数理解析研究所講究録 第2091巻 2018年 1-13

A test for high-dimensional covariance matrices via the extended cross-data-matrix methodology

Kohei Endo
Graduate School of Pure and Applied Sciences
University of Tsukuba

Kazuyoshi Yata
Institute of Mathematics
University of Tsukuba

Makoto Aoshima
Institute of Mathematics
University of Tsukuba

Abstract

In this paper, we consider two-sample tests for covariance matrices in high-dimensional settings. We introduce the extended cross-data-matrix (ECDM) methodology. We construct test statistics by using the ECDM methodology. We show that the ECDM test statistics have the consistency property and the asymptotic normality in high-dimensional settings. We propose a new test procedure based on the ECDM test statistics and evaluate its asymptotic size and power from theoretical and numerical aspects.

Keywords and phrases: Asymptotic normality; ECDM; HDLSS; Large p, small n

1 Introduction

High-dimension, low-sample-size (HDLSS) data situations occur in many areas of modern science such as genetic microarrays, medical imaging, text recognition, finance, chemometrics, and so on. In recent years, substantial work has been done on HDLSS asymptotic theory in which the sample size n is fixed or $n \to \infty$ while $n/p \to 0$ as the data dimension $p \to \infty$. Hall et al. [6] and Yata and Aoshima [11] explored several types of geometric representations of HDLSS data. Yata and Aoshima [11] developed the noise-reduction methodology and gave consistent estimators of both the eigenvalues and eigenvectors together with principal component (PC) scores in the HDLSS context. The HDLSS asymptotic theory was created under the assumption that either the population distribution is Gaussian or random variables in a certain sphered data matrix have the ρ-mixing dependency. However, Yata and Aoshima [10] developed the HDLSS asymptotic theory without such the assumptions. Moreover, they created a new principal component analysis (PCA) called the cross-data-matrix (CDM) methodology that is applicable to constructing an unbiased estimator in HDLSS non-parametric settings. Aoshima and Yata [2] developed a variety of inference for HDLSS data such as given-bandwidth confidence regions, two-sample tests, tests of the equality of two covariance matrices, classification, variable selection, regression, tests of the correlation coefficients and so on, and also discussed the sample size determination to ensure prespecified accuracy for each inference. Yata and Aoshima [12] improved the test of the correlation coefficients by using the *extended cross-data-matrix (ECDM) methodology* that is an extension of the CDM method. One of the advantages of the ECDM methodology is to produce an unbiased estimator having small asymptotic variance at a low computational cost.

In this paper, we consider two-sample tests for high-dimensional covariance matrices. Let $\boldsymbol{x}_{h1}, \dots, \boldsymbol{x}_{hn_h}$ be independent and identically distributed (i.i.d.) samples of a p-variate random variable from populations π_h $(h = 1, 2)$. We assume $n_1/n_2 \to \theta \in (0, \infty)$ and $n_h < p$ for $h = 1, 2$. We assume that $\boldsymbol{x}_{hj}$ has an unknown mean vector $\boldsymbol{\mu}_h$ and unknown covariance matrix $\boldsymbol{\Sigma}_h$ $(\geq \boldsymbol{O})$ for $h = 1, 2$. We consider a test problem as follows:

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

The test problem has been studied in the conventional low-dimensional settings. In particular, the likelihood ratio test (LRT) is commonly used and enjoys certain optimality under regularity conditions, see Anderson [1]. However, in the high-dimensional settings, the conventional test procedures such as the LRT perform poorly or are not even defined since the sample covariance matrix $\boldsymbol{S}$, the simplest estimator of the population covariance matrix $\boldsymbol{\Sigma}$, performs poorly. Li and Chen [8] gave a test statistic based on U-statistics for high-dimensional data. However, it requires high computational cost. Srivastava et al. [9] improved the statistic in terms of computational cost. In this paper, we shall produce statistics by the ECDM method and propose a new test procedure for (1).

The rest of the paper is organized as follows. In Section 2, we state assumptions required in the construction of a test procedure for (1). In Section 3, we introduce the ECDM methodology. In Section 4, we produce test statistics for (1) by using the ECDM methodology. We show that the ECDM test statistics have the consistency property and the asymptotic normality in high-dimensional settings. We propose a new test procedure based on the ECDM test statistics and evaluate its asymptotic size and power theoretically. Finally, in Section 5, we give simulation studies to check the performance of the proposed test procedure.

