Tests for high-dimensional covariance matrices and correlation matrices under the strongly spiked eigenvalue model

Aki Ishii[a], Kazuyoshi Yata[b] and Makoto Aoshima[b]

[a] Department of Information Sciences, Tokyo University of Science
[b] Institute of Mathematics, University of Tsukuba

Abstract

We consider the equality test of high-dimensional covariance matrices under the strongly spiked eigenvalue (SSE) model. We find the difference of covariance matrices by dividing high-dimensional eigenspaces into the first eigenspace and the others. We create a new test procedure on the basis of those high-dimensional eigenstructures. We precisely study the influence of spiked eigenvalues on a test statistic and consider a bias correction so that the proposed test procedure has a consistency property for the size.

Key words and phrases: HDLSS; Large p small n; Noise-reduction methodology; SSE model; Two-sample test

1 Introduction

In this paper, we consider the equality test of covariance matrices when the data dimension is much larger than the sample size. Suppose we have two classes $\pi_i,\ i=1,2$. We define independent $d\times n_i$ data matrices, $\boldsymbol{X}_i=[\boldsymbol{x}_{i1},...,\boldsymbol{x}_{in_i}],\ i=1,2$, for $\pi_i,\ i=1,2$. We assume that $\boldsymbol{x}_{ij},\ j=1,...,n_i$, are independent and identically distributed (i.i.d.) as a d-dimensional distribution with a mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. We assume $n_i\ge 4,\ i=1,2$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ is given by $\boldsymbol{\Sigma}_i=\boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i=\mathrm{diag}(\lambda_{1(i)},...,\lambda_{d(i)})$ having $\lambda_{1(i)}\ge\cdots\ge\lambda_{d(i)}(\ge 0)$ and $\boldsymbol{H}_i=[\boldsymbol{h}_{1(i)},...,\boldsymbol{h}_{d(i)}]$ is an orthogonal matrix of the corresponding eigenvectors. We assume $\lambda_{2(i)}>0$ for $i=1,2$, and $\lambda_{1(i)}$s are of multiplicity one in the sense that

$$\liminf_{d\to\infty}\lambda_{1(i)}/\lambda_{2(i)}>1\ \text{ for } i=1,2.$$

Let $\boldsymbol{X}_i-[\boldsymbol{\mu}_i,...,\boldsymbol{\mu}_i]=\boldsymbol{H}_i\boldsymbol{\Lambda}_i^{1/2}\boldsymbol{Z}_i$ 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 $\boldsymbol{Z}_i=[\boldsymbol{z}_{1(i)},...,\boldsymbol{z}_{d(i)}]^T$ and $\boldsymbol{z}_{s(i)}=(z_{s1(i)},...,z_{sn_i(i)})^T,\ s=1,...,d$, for $i=1,2$. Note that $E(z_{qj(i)}z_{sj(i)})=0\ (q\ne s)$ and $\mathrm{Var}(\boldsymbol{z}_{s(i)})=\boldsymbol{I}_{n_i}$, where $\boldsymbol{I}_{n_i}$ denotes the n_i-dimensional identity matrix. Also, note that if $\boldsymbol{X}_i$ is Gaussian, $z_{sj(i)}$s are i.i.d. as the standard normal distribution, $N(0,1)$. We assume that the fourth moments of each variable in $\boldsymbol{Z}_i$ are uniformly bounded for $i=1,2$. Also, we assume the following assumption:

(A-i) $E(z_{qj(i)}^2 z_{sj(i)}^2)=1$, $E(z_{qj(i)}z_{sj(i)}z_{tj(i)})=0$ and $E(z_{qj(i)}z_{sj(i)}z_{tj(i)}z_{uj(i)})=0$ for all $q\neq s,t,u$.

This kind of assumption was made by Aoshima and Yata [1], Bai and Saranadasa [3] and Chen and Qin [4]. We note that (A-i) naturally holds when $\boldsymbol{X}_i$ is Gaussian.

