Hiroshima Math. J.
47 (2017), 273–288

A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model

Aki Ishii

(Received April 27, 2016)
(Revised October 5, 2016)

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 consider a new two-sample test for high-dimensional data under the strongly spiked eigenvalue (SSE) model. We consider the distance-based two-sample test under the SSE model. We introduce the noise-reduction (NR) methodology and apply that to the two-sample test. Finally, we give simulation studies and demonstrate the new test procedure by using microarray data sets.

1. Introduction

Suppose that we have two independent $d \times n_i$ data matrices, $\boldsymbol{X}_i = [\boldsymbol{x}_{ij}, \dots, \boldsymbol{x}_{in_i}]$, $i = 1, 2$, where $\boldsymbol{x}_{ij}$, $j = 1, \dots, n_i$, are independent and identically distributed (i.i.d.) as a d-dimensional distribution (π_i) with a mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$ $(\geq \boldsymbol{O})$. We assume $n_i \geq 3$, $i = 1, 2$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{1(i)}, \dots, \lambda_{d(i)})$ having $\lambda_{1(i)} \geq \cdots \geq \lambda_{d(i)} (\geq 0)$ and $\boldsymbol{H}_i = [\boldsymbol{h}_{1(i)}, \dots, \boldsymbol{h}_{d(i)}]$ is an orthogonal matrix of the corresponding eigenvectors. Let $\boldsymbol{X}_i - [\boldsymbol{\mu}_i, \dots, \boldsymbol{\mu}_i] = \boldsymbol{H}_i \boldsymbol{\Lambda}_i^{1/2} \boldsymbol{Z}_i$ for $i = 1, 2$. Then, $\boldsymbol{Z}_i$ is a $d \times n_i$ sphered data matrix from a distribution with the zero mean and identity covariance matrix. Let $\boldsymbol{Z}_i = [\boldsymbol{z}_{1(i)}, \dots, \boldsymbol{z}_{d(i)}]^T$ and $\boldsymbol{z}_{j(i)} = (z_{j1(i)}, \dots, z_{jn_i(i)})^T$, $j = 1, \dots, d$, for $i = 1, 2$. Note that $E(z_{jk(i)} z_{j'k(i)}) = 0$ $(j \neq j')$ and $\mathrm{Var}(\boldsymbol{z}_{j(i)}) = \boldsymbol{I}_{n_i}$, where $\boldsymbol{I}_{n_i}$ is the n_i-dimensional identity matrix. We assume that the fourth moments of each variable in $\boldsymbol{Z}_i$ are uniformly bounded for $i = 1, 2$. Let $\boldsymbol{z}_{oj(i)} = \boldsymbol{z}_{j(i)} - (\bar{z}_{j(i)}, \dots, \bar{z}_{j(i)})^T$, $j = 1, \dots, d$; $i = 1, 2$, where $\bar{z}_{j(i)} = n_i^{-1} \sum_{k=1}^{n_i} z_{jk(i)}$. Also, note that if $\boldsymbol{X}_i$ is Gaussian, $z_{jk(i)}$s are i.i.d. as the standard normal distribution, $N(0, 1)$. We assume that $P(\lim_{d \to \infty} \|\boldsymbol{z}_{o1(i)}\| \neq 0) = 1$ for $i = 1, 2$, where $\|\cdot\|$ denotes the Euclidean norm. As necessary, we consider the following assumption for $z_{1k(i)}$s:

(A-i): $z_{1k(i)}$, $k = 1, \dots, n_i$, are i.i.d. as $N(0, 1)$ for $i = 1, 2$.

2010 *Mathematics Subject Classification*. Primary 62H15, secondary 34L20.

Key words and phrases. Asymptotic distribution, Distance-based two-sample test, HDLSS, Noise-reduction methodology, Microarray data.

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

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \qquad \text{vs.} \qquad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2. \tag{1}$$

We define $\bar{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$ and $\boldsymbol{S}_{in_i} = \sum_{j=1}^{n_i} (\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_{in_i})^T/(n_i - 1)$ for $i = 1, 2$. Then, Hotelling's T^2-statistic is defined by

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{2n_2})^T \boldsymbol{S}_n^{-1} (\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{2n_2}),$$

where $\boldsymbol{S}_n = \{(n_1 - 1)\boldsymbol{S}_{1n_1} + (n_2 - 1)\boldsymbol{S}_{2n_2}\}/(n_1 + n_2 - 2)$. However, $\boldsymbol{S}_n^{-1}$ does not exist in the HDLSS context such as $n_i/d \to 0$, $i = 1, 2$. In such situations, Dempster [10, 11], Srivastava [16] and Srivastava et al. [17] considered the test when π_1 and π_2 are Gaussian. Fujikoshi et al. [12] considered the Dempster's test statistic in the MANOVA context. When π_1 and π_2 are non-Gaussian, Bai and Saranadasa [4] and Cai et al. [7] considered the test under the homoscedasticity, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, and Chen and Qin [8] and Aoshima and Yata [1, 2] considered the test under the heteroscedasticity, $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$. We note that those two-sample tests were constructed under the eigenvalue condition as follows:

$$\frac{\lambda_{1(i)}^2}{\operatorname{tr}(\boldsymbol{\Sigma}_i^2)} \to 0 \qquad \text{as } d \to \infty \text{ for } i = 1, 2. \tag{2}$$

However, if (2) is not met, one cannot use those two-sample tests. See Aoshima and Yata [3] for the details. Aoshima and Yata [3] called (2) the "non-strongly spiked eigenvalue (NSSE) model". On the other hand, Aoshima and Yata [3] considered the "strongly spiked eigenvalue (SSE) model" as follows:

$$\liminf_{d \to \infty} \left\{ \frac{\lambda_{1(i)}^2}{\operatorname{tr}(\boldsymbol{\Sigma}_i^2)} \right\} > 0 \qquad \text{for } i = 1 \text{ or } 2. \tag{3}$$

For the SSE model, Katayama et al. [14] considered a one-sample test when the population distribution is Gaussian. Ishii et al. [13] considered the one-sample test for non-Gaussian cases. Ma et al. [15] considered a two-sample test for the factor model when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. Aoshima and Yata [3] gave two-sample tests by considering eigenstructures when $d \to \infty$ and $n_i \to \infty$, $i = 1, 2$. In this paper, we consider the divergence condition for d and n_is such as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$. For the divergence condition, we propose a two-sample test under the SSE model.

The rest of the paper is organized as follows. In §2, we consider the distance-based two-sample test under the SSE model. In §3, we introduce the noise-reduction (NR) methodology and provide asymptotic properties of the largest-eigenvalue estimation in the HDLSS context. We apply the NR method to the two-sample test and give a new test procedure for the SSE model. In §4, we give simulation studies and discuss the performance of the new test procedure. Finally, in §5, we demonstrate the new test procedure by using microarray data sets.