2 Assumptions

In this section, we introduce the basic assumptions for the test of hypotheses (1). The eigen-decomposition of $\boldsymbol{\Sigma}_h$ is given by $\boldsymbol{\Sigma}_h = \boldsymbol{H}_h \boldsymbol{\Lambda}_h \boldsymbol{H}_h^T$, where $\boldsymbol{\Lambda}_h = \text{diag}(\lambda_{h1}, ..., \lambda_{hp})$ is a diagonal matrix of eigenvalues, $\lambda_{h1} \geq \cdots \geq \lambda_{hp} \geq 0$, and $\boldsymbol{H}_h$ is an orthogonal matrix of the corresponding eigenvectors.

Now, we assume the following model:

$$\boldsymbol{x}_{hj} = \boldsymbol{\Gamma}_h \boldsymbol{w}_{hj} + \boldsymbol{\mu}_h, \quad \text{for } h = 1, 2,$$

where $\boldsymbol{\Gamma}_h = (\boldsymbol{\gamma}_{h1}, ..., \boldsymbol{\gamma}_{hq_h})$ is a $p \times q_h$ matrix for some $q_h > 0$ such that $\boldsymbol{\Gamma}_h \boldsymbol{\Gamma}_h^T = \boldsymbol{\Sigma}_h$, and $\boldsymbol{w}_{hj} = (w_{h1j}, ..., w_{hq_hj})^T$, $j = 1, ..., n_h$, are i.i.d. random vectors having $E(\boldsymbol{w}_{hj}) = \boldsymbol{0}$ and $\text{Var}(\boldsymbol{w}_{hj}) = \boldsymbol{I}_{q_h}$. Here, $\boldsymbol{I}_{q_h}$ denotes the identity matrix of dimension q_h. Let $\text{Var}(w_{hrj}^2) = M_{hr}$, $r = 1, ..., q_h$; $h = 1, 2$. We assume that $\limsup_{p\to\infty} M_{hr} < \infty$ for all h, r. Similar to Aoshima and Yata [3] and Bai and Saranadasa [4], we assume that

(A-i) $E(w_{hrj}^2 w_{hsj}^2) = E(w_{hrj}^2)E(w_{hsj}^2) = 1$ and $E(w_{hrj}w_{hsj}w_{htj}w_{huj}) = 0$ for all $r \neq s, t, u$.

We assume the following assumption instead of (A-i) as necessary:

(A-ii) $E(w_{hr_1j}^{\alpha_1} w_{hr_2j}^{\alpha_2} \cdots w_{hr_vj}^{\alpha_v}) = E(w_{hr_1j}^{\alpha_1})E(w_{hr_2j}^{\alpha_2}) \cdots E(w_{hr_vj}^{\alpha_v})$ for all $r_1 \neq r_2 \neq \cdots \neq r_v \in [1, q_h]$ and $\alpha_i \in [1, 4]$, $i = 1, ..., v$, where $v \leq 8$ and $\sum_{i=1}^{v} \alpha_i \leq 8$.

See Chen and Qin [5] about (A-ii). Note that (A-ii) implies (A-i). Note that (A-ii) is naturally satisfied when $\boldsymbol{x}_{hj}$ is Gaussian. We assume the following assumption for $\boldsymbol{\Sigma}_h$ as necessary:

(A-iii) $\dfrac{\text{tr}(\boldsymbol{\Sigma}_h^4)}{\text{tr}(\boldsymbol{\Sigma}_h^2)^2} \to 0$ as $p \to \infty$ for $h = 1, 2$.

Note that "$\text{tr}(\boldsymbol{\Sigma}_h^4)/\text{tr}(\boldsymbol{\Sigma}_h^2)^2 \to 0$ as $p \to \infty$ for $h = 1, 2$" is equivalent to "$\lambda_{h1}/\text{tr}(\boldsymbol{\Sigma}_h^2)^{1/2} \to 0$ as $p \to \infty$ for $h = 1, 2$". Let $m = \min\{p, n_1, n_2\}$ and $\Delta = ||\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2||_F^2 = \text{tr}\{(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)^2\}$. We assume one of the following two assumptions as necessary:

(A-iv) $\dfrac{\text{tr}(\boldsymbol{\Sigma}_h^2)}{n_h \Delta} \to 0$ as $m \to \infty$ for $h = 1, 2$;

(A-v) $\limsup_{m\to\infty} \left\{ \dfrac{n_h \Delta}{\text{tr}(\boldsymbol{\Sigma}_h^2)} \right\} < \infty$ for $h = 1, 2$.