We consider the equality test of covariance matrices as follows:

$$H_0:\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_2 \quad \text{vs.} \quad H_1:\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_2. \tag{1.1}$$

Schott [11] gave a test procedure based on the Frobenius norm when $d/n_i\to c_i\in[0,\infty)$. Srivastava and Yanagihara [13] considered a test procedure by using a Moore-Penrose inverse covariance matrix. Aoshima and Yata [1] gave a test procedure based on the quantity of tr$(\boldsymbol{\Sigma}_1-\boldsymbol{\Sigma}_2)$. They also discussed sample size determination so as to have a prespecified size and power simultaneously. Li and Chen [10] considered the test problem by using the quantity of tr$\{(\boldsymbol{\Sigma}_1-\boldsymbol{\Sigma}_2)^2\}$. The above references discussed asymptotic properties of their test procedures when $d\to\infty$ and $n_i\to\infty$ under the following eigenvalue condition:

$$\frac{\lambda_{1(i)}^2}{\text{tr}(\boldsymbol{\Sigma}_i^2)}\to 0 \ \text{ as } d\to\infty \text{ for } i=1,2. \tag{1.2}$$

Aoshima and Yata [2] called (1.2) the "non-strongly spiked eigenvalue (NSSE) model". On the other hand, Ishii, Yata and Aoshima [5] investigated asymptotic properties of the first principal component and considered the test problem (1.1) when $d\to\infty$ while n_is are fixed under the following eigenvalue condition:

$$\frac{\sum_{s=2}^d \lambda_{s(i)}^2}{\lambda_{1(i)}^2}=o(1) \text{ as } d\to\infty \text{ for } i=1,2. \tag{1.3}$$

Note that (1.3) implies the conditions that $\lambda_{2(i)}/\lambda_{1(i)}\to 0$ and $\lambda_{1(i)}^2/\text{tr}(\boldsymbol{\Sigma}_i^2)\to 1$ as $d\to\infty$. For a spiked model as

$$\lambda_{s(i)}=a_{s(i)}d^{\alpha_{s(i)}}\ (s=1,...,k_i) \quad \text{and} \quad \lambda_{s(i)}=c_{s(i)}\ (s=k_i+1,...,d) \tag{1.4}$$

with positive (fixed) constants, $a_{s(i)}$s, $c_{s(i)}$s and $\alpha_{s(i)}$s, and a positive (fixed) integer k_i, the condition (1.3) is met when $\alpha_{1(i)}>1/2$ and $\alpha_{1(i)}>\alpha_{2(i)}$. The condition (1.3) is generalized as

(A-ii) $\liminf_{d\to\infty}\left\{\frac{\lambda_{1(i)}^2}{\text{tr}(\boldsymbol{\Sigma}_i^2)}\right\}>0$ for $i=1$ and 2.

For the spiked model (1.4), (A-ii) is met when $\alpha_{1(i)}\geq 1/2$ for $i=1,2$. Aoshima and Yata [2] called (A-ii) the "strongly spiked eigenvalue (SSE) model" and showed that high-dimensional data often have the SSE model. They also provided a method to distinguish between the SSE model and the NSSE model. See Section 5 in Aoshima and Yata [2]. Ishii [6, 7] considered two-sample tests under (1.3) when $d\to\infty$ while n_is are fixed. The SSE model (A-ii) is quite difficult to handle because of the influence of strongly spiked noise. In order to handle huge noise, Aoshima and Yata [2] created a data-transformation technique for two-sample tests which transforms the SSE model to the NSSE model. In this paper, we give a new test procedure for (1.1) under the SSE model (A-ii) by using a new approach which is different from the data-transformation technique.

2 Performance of the earlier test statistic under the SSE model

In this section, we investigate the performance of the test statistic given by Li and Chen [10].