2. Distance-based two-sample test

In this section, we discuss asymptotic properties of the distance-based two-sample test for both the NSSE model and the SSE model.

Let

$$T_n = \|\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{2n_2}\|^2 - \sum_{i=1}^{2} \mathrm{tr}(\boldsymbol{S}_{in_i})/n_i.$$

Let $\boldsymbol{\mu}_{12} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Note that $E(T_n) = \|\boldsymbol{\mu}_{12}\|^2$ and

$$\mathrm{Var}(T_n) = 2\sum_{i=1}^{2} \frac{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)}{n_i(n_i-1)} + 4\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}{n_1 n_2} + 4\sum_{i=1}^{2} \frac{\boldsymbol{\mu}_{12}^T \boldsymbol{\Sigma}_i \boldsymbol{\mu}_{12}}{n_i}.$$

Bai and Saranadasa [4], Chen and Qin [8] and Aoshima and Yata [1] considered the statistics for high-dimensional data. We call the test with T_n the "distance-based two-sample test". By using Theorem 1 in Chen and Qin [8] or Theorem 4 in Aoshima and Yata [2], we can claim that as $d \to \infty$ and $n_i \to \infty$, $i = 1, 2$

$$\frac{T_n}{\mathrm{Var}(T_n)^{1/2}} \Rightarrow N(0, 1) \tag{4}$$

under H_0 in (1), (2) and the factor model given in Remark 2. Here, "$\Rightarrow$" denotes the convergence in distribution. However, we note that T_n does not hold (4) in the case of (3).

Now, we assume the following assumptions:

(A-ii): $\dfrac{\sum_{j=2}^{d} \lambda_{j(i)}^2}{\lambda_{1(i)}^2} = o(1)$ as $d \to \infty$ for $i = 1, 2$;

(A-iii): $\dfrac{\lambda_{1(1)}}{\lambda_{1(2)}} = 1 + o(1)$ and $\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)} = 1 + o(1)$ as $d \to \infty$.

Note that (A-ii) implies (3), that is (A-ii) is one of the SSE models. Also, note that (A-ii) implies the condition that $\lambda_{2(i)}/\lambda_{1(i)} \to 0$ as $d \to \infty$. In high-dimensional context, (A-iii) is much milder than $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$. In addition, one can check the validity of (A-iii). See §3.3.

Remark 1. *For a spiked model such as*

$$\lambda_{j(i)} = a_{ij} d^{\alpha_{ij}} \quad (j = 1, \ldots, m_i) \qquad \textit{and} \qquad \lambda_{j(i)} = c_{ij} \quad (j = m_i + 1, \ldots, d)$$

with positive and fixed constants, a_{ij}s, c_{ij}s and α_{ij}s, and a positive and fixed integer m_i, (A-ii) *holds under the conditions that $\alpha_{i1} > 1/2$ and $\alpha_{i1} > \alpha_{i2}$. See Yata and Aoshima* [19] *for the details.*

Let $n_{\min} = \min\{n_1, n_2\}$. Under (A-ii) and (A-iii), we have the following result.

Lemma 1. *Under H_0 in* (1), (A-ii) *and* (A-iii)*, it holds that*

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

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

Let $c_n = 1/n_1 + 1/n_2$. From Lemma 1, under H_0 in (1), (A-ii) and (A-iii), we have that

$$\frac{1}{\lambda_{1(1)} c_n}\left(T_n + \lambda_{1(1)} \sum_{i=1}^{2} \frac{\|z_{o1(i)}\|^2}{n_i(n_i-1)}\right) = c_n^{-1}(\bar{z}_{1(1)} - \bar{z}_{1(2)})^2 + o_p(1) \tag{5}$$

as $d \to \infty$ either when $n_{\min}$ is fixed or $n_{\min} \to \infty$. Note that $E(z_{1k(i)}^4)$'s are bounded when $n_{\min} \to \infty$. Then, it holds that

$$c_n^{-1/2}(\bar{z}_{1(1)} - \bar{z}_{1(2)}) \Rightarrow N(0,1)$$

as $n_{\min} \to \infty$ by Lyapunov's central limit theorem. Hence, from (5) it holds that as $d \to \infty$ and $n_{\min} \to \infty$

$$\frac{1}{\lambda_{1(1)} c_n}\left(T_n + \lambda_{1(1)} \sum_{i=1}^{2} \frac{\|z_{o1(i)}\|^2}{n_i(n_i-1)}\right) \Rightarrow \chi_1^2 \tag{6}$$

under H_0 in (1), (A-ii) and (A-iii), where χ_k^2 denotes a random variable distributed as the χ^2 distribution with k degrees of freedom. On the other hand, under (A-i), we note that $c_n^{-1/2}(\bar{z}_{1(1)} - \bar{z}_{1(2)})$ is distributed as $N(0,1)$ even when $n_{\min}$ is fixed. Hence, from (5) we have (6) as $d \to \infty$ when $n_{\min}$ is fixed under H_0 in (1), (A-i) to (A-iii).

In order to construct a test procedure for (1) under the SSE model, (A-ii), it is necessary to estimate $\lambda_{1(1)}$ and $\|z_{o1(i)}\|^2$, $i = 1, 2$ in (6).

3. Two-sample test for SSE model

In this section, we propose a two-sample test for the SSE model. We first introduce the noise-reduction (NR) methodology and provide asymptotic properties of the largest-eigenvalue estimation.

3.1. Noise-reduction methodology. Yata and Aoshima [19] proposed a method for eigenvalue estimation called the *noise-reduction (NR) methodology* that was brought by a geometric representation of the sample covariance matrix.

We consider the following assumption for $i = 1, 2$:

(A-iv): $\dfrac{\sum_{r,s\geq 2}^{d} \lambda_{r(i)}\lambda_{s(i)} E\{(z_{rk(i)}^2 - 1)(z_{sk(i)}^2 - 1)\}}{n_i \lambda_{1(i)}^2} = o(1)$ as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$.

Remark 2. *For several statistical inference of high-dimensional data, Aoshima and Yata* [2], *Bai and Saranadasa* [4] *and Chen and Qin* [8] *assumed a general factor model as follows:*

$$\boldsymbol{x}_{ij} = \boldsymbol{\Gamma}_i \boldsymbol{w}_{ij} + \boldsymbol{\mu}_i$$

for $j = 1, \ldots, n_i$, where $\boldsymbol{\Gamma}_i$ is a $d \times q_i$ matrix for some $q_i > 0$ such that $\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T = \boldsymbol{\Sigma}_i$, and $\boldsymbol{w}_{ij}$, $j = 1, \ldots, n_i$, are i.i.d. random vectors having $E(\boldsymbol{w}_{ij}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{w}_{ij}) = \boldsymbol{I}_{q_i}$. As for $\boldsymbol{w}_{ij} = (w_{1j(i)}, \ldots, w_{q_i j(i)})^T$, assume that $E(w_{rj(i)}^2 w_{sj(i)}^2) = 1$ and $E(w_{rj(i)} w_{sj(i)} w_{tj(i)} w_{uj(i)}) = 0$ for all $r \neq s, t, u$.

