 0, 1, \ldots, 1)^T \) for \( H_1 \) whose last \( \eta \) elements are 1. We set \( \eta = [1.5\lambda_{1(1)}] \) for (a), \( \eta = [1.4\lambda_{1(1)}] \) for (b) and \( \eta = [6\lambda_{1(1)}] \) for (c), where \( \lfloor x \rfloor \) denotes the smallest integer \( \geq x \). For each case we checked the performance as follows: We defined \( P_r = 1 \) (or 0) when \( H_0 \) was falsely rejected (or not) for \( r = 1, \ldots, 2000, \)
Figure 4. The performances of the three test procedures by (4.5), (8) in Ishii (2017) and (5.5) in Aoshima and Yata (2017). Independent pseudo-random observations were generated from $N_{p-p^*}(0, I_{p-p^*})$ for $(z_{1j(i)}, \ldots, z_{p-p^*+1j(i)})^T$, $j = 1, \ldots, n_1$ and from $p^*$-variate $t$-distribution, $t_{p^*}(0, I_{p^*}, \nu)$ for $(z_{p^*+1j(i)}, \ldots, z_{pj(i)})^T$, $j = 1, \ldots, n_1$, $i = 1, 2$. The values of $\alpha$ are denoted by the dashed lines in the left panels and the values of $1 - \beta$ are denoted by the dashed lines in the right panels.

and defined $\bar{\alpha} = \sum_{r=1}^{2000} P_r/2000$ to estimate the size. We also defined $P_r = 1$ (or 0) when $H_1$ was falsely rejected (or not) for $r = 1, \ldots, 2000$, and defined $1 - \bar{\beta} = 1 - \sum_{r=1}^{2000} P_r/2000$ to estimate the power. Note that their standard deviations are less than 0.011. In Fig. 4, we plotted $\bar{\alpha}$ (left panel) and $1 - \bar{\beta}$ (right panel) for the test procedure by (4.5) in each of (a), (b) and (c). We also
plotted them for the test procedure by (8) in Ishii (2017) in (a) and (b), and for the test procedure by (5.5) in Aoshima and Yata (2017) in (c). One can observe from (a), (b) and (c) that the test procedure by (4.5) gave good performances for large $p$ even when $n_i$s are fixed. Contrary to that, the test procedure by Ishii (2017) gave a bad performance for (a) with respect to the power when $p$ is large. This is probably because (C-i) is not met in (a) when $\nu$ is small. On the other hand, it gave a good performance for (b) when $p$ is large. The test procedure by Aoshima and Yata (2017) gave a good performance when both $p$ and $n_i$s are large. We recommend to use the test procedure by (4.5) when the data is non-Gaussian and the sample size is quite small.

7. Actual data analyses

We used acute myeloid leukemia data with 22283 ($= p$) genes consisting of four classes: acute promyelocytic leukemia (APL) with t(15;17) (10 samples), acute myelogenous leukemia (AML) with inv(16) (4 samples), monocytic leukemia (ML) (7 samples) and nonmonocytic leukemia (NL) (22 samples). See Gutierrez et al. (2005) for the details. The data set is available at NCBI Gene Expression Omnibus. First, we checked (A-i) for each class. As for each class, we divided the sample into two groups: the first $[n_i/2]$ samples and the remaining samples. Then, we constructed the cross data matrix $S_{D(i)}$ for each class. We calculated $\hat{\lambda}_{1(i)}$ and estimated $\delta = \sum_{s=2}^{p} \lambda_{s(i)}^2 / \lambda_{1(i)}^2$ by $\hat{\delta} = \{\text{tr}(S_{D(i)}S_{D(i)}^T)\} / \hat{\lambda}_{1(i)}^2$. We had $\hat{\delta} = 0.013$ for APL, $\hat{\delta} = 0$ for AML, $\hat{\delta} = 0.171$ for ML and $\hat{\delta} = 0.034$ for NL. From these observations we concluded that each class satisfies (A-i). In addition, from Lemma 5.1, we could confirm that each class satisfies (A-ii) with the level of significance 0.05. We also checked (A-iii) for six pairs out of the four classes and tested (5.1) by using the test procedure (5.2) with the level of significance 0.05. We had P-values as 0.481 for (APL, AML), 0.187 for (APL, ML), 0.902 for (APL, NL), 0.52 for (AML, ML), 0.746 for (AML, NL) and 0.920 for (ML, NL). From these observations, we applied the two-sample test procedure (4.2) to all the cases. We tested (1.1) with the level of significance 0.05. Then, $H_0$ in (1.1) was rejected for (APL, ML), (APL, NL), (AML, NL) and (ML, NL). The results are summarized in Table 1.

