



Limit theorem associated with Wishart matrices with application to hypothesis testing for common principal components

Koji Tsukuda^{a,*}, Shun Matsuura^b

^a Faculty of Mathematics, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan

^b Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan

ARTICLE INFO

Article history:

Received 21 July 2020

Received in revised form 24 August 2021

Accepted 24 August 2021

Available online 3 September 2021

AMS 2020 subject classifications:

60F05

62F05

62H25

Keywords:

Asymptotic test

Central limit theorem

Common principal components model

High-dimension

Wishart distribution

ABSTRACT

This paper describes the derivation of a new property of the Wishart distribution when the degrees of freedom and the sizes of scale matrices grow simultaneously. In particular, the asymptotic normality of the trace of the product of four independent Wishart matrices is demonstrated for a high-dimensional asymptotic regime. As an application of the result, a statistical test procedure for the common principal components hypothesis is proposed. For this problem, the proposed test statistic is asymptotically normal under the null hypothesis and diverges to positive infinity in probability under the alternative hypothesis.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

This paper investigates the asymptotic normality of the trace of products of four independent Wishart matrices in a high-dimensional setting, and proposes a statistical test procedure for the common principal components (CPC) hypothesis on two covariance matrices. Here, the CPC hypothesis, first considered by Flury [8] in 1984, means several covariance matrices can be simultaneously diagonalized.

A classical setting in multivariate analysis is that population distributions are normal and that the number of observed variables is much less than the number of individuals in a sample. In 1928, John Wishart derived the probability density function of sample covariance matrices calculated from a sample of size N from the p -dimensional normal population ($p < N - 1$). It determines the so-called Wishart distribution, the distribution of a scatter matrix $\sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^T$ calculated from i.i.d. (independent and identically distributed) p -dimensional zero-mean normal vectors $\{\mathbf{v}_i\}_{i=1}^n$, where $n = N - 1$ and T denotes the transpose. A number of studies have investigated its asymptotic properties under the traditional multivariate analysis setting: $n \rightarrow \infty$ with fixed p . As the observed variables have increased with the development of information technology, however, multivariate statistical methods have been developed to deal with this situation. In particular, when few variables are observed, the likelihood ratio test is quite useful for testing hypotheses about population covariance

* Corresponding author.

E-mail address: tsukuda@math.kyushu-u.ac.jp (K. Tsukuda).

matrices. On the other hand, when more variables are observed than the number of individuals in the samples (the so-called high-dimensional setting), the likelihood ratio test is unavailable in many cases because scatter matrices are not of full-rank. In one-sample testing problems such as “the population covariance matrix is an identity matrix”, “the covariance population matrix is spherical”, and “the covariance matrix is diagonal”, alternative test procedures that cleverly use the trace of some functions of a scatter matrix have been proposed. Such procedures are considered to be effective in high-dimensional settings; see, e.g., Chen et al. [6], Hyodo et al. [16,17], Ishii et al. [19], Srivastava [24], Srivastava et al. [26], Yamada et al. [30], and Yata et al. [31].

In two-sample testing problems for covariance matrices, hypotheses such as “two covariance population matrices are identical”, “two population covariance matrices are proportional”, and “two population covariance matrices can be simultaneously diagonalized” (the CPC hypothesis) have been considered. These three hypotheses are typical in two-sample problems in multivariate analysis. Indeed, they correspond to Flury’s hierarchical model, where the likelihood ratio is used for testing and model selection; see, e.g., Flury [10] and Phillips and Arnold [22]. As the likelihood ratio test for testing the above three hypotheses is unavailable in a high-dimensional setting, alternative test procedures have been proposed for the former two hypotheses (equality and proportionality); see, e.g., Aoshima and Yata [1], Ishii et al. [18], Li and Chen [20], Liu et al. [21], Schott [23], Srivastava and Yanagihara [25], Srivastava et al. [26], Tsukuda and Matsuura [28] and Xu et al. [29]. In contrast, for testing the CPC hypothesis, Flury [8] and subsequent studies, such as those of Boente et al. [4], Boik [5] and Hallin et al. [13,14] have not considered high-dimensional settings. Therefore, in this paper, we propose a test procedure for the CPC hypothesis in a high-dimensional setting. Approaches based on a statistic that use the trace of some functions of scatter matrices are considered to be promising for two-sample high-dimensional problems; see, e.g., Aoshima and Yata [1], Schott [23], Srivastava and Yanagihara [25], Srivastava et al. [26], and Sugiyama et al. [27]. Hence, we take a similar approach with the trace operator. High-dimensional data have been collected in various fields such as evolutionary biology; see, e.g., Collyer et al. [7]. In evolutionary biology, CPC analysis is widespread (Houle et al. [15]), and testing the CPC hypothesis is crucial for conducting CPC analysis. Thus, this paper aims to contribute to making CPC analysis more widely available.

The remainder of this paper is organized as follows. In Section 2, we present the main result with an outline of its proof. An application of the limit theorem for testing the CPC hypothesis is described in Section 3. Concluding remarks are provided in Section 4.