2.1 The earlier test procedure for (1.1)

For (1.1), Li and Chen [10] assumed that

$$\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l)=o\{\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j)\mathrm{tr}(\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_l)\} \tag{2.1}$$

for any i,j,k and $l\in\{1,2\}$. Note that (2.1) is one of the NSSE models. They proposed a test statistic as follows:

$$U_{n_1,n_2}=A_{n_1}+A_{n_2}-2\mathrm{tr}\left(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}\right),$$

where $\boldsymbol{S}_{in_i}$ is the sample covariance matrix having $E(\boldsymbol{S}_{in_i})=\boldsymbol{\Sigma}_i$ and

$$\begin{aligned}A_{n_i}=&\frac{1}{n_i(n_i-1)}\sum_{j\neq k}^{n_i}(\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik})^2-\frac{2}{n_i(n_i-1)(n_i-2)}\sum_{j\neq k\neq l}^{n_i}\boldsymbol{x}_{ik}^T\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{il}\\&+\frac{1}{n_i(n_i-1)(n_i-2)(n_i-3)}\sum_{j\neq k\neq l\neq l'}^{n_i}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik}\boldsymbol{x}_{il}^T\boldsymbol{x}_{il'}.\end{aligned}$$

Note that U_{n_1,n_2} is an unbiased estimator of $||\boldsymbol{\Sigma}_1-\boldsymbol{\Sigma}_2||_F^2=\mathrm{tr}\{(\boldsymbol{\Sigma}_1-\boldsymbol{\Sigma}_2)^2\}(=\Delta,\ \text{say})$. Let

$$m=\min\{d,n_{\min}\},\ \ \text{where}\ \ n_{\min}=\min\{n_1,n_2\}. \tag{2.2}$$

In this paper, we consider the divergence condition as

$$d\to\infty, n_1\to\infty\ \ \text{and}\ \ n_2\to\infty,$$

which is equivalent to $m\to\infty$. Note that

$$\mathrm{Var}(U_{n_1,n_2})=\sum_{i=1}^{2}\left(\frac{4\mathrm{tr}(\boldsymbol{\Sigma}_i^2)^2}{n_i^2}\{1+o(1)\}+O\Big(\frac{\mathrm{tr}\big\{\big(\boldsymbol{\Sigma}_i(\boldsymbol{\Sigma}_1-\boldsymbol{\Sigma}_2)\big)^2\big\}}{n_i}\Big)\right)+\frac{8\mathrm{tr}\{(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\}}{n_1n_2} \tag{2.3}$$

as $m\to\infty$ under (A-i), so that

$$\mathrm{Var}(U_{n_1,n_2})=(2\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1+2\mathrm{tr}(\boldsymbol{\Sigma}_2^2)/n_2)^2\{1+o(1)\}\ \ \text{under}\ H_0.$$

See Section 2 in Li and Chen [10] for the details. Let

$$T_{\text{LC}} = \frac{U_{n_1,n_2}}{2A_{n_1}/n_2 + 2A_{n_2}/n_1}.$$

They showed that

$$T_{\text{LC}} \Rightarrow N(0,1) \text{ as } m \to \infty$$

under H_0, (2.1) and some regularity conditions. Here, " $\Rightarrow$ " denotes the convergence in distribution and $N(0,1)$ denotes a random variable distributed as the standard normal distribution. We note that T_{LC} converges to $N(0,1)$ under the NSSE model, however does not so under the SSE model. In order to overcome this inconvenience, we first modify T_{LC} under (1.3) in Section 2.2 and newly construct a different test procedure for the SSE model (A-ii) in Section 3.

2.2 Modification of T_{LC} under a SSE model

We assume (1.3) as a SSE model. Let

$$K = 2\lambda_{1(1)}^2/n_1 + 2\lambda_{1(2)}^2/n_2.$$

From (2.3), we note that Var$(U_{n_1,n_2}) = K^2\{1+o(1)\}$ as $m \to \infty$ under (A-i), (1.3) and H_0. We assume the following assumption for the first (normalized) principal component (PC) scores:

(A-iii) $z_{1j(i)}$, $j = 1, ..., n_i$, are i.i.d. as $N(0,1)$ for $i = 1, 2$.