8. Conclusion

As pointed out in Aoshima and Yata (2017), we should choose a suitable test procedure reflected by the eigenstructures. In this paper, we focused on

| $(n_1, n_2)$ | (APL, AML) | (APL, ML) | (APL, NL) | (AML, ML) | (AML, NL) | (ML, NL) |
|--------------|------------|------------|------------|------------|------------|------------|
| $(n_1, n_2)$ | (10, 4)    | (10, 7)    | (10, 22)   | (4, 7)     | (4, 22)    | (7, 22)    |
| $f_{n_1, n_2}(0.05)$ | 5.94       | 4.91       | 4.38       | 6.05       | 5.76       | 4.71       |
| $F$          | 2.39       | 13.22      | 12.26      | 4.31       | 6.04       | 19.37      |

Table 1. The upper 0.05 point, $f_{n_1, n_2}(0.05)$, of (4.3) and the value of $F$ given by (4.2) for all the pairs from Gutierrez et al. (2005)’s data sets having $p = 22283$. 
the SSE model. In high-dimensional settings, it is unrealistic to assume the equality of the covariance matrices between the two classes. However, when analysing microarray data sets, we sometimes observe that the two covariance matrices share the first eigenspace at least. In such situations, we positively make use of the common eigenspace as a ground to compare the two class means. From this point of view, we provided a two-sample test procedure by using the CDM method. Also, we discussed how to check the validity of the assumptions. Through the simulation studies, the proposed test procedure by (4.5) gave a good performance when the dimension is large while the sample-size is quite small.

Appendix

Let \( P_n = I_n - n^{-1}1_n1_n^T \), where \( 1_n = (1, \ldots, 1)^T \).

**Proofs of Theorem 3.1, Corollaries 3.1 and 3.2.** Let \( e_{(j)} = (e_{(j)1}, \ldots, e_{(j)n_{(j)}})^T, j = 1, 2, \) be arbitrary unit \( n_{(j)} \)-vectors. From (3.1) and (3.2) we have under (A-i) that as \( p \to \infty \)

\[
\lambda_1^{-1} e_{(1)}^T S_{D(1)} e_{(2)} = e_{(1)}^T \frac{z_{o(1)}^T z_{o(2)}^T}{\sqrt{(n_{(1)} - 1)(n_{(2)} - 1)}} e_{(2)} + o_p(1).
\]

Now we consider the first singular value of \( S_{D(1)} \). Then, it holds that as \( p \to \infty \)

\[
\frac{\lambda_1}{\lambda_1} = \max(\lambda_{(1)}^T (\lambda_{o(1)}^T / \sqrt{n_{(1)} - 1})(\lambda_{o(2)}^T / \sqrt{n_{(2)} - 1}) e_{(2)} + o_p(1))
\]

(A.1)

\[
= \|z_{o(1)}^T / \sqrt{n_{(1)} - 1}\| \|z_{o(2)}^T / \sqrt{n_{(2)} - 1}\| + o_p(1).
\]

Note that \( \|z_{o(i)}^T / \sqrt{n_{(i)} - 1}\| = 1 + o_p(1), i = 1, 2, \) when \( p \to \infty \) and \( n \to \infty \). From (3.3) and (A.1), it concludes the result of Theorem 3.1.

Note that \( \|z_{o(i)}^T \|^2 = \sum_{k=1}^{n_{(i)}} z_{(i)1k}^2 - n_{(i)}(\bar{z}_{(i)1})^2 \) is distributed as \( \chi^2_{n_{(i)} - 1} \) for \( i = 1, 2 \), if \( z_{(i)1k}, k = 1, \ldots, n_{(i)} \), are i.i.d. as \( N(0, 1) \). Thus we can conclude the result of Corollary 3.1.