2. Limit theorem

2.1. Problem setting and assumption

Let n_a, n_b, n_c, n_d and p be positive integers and $\Sigma_a, \Sigma_b, \Sigma_c$, and Σ_d be positive-definite matrices. Consider four independent Wishart matrices

$$\mathbf{T}_a(n_a) \sim \mathcal{W}_p(n_a, \Sigma_a), \quad \mathbf{T}_b(n_b) \sim \mathcal{W}_p(n_b, \Sigma_b), \quad \mathbf{T}_c(n_c) \sim \mathcal{W}_p(n_c, \Sigma_c), \quad \mathbf{T}_d(n_d) \sim \mathcal{W}_p(n_d, \Sigma_d),$$

where $\mathbf{T} \sim \mathcal{W}_p(n, \Sigma)$ denotes that a random $p \times p$ matrix \mathbf{T} follows the p -dimensional Wishart distribution with n degrees of freedom and scale matrix Σ . Letting

$$r_p = r_{p, n_a, n_b, n_c, n_d} = p^2 \sqrt{n_a n_b n_c n_d},$$

we will study the asymptotic behavior of

$$M = \frac{1}{r_p} \text{tr}(\mathbf{T}_a(n_a) \mathbf{T}_b(n_b) \mathbf{T}_c(n_c) \mathbf{T}_d(n_d))$$

under the following high-dimensional asymptotic regime:

$$n_a, n_b, n_c, n_d \asymp p^\delta, \quad 0 < \delta < 1, \quad (1)$$

where $n \asymp p^\delta$ denotes that there exist positive constants C_1 and C_2 such that $C_1 n \leq p^\delta \leq C_2 n$ for all sufficiently large p .

To provide our limit theorem, we make the following assumption, where Σ_{ij} is one of scale matrices $\Sigma_a, \Sigma_b, \Sigma_c, \Sigma_d$ for $j \in \{1, \dots, 16\}$ and $\prod_{j=1}^k \Sigma_{ij} = \Sigma_{i_1} \cdots \Sigma_{i_k}$ for $k \in \{2, \dots, 16\}$.

Assumption 1. As $p \rightarrow \infty$ together with (1), there exists

$$\lim_{p \rightarrow \infty} \left\{ \frac{1}{p} \text{tr} \left(\prod_{j=1}^k \Sigma_{ij} \right) \right\}, \quad k \in \{2, \dots, 16\},$$

for $i_1, \dots, i_{16} \in \{a, b, c, d\}$. For $k = 2$, the limit is positive.

In particular, $\lim_{p \rightarrow \infty} \{p^{-1} \text{tr}(\Sigma_{i_1} \Sigma_{i_2})\}$ and $\lim_{p \rightarrow \infty} \{p^{-1} \text{tr}(\Sigma_{i_1} \Sigma_{i_2} \Sigma_{i_3} \Sigma_{i_4})\}$ are denoted by $\sigma_{i_1 i_2}$ and $\sigma_{i_1 i_2 i_3 i_4}$ for $i_1, \dots, i_4 \in \{a, b, c, d\}$, respectively.

Remark 1. A simple sufficient condition for [Assumption 1](#) is that the maximum eigenvalues of the matrices $\Sigma_a, \Sigma_b, \Sigma_c, \Sigma_d$ are asymptotically bounded as $p \rightarrow \infty$ together with (1).

Obviously, due to independence, it holds that $E[M] = r_p^{-1} n_a n_b n_c n_d \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d)$. Moreover, the following proposition describes the asymptotic behavior of $\text{Var}[M]$ under our asymptotic regime.

Proposition 1. Under [Assumption 1](#), it holds, as $p \rightarrow \infty$ together with (1),

$$\begin{aligned} \text{Var}[M] = & \sigma_{ab}\sigma_{ad}\sigma_{bc}\sigma_{cd} \\ & + \frac{n_a + 1}{p} \sigma_{bc}\sigma_{cd}\sigma_{abad} + \frac{n_b + 1}{p} \sigma_{ad}\sigma_{cd}\sigma_{abcb} + \frac{n_c + 1}{p} \sigma_{ab}\sigma_{ad}\sigma_{bcd} + \frac{n_d + 1}{p} \sigma_{ab}\sigma_{bc}\sigma_{adcd} + o(p^{\delta-1}). \end{aligned}$$

Proof. As it is shown in the supplementary material, it holds that

$$\begin{aligned} & \text{Var}[\text{tr}(\mathbf{T}_a \mathbf{T}_b \mathbf{T}_c \mathbf{T}_d)] \\ &= n_a n_b n_c n_d \left[(n_a n_b + n_a n_c + n_a n_d + n_b n_c + n_b n_d + n_c n_d + 1) \{\text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d)\}^2 \right. \\ & \quad + (n_a n_b n_c + n_a n_b n_d + n_a n_c n_d + n_b n_c n_d + n_a + n_b + n_c + n_d) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d \Sigma_a \Sigma_b \Sigma_c \Sigma_d) \\ & \quad + (n_a + 1)(n_b + 1)(n_c + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d \Sigma_c \Sigma_b \Sigma_a \Sigma_d) + (n_a + 1)(n_b + 1)(n_d + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_b \Sigma_a \Sigma_d \Sigma_c \Sigma_d) \\ & \quad + (n_a + 1)(n_c + 1)(n_d + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_a \Sigma_d \Sigma_c \Sigma_b \Sigma_c \Sigma_d) + (n_b + 1)(n_c + 1)(n_d + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d \Sigma_a \Sigma_d \Sigma_c \Sigma_b) \\ & \quad + (n_a + 1)(n_b + 1) \text{tr}(\Sigma_c \Sigma_d) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_b \Sigma_a \Sigma_d) + (n_a + 1)(n_c + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_a \Sigma_d) \text{tr}(\Sigma_b \Sigma_c \Sigma_d \Sigma_c) \\ & \quad + (n_a + 1)(n_d + 1) \text{tr}(\Sigma_b \Sigma_c) \text{tr}(\Sigma_a \Sigma_b \Sigma_a \Sigma_d \Sigma_c \Sigma_d) + (n_b + 1)(n_c + 1) \text{tr}(\Sigma_a \Sigma_d) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_d \Sigma_c \Sigma_b) \\ & \quad + (n_b + 1)(n_d + 1) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_b) \text{tr}(\Sigma_a \Sigma_d \Sigma_c \Sigma_d) + (n_c + 1)(n_d + 1) \text{tr}(\Sigma_a \Sigma_b) \text{tr}(\Sigma_a \Sigma_d \Sigma_c \Sigma_b \Sigma_c \Sigma_d) \\ & \quad + (n_a + 1) \text{tr}(\Sigma_b \Sigma_c) \text{tr}(\Sigma_c \Sigma_d) \text{tr}(\Sigma_a \Sigma_b \Sigma_a \Sigma_d) + (n_b + 1) \text{tr}(\Sigma_a \Sigma_d) \text{tr}(\Sigma_c \Sigma_d) \text{tr}(\Sigma_a \Sigma_b \Sigma_c \Sigma_b) \\ & \quad + (n_c + 1) \text{tr}(\Sigma_a \Sigma_b) \text{tr}(\Sigma_a \Sigma_d) \text{tr}(\Sigma_b \Sigma_c \Sigma_d \Sigma_c) + (n_d + 1) \text{tr}(\Sigma_a \Sigma_b) \text{tr}(\Sigma_b \Sigma_c) \text{tr}(\Sigma_a \Sigma_d \Sigma_c \Sigma_d) \\ & \quad \left. + \text{tr}(\Sigma_a \Sigma_b) \text{tr}(\Sigma_a \Sigma_d) \text{tr}(\Sigma_b \Sigma_c) \text{tr}(\Sigma_c \Sigma_d) \right]. \end{aligned}$$