We note that (A-iii) is a Gaussian assumption only for the first PC scores. Thus, (A-iii) is much milder than the Gaussian assumption for $\boldsymbol{X}_i$ because $z_{sj(i)}$, $j = 1, ..., n_i$; $s = 1, ..., d$, are i.i.d. as $N(0,1)$ when $\boldsymbol{X}_i$ is Gaussian. Note that $E\{(z_{1j(i)}^2 - 1)^2\} = 2$ for all i, j, under (A-iii). Let

$$W = \sum_{j=1}^{n_1} \frac{\lambda_{1(1)}(z_{1j(1)}^2 - 1)}{n_1} - \sum_{k=1}^{n_2} \frac{\lambda_{1(2)}(z_{1k(2)}^2 - 1)}{n_2}.$$

Then, we have that Var$(W) = E(W^2) = K$. We have the following result.

Lemma 2.1 (Ishii, Yata and Aoshima [8]). *Under (A-i), (A-iii) and (1.3), it holds that as* $m \to \infty$

$$\begin{aligned} U_{n_1,n_2} &= \big(W + \lambda_{1(1)} - \lambda_{1(2)}\big)^2 + \Delta - (\lambda_{1(1)} - \lambda_{1(2)})^2 - K + o_p(K) \\ &\quad + O_p[n_{\max}^{1/2}\{1 - (\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)})^2\}K] + o_p\{(K\Delta)^{1/2}\}, \end{aligned}$$

where $n_{\max} = \max\{n_1, n_2\}$.

From Lemma 2.1, under (A-i), (A-iii), (1.3) and H_0, it holds that as $m \to \infty$

$$U_{n_1,n_2} = W^2 - K + o_p(K).$$

Let $T_{n_1,n_2} = U_{n_1,n_2}/K + 1$. Then, we have an asymptotic distribution of T_{n_1,n_2} under H_0.

Proposition 2.1 (Ishii, Yata and Aoshima [8]). *Under (A-i), (A-iii), (1.3) and H_0, it holds that $T_{n_1,n_2} \Rightarrow \chi_1^2$ as $m \to \infty$. Here, χ_ν^2 denotes a random variable distributed as a χ^2 distribution with ν degrees of freedom.*

Since $\lambda_{1(i)}$s are unknown, we need to estimate them. It is well known that the sample eigenvalues involve too much noise for high-dimensional data. See Ishii, Yata and Aoshima [5], Jung and Marron [9] and Shen et al. [12] for the details. We consider estimating $\lambda_{1(i)}$s by using the noise-reduction (NR) methodology given by Yata and Aoshima [15]. Let $\overline{\boldsymbol{X}}_i = [\bar{\boldsymbol{x}}_i, ..., \bar{\boldsymbol{x}}_i]$ and $\bar{\boldsymbol{x}}_i = n_i^{-1}\sum_{j=1}^{n_i}\boldsymbol{x}_{ij}$ for $i = 1, 2$. We denote the dual matrix of $\boldsymbol{S}_{in_i}$ by $\boldsymbol{S}_{iD}$ and define its eigen-decomposition as follows:

$$\boldsymbol{S}_{iD} = (n_i - 1)^{-1}(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)^T(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i) = \sum_{s=1}^{n_i-1}\hat{\lambda}_{s(i)}\hat{\boldsymbol{u}}_{s(i)}\hat{\boldsymbol{u}}_{s(i)}^T, \tag{2.4}$$

where $\hat{\lambda}_{1(i)} \geq \cdots \geq \hat{\lambda}_{n_i-1(i)}$ (≥ 0) and $\hat{\boldsymbol{u}}_{s(i)}$ denotes a unit eigenvector corresponding to the eigenvalue $\hat{\lambda}_{s(i)}$. Note that $\boldsymbol{S}_{in_i}$ and $\boldsymbol{S}_{iD}$ share non-zero eigenvalues. If one uses the NR method, $\lambda_{j(i)}$s are estimated by

$$\tilde{\lambda}_{j(i)} = \hat{\lambda}_{j(i)} - \frac{\text{tr}(\boldsymbol{S}_{iD}) - \sum_{s=1}^{j}\hat{\lambda}_{s(i)}}{n_i - 1 - j} \quad (j = 1, ..., n_i - 2). \tag{2.5}$$