Then, from Lemma 1 *given by Yata and Aoshima* [21], *we claim that* (A-iv) *holds under* (A-ii) *for the factor model. Also, we note that the factor model naturally holds when π_i is Gaussian.*

Let $\hat{\lambda}_{1(i)} \geq \cdots \geq \hat{\lambda}_{d(i)} \geq 0$ be the eigenvalues of $\boldsymbol{S}_{in_i}$ for $i = 1, 2$. Let us write the eigen-decomposition of $\boldsymbol{S}_{in_i}$ as $\boldsymbol{S}_{in_i} = \sum_{j=1}^{d} \hat{\lambda}_{j(i)} \hat{\boldsymbol{h}}_{j(i)} \hat{\boldsymbol{h}}_{j(i)}^T$, where $\hat{\boldsymbol{h}}_{j(i)}$ denotes a unit eigenvector corresponding to $\hat{\lambda}_{j(i)}$. By using the NR method, $\lambda_{j(i)}$s are estimated by

$$\tilde{\lambda}_{j(i)} = \hat{\lambda}_{j(i)} - \frac{\mathrm{tr}(\boldsymbol{S}_{in_i}) - \sum_{s=1}^{j} \hat{\lambda}_{s(i)}}{n_i - 1 - j} \qquad (j = 1, \ldots, n_i - 2). \tag{7}$$

Note that $\tilde{\lambda}_{j(i)} \geq 0$ w.p.1 for $j = 1, \ldots, n_i - 2$. Yata and Aoshima [19, 21] and Ishii et al. [13] showed that $\tilde{\lambda}_{j(i)}$ has several consistency properties in high-dimensional context. Ishii et al. [13] gave the following result when n_i is fixed or $n_i \to \infty$.

Theorem 1 ([13]). *Under* (A-ii) *and* (A-iv), *it holds that as $d \to \infty$*

$$\frac{\tilde{\lambda}_{1(i)}}{\lambda_{1(i)}} = \begin{cases} \|\boldsymbol{z}_{o1(i)}\|^2/(n_i - 1) + o_p(1) & \text{when } n_i \text{ is fixed,} \\ 1 + o_p(1) & \text{when } n_i \to \infty \end{cases}$$

for $i = 1, 2$. Under (A-i), (A-ii) *and* (A-iv), *it holds that as $d \to \infty$ when n_i is fixed*

$$(n_i - 1)\frac{\tilde{\lambda}_{1(i)}}{\lambda_{1(i)}} \Rightarrow \chi_{n_i - 1}^2 \qquad \text{for } i = 1, 2.$$

REMARK 3. *Under* (A-ii) *and* (A-iv), *it holds that as* $d \to \infty$ *either when* n_i *is fixed or* $n_i \to \infty$

$$\frac{\hat{\lambda}_{1(i)}}{\lambda_{1(i)}} = \frac{\|z_{o1(i)}\|^2}{n_i - 1} + \frac{\sum_{s=2}^{d} \lambda_{s(i)}}{\lambda_{1(i)}(n_i - 1)} + o_p(1) \qquad \textit{for } i = 1, 2.$$

If $\sum_{s=2}^{d} \lambda_{s(i)}/(\lambda_{1(i)} n_i) \to \infty$ as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$, $\hat{\lambda}_{1(i)}$ is strongly inconsistent in the sense that $\lambda_{1(i)}/\hat{\lambda}_{1(i)} = o_p(1)$. We emphasize that one can remove the bias term of $\hat{\lambda}_{1(i)}$ by using the NR method.

3.2. Test procedure for (1). In this section, we apply the NR method to the distance-based two-sample test for the SSE model and give a new test procedure in the HDLSS context.

Let $v = n_1 + n_2 - 2$. From Theorem 1 we have the following result.

LEMMA 2. *Under* (A-i) *to* (A-iv), *it holds that as* $d \to \infty$ *when* v *is fixed*

$$\frac{\sum_{i=1}^{2} (n_i - 1)\tilde{\lambda}_{1(i)}}{\lambda_{1(1)}} \Rightarrow \chi_v^2.$$

Under (A-ii) *to* (A-iv), *it holds that as* $d \to \infty$ *and* $v \to \infty$

$$\frac{\sum_{i=1}^{2} (n_i - 1)\tilde{\lambda}_{1(i)}}{v\lambda_{1(1)}} = 1 + o_p(1).$$

In addition, from Theorem 1, we can estimate

$$\lambda_{1(1)} \sum_{i=1}^{2} \frac{\|z_{o1(i)}\|^2}{n_i(n_i - 1)}$$

in (6) by $\sum_{i=1}^{2} \tilde{\lambda}_{1(i)}/n_i$. Hence, we consider a test statistic for (1) by

$$F_0 = u_n \frac{T_n + \sum_{i=1}^{2} \tilde{\lambda}_{1(i)}/n_i}{\sum_{i=1}^{2} (n_i - 1)\tilde{\lambda}_{1(i)}},$$

where $u_n = v/c_n$. Let F_{k_1,k_2} denotes a random variable distributed as the F distribution with degrees of freedom, k_1 and k_2. Then, by combining Lemmas 1 with 2, we have the following results.

THEOREM 2. *Under* (A-i) *to* (A-iv), *it holds that as* $d \to \infty$

$$F_0 \Rightarrow \begin{cases} F_{1,v} & \textit{when } v \textit{ is fixed}, \\ \chi_1^2 & \textit{when } v \to \infty. \end{cases}$$

Corollary 1. *Under* (A-ii) *to* (A-iv), *it holds that as* $d \to \infty$ *and* $n_{\min} \to \infty$

$$F_0 \Rightarrow \chi_1^2 \qquad \textit{under } H_0 \textit{ in } (1).$$

Note that $\nu \to \infty$ as $n_i \to \infty$ for $i = 1$ or 2. From Theorem 2 F_0 is asymptotically distributed as χ_1^2 under (A-i) and some conditions. On the other hand, from Corollary 1, one can claim the result without (A-i) if $n_{\min} \to \infty$ (i.e., $n_i \to \infty$ for $i = 1, 2$).