From Corollary 3.1 and (3.5), we can conclude the result of Corollary 3.2. \( \Box \)

**Proof of Lemma 4.1.** We assume \( \mu_1 = \mu_2 = 0 \) without loss of generality. Let \( \beta_{st} = (\lambda_{s(1)} \lambda_{t(2)})^{1/2} \times h_{s(1)}^T h_{t(2)} \) for all \( s, t \). We write that

\[
X_1^T X_2 = \beta_{11} z_{1(1)}^T z_{1(2)}^T + \sum_{s,t \geq 2} \beta_{st} z_{s(1)}^T z_{t(2)}^T
\]

\[
+ \sum_{s=2}^p \beta_{s1} z_{s(1)}^T z_{(1)(2)}^T + \sum_{t=2}^p \beta_{1t} z_{(1)(2)}^T z_{(1)(2)}^T.
\]
Let $\Sigma_{is} = \sum_{s=2}^{p} \lambda_{s(i)} h_{s(i)} h_{s(i)}^T$ for $i = 1, 2$. Here, we have that

$$E \left\{ \left( \sum_{s=2}^{p} \beta_{s1} z_{sk(1)} z_{k'(2)} \right)^2 \right\} = \lambda_{1(2)} h_{1(2)}^T \Sigma_{1s} h_{1(2)} \leq \lambda_{1(2)} \lambda_{2(1)},$$

$$E \left\{ \left( \sum_{t=2}^{p} \beta_{t1} z_{1k(1)} z_{tk'(2)} \right)^2 \right\} = \lambda_{1(1)} h_{1(1)}^T \Sigma_{2s} h_{1(1)} \leq \lambda_{1(1)} \lambda_{2(2)},$$

and

$$E \left\{ \left( \sum_{s,t \geq 2}^{p} \beta_{st} z_{sk(1)} z_{tk'(2)} \right)^2 \right\} = \text{tr}(\Sigma_{1s} \Sigma_{2s}) \leq \{\text{tr}(\Sigma_{1s}^2) \text{tr}(\Sigma_{2s}^2)\}^{1/2}$$

for all $k, k'$. Then, by using Marcov’s inequality, for any $\tau > 0$, it holds under (A-i) that

$$P \left\{ \sum_{k=1}^{n_1} \sum_{k' = 1}^{n_2} \left( \frac{\sum_{s=2}^{p} \beta_{s1} z_{sk(1)} z_{k'(2)}}{(n_1 n_2)^{1/2} \lambda_{1(1)}} \right)^2 > \tau \right\} \leq \frac{\lambda_{1(2)} \lambda_{2(1)}}{\tau \lambda_{1(1)}^2} \to 0,$$

$$P \left\{ \sum_{k=1}^{n_1} \sum_{k' = 1}^{n_2} \left( \frac{\sum_{t=2}^{p} \beta_{t1} z_{1k(1)} z_{tk'(2)}}{(n_1 n_2)^{1/2} \lambda_{1(1)}} \right)^2 > \tau \right\} \leq \frac{\lambda_{1(1)} \lambda_{2(2)}}{\tau \lambda_{1(1)}^2} \to 0,$$

and

$$P \left\{ \sum_{k=1}^{n_1} \sum_{k' = 1}^{n_2} \left( \frac{\sum_{s,t \geq 2}^{p} \beta_{st} z_{sk(1)} z_{tk'(2)}}{(n_1 n_2)^{1/2} \lambda_{1(1)}} \right)^2 > \tau \right\} \leq \frac{\sqrt{\text{tr}(\Sigma_{1s}^2) \text{tr}(\Sigma_{2s}^2)}}{\tau \lambda_{1(1)}^2} \to 0$$

as $p \to \infty$ either when $n_i$ is fixed or $n_i \to \infty$ for $i = 1, 2$. Let $\nu_i = n_i - 1$. Then, it holds that

$$\frac{P_{n_1} X_{1}^T X_2 P_{n_2}}{(\nu_1 \nu_2)^{1/2} \lambda_{1(1)}^2} = \frac{P_{n_1} \beta_{11} z_{1(1)} z_{1(2)}^T P_{n_2}}{(\nu_1 \nu_2)^{1/2} \lambda_{1(1)}^2} + o_p(1).$$