Note that $\tilde{\lambda}_{j(i)} \geq 0$ w.p.1 for $j = 1, ..., n_i - 2$. See Appendix A for asymptotic properties of $\tilde{\lambda}_{1(i)}$s. Let

$$\widetilde{T}_{n_1,n_2} = U_{n_1,n_2}/\widetilde{K} + 1 \ \text{ with } \ \widetilde{K} = 2\tilde{\lambda}_{1(1)}^2/n_1 + 2\tilde{\lambda}_{1(2)}^2/n_2.$$

Then, we have the following result.

Theorem 2.1 (Ishii, Yata and Aoshima [8]). *Under (A-i), (A-iii), (1.3) and H_0, it holds that as $m \to \infty$*

$$\widetilde{T}_{n_1,n_2} \Rightarrow \chi_1^2.$$

We consider testing (1.1) for a given $\alpha \in (0, 1/2)$ by

$$\text{rejecting } H_0 \Longleftrightarrow \widetilde{T}_{n_1,n_2} \geq c_1(\alpha), \tag{2.6}$$

where $c_1(\alpha)$ denotes the upper α point of χ_1^2. Then, under (A-i), (A-iii) and (1.3), it holds that as $m \to \infty$

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

Although (1.3) is one of the SSE models, it is limited in actual data analyses. In Section 3, we give a new test procedure in general under the SSE model (A-ii).

3 New test procedure for the SSE model

In this section, we construct a new test procedure under the SSE model (A-ii). From the fact that $\mathrm{tr}(\boldsymbol{\Sigma}_{1\star}\boldsymbol{\Sigma}_{2\star}) \le \{\mathrm{tr}(\boldsymbol{\Sigma}_{1\star}^2)\mathrm{tr}(\boldsymbol{\Sigma}_{2\star}^2)\}^{1/2} \le \mathrm{tr}(\boldsymbol{\Sigma}_{1\star}^2) + \mathrm{tr}(\boldsymbol{\Sigma}_{2\star}^2)$, we note that $\boldsymbol{h}_{1(i)}^T \boldsymbol{\Sigma}_{i'\star} \boldsymbol{h}_{1(i)} = o(\lambda_{1(i')})$ for $i \ne i'$ and $||\boldsymbol{\Sigma}_{1\star} - \boldsymbol{\Sigma}_{2\star}||_F^2 = o(\lambda_{1(1)}^2 + \lambda_{1(2)}^2)$ as $d \to \infty$ under (1.3). Then, it holds that as $d \to \infty$

$$\Delta = (\lambda_{1(1)} - \lambda_{1(2)})^2 + 2\lambda_{1(1)}\lambda_{1(2)}\{1 - (\boldsymbol{h}_{1(1)}^T \boldsymbol{h}_{1(2)})^2\} + o(\lambda_{1(1)}^2 + \lambda_{1(2)}^2). \tag{3.1}$$