From this equation and [Assumption 1](#), the conclusion follows. \square

To close this subsection, let us define four independent i.i.d. p -dimensional random sequences $\{\mathbf{x}_i\}_{i=1}^{n_a}, \{\mathbf{y}_i\}_{i=1}^{n_b}, \{\mathbf{z}_i\}_{i=1}^{n_c}$ and $\{\mathbf{w}_i\}_{i=1}^{n_d}$ satisfying

$$\mathbf{T}_a(n_a) = \sum_{i=1}^{n_a} \mathbf{x}_i \mathbf{x}_i^\top, \quad \mathbf{T}_b(n_b) = \sum_{i=1}^{n_b} \mathbf{y}_i \mathbf{y}_i^\top, \quad \mathbf{T}_c(n_c) = \sum_{i=1}^{n_c} \mathbf{z}_i \mathbf{z}_i^\top, \quad \mathbf{T}_d(n_d) = \sum_{i=1}^{n_d} \mathbf{w}_i \mathbf{w}_i^\top,$$

where

$$\begin{aligned} \mathbf{x}_i &\sim \mathcal{N}_p(\mathbf{0}_p, \Sigma_a), \quad i \in \{1, \dots, n_a\}, \quad \mathbf{y}_i \sim \mathcal{N}_p(\mathbf{0}_p, \Sigma_b), \quad i \in \{1, \dots, n_b\}, \\ \mathbf{z}_i &\sim \mathcal{N}_p(\mathbf{0}_p, \Sigma_c), \quad i \in \{1, \dots, n_c\}, \quad \mathbf{w}_i \sim \mathcal{N}_p(\mathbf{0}_p, \Sigma_d), \quad i \in \{1, \dots, n_d\}. \end{aligned}$$

Moreover, for later discussion, when $n_a, n_b, n_c, n_d \geq 2$, let us denote

$$\begin{aligned} \mathbf{T}_a(h) &= \sum_{i=1}^h \mathbf{x}_i \mathbf{x}_i^\top, \quad h \in \{1, \dots, n_a - 1\}, \quad \mathbf{T}_b(h) = \sum_{i=1}^h \mathbf{y}_i \mathbf{y}_i^\top, \quad h \in \{1, \dots, n_b - 1\}, \\ \mathbf{T}_c(h) &= \sum_{i=1}^h \mathbf{z}_i \mathbf{z}_i^\top, \quad h \in \{1, \dots, n_c - 1\}, \quad \mathbf{T}_d(h) = \sum_{i=1}^h \mathbf{w}_i \mathbf{w}_i^\top, \quad h \in \{1, \dots, n_d - 1\}. \end{aligned}$$

2.2. Main result

The main result of this paper is the following theorem.

Theorem 1. Under [Assumption 1](#), it holds that, as $p \rightarrow \infty$ together with (1),

$$M - E[M] \xrightarrow{d} \mathcal{N}(0, \sigma_{ab}\sigma_{ad}\sigma_{bc}\sigma_{cd}),$$

where \xrightarrow{d} denotes convergence in distribution.

Proof. Define a sequence $\{\mathbf{u}_i\}_{i=1}^{n_a+n_b+n_c+n_d}$ by

$$\mathbf{u}_i = \mathbf{x}_i, \quad i \in \{1, \dots, n_a\}, \quad \mathbf{u}_{n_a+i} = \mathbf{y}_i, \quad i \in \{1, \dots, n_b\}, \quad \mathbf{u}_{n_a+n_b+i} = \mathbf{z}_i, \quad i \in \{1, \dots, n_c\},$$

$$\mathbf{u}_{n_a+n_b+n_c+i} = \mathbf{w}_i, \quad i \in \{1, \dots, n_d\}.$$

Moreover, introduce a filtration $\{\mathcal{F}_h\}_{h=1}^{n_a+n_b+n_c+n_d}$, where $\mathcal{F}_h = \mathcal{F}_{p,h}$ is the σ -field generated by $\mathbf{u}_1, \dots, \mathbf{u}_h$ for $h \in \{1, \dots, n_a + n_b + n_c + n_d\}$. Consider a martingale difference array $\{D_h\}_{h=1}^{n_a+n_b+n_c+n_d}$ defined by

$$D_h = D_{p,h} = E_h[M] - E_{h-1}[M], \quad h \in \{1, \dots, n_a + n_b + n_c + n_d\},$$

where we use the notation $E_0[\cdot] = E[\cdot]$ and $E_h[\cdot] = E[\cdot | \mathcal{F}_h]$, $h \in \{1, \dots, n_a + n_b + n_c + n_d\}$, for simplicity. From the definition, it holds that

$$\sum_{h=1}^{n_a+n_b+n_c+n_d} D_h = M - E[M] = \frac{1}{r_p} \text{tr}(\mathbf{T}_a(n_a)\mathbf{T}_b(n_b)\mathbf{T}_c(n_c)\mathbf{T}_d(n_d)) - \frac{n_a n_b n_c n_d}{r_p} \text{tr}(\boldsymbol{\Sigma}_a \boldsymbol{\Sigma}_b \boldsymbol{\Sigma}_c \boldsymbol{\Sigma}_d).$$