3 ECDM methodology

The ECDM methodology was developed by Yata and Aoshima [12, 13] as an extension of the CDM method due to Yata and Aoshima [10]. Throughout this section, we omit the subscript with regard to the population. Let $n_{(1)} = \lceil n/2 \rceil$ and $n_{(2)} = n - n_{(1)}$, where $\lceil x \rceil$ denotes the smallest integer $\geq x$. Let

$$\boldsymbol{V}_{n(1)(k)} = \begin{cases} \{\lfloor k/2 \rfloor - n_{(1)} + 1, ..., \lfloor k/2 \rfloor\} & \text{if } \lfloor k/2 \rfloor \geq n_{(1)}, \\ \{1, ..., \lfloor k/2 \rfloor\} \cup \{\lfloor k/2 \rfloor + n_{(2)} + 1, ..., n\} & \text{otherwise;} \end{cases}$$

$$\boldsymbol{V}_{n(2)(k)} = \begin{cases} \{\lfloor k/2 \rfloor + 1, ..., \lfloor k/2 \rfloor + n_{(2)}\} & \text{if } \lfloor k/2 \rfloor \leq n_{(1)}, \\ \{1, ..., \lfloor k/2 \rfloor - n_{(1)}\} \cup \{\lfloor k/2 \rfloor + 1, ..., n\} & \text{otherwise} \end{cases}$$

for $k = 3, ..., 2n-1$, where $\lfloor x \rfloor$ denotes the largest integer $\le x$. Let $\#\boldsymbol{S}$ denote the number of elements in a set $\boldsymbol{S}$. Note that $\#\boldsymbol{V}_{n(l)(k)} = n_{(l)}$, $l = 1, 2$, $\boldsymbol{V}_{n(1)(k)} \cap \boldsymbol{V}_{n(2)(k)} = \emptyset$ and $\boldsymbol{V}_{n(1)(k)} \cup \boldsymbol{V}_{n(2)(k)} = \{1, ..., n\}$ for $k = 3, ..., 2n-1$. Also, note that $i \in \boldsymbol{V}_{n(1)(i+j)}$ and $j \in \boldsymbol{V}_{n(2)(i+j)}$ for $i < j$ $(\le n)$. Let

$$\overline{\boldsymbol{x}}_{n(1)(k)} = n_{(1)}^{-1} \sum_{j \in \boldsymbol{V}_{n(1)(k)}} \boldsymbol{x}_j \quad \text{and} \quad \overline{\boldsymbol{x}}_{n(2)(k)} = n_{(2)}^{-1} \sum_{j \in \boldsymbol{V}_{n(2)(k)}} \boldsymbol{x}_j$$

for $k = 3, ..., 2n-1$. Then, Yata and Aoshima [12] gave an estimator of $\mathrm{tr}(\boldsymbol{\Sigma}^2)$ by

$$W_n = \frac{2u_n}{n(n-1)} \sum_{i<j}^{n} \left\{ (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{n(1)(i+j)})^T (\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{n(2)(i+j)}) \right\}^2, \tag{2}$$

where $u_n = n_{(1)} n_{(2)} / \{(n_{(1)} - 1)(n_{(2)} - 1)\}$. Note that $E(W_n) = \mathrm{tr}(\boldsymbol{\Sigma}^2)$. Let $m_0 = \min\{p, n\}$. Aoshima and Yata [3] and Yata and Aoshima [13] gave the following result.