We consider (3.1) as a starting point to handle the SSE model (A-ii). We give a test statistic based on (3.1) and show that it holds an asymptotic null distribution even when (1.3) is not met. By using the NR method, we estimate the first eigenvector as

$$\tilde{\boldsymbol{h}}_{1(i)} = \{(n_i - 1)\tilde{\lambda}_{1(i)}\}^{-1/2} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i)\hat{\boldsymbol{u}}_{1(i)}$$

for $i = 1, 2$, where $\hat{\boldsymbol{u}}_{1(i)}$ is given in (2.4). Note that $\tilde{\boldsymbol{h}}_{1(i)} = (\hat{\lambda}_{1(i)}/\tilde{\lambda}_{1(i)})^{1/2}\hat{\boldsymbol{h}}_{1(i)}$, where $\hat{\boldsymbol{h}}_{1(i)}$ is the first (unit) eigenvector of $\boldsymbol{S}_{in_i}$. Let $\delta_i = \mathrm{tr}(\boldsymbol{\Sigma}_{i\star}^2),\ i = 1, 2$. From Lemma A.1 in Appendix A, it holds that as $m \to \infty$

$$|\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)}| = 1 + O_p\left(\frac{\delta_1^{1/2}}{n_1\lambda_{1(1)}} + \frac{\delta_2^{1/2}}{n_2\lambda_{1(2)}} \right) = 1 + o_p(n_{\min}^{-1}) \tag{3.2}$$

under (A-i), (1.3) and H_0, where $n_{\min}$ and m are given in (2.2). Then, from (3.1) we consider the following test statistic:

$$T_{\text{NR}} = \frac{(\tilde{\lambda}_{1(1)} - \tilde{\lambda}_{1(2)})^2 + 2\tilde{\lambda}_{1(1)}\tilde{\lambda}_{1(2)}\big\{1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\}\big\}}{\sum_{i=1}^2 2\tilde{\lambda}_{1(i)}^2/(n_i - 1)}.$$

From Lemma A.1 and (3.2), it holds that as $m \to \infty$

$$\begin{aligned} &\frac{\tilde{\lambda}_{1(1)} - \tilde{\lambda}_{1(2)}}{\Big\{ \sum_{i=1}^2 2\lambda_{1(i)}^2/(n_i - 1)\Big\}^{1/2}} = \frac{W}{K^{1/2}} + o_p(1) \Rightarrow N(0,1), \\ \text{and}\quad &\frac{\tilde{\lambda}_{1(1)}\tilde{\lambda}_{1(2)}\big\{1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\}\big\}}{\Big\{ \sum_{i=1}^2 2\lambda_{1(i)}^2/(n_i - 1)\Big\}} = o_p(1) \end{aligned} \tag{3.3}$$

under (A-i), (A-iii), (1.3) and H_0. Then, we have the following result.

Proposition 3.1 (Ishii, Yata and Aoshima [8])**.** *Under (A-i), (A-iii), (1.3) and H_0, it holds that $T_{\text{NR}} \Rightarrow \chi_1^2$ as $m \to \infty$.*

From (3.3) it holds that $T_{\text{NR}} = W^2/K + o_p(1)$ as $m \to \infty$ under (A-i), (A-iii), (1.3) and H_0. Thus, from Lemmas 2.1 and A.1, T_{NR} is asymptotically equivalent to $\widetilde{T}_{n_1,n_2}$ under (1.3) and H_0.

However, as for T_{NR}, one can consider it in general under the SSE model (A-ii) as follows: If $\limsup_{d\to\infty} \lambda_{1(i)}/\lambda_{2(i)} < \infty$ for some i, it holds from Lemma A.1 that

$$1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\} = O_p(n_{\min}^{-1}) \tag{3.4}$$

under H_0, (A-i) and (A-ii), so that (3.3) does not hold. Thus, one cannot ignore the bias of $1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\}$ in T_{NR} especially when $\lambda_{2(i)}$ is close to $\lambda_{1(i)}$. In order to reduce the bias, we consider modifying T_{NR}. Let $\eta = \delta_1^{1/2}/\lambda_{1(1)} + \delta_2^{1/2}/\lambda_{1(2)}$. From Lemma A.1, it follows that

$$\big\{1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\}\big\}^{1+\eta} = o_p(n_{\min}^{-1}) \tag{3.5}$$

under H_0, (A-i) and (A-ii). By using the cross-data-matrix (CDM) method by Yata and Aoshima [14], we have a consistent estimator of δ_i, $\hat{\delta}_i$. See (3.6) in Appendix A for the details. We estimate η by

$$\hat{\eta} = \hat{\delta}_1^{1/2}/\tilde{\lambda}_{1(1)} + \hat{\delta}_2^{1/2}/\tilde{\lambda}_{1(2)}.$$