For a given $\alpha \in (0, 1/2)$ we test (1) by

$$\text{rejecting } H_0 \Leftrightarrow F_0 > F_{1,\nu}(\alpha), \tag{8}$$

where $F_{k_1,k_2}(\alpha)$ denotes the upper α point of the F distribution with degrees of freedom, k_1 and k_2. Note that $F_{1,\nu}(\alpha) \to \chi_1^2(\alpha)$ as $\nu \to \infty$, where $\chi_k^2(\alpha)$ denotes the upper α point of χ^2 distribution with k degrees of freedom. Then, under the conditions in Theorem 2 (or Corollary 1), it holds that

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

as $d \to \infty$ either when ν is fixed or $\nu \to \infty$. Hence, one can use the test procedure by (8) even when n_is are fixed.

Next, we consider the power of the test by (8). We consider the following assumption under H_1 in (1):

(A-v): $\dfrac{n_{\min}\boldsymbol{\mu}_{12}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{12}}{\lambda_{1(1)}^2} \to 0$, $i = 1, 2$, as $d \to \infty$ either when $n_{\min}$ is fixed or $n_{\min} \to \infty$.

Here, we have the following result.

Lemma 3. *Under* (A-ii) *to* (A-v), *it holds that*

$$\frac{T_n + \sum_{i=1}^2 \tilde{\lambda}_{1(i)}/n_i}{c_n\lambda_{1(1)}} = \frac{(\bar{z}_{1(1)} - \bar{z}_{1(2)})^2}{c_n} + \frac{\|\boldsymbol{\mu}_{12}\|^2}{c_n\lambda_{1(1)}} + o_p(1).$$

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

Then, we have the following results.

Theorem 3. *Under* (A-i) *to* (A-v), *the test by* (8) *has that*

$$\textit{Power} = 1 - F_{\chi_1^2}\left(\chi_1^2(\alpha) - \frac{\|\boldsymbol{\mu}_{12}\|^2}{c_n\lambda_{1(1)}}\right) + o(1)$$

as $d \to \infty$ *and* $\nu \to \infty$, *where* $F_{\chi_1^2}(\cdot)$ *denotes the cumulative distribution function of* χ_1^2.

Corollary 2. *Assume that*

$$\frac{\|\boldsymbol{\mu}_{12}\|^2}{c_n \lambda_{1(1)}} \to \infty \qquad \text{as } d \to \infty \text{ either when } n_{\min} \text{ is fixed or } n_{\min} \to \infty.$$

Then, under (A-ii) *to* (A-v), *the test by* (8) *has that*

$$\textit{Power} = 1 + o(1)$$

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

Remark 4. *When* $d \to \infty$ *and* $n_{\min} \to \infty$, *we can claim Theorem* 3 *without* (A-i).

3.3. How to check (A-iii). When (A-iii) is met, one can use the test procedure by (8). However, (A-iii) is not a general condition for high-dimensional settings, so that it is necessary to check the validity in actual data analyses. We consider the following test:

$$H_0 : (\lambda_{1(1)}, \boldsymbol{h}_{1(1)}) = (\lambda_{1(2)}, \boldsymbol{h}_{1(2)}) \qquad \text{vs.} \qquad H_1 : (\lambda_{1(1)}, \boldsymbol{h}_{1(1)}) \neq (\lambda_{1(2)}, \boldsymbol{h}_{1(2)}). \quad (9)$$

Note that (A-iii) is met under H_0 in (9). Let $\tilde{\boldsymbol{h}}_{1(i)} = (\hat{\lambda}_{1(i)}^{1/2}/\tilde{\lambda}_{1(i)}^{1/2})\hat{\boldsymbol{h}}_{1(i)}$ for $i = 1, 2$. Let $\tilde{h} = \max\{|\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)}|, |\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)}|^{-1}\}$. Note that $\tilde{h} \geq 1$ w.p.1. Then, Ishii et al. [13] gave the following test statistic:

$$F_1 = \frac{\tilde{\lambda}_{1(1)}}{\tilde{\lambda}_{1(2)}} \tilde{h}_*,$$

where

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

From Theorem 4.1 in Ishii et al. [13], under (A-i), (A-ii) and (A-iv), it holds that

$$F_1 \Rightarrow F_{\nu_1, \nu_2} \quad \text{under } H_0 \text{ in } (9)$$

as $d \to \infty$ when n_is are fixed, where $\nu_i = n_i - 1$ for $i = 1, 2$. For a given $\alpha \in (0, 1/2)$ we test (9) by

$$\text{rejecting } H_0 \Leftrightarrow F_1 \notin [\{F_{\nu_2, \nu_1}(\alpha/2)\}^{-1}, F_{\nu_1, \nu_2}(\alpha/2)]. \qquad (10)$$

Then, under (A-i), (A-ii) and (A-iv), it holds that

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

as $d \to \infty$ when n_is are fixed. Hence, by using (10), one can check whether (A-iii) holds or not.

4. Simulation studies

We used computer simulations to study the performance of the test procedure by (8). We also checked the performance of the test procedure by

$$\text{rejecting } H_0 \Leftrightarrow T_n/\hat{K}^{1/2} > z_\alpha, \tag{11}$$

where z_α is a constant such that $P(N(0,1) > z_\alpha) = \alpha$ and

$$\hat{K} = 2\sum_{i=1}^{2} \frac{W_{in_i}}{n_i(n_i-1)} + 4\frac{\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})}{n_1 n_2} \qquad \text{with}$$

$$W_{in_i} = \frac{\sum_{j\neq k}^{n_i}(\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik})^2}{n_i(n_i-1)} - \frac{2\sum_{j\neq k\neq l}^{n_i}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^T\boldsymbol{x}_{il}}{n_i(n_i-1)(n_i-2)} + \frac{\sum_{j\neq k\neq l\neq m}^{n_i}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik}\boldsymbol{x}_{il}^T\boldsymbol{x}_{im}}{n_i(n_i-1)(n_i-2)(n_i-3)}.$$

Here, W_{in_i} is an unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$ given by Chen et al. [9]. See Srivastava et al. [18] for the details of W_{in_i}. Note that Aoshima and Yata [1] and Yata and Aoshima [20] gave a different unbiased estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$. From Theorems 1 and 2 in Chen and Qin [8] or Corollary 1 in Aoshima and Yata [3], under (2) and the factor model given in Remark 2, the test procedure by (11) has size $= \alpha + o(1)$ as $d \to \infty$ and $n_i \to \infty$, $i = 1, 2$. If (3) is met or n_is are fixed, we cannot claim "size$= \alpha + o(1)$" for the test procedure by (11). We set $\alpha = 0.05$, $\boldsymbol{\mu}_1 = \boldsymbol{0}$ and

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \boldsymbol{\Sigma}_{(1)} & \boldsymbol{O}_{2,d-2} \\ \boldsymbol{O}_{d-2,2} & c_i\boldsymbol{\Sigma}_{(2)} \end{pmatrix}, \qquad i = 1, 2, \tag{12}$$

where $\boldsymbol{O}_{k,l}$ is the $k \times l$ zero matrix, $\boldsymbol{\Sigma}_{(1)} = \mathrm{diag}(d^\beta, d^{1/2})$, $\boldsymbol{\Sigma}_{(2)} = (0.3^{|i-j|^{1/2}})$ and $(c_1, c_2) = (1, 1.5)$. Note that (A-ii) is met for $\beta > 1/2$. Also, note that (A-iii) is met.