Theorem 3.1 ([3, 13]). *Assume (A-i). Then, it holds that as* $m_0 \to \infty$

$$Var\left(\frac{W_n}{tr(\boldsymbol{\Sigma}^2)}\right) = \left(\frac{4}{n^2} + \frac{8tr(\boldsymbol{\Sigma}^4) + 4\sum_{r=1}^{q}(M_r - 2)(\boldsymbol{\gamma}_r^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_r)^2}{tr(\boldsymbol{\Sigma}^2)^2 n}\right)\{1 + o(1)\}.$$

Remark 1. When $\boldsymbol{x}_j$ is Gaussian, it holds that as $m_0 \to \infty$

$$\mathrm{Var}\left(\frac{W_n}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\right) = \left(\frac{4}{n^2} + \frac{8\mathrm{tr}(\boldsymbol{\Sigma}^4)}{\mathrm{tr}(\boldsymbol{\Sigma}^2)^2 n}\right)\{1 + o(1)\}.$$

4 Two-sample tests for covariance matrices

In this section, we consider two-sample tests for covariance matrices using the ECDM methodology. Let us consider the following hypotheses which are equivalent to (1):

$$H_0 : \Delta = 0 \quad \text{vs.} \quad H_1 : \Delta > 0.$$

Note that $\Delta = \mathrm{tr}(\boldsymbol{\Sigma}_1^2) + \mathrm{tr}(\boldsymbol{\Sigma}_2^2) - 2\mathrm{tr}(\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2)$. Li and Chen [8] proposed a test statistic as follows:

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

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

$$A_{n_h} = \frac{1}{n_h(n_h-1)}\sum_{j\neq k}^{n_h}(\boldsymbol{x}_{hj}^T\boldsymbol{x}_{hk})^2 - \frac{2}{n_h(n_h-1)(n_h-2)}\sum_{j\neq k\neq l}^{n_h}\boldsymbol{x}_{hk}^T\boldsymbol{x}_{hj}\boldsymbol{x}_{hj}^T\boldsymbol{x}_{hl}$$
$$+\frac{1}{n_h(n_h-1)(n_h-2)(n_h-3)}\sum_{j\neq k\neq l\neq l'}^{n_h}\boldsymbol{x}_{hj}^T\boldsymbol{x}_{hk}\boldsymbol{x}_{hl}^T\boldsymbol{x}_{hl'}.$$

Under (A-ii), (A-iii) and some regularity conditions, they showed the following asymptotic result:

$$\frac{U-\Delta}{2\mathrm{tr}(\boldsymbol{\Sigma}_1^2)/n_1 + 2\mathrm{tr}(\boldsymbol{\Sigma}_2^2)/n_2} \Rightarrow N(0,1) \text{ as } m\to\infty. \tag{3}$$

Here, " $\Rightarrow$ " denotes the convergence in distribution and $N(0,1)$ denotes a random variable distributed as the standard normal distribution. However, the computational cost of A_{n_h} is of the order, $O(pn_h^4)$, which is inappropriate for practical use. On the other hand, Srivastava et al. [9] modified A_{n_h} as

$$\frac{(n_h-1)(n_h-2)\mathrm{tr}(\boldsymbol{M}_h^2) - n_h(n_h-1)\mathrm{tr}(\boldsymbol{D}_h^2) + \mathrm{tr}(\boldsymbol{D}_h)^2}{n_h(n_h-1)(n_h-2)(n_h-3)} \quad (= A_{n_h}^*, \text{ say}),$$

where $\boldsymbol{Y}_h = (\boldsymbol{y}_{h1},\dots,\boldsymbol{y}_{hn_h})$, $\boldsymbol{y}_{hj} = \boldsymbol{x}_{hj} - \overline{\boldsymbol{x}}_h, j = 1,\dots,n_h$, $\overline{\boldsymbol{x}}_h = n_h^{-1}\sum_{j=1}^{n_h}\boldsymbol{x}_{hj}$, $\boldsymbol{M}_h = \boldsymbol{Y}_h^T\boldsymbol{Y}_h$, $\boldsymbol{D}_h = \mathrm{diag}(\boldsymbol{y}_{h1}^T\boldsymbol{y}_{h1},\dots,\boldsymbol{y}_{hn_h}^T\boldsymbol{y}_{hn_h})$, for $h=1,2$. The computational cost of $A_{n_h}^*$ is of the order, $O(pn_h^2)$. Let

$$\sigma_U = 2\Big(\frac{1}{n_1-1}+\frac{1}{n_2-1}\Big)\frac{(n_1-1)\mathrm{tr}(\boldsymbol{\Sigma}_1^2)+(n_2-1)\mathrm{tr}(\boldsymbol{\Sigma}_2^2)}{n_1+n_2-2}$$

and

$$\widehat{\sigma}_U = 2\Big(\frac{1}{n_1-1}+\frac{1}{n_2-1}\Big)\frac{(n_1-1)A_{n_1}^*+(n_2-1)A_{n_2}^*}{n_1+n_2-2}.$$

Then, they gave a test procedure for (1) by

$$\text{rejecting } H_0 \Longleftrightarrow \frac{U}{\widehat{\sigma}_U} > z_\alpha, \tag{4}$$

where z_α is a constant such that $P\{N(0,1) > z_\alpha\} = \alpha$. Then, under (A-ii) and (A-iii), it holds that as $m\to\infty$

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

Also, we have the following result.