Define $\{Q_h\}_{h=1}^{n_a+n_b+n_c+n_d}$ by

$$Q_h = Q_{p,h} = E_{h-1}[D_h^2], \quad h \in \{1, \dots, n_a + n_b + n_c + n_d\}.$$

Then, it holds that, for $h \in \{1, \dots, n_a + n_b + n_c + n_d\}$,

$$Q_h = E_{h-1}[(E_h[M])^2] - (E_{h-1}[M])^2 \text{ a.s.},$$

and so

$$\begin{aligned} E \left[\sum_{h=1}^{n_a+n_b+n_c+n_d} Q_h \right] &= E \left[\sum_{h=1}^{n_a+n_b+n_c+n_d} \{E_{h-1}[(E_h[M])^2] - (E_{h-1}[M])^2\} \right] = \sum_{h=1}^{n_a+n_b+n_c+n_d} \{E[(E_h[M])^2] - E[(E_{h-1}[M])^2]\} \\ &= E[(E_{n_a+n_b+n_c+n_d}[M])^2] - E[(E_0[M])^2] = E[M^2] - (E[M])^2 = \text{Var}[M]. \end{aligned} \quad (2)$$

As it is shown in the supplementary material, the following two lemmas hold:

Lemma 1. Under [Assumption 1](#), it holds that, as $p \rightarrow \infty$ together with [\(1\)](#),

$$\text{Var} \left[\sum_{h=1}^{n_a+n_b+n_c+n_d} Q_h \right] \rightarrow 0.$$

Lemma 2. Under [Assumption 1](#), it holds that, as $p \rightarrow \infty$ together with [\(1\)](#),

$$\sum_{h=1}^{n_a+n_b+n_c+n_d} E[D_h^4] \rightarrow 0.$$

From [\(2\)](#), [Proposition 1](#), and [Lemma 1](#), it follows that

$$\sum_{h=1}^{n_a+n_b+n_c+n_d} Q_h \xrightarrow{p} \sigma_{ab}\sigma_{ad}\sigma_{bc}\sigma_{cd},$$

where \xrightarrow{p} denotes convergence in probability. Moreover, from [Proposition 1](#) and [Lemma 2](#), it follows that

$$\frac{\sum_{h=1}^{n_a+n_b+n_c+n_d} E[D_h^4]}{(\text{Var}[M])^2} \rightarrow 0. \quad (3)$$

Therefore, the conclusion follows from the martingale central limit theorem (see, e.g., Corollary 3.1. of [\[11\]](#), pp. 58–59), where the Lindeberg condition follows from [\(3\)](#). This completes the proof. \square

3. Testing for common principal components model

3.1. Problem setting

In the CPC model, the covariance matrices of (more than) two populations can be simultaneously diagonalized using a common orthogonal matrix. The CPC model was first introduced in [\[8\]](#), and the fundamental asymptotic theory of statistical inference was established in [\[9\]](#). Flury [\[8,9\]](#) considered the population distributions to be normal. Later, the CPC model was discussed by Hallin et al. [\[12\]](#) for elliptical and possibly heterokurtic distributions other than the normal distribution. As introduced in [Section 1](#), tests for the CPC model have been studied by Boente et al. [\[4\]](#), Boik [\[5\]](#), and Hallin et al. [\[13,14\]](#), but no CPC tests under the high-dimensional setting have been reported in the literature. In this section, we consider the problem using [Theorem 1](#).

Denote the spectral decompositions of two population covariance matrices Σ_x and Σ_y as $\Sigma_x = \mathbf{U}_x \mathbf{\Lambda}_x \mathbf{U}_x^\top$ and $\Sigma_y = \mathbf{U}_y \mathbf{\Lambda}_y \mathbf{U}_y^\top$, respectively. The CPC model means that there exist spectral decompositions satisfying $\mathbf{U}_x = \mathbf{U}_y$. This model can equivalently be expressed as $\Sigma_x \Sigma_y = \Sigma_y \Sigma_x$.

Henceforth, let $p, m (\geq 2), n (\geq 2)$ be positive integers, and let Σ_x and Σ_y be $p \times p$ positive-definite matrices. Suppose that we have a random sample of size $M = 4m$ from $\mathcal{N}_p(\mu_x, \Sigma_x)$, and split the sample randomly into four subsamples of size m . In the same way, suppose that we have a random sample of size $N = 4n$ from $\mathcal{N}_p(\mu_y, \Sigma_y)$, and split the sample randomly into four subsamples of size n . Under this setting, we wish to test

$$\mathcal{H}_0 : \Sigma_x \Sigma_y = \Sigma_y \Sigma_x, \quad \mathcal{H}_1 : \Sigma_x \Sigma_y \neq \Sigma_y \Sigma_x.$$

We consider the asymptotic regime $p \rightarrow \infty$ together with

$$m, n \asymp p^\delta, \quad 0 < \delta < 1. \quad (4)$$

When discussing the power of the test, the regime is limited to $1/2 < \delta < 1$ in order to guarantee its consistency. The following assumption is posed on the covariance matrices.

Assumption 2. As $p \rightarrow \infty$ together with (4), there exists

$$\lim_{p \rightarrow \infty} \left\{ \frac{1}{p} \text{tr} \left(\prod_{j=1}^k \Sigma_{i_j} \right) \right\}, \quad k \in \{2, \dots, 16\},$$

for $i_1, \dots, i_{16} \in \{x, y\}$. For $k \in \{2, 3, 4, 5\}$, the limits are positive.