First, we considered the case when $d \to \infty$ while n_is are fixed. We set $d = 2^s$, $s = 3, \ldots, 11$ and $(n_1, n_2) = (10, 15)$. Independent pseudo-random observations were generated from $\pi_i : N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1, 2$. We considered two cases for β in (12): (a) $\beta = 1$ and (b) $\beta = 2/3$. We considered the following cases for $\boldsymbol{\mu}_2$: (i) $\boldsymbol{\mu}_2 = \boldsymbol{0}$ and (ii) $\boldsymbol{\mu}_2 = (0, \ldots, 0, 1, \ldots, 1)^T$ whose last $\lceil d^\beta \rceil$ elements are 1, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Note that $\boldsymbol{\mu}_2 = (1, \ldots, 1)^T$ when $\beta = 1$. We considered a naive estimation of F_0 as

$$\hat{F}_0 = u_n \frac{T_n + \sum_{i=1}^2 \hat{\lambda}_{1(i)}/n_i}{\sum_{i=1}^2 (n_i - 1)\hat{\lambda}_{1(i)}}$$

and checked the performance of the test procedure given by

$$\text{rejecting } H_0 \Leftrightarrow \hat{F}_0 > F_{1,\nu}(\alpha). \tag{13}$$

For each case, we checked the performance of the test procedures given by (8), (11) and (13) and observed the results with $2000(=\mathrm{R}, \mathrm{say})$ repetitions. We defined $P_r = 1$ (or 0) when H_0 was falsely rejected (or not) for $r = 1, \dots, 2000$ for (a) and defined $\bar{\alpha} = \sum_{r=1}^{R} P_r/R$ to estimate the size. We also defined $P_r = 1$ (or 0) when H_1 was falsely rejected (or not) for $r = 1, \dots, 2000$ for (b) and (c) and defined $1 - \bar{\beta} = 1 - \sum_{r=1}^{R} P_r/R$ to estimate the power. Note that their standard deviations are less than 0.011. In Fig. 1, we plotted $\bar{\alpha}$ and $1 - \bar{\beta}$ for (a) and (b). We observed that the test procedure by (8) gives better performances compared to (11) regarding the size. The size by (11) did not become close to α. This is probably because T_n does not hold the asymptotic normality when (2) is not met. On the other hand, (11) gave better performances compared to (8) regarding the power. This is because (11) cannot control the size when (3) is met. The test procedure by (13) gave quite bad performances for (b). The power was much lower than the power of (8). The main reason must be that the bias of $\hat{\lambda}_{1(i)}$ is getting larger as d increases. From Remark 3 $\hat{\lambda}_{1(i)}$ is strongly inconsistent in the sense that $\lambda_{1(i)}/\hat{\lambda}_{1(i)} = o_p(1)$ for (b).

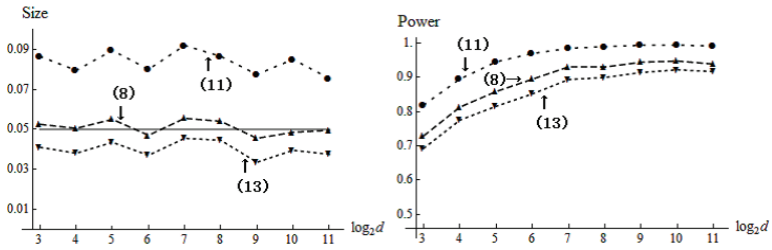
Next, we considered the case when $n_i \to \infty$, $i = 1, 2$. We considered two cases of d: (a) $d = 200$ and (b) $d = 1000$. We set $n_1 = 4s$, $s = 2, \dots, 10$, $n_2 = 1.5n_1$ and $\beta = 3/4$ in (12). We considered two cases of $\boldsymbol{\mu}_2$: (i) $\boldsymbol{\mu}_2 = \mathbf{0}$ and (ii) $\boldsymbol{\mu}_2 = (0, \dots, 0, 1, \dots, 1)^T$ whose last $\lceil 5c_n\lambda_{1(1)} \rceil$ elements are 1. Note that $\|\boldsymbol{\mu}_{12}\|^2 = \lceil 5c_n\lambda_{1(1)} \rceil$ for (ii). Then, it holds that

$$F_{\chi_1^2}\{\chi_1^2(\alpha) - \|\boldsymbol{\mu}_{12}\|^2/(c_n\lambda_{1(1)})\} = 0$$

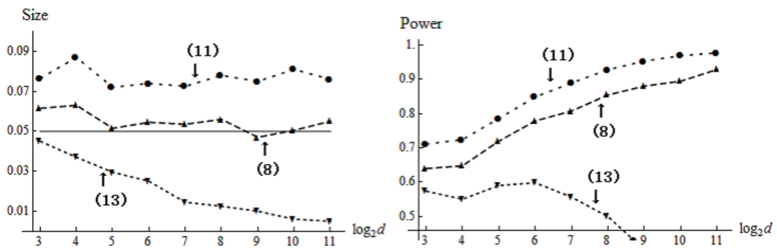
for (ii). Thus from Theorem 3 the test by (8) has $Power = 1 + o(1)$ as $d \to \infty$ and $n_i \to \infty$, $i = 1, 2$. We also checked the performance of the test procedure by

$$\text{rejecting } H_0 \Leftrightarrow \hat{T}_*/\hat{K}_*^{1/2} > z_\alpha, \tag{14}$$