Theorem 4.1. *Assume (A-ii) and (A-iii). The test by (4) has that as* $m \to \infty$

$$Power = \Phi\Big(\frac{\Delta}{\sigma_U} - z_\alpha\Big) + o(1), \tag{5}$$

where $\Phi(\cdot)$ *denotes the c.d.f. of* $N(0,1)$.

In this paper, we give a more powerful test statistic compared with U.

4.1 A test statistic based on the ECDM methodology

Now, by using the ECDM methodology, we estimate Δ by

$$T = W_{n_1} + W_{n_2} - 2\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}),$$

where W_{n_h}s are given by (2). Note that $E(T) = \Delta$. Also, note that the computational cost of W_{n_h} is of the order, $O(pn_h^2)$. Let

$$\sigma^2 = \sum_{h=1}^{2}\Big(\frac{4}{n_h^2}\mathrm{tr}(\boldsymbol{\Sigma}_h^2)^2 + \frac{8}{n_h}\mathrm{tr}\{(\boldsymbol{\Sigma}_h^2 - \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2\} + \frac{4}{n_h}\sum_{r=1}^{q_h}(M_{hr} - 2)\{\boldsymbol{\gamma}_{hr}^T(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\boldsymbol{\gamma}_{hr}\}^2\Big) + \frac{8}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2.$$

Then, we have the following result.

Lemma 4.1. *Assume (A-i). It holds that as* $m \to \infty$

$$Var(T) = \sigma^2\{1 + o(1)\}.$$

From Lemma 4.1 we have the following result.

Theorem 4.2. *Assume (A-i) and (A-iv). It holds that as* $m \to \infty$

$$\frac{T}{\Delta} = 1 + o_P(1).$$

Next, we consider the case when (A-iv) is not met. Let

$$\sigma_T^2 = \frac{4}{n_1^2}\mathrm{tr}(\boldsymbol{\Sigma}_1^2)^2 + \frac{4}{n_2^2}\mathrm{tr}(\boldsymbol{\Sigma}_2^2)^2 + \frac{8}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)^2$$

and

$$\widehat{\sigma}_T^2 = \frac{4}{n_1^2}W_{n_1}^2 + \frac{4}{n_2^2}W_{n_2}^2 + \frac{8}{n_1 n_2}\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})^2.$$

We have the following result.

Lemma 4.2. *Assume (A-i), (A-iii) and (A-v). It holds that as $m \to \infty$*

$$\sigma^2 = \sigma_T^2\{1 + o(1)\}.$$

Note that as $m \to \infty$

$$\frac{\widehat{\sigma}_T^2}{\sigma_T^2} = 1 + o_P(1)$$

under (A-i). From Lemma 4.2 we have the following result.

Theorem 4.3. *Assume (A-ii), (A-iii) and (A-v). It holds that as $m \to \infty$*

$$\frac{T - \Delta}{\widehat{\sigma}_T} = \frac{T - \Delta}{\sigma_T} + o_P(1) \Rightarrow N(0, 1).$$

Now, we consider a more powerful test statistic. Let

$$c = \max\Big\{\frac{\mathrm{tr}(\boldsymbol{\Sigma}_1)}{\mathrm{tr}(\boldsymbol{\Sigma}_2)}, \frac{\mathrm{tr}(\boldsymbol{\Sigma}_2)}{\mathrm{tr}(\boldsymbol{\Sigma}_1)}\Big\} \quad \text{and} \quad \hat{c} = \max\Big\{\frac{\mathrm{tr}(\boldsymbol{S}_{1n_1})}{\mathrm{tr}(\boldsymbol{S}_{2n_2})}, \frac{\mathrm{tr}(\boldsymbol{S}_{2n_2})}{\mathrm{tr}(\boldsymbol{S}_{1n_1})}\Big\}.$$

Then, under (A-i), we have that as $m \to \infty$

$$\hat{c} = c + o_P(1). \tag{6}$$

We propose the following test statistic:

$$T_* = \hat{c}T.$$

Note that $T_* \geq T$ w.p.1. Also, note that as $m \to \infty$

$$T_* = T\{1 + o_P(1)\}$$

under (A-i) and H_0. Then, from Theorems 4.2 and 4.3, we have the following results.