In particular, $\lim_{p \rightarrow \infty} \{p^{-1} \text{tr}(\Sigma_{i_1} \Sigma_{i_2})\}$ and $\lim_{p \rightarrow \infty} \{p^{-1} \text{tr}(\Sigma_{i_1} \Sigma_{i_2} \Sigma_{i_3} \Sigma_{i_4})\}$ are denoted by $\sigma_{i_1 i_2}$ and $\sigma_{i_1 i_2 i_3 i_4}$ for $i_1, \dots, i_4 \in \{x, y\}$, respectively.

3.2. Test procedure

As

$$\begin{aligned} \Sigma_x \Sigma_y = \Sigma_y \Sigma_x &\iff \text{tr} \left\{ (\Sigma_x \Sigma_y - \Sigma_y \Sigma_x) (\Sigma_x \Sigma_y - \Sigma_y \Sigma_x)^\top \right\} = 0 \\ &\iff \text{tr} \left\{ (\Sigma_x \Sigma_y - \Sigma_y \Sigma_x) (\Sigma_y \Sigma_x - \Sigma_x \Sigma_y) \right\} = 0 \\ &\iff \text{tr} (\Sigma_x \Sigma_x \Sigma_y \Sigma_y) - \text{tr} (\Sigma_x \Sigma_y \Sigma_x \Sigma_y) = 0, \end{aligned}$$

we can equivalently transform $\mathcal{H}_0 : \Sigma_x \Sigma_y = \Sigma_y \Sigma_x$ into $\mathcal{H}_0 : \theta = 0$, where

$$\theta = \theta_p = \sigma_{xxyy}(p) - \sigma_{xyxy}(p), \quad \sigma_{xxyy}(p) = \frac{1}{p} \text{tr} (\Sigma_x \Sigma_x \Sigma_y \Sigma_y), \quad \sigma_{xyxy}(p) = \frac{1}{p} \text{tr} (\Sigma_x \Sigma_y \Sigma_x \Sigma_y).$$

Moreover, \mathcal{H}_1 can be equivalently transformed into $\mathcal{H}_1 : \theta > 0$.

Let us denote the scatter matrices calculated from the centered split subsamples as

$$\mathbf{T}_{x1}, \mathbf{T}_{x2}, \mathbf{T}_{x3}, \mathbf{T}_{x4}, \mathbf{T}_{y1}, \mathbf{T}_{y2}, \mathbf{T}_{y3}, \mathbf{T}_{y4}.$$

In this case, it holds that

$$\mathbf{T}_{xk} \sim \mathcal{W}_p(m-1, \Sigma_x), \quad k \in \{1, 2, 3, 4\},$$

and

$$\mathbf{T}_{yk} \sim \mathcal{W}_p(n-1, \Sigma_y), \quad k \in \{1, 2, 3, 4\}.$$

Clearly,

$$\hat{\theta} = \frac{1}{(m-1)^2(n-1)^2 p} \left\{ \text{tr} (\mathbf{T}_{x1} \mathbf{T}_{x2} \mathbf{T}_{y1} \mathbf{T}_{y2}) - \text{tr} (\mathbf{T}_{x3} \mathbf{T}_{y3} \mathbf{T}_{x4} \mathbf{T}_{y4}) \right\}$$

is an unbiased estimator of $\theta = \theta_p = \sigma_{xxyy}(p) - \sigma_{xyxy}(p)$. As for the variance of $p^{-1}(m-1)(n-1)\hat{\theta}$, it follows from Proposition 1 that

$$\begin{aligned} \text{Var} \left[\frac{(m-1)(n-1)}{p} \hat{\theta} \right] &= \text{Var} \left[\frac{1}{(m-1)(n-1)p^2} \left\{ \text{tr} (\mathbf{T}_{x1} \mathbf{T}_{x2} \mathbf{T}_{y1} \mathbf{T}_{y2}) - \text{tr} (\mathbf{T}_{x3} \mathbf{T}_{y3} \mathbf{T}_{x4} \mathbf{T}_{y4}) \right\} \right] \\ &= \text{Var} \left[\frac{1}{(m-1)(n-1)p^2} \text{tr} (\mathbf{T}_{x1} \mathbf{T}_{x2} \mathbf{T}_{y1} \mathbf{T}_{y2}) \right] + \text{Var} \left[\frac{1}{(m-1)(n-1)p^2} \text{tr} (\mathbf{T}_{x3} \mathbf{T}_{y3} \mathbf{T}_{x4} \mathbf{T}_{y4}) \right] \\ &= \sigma_{xx} \sigma_{yy} \sigma_{xy}^2 + \sigma_{xy}^4 + \frac{2m}{p} \sigma_{xy} \sigma_{yy} \sigma_{xxxxy} + \frac{2n}{p} \sigma_{xx} \sigma_{xy} \sigma_{xyyy} + \frac{2(m+n)}{p} \sigma_{xy}^2 \sigma_{xyxy} + o(p^{\delta-1}). \end{aligned}$$

The following proposition establishes the asymptotic behavior of $\hat{\theta}$ under our asymptotic regime.