where $\hat{T}_*$ and $\hat{K}_*$ are given in Section 5.2 of Aoshima and Yata [3]. We set $k_1 = k_2 = 2$ in $\hat{T}_*$ and $\hat{K}_*$. From Theorem 6 in Aoshima and Yata [3], under (3) and some regularity conditions, the test procedure by (14) has size $= \alpha + o(1)$ as $d \to \infty$ and $n_i \to \infty$, $i = 1, 2$. Let $d_* = \lceil d^{1/2} \rceil$. We considered a non-Gaussian distribution for $i = 1, 2$, as follows: $(z_{1j(i)}, \dots, z_{d-d_*j(i)})^T$, $j = 1, \dots, n_i$, are i.i.d. as $N_{d-d_*}(\mathbf{0}, \boldsymbol{I}_{d-d_*})$ and $(z_{d-d_*+1j(i)}, \dots, z_{dj(i)})^T$, $j = 1, \dots, n_i$, are i.i.d. as the d_*-variate t-distribution, $t_{d_*}(\mathbf{0}, \boldsymbol{I}_{d_*}, 10)$, with mean zero, covariance matrix $\boldsymbol{I}_{d_*}$ and degrees of freedom 10, where $(z_{1j(i)}, \dots, z_{d-d_*j(i)})^T$ and $(z_{d-d_*+1j(i)}, \dots, z_{dj(i)})^T$ are independent for each j. Note that (A-iv) holds from the fact that $\sum_{r,s\geq 2}^{d} \lambda_{r(i)}\lambda_{s(i)} E\{(z_{rk(i)}^2 - 1)(z_{sk(i)}^2 - 1)\} = 2\sum_{s=2}^{d-d_*} \lambda_{s(i)}^2 + O(\sum_{r,s\geq d-d_*+1}^{d} \lambda_{r(i)}\lambda_{s(i)}) = o(\lambda_{1(i)}^2)$ for $i = 1, 2$. Similar to Fig. 1, we calculated



(a) $\beta = 1$ in (12)



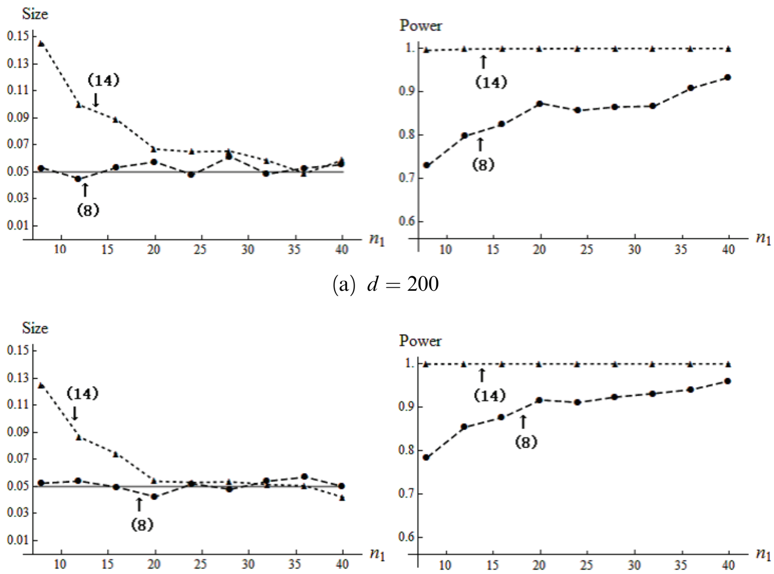
(b) $\beta = 2/3$ in (12)

Fig. 1. The test procedures given by (8), (11) and (13) for $d = 2^s$, $s = 3, \ldots, 11$ and $(n_1, n_2) = (10, 15)$ when (a) $\beta = 1$ and (b) $\beta = 2/3$. The values of $\bar{\alpha}$ are denoted by the dashed lines in the left panels and the values of $1 - \bar{\beta}$ are denoted by the dashed lines in the right panels. When d is large, $1 - \bar{\beta}$ of (13) was too low to describe in the right panel of (b).

$\bar{\alpha}$ and $1 - \bar{\beta}$ for the test procedures given by (8) and (14). In Fig. 2, we plotted $\bar{\alpha}$ and $1 - \bar{\beta}$ for (a) and (b). We observed that the test procedure by (8) gives better performances compared to (14) regarding the size, especially when n_is are small. On the other hand, the test procedure by (14) became close to α as n_is increase. In addition, (14) gave better performances compared to (8) regarding the power. This is probably because the asymptotic variance of $\hat{T}_*$ is smaller than $\mathrm{Var}(T_n)$ for the high-dimensional settings. See Section 5.1 in Aoshima and Yata [3] for the details. Hence, we recommend to use the test procedure by (14) when n_is are not small and (3) holds. If n_is are small (e.g. n_is are about 10), we recommend to use the test procedure by (8) for the SSE model. We emphasize that high-dimensional data often have the SSE model. Also, the sample size is often quite small. See § 5 for example.

5. Demonstration

In this section, we use two high-dimensional gene expression data sets that have the SSE model. We demonstrate the proposed test procedure by (8). We analyzed the following data sets: (I) Huntington's disease data with 22283 $(= d)$ genes consisting of π_1: huntington's disease patients $(n_1 = 17)$ and



(a) $d = 200$

(b) $d = 1000$

Fig. 2. The test procedures given by (8) and (14) for $n_1 = 4s$, $s = 2, \dots, 10$, $n_2 = 1.5n_1$ and $\beta = 3/4$ when (a) $d = 200$ and (b) $d = 1000$. The values of $\bar{\alpha}$ are denoted by the dashed lines in the left panels and the values of $1 - \bar{\beta}$ are denoted by the dashed lines in the right panels.

π_2: healthy controls ($n_2 = 14$) given by Borovecki et al. [5]; and (II) ovarian cancer data with 54675 ($= d$) genes consisting of π_1: normal ovarian samples ($n_1 = 12$) and π_2: ovarian cancer samples ($n_2 = 12$) given by Bowen et al. [6]. One can obtain these data sets from NCBI Gene Expression Omnibus. We standardized each sample so as to have the unit variance. Then, it holds that $\mathrm{tr}(\boldsymbol{S}_{in_i}) = d$.

First, we confirmed that the data sets satisfy (A-ii). Let $\delta = \sum_{j=2}^{d} \lambda_{j(i)}^2 / \lambda_{1(i)}^2$. We considered an estimator of δ by $\tilde{\delta} = (W_{n_i} - \tilde{\lambda}_{1(i)}^2)/\tilde{\lambda}_{1(i)}^2$ having W_{n_i} by (4) in Aoshima and Yata [2], where W_{n_i} is an unbiased and consistent estimator of $\mathrm{tr}(\boldsymbol{\Sigma}_i^2)$. We had $\tilde{\delta} = -0.39$ for huntington's disease, $\tilde{\delta} = -0.334$ for healthy controls, $\tilde{\delta} = 0.273$ for normal ovarian samples and $\tilde{\delta} = -0.115$ for ovarian cancer samples. From these observations we concluded that these data sets satisfied (A-ii). In addition, from Remark 3.1 given in Ishii et al. [13], by using Jarque-Bera test, we could confirm that these data sets satisfy (A-i) with the level of significance 0.05.