Corollary 4.1. *Assume (A-i) and (A-iv). It holds that as $m \to \infty$*

$$\frac{T_*}{c\Delta} = 1 + o_P(1).$$

Corollary 4.2. *Assume (A-ii), (A-iii) and (A-v). It holds that as $m \to \infty$*

$$\frac{T_*/c - \Delta}{\widehat{\sigma}_T} \Rightarrow N(0, 1).$$

4.2 A more powerful two-sample test

We give a test procedure for (1) by

$$\text{rejecting } H_0 \Longleftrightarrow \frac{T_*}{\widehat{\sigma}_T} > z_\alpha. \tag{7}$$

Then, we have the following results.

Theorem 4.4. *Assume (A-ii) and (A-iii). The test by (7) has that as $m \to \infty$*

$$\textit{Size} = \alpha + o(1) \quad \textit{and} \quad \textit{Power} = \Phi\Big(\frac{\Delta}{\sigma_T} - z_\alpha/c\Big) + o(1).$$

Corollary 4.3. *Assume (A-iv) under H_1. Assume also (A-i). Then, the test by (7) has that as $m \to \infty$*

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

Remark 2. We consider testing (1) by (7) with T instead of T_*. Then, it has (5) as $m \to \infty$.

From Theorems 4.1 and 4.4, when $c > 1$, the asymptotic power of (7) is greater than that of (4). Thus, we recommend to use the test by (7).

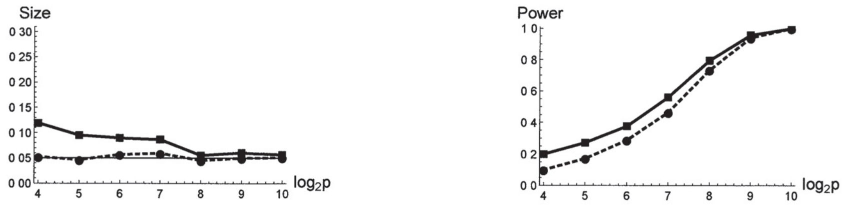
5 Simulation studies

In this section, we summarize simulation studies of the findings. We used computer simulations to study performances of the test procedures by (4) and (7). Independent pseudo-random normal observations were generated from $\pi_h : N_p(\mathbf{0}, \mathbf{\Sigma}_h)$ for $h = 1, 2$. We set $\alpha = 0.05$ and $\mathbf{\Sigma}_1 = \boldsymbol{B}(0.3^{|i-j|^{1/3}})\boldsymbol{B}$, where

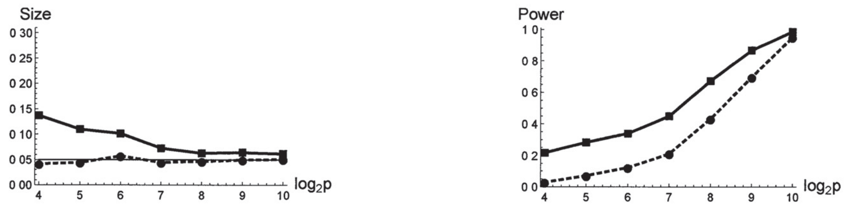
$$\boldsymbol{B} = \mathrm{diag}[\{0.5 + 1/(p+1)\}^{1/2}, ..., \{0.5 + p/(p+1)\}^{1/2}].$$

As for the alternative hypothesis, we set $\mathbf{\Sigma}_2 = 1.2\boldsymbol{B}(0.4^{|i-j|^{1/3}})\boldsymbol{B}$. We considered four cases for p and n_is:

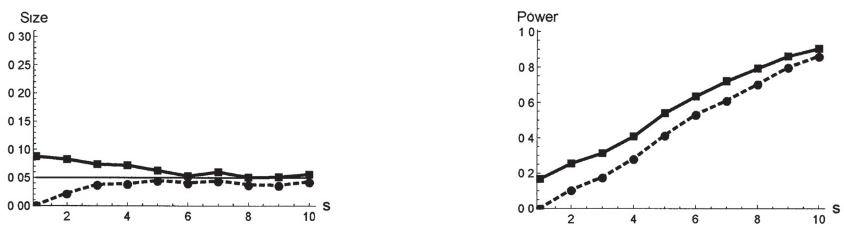
$$\begin{array}{ll}
\text{(I)} & p = 2^s \ (s = 4, ..., 10), \ n_1 = n_2 = 4\lceil (p/2)^{1/2} \rceil; \\
\text{(II)} & p = 2^s \ (s = 4, ..., 10), \ n_1 = 2\lceil (p/2)^{1/2} \rceil, \ n_2 = 4\lceil (p/2)^{1/2} \rceil; \\
\text{(III)} & p = 1000, \ n_1 = n_2 = 5s \ (s = 1, ..., 10); \\
\text{(IV)} & p = 1000, \ n_1 = 4s, \ n_2 = 8s \ (s = 1, ..., 10).
\end{array}$$



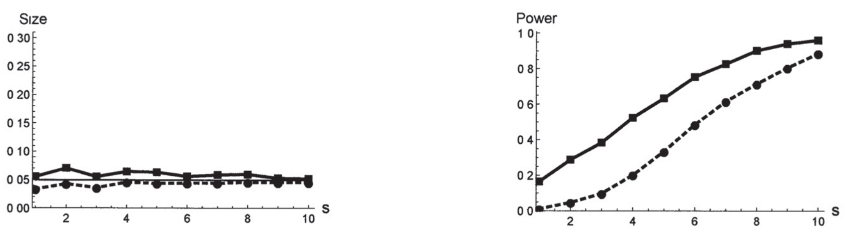
(I) $p = 2^s$ $(s = 4, ..., 10)$, $n_1 = n_2 = 4\lceil (p/2)^{1/2} \rceil$



(II) $p = 2^s$ $(s = 4, ..., 10)$, $n_1 = 2\lceil (p/2)^{1/2} \rceil$, $n_2 = 4\lceil (p/2)^{1/2} \rceil$



(III) $p = 1000$, $n_1 = n_2 = 5s$ $(s = 1, ..., 10)$



(IV) $p = 1000$, $n_1 = 4s$, $n_2 = 8s$ $(s = 1, ..., 10)$

Figure 1: The performances of the test procedures by (4) and (7). For each panel, the value of (4) is denoted by the dashed line and the value of (7) is denoted by the solid line.

For each case, we checked the performance by 2000 replications. We defined $P_r = 1$ (or 0) when H_0 was falsely rejected (or not) for $r = 1, ..., 2000$, and defined $\overline{\alpha} = \sum_{r=1}^{2000} P_r/2000$ to estimate the size. We also defined $P_r = 1$ (or 0) when H_1 was falsely rejected (or not) for $r = 1, ..., 2000$, and defined $1 - \overline{\beta} = 1 - \sum_{r=1}^{2000} P_r/2000$ to estimate the power. Note that their standard deviations are less than 0.011. In Figure 1, we plotted $\overline{\alpha}$ (left panel) and $1 - \overline{\beta}$ (right panel) in case of (I) to (IV). We observed that both the test procedures gave preferable performances for the size in (I) to (IV). However, the test procedure by (7) gave better performance compared to (4) with respect to the power. See Section 4.2 for the theoretical reason.

Appendix

Proof of Theorem 4.1. Note that $\mathrm{tr}(\boldsymbol{\Sigma}_1^2) = \mathrm{tr}(\boldsymbol{\Sigma}_2^2)\{1 + o(1)\}$ as $p \to \infty$ under (A-v). Also, note that

$$\sum_{r=1}^{q_h}\{\gamma_{hr}^T(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\gamma_{hr}\}^2 \le \mathrm{tr}[\{\boldsymbol{\Sigma}_h(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\}^2] \le \lambda_{h1}^2\Delta = o(\mathrm{tr}(\boldsymbol{\Sigma}_h^2)\Delta) \tag{8}$$

under (A-iii) for $h = 1, 2$. Then, it holds that $\sigma/\sigma_U = 1 + o(1)$ under (A-iii) and (A-v). Thus, from Theorems 1 and 2 in Li and Chen [8], under (A-ii), (A-iii) and (A-v), we have that as $m \to \infty$

$$\begin{aligned} P\Big(\frac{U}{\widehat{\sigma}_U} > z_\alpha\Big) &= P\Big(\frac{U - \Delta}{\sigma_U} > z_\alpha - \frac{\Delta}{\sigma_U} + o_P(1)\Big) \\ &= \Phi\Big(\frac{\Delta}{\sigma_U} - z_\alpha\Big) + o(1). \end{aligned}$$