Note that $X_1 P_{n_1} = (X_i - \overline{X}_i)$ and $\hat{\mathbf{u}}_{1n_1}^T P_{n_1} = \hat{\mathbf{u}}_{1n_1}$ for $i = 1, 2$, when $(X_1 - \overline{X}_1)^T (X_2 - \overline{X}_2) \neq O$. Then, under (A-i) and (A-iii), we have that

$$\frac{\hat{\mathbf{u}}_{1n_1}^T (X_1 - \overline{X}_1)^T (X_2 - \overline{X}_2) \hat{\mathbf{u}}_{1n_2}}{(\nu_1 \nu_2)^{1/2} \lambda_{1(1)}} = \frac{\hat{\mathbf{u}}_{1n_1}^T \beta_{11} z_{o1(1)} z_{o1(2)}^T \hat{\mathbf{u}}_{1n_2}}{(\nu_1 \nu_2)^{1/2} \lambda_{1(1)}} + o_p(1)$$

$$= \frac{\|z_{o1(1)}\| \|z_{o1(2)}\|}{(\nu_1 \nu_2)^{1/2} \lambda_{1(1)}} + o_p(1)$$

as $p \to \infty$ either when $n_i$ is fixed or $n_i \to \infty$ for $i = 1, 2$. Note that $\|z_{o1(i)}\|/\nu_i^{1/2} = 1 + o_p(1)$ when $n_i \to \infty$ for $i = 1, 2$. Then, we can claim the result. $\square$

**Proof of Theorem 4.1.** Under (A-ii) for each $\pi_i$, $u_n(\bar{z}_{1(i)} - \tilde{z}_{1(2)})^2$ is distributed as $\chi_1^2$. We note that $\bar{z}_{1(i)}$ and $z_{o1(i)}$ ($i = 1, 2$) are independent under
(A-ii). Then, by combining Corollary 3.1 with (4.1) and Lemma 4.1, it concludes the result. □

**Proof of Lemma 4.2.** We write that

\[
T_n = \sum_{i=1}^{2} \left\{ \| x_{in_i} - \mu_i \|^2 - \frac{\text{tr}(S_{in_i})}{n_i} \right\} - 2(\bar{x}_{1n_1} - \mu_1)^T(\bar{x}_{2n_2} - \mu_2) + 2\mu_{12}^T((\bar{x}_{1n_1} - \mu_1) - (\bar{x}_{2n_2} - \mu_2)) + \| \mu_{12} \|^2.
\]
(A.2)

Under (A-iv), by using Chebyshev’s inequality, for any \( \tau > 0 \) we have that

\[
P(\mu_{12}^T(\bar{x}_{in_i} - \mu_i) > \tau \lambda_{1(i)}/n_{\min}) = O \left( \frac{n_{\min} \mu_{12}^T \Sigma_i \mu_{12}}{\tau^2 \lambda_{1(i)}^2} \right) \rightarrow 0.
\]
(A.3)

Then, by combining (A.3) with (A.2) and (4.1), it concludes the result. □

**Proof of Theorem 4.2.** Note that \( f_{n_1,n_2}(\alpha) \rightarrow \chi_1^2(\alpha) \) as \( n_{\min} \rightarrow \infty \). From Remark 4.1 and Lemmas 4.1 and 4.2, under (A-i), (A-iii) and (A-iv), we have that

\[
P \left( \frac{u_n T_n}{\lambda_{1(n)}} + 1 > f_{n_1,n_2}(\alpha) \right) = P \left( \chi_1^2 > \chi_1^2(\alpha) - \frac{u_n \| \mu_{12} \|^2}{\lambda_{1(1)}} + o_p(1) \right)
= 1 - F_{\chi_1^2} \left( \chi_1^2(\alpha) - \frac{u_n \| \mu_{12} \|^2}{\lambda_{1(1)}} \right) + o(1).
\]

It concludes the result. □

**Proof of Lemma 5.1.** We note that \( \hat{u}_{(j)11}^T \mathbf{1}_{n_{(j)1}} = 0 \), \( i = 1, 2 \), with probability tending to 1 under \( P(\lim_{p \rightarrow \infty} \| z_{o(j)11} \| \neq 0) = 1 \) for \( i = 1, 2 \) and \( j = 1, 2 \). Also, note that \( z_{o(j)11}^T \mathbf{1}_{n_{(j)1}} = 0 \) for \( i = 1, 2 \) and \( j = 1, 2 \). Thus, similar to Theorem 3.1, we have the result. □