Next, we tested (9) by (10) with $\alpha = 0.05$. We calculated that $F_1 = 1.97$ for huntington's disease data and $F_1 = 1.31$ for ovarian cancer data. Then, H_0 in (9) was accepted by (10) both for (I) and (II). Hence, we concluded that these data sets satisfied (A-iii).

Finally, we tested (1) by (8) with $\alpha = 0.05$. We calculated that $F_0 = 77.87$ for (I) and $F_0 = 19.78$ for (II). Then, H_0 in (1) was rejected by the test procedure (8) both for (I) and (II).

Appendix A

A.1. Proof of Lemma 1. By using Chebyshev's inequality, for any $\tau > 0$, under (A-ii), we have that for $i = 1, 2$

$$P\left(\left|\sum_{j\neq j'}^{n_i}\sum_{s=2}^{d}\frac{\lambda_{s(i)}z_{sj(i)}z_{sj'(i)}}{n_i(n_i-1)}\right| > \tau\lambda_{1(i)}/n_i\right) = O\left(\frac{\sum_{s=2}^{p}\lambda_{s(i)}^2}{\tau^2\lambda_{1(i)}^2}\right) \to 0 \tag{15}$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$. We write that

$$\|\bar{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i\|^2 - \frac{\mathrm{tr}(\boldsymbol{S}_{in_i})}{n_i} = \sum_{s=1}^{d}\lambda_{s(i)}\left(\bar{z}_{s(i)}^2 - \frac{\|\boldsymbol{z}_{os(i)}\|^2}{n_i(n_i-1)}\right).$$

Here, $\bar{z}_{s(i)}^2 - \|\boldsymbol{z}_{os(i)}\|^2/\{n_i(n_i-1)\} = \sum_{j\neq j'}^{n_i} z_{sj(i)}z_{sj'(i)}/\{n_i(n_i-1)\}$ for all i, s. Then, from (15) under (A-ii), we have that

$$\frac{\|\bar{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i\|^2 - \mathrm{tr}(\boldsymbol{S}_{in_i})/n_i}{\lambda_{1(i)}} = \bar{z}_{1(i)}^2 - \frac{\|\boldsymbol{z}_{o1(i)}\|^2}{n_i(n_i-1)} + o_p(n_i^{-1}) \tag{16}$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$. Let $\beta_{st} = (\lambda_{s(1)}\lambda_{t(2)})^{1/2} \times \boldsymbol{h}_{s(1)}^T\boldsymbol{h}_{t(2)}$ for all s, t. Then, we write that

$$\begin{aligned}(\bar{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)^T(\bar{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) &= \sum_{s,t\geq 1}^{d}\beta_{st}\bar{z}_{s(1)}\bar{z}_{t(2)} = \beta_{11}\bar{z}_{1(1)}\bar{z}_{1(2)} + \sum_{s=2}^{d}\beta_{s1}\bar{z}_{s(1)}\bar{z}_{1(2)} \\ &\quad + \sum_{t=2}^{d}\beta_{1t}\bar{z}_{1(1)}\bar{z}_{t(2)} + \sum_{s,t\geq 2}^{d}\beta_{st}\bar{z}_{s(1)}\bar{z}_{t(2)}.\end{aligned} \tag{17}$$

Let $\boldsymbol{\Sigma}_{i*} = \sum_{s=2}^{d}\lambda_{s(i)}\boldsymbol{h}_{s(i)}\boldsymbol{h}_{s(i)}^T$ for $i = 1, 2$. Here, we have that

$$E\left\{\left(\sum_{s=2}^{d}\beta_{s1}\bar{z}_{s(1)}\bar{z}_{1(2)}\right)^2\right\} = \frac{\lambda_{1(2)}\boldsymbol{h}_{1(2)}^T\boldsymbol{\Sigma}_{1*}\boldsymbol{h}_{1(2)}}{n_1n_2} \leq \frac{\lambda_{1(2)}\lambda_{2(1)}}{n_1n_2};$$

$$E\left\{\left(\sum_{t=2}^{d}\beta_{1t}\bar{z}_{1(1)}\bar{z}_{t(2)}\right)^2\right\} = \frac{\lambda_{1(1)}\boldsymbol{h}_{1(1)}^T\boldsymbol{\Sigma}_{2*}\boldsymbol{h}_{1(1)}}{n_1n_2} \leq \frac{\lambda_{1(1)}\lambda_{2(2)}}{n_1n_2};$$

$$E\left\{\left(\sum_{s,t\geq 2}^{d}\beta_{st}\bar{z}_{s(1)}\bar{z}_{t(2)}\right)^2\right\} = \frac{\mathrm{tr}(\boldsymbol{\Sigma}_{1*}\boldsymbol{\Sigma}_{2*})}{n_1n_2} \leq \frac{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{1*}^2)\,\mathrm{tr}(\boldsymbol{\Sigma}_{2*}^2)}}{n_1n_2}.$$

Then, by using Chebyshev's inequality, for any $\tau > 0$, under (A-ii) and (A-iii), it holds that

$$P\left(\left|\sum_{s=2}^{d}\beta_{s1}\bar{z}_{s(1)}\bar{z}_{1(2)}\right| > \tau\lambda_{1(1)}/n_{\min}\right) \le \frac{\lambda_{1(2)}\lambda_{2(1)}}{\tau^2\lambda_{1(1)}^2} \to 0;$$

$$P\left(\left|\sum_{t=2}^{d}\beta_{1t}\bar{z}_{1(1)}\bar{z}_{t(2)}\right| > \tau\lambda_{1(1)}/n_{\min}\right) \le \frac{\lambda_{1(1)}\lambda_{2(2)}}{\tau^2\lambda_{1(1)}^2} \to 0;$$

$$P\left(\left|\sum_{s,t\ge 2}^{d}\beta_{st}\bar{z}_{s(1)}\bar{z}_{t(2)}\right| > \tau\lambda_{1(1)}/n_{\min}\right) \le \frac{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{1*}^2)\,\mathrm{tr}(\boldsymbol{\Sigma}_{2*}^2)}}{\tau^2\lambda_{1(1)}^2} \to 0$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$. Note that $\bar{z}_{1(1)}\bar{z}_{1(2)} = O_p(n_{\min}^{-1})$. Hence, from (17), under (A-ii) and (A-iii), we have that

$$\begin{aligned}\frac{(\bar{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)^T(\bar{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2)}{\lambda_{1(1)}} &= \frac{\beta_{11}\bar{z}_{1(1)}\bar{z}_{1(2)}}{\lambda_{1(1)}} + o_p(n_{\min}^{-1}) \\ &= \bar{z}_{1(1)}\bar{z}_{1(2)} + o_p(n_{\min}^{-1}) \qquad (18)\end{aligned}$$

as $d \to \infty$ either when n_i is fixed or $n_i \to \infty$ for $i = 1, 2$. Here, we write that

$$\begin{aligned}\|\bar{\boldsymbol{x}}_{1n_1} - \bar{\boldsymbol{x}}_{2n_2}\|^2 = \sum_{i=1}^{2}\|\bar{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i\|^2 - 2(\bar{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1)^T(\bar{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2) \\ + 2\boldsymbol{\mu}_{12}^T\{(\bar{\boldsymbol{x}}_{1n_1} - \boldsymbol{\mu}_1) - (\bar{\boldsymbol{x}}_{2n_2} - \boldsymbol{\mu}_2)\} + \|\boldsymbol{\mu}_{12}\|^2. \qquad (19)\end{aligned}$$

Then, by combining (16) and (18) with (19) under H_0 in (1), we can conclude the result.