Proposition 2. Under [Assumption 2](#), it holds that, as $p \rightarrow \infty$ together with (4),

$$\frac{(m-1)(n-1)}{p} (\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \sigma_{xx}\sigma_{yy}\sigma_{xy}^2 + \sigma_{xy}^4).$$

Proof. The left-hand side is equal to

$$\left\{ \frac{1}{(m-1)(n-1)p^2} \text{tr}(\mathbf{T}_{x1}\mathbf{T}_{x2}\mathbf{T}_{y1}\mathbf{T}_{y2}) - \frac{(m-1)(n-1)}{p} \sigma_{xxyy}(p) \right\} - \left\{ \frac{1}{(m-1)(n-1)p^2} \text{tr}(\mathbf{T}_{x3}\mathbf{T}_{y3}\mathbf{T}_{x4}\mathbf{T}_{y4}) - \frac{(m-1)(n-1)}{p} \sigma_{xyxy}(p) \right\}. \quad (5)$$

From [Theorem 1](#), it follows that

$$\frac{1}{(m-1)(n-1)p^2} \text{tr}(\mathbf{T}_{x1}\mathbf{T}_{x2}\mathbf{T}_{y1}\mathbf{T}_{y2}) - \frac{(m-1)(n-1)}{p} \sigma_{xxyy}(p) \xrightarrow{d} \mathcal{N}(0, \sigma_{xx}\sigma_{yy}\sigma_{xy}^2)$$

and that

$$\frac{1}{(m-1)(n-1)p^2} \text{tr}(\mathbf{T}_{x3}\mathbf{T}_{y3}\mathbf{T}_{x4}\mathbf{T}_{y4}) - \frac{(m-1)(n-1)}{p} \sigma_{xyxy}(p) \xrightarrow{d} \mathcal{N}(0, \sigma_{xy}^4).$$

As the first and second terms of (5) are independent, the conclusion follows. \square

When $\mathcal{H}_0 : \theta = 0$ is true, [Proposition 2](#) implies that

$$\frac{(m-1)(n-1)}{p} \hat{\theta} \xrightarrow{d} \mathcal{N}(0, \sigma_{xx}\sigma_{yy}\sigma_{xy}^2 + \sigma_{xy}^4).$$

Hence, constructing a consistent estimator of $\sigma_{xx}\sigma_{yy}\sigma_{xy}^2 + \sigma_{xy}^4$ enables us to propose a test procedure. Let us denote the scatter matrices calculated from the two centered samples before splitting by \mathbf{T}_x and \mathbf{T}_y . Define

$$\hat{\sigma}_{xx} = \frac{1}{p(M-2)(M+1)} \left\{ \text{tr}(\mathbf{T}_x \mathbf{T}_x) - \frac{(\text{tr} \mathbf{T}_x)^2}{M-1} \right\}, \quad \hat{\sigma}_{yy} = \frac{1}{p(N-2)(N+1)} \left\{ \text{tr}(\mathbf{T}_y \mathbf{T}_y) - \frac{(\text{tr} \mathbf{T}_y)^2}{N-1} \right\},$$

$$\hat{\sigma}_{xy} = \frac{1}{p(M-1)(N-1)} \text{tr}(\mathbf{T}_x \mathbf{T}_y).$$

Remark 2. The estimators $\hat{\sigma}_{xx}$ and $\hat{\sigma}_{yy}$ were originally introduced by Bai and Saranadasa [3]. It is known that, under [Assumption 2](#), as $p \rightarrow \infty$ together with (4),

$$\hat{\sigma}_{xx} - \sigma_{xx}(p) = O_p(p^{-\delta}), \quad \hat{\sigma}_{yy} - \sigma_{yy}(p) = O_p(p^{-\delta}), \quad \hat{\sigma}_{xy} - \sigma_{xy}(p) = O_p(p^{-\delta}),$$

and it holds that

$$\hat{\sigma}_{xx}\hat{\sigma}_{yy}\hat{\sigma}_{xy}^2 + \hat{\sigma}_{xy}^4 - \{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4\} = O_p(p^{-\delta}). \quad (6)$$

Consider the test statistic T defined by

$$T = \frac{(m-1)(n-1)}{p} \frac{\hat{\theta}}{\sqrt{\hat{\sigma}_{xx}\hat{\sigma}_{yy}\hat{\sigma}_{xy}^2 + \hat{\sigma}_{xy}^4}}.$$

Using this, we propose the following test procedure (approximate significance level is α):

if $T > \Phi^{-1}(1 - \alpha)$, reject \mathcal{H}_0 ; where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution. This test procedure is justified using the following proposition.

Proposition 3. Under [Assumption 2](#), as $p \rightarrow \infty$ together with (4), for any $x \in \mathbb{R}$,

$$\Pr(T > x + \Delta) = 1 - \Phi(x) + o(1),$$

where

$$\Delta = \Delta_p = \frac{(m-1)(n-1)}{p} \frac{\theta}{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}}.$$

Proof. It follows from (6) that

$$\begin{aligned} T &= \frac{(m-1)(n-1)}{p} \frac{\hat{\theta}}{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}} \\ &= \hat{\theta} \frac{(m-1)(n-1)}{p^{1+\delta}} p^\delta \left\{ \frac{1}{\sqrt{\hat{\sigma}_{xx}\hat{\sigma}_{yy}\hat{\sigma}_{xy}^2 + \hat{\sigma}_{xy}^4}} - \frac{1}{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}} \right\} \xrightarrow{p} 0 \end{aligned}$$

because $|\hat{\theta}| \xrightarrow{p} |\sigma_{xxyy} - \sigma_{xyxy}| (< \infty)$ and $(m-1)(n-1)/p^{1+\delta} \asymp p^{\delta-1} \rightarrow 0$. Hence, Proposition 2 implies that

$$T - \Delta = \frac{(m-1)(n-1)}{p} \frac{\hat{\theta} - \theta}{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}} + o_p(1) \xrightarrow{d} \mathcal{N}(0, 1).$$

This completes the proof. \square

The following corollaries, which are direct consequences of Proposition 3 and its proof, show that the type-I error rate of the test is asymptotically α and that the test is consistent when $1/2 < \delta < 1$, which guarantees $\Delta \rightarrow \infty$.

Corollary 1. Under Assumption 2, when \mathcal{H}_0 is true, as $p \rightarrow \infty$ together with (4), $\Pr(T > \Phi^{-1}(1 - \alpha)) \rightarrow \alpha$.