**Proofs of Lemmas 5.2 and 5.3.** Let \( \nu_{(j)i} = n_{(j)i} - 1 \) for \( i = 1, 2 \) and \( j = 1, 2 \). Similar to Lemma 4.1, under (A-i), we have that

\[
\frac{\hat{u}_{(j)11}^T(X_{(j)1} - \bar{X}_{(j)1})^T(X_{(j)2} - \bar{X}_{(j)2})\hat{u}_{(j)12}}{\{\nu_{(j)1}\nu_{(j)2}\lambda_{1(1)}\lambda_{1(2)}\}^{1/2}} = \frac{\hat{u}_{(j)11}^T\beta_{11}z_{o(j)11}z_{o(j)12}^T\hat{u}_{(j)12}}{\{\nu_{(j)1}\nu_{(j)2}\lambda_{1(1)}\lambda_{1(2)}\}^{1/2}} + o_p(1)
\]
(A.4)

as \( p \rightarrow \infty \) when \( n_{(j)i} \) is fixed for \( i = 1, 2 \) and \( j = 1, 2 \). Note that \( \hat{h}_{(j)1i} = \{\nu_{(j)i}\lambda_{1(1)}\}^{-1/2}(X_{(j)1} - \bar{X}_{(j)1})\hat{u}_{(j)11} \) for \( i = 1, 2 \) and \( j = 1, 2 \). By combining (A.4) with Theorem 3.1 and Lemma 5.1 for each \( \pi_i \), we can conclude the result of Lemma 5.2. The result of Lemma 5.3 can be obtained straightforwardly from Lemma 5.2. □
Proof of Theorem 5.1. Note that $z_{o(j)1i}$s are independent for $i = 1, 2$ and $j = 1, 2$. Under (A-ii) for each class, we have as $p \to \infty$ that

$$\frac{\|z_{o(j)1i}\|^2/(n_{o(j)1i} - 1)}{\|z_{o(j)1i'}\|^2/(n_{o(j)1i'} - 1)} \to F_{n_{o(j)1i}-1,n_{o(j)1i'}-1}$$

for $i = 1, 2; i \neq i'$ and $j = 1, 2$. By combining (A.5) with Lemma 5.3, we can get the result. □

Acknowledgements

I would like to express my sincere gratitude to the Chief Editor, Professor Makoto Aoshima. I would also like to thank two anonymous referees for their valuable and kind comments.

References

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

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

Aoshima, M. and Yata, K. (2015). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions, *Methodology and Computing in Applied Probability*, 17, 419–439.

Aoshima, M. and Yata, K. (2017). Two-sample tests for high-dimension, strongly spiked eigenvalue models, *Statistica Sinica*, in press, arXiv: 1602.02491.

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

Cai, T. T., Liu, W. and Xia, Y. (2014). Two sample test of high dimensional means under dependence, *J. R. Stat. Soc. Ser. B*, 76, 349–372.

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

Gutiérrez, N. C., López-Pérez, R., Hernández, J. M., Isidro, I., González, B., Delgado, M., Fermiñán, E., García, J. L., Vázquez, L., González, M. and San Miguel, J. F. (2005). Gene expression profile reveals deregulation of genes with relevant functions in the different subclasses of acute myeloid leukemia, *Leukemia*, 19, 402–409.

Hall, P., Marron, J. S. and Neeman, A. (2005). Geometric representation of high dimension, low sample size data, *J. R. Stat. Soc. Ser. B*, 67, 427–444.

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

Ishii, A., Yata, K. and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context, *J. Stat. Plann. inference*, 170, 186–199.

Ishii, A. (2017). A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model, *Hiroshima Math. J.*, in press.

Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context, *Annals of Statistics*, 37, 4104–4130.

Jung, S., Sen A. and Marron, J. S. (2012). Boundary behavior in high dimension, low sample size asymptotics of PCA, *J. Multivar. Anal.*, 109, 190–203.

Yata, K. and Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix, *J. Multivar. Anal.*, 101, 2060–2077.
Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations, *J. Multivar. Anal.*, **105**, 193–215.

Yata, K. and Aoshima, M. (2013). PCA consistency for the power spiked model in high-dimensional settings, *J. Multivar. Anal.*, **122**, 334–354.