We provide the following new test statistic:

$$T_{\text{NR}}^{\star} = \frac{(\tilde{\lambda}_{1(1)} - \tilde{\lambda}_{1(2)})^2 + 2\tilde{\lambda}_{1(1)}\tilde{\lambda}_{1(2)}\{1 - \min\{1, (\tilde{\boldsymbol{h}}_{1(1)}^T \tilde{\boldsymbol{h}}_{1(2)})^2\}\}^{1+\hat{\eta}}}{\sum_{i=1}^2 2\tilde{\lambda}_{1(i)}^2/(n_i - 1)}.$$

Then, we have the following result.

Proposition 3.2 (Ishii, Yata and Aoshima [8])**.** *Under (A-i) to (A-iii) and H_0, it holds that $T_{NR}^{\star} \Rightarrow \chi_1^2$ as $m \to \infty$.*

Note that one can use $T_{\text{NR}}^{\star}$ even when (1.3) is not met.

Appendix A

Estimation of several parameters in the new test procedure

In this section, we give asymptotic properties of the estimators for the parameters in the new test procedure.

A.1 Estimation of $\lambda_{1(i)}$ and $\boldsymbol{h}_{1(i)}$

Let $s_{1(i)} = \sum_{j=1}^{n_i}(z_{1j(i)} - \bar{z}_{1(i)})^2/(n_i - 1)$ for $i = 1, 2$, where $\bar{z}_{1(i)} = n_i^{-1}\sum_{j=1}^{n_i} z_{1j(i)}$.

Lemma A.1 (Ishii, Yata and Aoshima [8]). *Under (A-i) and (A-ii), it holds that as* $m \to \infty$

$$\frac{\tilde{\lambda}_{1(i)}}{\lambda_{1(i)}} = s_{1(i)} + O_p\left(\delta_i^{1/2}/(n_i\lambda_{1(i)})\right) = 1 + o_p(1) \text{ for } i = 1, 2, \text{ and}$$

$$\begin{aligned}|\tilde{\boldsymbol{h}}_{1(1)}^T\tilde{\boldsymbol{h}}_{1(2)}| &= |\boldsymbol{h}_{1(1)}^T\boldsymbol{h}_{1(2)}| + O_p\left(\frac{\delta_1^{1/2}}{n_1\lambda_{1(1)}} + \frac{\delta_2^{1/2}}{n_2\lambda_{1(2)}}\right)\\ &\quad + O_p\left(\{1 - (\boldsymbol{h}_{1(1)}^T\boldsymbol{h}_{1(2)})^2\}^{1/2}/n_{\min}^{1/2}\right)\\ &= |\boldsymbol{h}_{1(1)}^T\boldsymbol{h}_{1(2)}| + o_p(1).\end{aligned}$$

In addition, under (A-i) to (A-iii), it holds that as $m \to \infty$

$$\sqrt{\frac{n_i - 1}{2}}\left(\frac{\tilde{\lambda}_{1(i)}}{\lambda_{1(i)}} - 1\right) \Rightarrow N(0, 1).$$