On the other hand, from (8), under (A-ii) and (A-iv), it holds that $\mathrm{Var}(U)/\Delta^2 = o(1)$, so that $U/\Delta = 1 + o_P(1)$. Then, we have that

$$P\Big(\frac{U}{\widehat{\sigma}_U} > z_\alpha\Big) = P\Big(\frac{\Delta}{\sigma_U}\{1 + o_P(1)\} > z_\alpha\Big) \to 1$$

from the fact that $\sigma_U/\Delta = o(1)$ under (A-iv). Thus, by considering the convergent subsequence of Δ/σ_U, we can conclude the result. □

Proof of Lemma 4.1. Assume (A-i). Recall that

$$T = W_{n_1} + W_{n_2} - 2\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}) \ \text{ and } \ U = A_{n_1} + A_{n_2} - 2\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}).$$

Note that $E(T) = \Delta$. By noting that $\mathrm{Cov}(W_{n_1}, W_{n_2}) = 0$, it holds that

$$\begin{aligned}\mathrm{Var}(T) =& \mathrm{Var}(W_{n_1}) + \mathrm{Var}(W_{n_2}) + 4\mathrm{Var}(\mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})) \\ & -4\mathrm{Cov}(W_{n_1}, \mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})) - 4\mathrm{Cov}(W_{n_2}, \mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})).\end{aligned}$$

From Theorem 3.1 and (6.2) in [8], we can claim that $\mathrm{Var}(W_{n_h}) = \mathrm{Var}(A_{n_h})\{1+o(1)\}$ as $m \to \infty$ for $h = 1, 2$. Also, we can claim that $\mathrm{Cov}(W_{n_h}, \mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2})) = \mathrm{Cov}(A_{n_h}, \mathrm{tr}(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}))\{1+o(1)\}$ for $h = 1, 2$. Then, from (2.5) and (6.1) in [8], we can conclude the result. □

Proof of Theorem 4.2. From (8), under (A-i) and (A-iv), it holds that $\mathrm{Var}(T)/\Delta^2 = o(1)$, so that $T/\Delta = 1 + o_P(1)$. It concludes the result. □

Proof of Lemma 4.2. From (8), we can conclude the result. □

Proof of Theorem 4.3. Similarly to Proof of Lemma 3.1 in Ishii et al. [7], under (A-i), we can claim that $W_{n_h} = A_{n_h} + o_P(\sigma)$ as $m \to \infty$ for $h = 1, 2$. From (8), it holds that $\sigma = \sigma_T\{1+o(1)\}$ under (A-iii) and (A-v). Then, in a way similar to Proof of Theorem 1 in [8], we can conclude the result. □

Proofs of Corollary 4.1 and Corollary 4.2. From Theorems 4.2 and 4.3, by using Slutsky's theorem, we can conclude the results. □

Proof of Theorem 4.4. Similarly to Proof of Theorem 4.1, by using Corollary 4.2, we can conclude the result. □

Proof of Corollary 4.3. By using Corollary 4.1, we can conclude the result. □

Acknowledgements

The research of the second author was partially supported by Grant-in-Aid for Young Scientists (B), Japan Society for the Promotion of Science (JSPS), under Contract Number 26800078. The 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] Anderson T. W. (2003). An introduction to multivariate statistical analysis 3rd ed. Wiley.

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

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

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

[5] 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.

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

[7] Ishii A., Yata K., Aoshima M. (2017). Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model, submitted.

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

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

[10] 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.

[11] 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.

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

[13] Yata K., Aoshima M. (2016). High-dimensional inference on covariance structures via the extended cross-data-matrix methodology. J. Multivariate Anal. 151, 151-166.

Institute of Mathematics
University of Tsukuba
Ibaraki 305-8571
Japan
E-mail address: yata@math.tsukuba.ac.jp

筑波大学・数理物質科学研究科　遠藤紘平
筑波大学・数理物質系　矢田和善
筑波大学・数理物質系　青嶋　誠