Corollary 2. Under Assumption 2, when \mathcal{H}_1 is true, if $\sigma_{xxyy} - \sigma_{xyxy} > 0$, then as $p \rightarrow \infty$ together with $m, n \asymp p^\delta$, $1/2 < \delta < 1$, for any positive constant C , $\Pr(T > C) \rightarrow 1$.

From Proposition 3, by substituting $x = \Phi^{-1}(1 - \alpha) - \Delta$ and neglecting $o(1)$ terms, for large p , we obtain

$$\Pr(T > \Phi^{-1}(1 - \alpha)) \approx 1 - \Phi(\Phi^{-1}(1 - \alpha) - \Delta),$$

which gives the asymptotic power when \mathcal{H}_1 is true. Here the symbol \approx is used to denote that $o(1)$ terms are neglected. Using this formula, let us discuss an approximate method for sample size determination. To ensure the asymptotic power $1 - \beta$ ($0 < \beta < 1$), it is sufficient to determine M and N such that

$$\begin{aligned} 1 - \Phi\left(\Phi^{-1}(1 - \alpha) - \frac{(m-1)(n-1)}{p} \frac{\theta}{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}}\right) &> 1 - \beta \\ \iff (m-1)(n-1) &> p \frac{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}}{\theta} (\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\beta)) \\ \iff (M-4)(N-4) &> 16p \frac{\sqrt{\sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4}}{\sigma_{xxyy}(p) - \sigma_{xyxy}(p)} (\Phi^{-1}(1 - \alpha) - \Phi^{-1}(\beta)). \end{aligned}$$

Although the asymptotic variance of $p^{-1}(m-1)(n-1)(\hat{\theta} - \theta)$ is $\sigma_{xx}\sigma_{yy}\sigma_{xy}^2 + \sigma_{xy}^4$, the terms of order $p^{\delta-1}$ in the variance may not be negligible in realistic settings. To deal with this issue, we also consider the test statistic

$$\tilde{T} = \frac{(m-1)(n-1)}{p} \frac{\hat{\theta}}{\sqrt{\hat{\sigma}_{xx}\hat{\sigma}_{yy}\hat{\sigma}_{xy}^2 + \hat{\sigma}_{xy}^4 + \frac{2m}{p}\hat{\sigma}_{xy}\hat{\sigma}_{yy}\hat{\sigma}_{xxyy} + \frac{2n}{p}\hat{\sigma}_{xx}\hat{\sigma}_{xy}\hat{\sigma}_{xyyy} + \frac{2(m+n)}{p}\hat{\sigma}_{xy}^2\hat{\sigma}_{xyxy}}},$$

where $\hat{\sigma}_{xxyy}$, $\hat{\sigma}_{xyyy}$, and $\hat{\sigma}_{xyxy}$ are unbiased estimators of $\sigma_{xxyy}(p)$, $\sigma_{xyyy}(p)$, and $\sigma_{xyxy}(p)$, respectively, calculated from split subsamples. From Proposition 1, $\text{Var}[\hat{\sigma}_{xxyy}]$, $\text{Var}[\hat{\sigma}_{xyyy}]$, $\text{Var}[\hat{\sigma}_{xyxy}]$ are $O(p^{2-4\delta})$, and hence

$$\hat{\sigma}_{xxyy} - \sigma_{xxyy}(p) = O_p(p^{1-2\delta}), \quad \hat{\sigma}_{xyyy} - \sigma_{xyyy}(p) = O_p(p^{1-2\delta}), \quad \hat{\sigma}_{xyxy} - \sigma_{xyxy}(p) = O_p(p^{1-2\delta}).$$

This implies the asymptotic equivalence of T and \tilde{T} under \mathcal{H}_0 . In the simulation study presented in the next subsection, the asymptotic power of the test using \tilde{T} will be calculated by $1 - \Phi(\Phi^{-1}(1 - \alpha) - \Delta)$, where

$$\tilde{\Delta} = \frac{(m-1)(n-1)}{p} \frac{\theta}{\sqrt{\Gamma}}$$

and

$$\begin{aligned} \Gamma &= \sigma_{xx}(p)\sigma_{yy}(p)(\sigma_{xy}(p))^2 + (\sigma_{xy}(p))^4 \\ &\quad + \frac{2m}{p}\sigma_{xy}(p)\sigma_{yy}(p)\sigma_{xxyy}(p) + \frac{2n}{p}\sigma_{xx}(p)\sigma_{xy}(p)\sigma_{xyyy}(p) + \frac{2(m+n)}{p}(\sigma_{xy}(p))^2\sigma_{xyxy}(p). \end{aligned}$$

Remark 3. There is another natural unbiased estimator of θ other than $\hat{\theta}$. For instance,

$$\frac{1}{p(M-2)(M+1)(N-2)(N+1)} \left\{ \text{tr}(\mathbf{T}_x \mathbf{T}_x \mathbf{T}_y \mathbf{T}_y) - \frac{1}{N-1} \text{tr}(\mathbf{T}_x \mathbf{T}_x \mathbf{T}_y) \text{tr} \mathbf{T}_y - \frac{1}{M-1} \text{tr}(\mathbf{T}_x \mathbf{T}_y \mathbf{T}_y) \text{tr} \mathbf{T}_x \right. \\ \left. + \frac{1}{(M-1)(N-1)} \text{tr}(\mathbf{T}_x \mathbf{T}_y) \text{tr} \mathbf{T}_x \text{tr} \mathbf{T}_y - \frac{MN-M-N+3}{(M-1)(N-1)} \text{tr}(\mathbf{T}_x \mathbf{T}_y \mathbf{T}_x \mathbf{T}_y) + \frac{M+N-1}{(M-1)(N-1)} \text{tr}(\mathbf{T}_x \mathbf{T}_y) \text{tr}(\mathbf{T}_x \mathbf{T}_y) \right\}$$

is an unbiased estimator of θ . Deriving the asymptotic behavior of this quantity is a possible future direction of study.