A.2 Estimation of δ_i

First, we consider estimation of δ_i. Aoshima and Yata [2] gave an estimator of δ_i by using the CDM method: Let $n_{i1} = \lceil n_i/2 \rceil$ and $n_{i2} = n_i - n_{i1}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let $\boldsymbol{X}_{i1} = [\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_{i1}}]$ and $\boldsymbol{X}_{i2} = [\boldsymbol{x}_{in_{i1}+1}, ..., \boldsymbol{x}_{in_i}]$. We define

$$\boldsymbol{S}_{iC} = \{(n_{i1} - 1)(n_{i2} - 1)\}^{-1/2}(\boldsymbol{X}_{i1} - \overline{\boldsymbol{X}}_{i1})^T(\boldsymbol{X}_{i2} - \overline{\boldsymbol{X}}_{i2}),$$

where $\overline{\boldsymbol{X}}_{ij} = [\overline{\boldsymbol{x}}_{ij}, ..., \overline{\boldsymbol{x}}_{ij}]$ with $\overline{\boldsymbol{x}}_{i1} = \sum_{l=1}^{n_{i1}} \boldsymbol{x}_{il}/n_{i1}$ and $\overline{\boldsymbol{x}}_{i2} = \sum_{l=n_{i1}+1}^{n_i} \boldsymbol{x}_{il}/n_{i2}$. We estimate $\lambda_{j(i)}$ by the j-th singular value, $\acute{\lambda}_{j(i)}$, of $\boldsymbol{S}_{iC}$, where $\acute{\lambda}_{1(i)} \geq \cdots \geq \acute{\lambda}_{n_{i(2)}-1(i)} \geq 0$. Also, we estimate $\text{tr}(\boldsymbol{\Sigma}_i^2)$ by $\text{tr}(\boldsymbol{S}_{iC}\boldsymbol{S}_{iC}^T)$. Note that $E\{\text{tr}(\boldsymbol{S}_{iC}\boldsymbol{S}_{iC}^T)\} = \text{tr}(\boldsymbol{\Sigma}_i^2)$. Finally, δ_i is estimated by

$$\hat{\delta}_i = \text{tr}(\boldsymbol{S}_{iC}\boldsymbol{S}_{iC}^T) - \acute{\lambda}_{1(i)}^2. \tag{3.6}$$

From Lemma S2.1 in Aoshima and Yata [2], we have the following result.

Lemma A.2 (Aoshima and Yata [2]). *Under (A-i) and (A-ii), it holds that* $\hat{\delta}_i/\delta_i = 1 + o_p(1)$ *as* $m \to \infty$ *for* $i = 1, 2$.

Acknowledgements

Research of the first author was partially supported by Grant-in-Aid for Young Scientists, Japan Society for the Promotion of Science (JSPS), under Contract Number 18K18015. Research of the second author was partially supported by Grant-in-Aid for Scientific Research (C), JSPS, under Contract Number 18K03409. Research of the third author was partially supported by Grants-in-Aid for Scientific Research (A) and Challenging Research (Exploratory), JSPS, under Contract Numbers 15H01678 and 17K19956.

References

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

[2] Aoshima, M., Yata, K., 2018. Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Stat. Sin.* 28, 43-62.

[3] Bai, Z., Saranadasa, H., 1996. Effect of high dimension: By an example of a two sample problem. *Stat. Sin.* 6, 311-329.

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

[5] Ishii, A., Yata, K., 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. Plan. Inference* 170, 186-199.

[6] Ishii, A., 2017a. A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model. *Hiroshima Math. J.* 47, 273-288.

[7] Ishii, A., 2017b. A high-dimensional two-sample test for non-Gaussian data under a strongly spiked eigenvalue model. *J. Japan Statist. Soc.* 47, 273-291.

[8] Ishii, A., Yata, K., Aoshima, M., 2018. Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model. *J. Stat. Plan. Inference*, revised.

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

[10] Li, J., Chen, S.X., 2012. Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* 40, 908-940.

[11] Schott, J.R., 2007. A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Comput. Statist. Data Anal.* 51, 6535-6542.

[12] Shen, D., Shen, H., Zhu, H., Marron, J.S., 2016. The statistics and mathematics of high dimension low sample size asymptotics. *Stat. Sin.* 26, 1747-1770.

[13] 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.

[14] 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.

[15] 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.