A.2. Proof of Lemma 2. Under (A-i), we note that $\boldsymbol{z}_{o1(1)}$ and $\boldsymbol{z}_{o1(2)}$ are independent, and $\|\boldsymbol{z}_{o1(i)}\|^2$ is distributed as $\chi_{n_i-1}^2$ for $i = 1, 2$. Hence, from Theorem 1 we can conclude the result.

A.3. Proofs of Theorem 2 and Corollary 1. Under (A-i), we note that $\bar{z}_{1(i)}$ and $\boldsymbol{z}_{o1(i)}$ are independent for $i = 1, 2$. By combining (6) with Theorem 1 and Lemma 2, we can conclude the results.

A.4. Proof of Lemma 3. By using Chebyshev's inequality, for any $\tau > 0$, under (A-v), we have that for $i = 1, 2$

$$P(|\boldsymbol{\mu}_{12}^T(\bar{\boldsymbol{x}}_{in_i} - \boldsymbol{\mu}_i)| > \tau\lambda_{1(i)}/n_{\min}) = O\left(\frac{n_{\min}\boldsymbol{\mu}_{12}^T\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{12}}{\tau^2\lambda_{1(i)}^2}\right) \to 0 \qquad (20)$$

as $d \to \infty$ either when $n_{\min}$ is fixed or $n_{\min} \to \infty$. Then, by combining (19) with (16), (18), (20) and Theorem 1, under (A-ii) to (A-v), we have that

$$\frac{T_n + \sum_{i=1}^{2} \tilde{\lambda}_{1(i)}/n_i}{\lambda_{1(1)}} = (\bar{z}_{1(1)} - \bar{z}_{1(2)})^2 + \frac{\|\boldsymbol{\mu}_{12}\|^2}{\lambda_{1(1)}} + o_p(n_{\min}^{-1})$$

as $d \to \infty$ either when $n_{\min}$ is fixed or $n_{\min} \to \infty$ for $i = 1, 2$. Hence, we can claim the result.

A.5. Proof of Theorem 3. Note that $F_{1,\nu}(\alpha) \to \chi_1^2(\alpha)$ as $\nu \to \infty$. From Lemmas 2 and 3, under (A-i) to (A-v), we have that as $d \to \infty$ and $\nu \to \infty$

$$\begin{aligned} P\left(u_n \frac{T_n + \sum_{i=1}^{2} \tilde{\lambda}_{1(i)}/n_i}{\sum_{i=1}^{2}(n_i - 1)\tilde{\lambda}_{1(i)}} > F_{1,\nu}(\alpha)\right) &= P\left(\chi_1^2 > \chi_1^2(\alpha) - \frac{\|\boldsymbol{\mu}_{12}\|^2}{c_n \lambda_{1(1)}} + o_p(1)\right) \\ &= 1 - F_{\chi_1^2}\left(\chi_1^2(\alpha) - \frac{\|\boldsymbol{\mu}_{12}\|^2}{c_n \lambda_{1(1)}}\right) + o(1). \end{aligned}$$

It concludes the result.

A.6. Proof of Corollary 2. From Lemma 3 the result is obtained straightforwardly.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Professor Makoto Aoshima, for his enthusiastic guidance and helpful support to my research project. I would also like to thank Professor Kazuyoshi Yata for his valuable suggestions.

References

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

[2] M. Aoshima and K. Yata, Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions, Methodol. Comput. Appl. Probab., **17** (2015), 419–439.

[3] M. Aoshima and K. Yata, Two-sample tests for high-dimension, strongly spiked eigenvalue models, Statist. Sinica (2017), in press.

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

[5] F. Borovecki, L. Lovrecic, J. Zhou, H. Jeong, F. Then, H. D. Rosas, S. M. Hersch, P. Hogarth, B. Bouzou, R. V. Jensen and D. Krainc, Genome-wide expression profiling

of human blood reveals biomarkers for Huntington's disease, Proc. Natl. Acad. Sci. USA, **102** (2005), 11023–11028.

[6] N. J. Bowen, L. D. Walker, L. V. Matyunina, S. Logani, K. A. Totten, B. B. Benigno and F. M. John, Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells, BMC Medical Genomics, **2** (2009), 71.

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

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

[9] S. X. Chen, L.-X. Zhang and P.-S. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc., **105** (2010), 810–819.

[10] A. P. Dempster, A high dimensional two sample significance test, Ann. Math. Statist., **29** (1958), 995–1010.

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

[12] Y. Fujikoshi, T. Himeno and H. Wakaki, Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size, J. Japan Statist. Soc., **34** (2004), 19–26.

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

[14] S. Katayama, Y. Kano and M. S. Srivastava, Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension, J. Multivariate Anal., **116** (2013), 410–421.

[15] Y. Ma, W. Lan and H. Wang, A high dimensional two-sample test under a low dimensional factor structure, J. Multivariate Anal., **140** (2015), 162–170.

[16] M. S. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc., **37** (2007), 53–86.

[17] M. S. Srivastava, S. Katayama and Y. Kano, A two sample test in high dimensional data, J. Multivariate Anal., **114** (2013), 349–358.

[18] M. S. Srivastava, H. Yanagihara and T. Kubokawa, Tests for covariance matrices in high dimension with less sample size, J. Multivariate Anal., **130** (2014), 289–309.

[19] K. Yata and M. Aoshima, Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations, J. Multivariate Anal., **105** (2012), 193–215.

[20] K. Yata and M. Aoshima, Correlation tests for high-dimensional data using extended cross-data-matrix methodology, J. Multivariate Anal., **117** (2013), 313–331.

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

Aki Ishii
Department of Information Sciences
Tokyo University of Science
2641 *Yamazaki, Noda-shi, Chiba* 278-8510, *Japan*
E-mail: a.ishii@rs.tus.ac.jp