Remark 4. As the Wishart distribution is considered in Proposition 1 and Theorem 1, the population distributions are assumed to be normal in the problem setting of this section. Therefore, it is expected that the proposed test procedures will not work well for population distributions other than the multivariate normal distribution. Indeed, as stated in the concluding remarks, the proposed tests show poor performances when simulations are conducted with the multivariate t-distribution.

3.3. Simulation study

In this subsection, we present the results of a simulation study. The purpose is to evaluate the non-asymptotic performance of the proposed procedure. Let us describe the setting. We draw two random samples of size $M = 4m$ and $N = 4n$ from $\mathcal{N}_p(\mathbf{0}, \Sigma_x)$ and $\mathcal{N}_p(\mathbf{0}, \Sigma_y)$, respectively. The two samples are randomly split into four equal-size subsamples. Let Σ_x be a diagonal matrix with diagonal elements $2 - 2i/(p+1)$, $i \in \{1, \dots, p\}$. For Σ_y , we consider the following three cases:

- (i) $\Sigma_y = \Sigma_x$,
- (ii) $\Sigma_y = \mathbf{U} \mathbf{A} \mathbf{U}'$, where the column vectors of \mathbf{U} are the eigenvectors of the scatter matrix generated from a random sample of size $2p$ from $\mathcal{N}_p(\mathbf{0}_p, \mathbf{I}_p)$ and \mathbf{A} is a diagonal matrix whose diagonal elements independently follow the chi-squared distribution with one degree of freedom, where \mathbf{U} and \mathbf{A} are fixed for each value of p while simulations,
- (iii) $\Sigma_y = \mathbf{U} \mathbf{A} \mathbf{U}'$, where the column vectors of \mathbf{U} are generated in the same way as in Case (ii) and \mathbf{A} is a diagonal matrix whose diagonal elements independently follow the log-normal distribution with parameter $(0, 1)$, where \mathbf{U} and \mathbf{A} are fixed for each value of p while simulations.

We examine the sizes of the tests using the test statistics T and \tilde{T} in Case (i) and the powers of these tests in Cases (ii) and (iii). We set $\alpha = 0.05$. Note that $\hat{\sigma}_{xyxy}$, $\hat{\sigma}_{xxxy}$, and $\hat{\sigma}_{xyyy}$ must be calculated to compute T . In the simulations, $\hat{\sigma}_{xyxy}$ is obtained from two subsamples of X with sample sizes $M/2$ and two subsamples of Y with sample sizes $N/2$, $\hat{\sigma}_{xxxy}$ is obtained from three subsamples of X with sample sizes $\lfloor M/3 \rfloor$ and whole sample of Y (with sample size N), and $\hat{\sigma}_{xyyy}$ is obtained from whole sample of X (with sample size M) and three subsamples of Y with sample sizes $\lfloor N/3 \rfloor$. Tables 1–3 give the results for $p \in \{50, 100, 200, 500, 1000, 2000, 3000\}$ and $m = n = \lfloor p^{0.8}/4 \rfloor$ in Cases (i)–(iii), respectively. Tables 4–6 give the results for $p \in \{50, 100, 200, 500, 1000, 2000, 3000\}$ and $m = n = \lfloor p^{0.8}/2 \rfloor$ in Cases (i)–(iii), respectively. Note that each empirical size or power result is based on 1000 numerical simulations. For reference, Tables 1–6 also present the asymptotic sizes and powers derived in the previous subsection together with the values of Δ and $\tilde{\Delta}$. Tables 1 and 4 show that the size of the test using T is much larger than $\alpha = 0.05$, while the size of the test using \tilde{T} is close to $\alpha = 0.05$, especially for $m = n = \lfloor p^{0.8}/4 \rfloor$. Tables 2, 3, 5, and 6 indicate that the powers of both T and \tilde{T} become larger as p (and m, n) increases. In particular, for $m = n = \lfloor p^{0.8}/2 \rfloor$, the power of the test using \tilde{T} is close to 1, while its size is close to $\alpha = 0.05$ compared with T when p is large. Comparing the simulated size or power with the asymptotic size or power for each case, they are close to some extent, but there can be non-negligible differences between them. This might be because the test statistics T and \tilde{T} do not necessarily follow the normal distribution for finite samples (although their asymptotic normality was proved in the previous subsection) and because the approximation accuracy for the variance of $p^{-1}(m-1)(n-1)(\hat{\theta} - \theta)$ is insufficient. Further approximate correction procedures for finite samples might be needed to make the distribution of the test statistic closer to a normal distribution and its size closer to the significance level α ; these are topics for future research.

4. Concluding remarks

In this paper, we have proved the asymptotic normality of the trace of the product of four independent Wishart matrices under the high-dimensional asymptotic regime $p \rightarrow \infty$ with (4) supposing Assumption 1 and proposed a test of the CPC hypothesis when the number of observed variables is larger than sample sizes. We conclude this paper with the following discussion on the assumptions in our limit theorem.

- In Section 3, we assumed that random samples were drawn from multivariate normal distributions. To study how the violation of this assumption affects the sizes of the test statistics T and \tilde{T} , we conducted numerical simulations for the multivariate t-distributions with three degrees of freedom and 10 degrees of freedom under Case (i) (in Section 3.3). We saw from the results that the sizes of both T and \tilde{T} are much greater than $\alpha = 0.05$, especially for the cases of the multivariate t-distribution with three degrees of freedom and $m = n = \lfloor p^{0.8}/2 \rfloor$. Hence, some modification of the test statistics will be needed for the proposed test to apply to non-normal distributions.

Table 1

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} in Case (i) with $m = n = \lfloor p^{0.8}/4 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	5	0.093	0.05	0	0.056	0.05	0
100	9	0.111	0.05	0	0.067	0.05	0
200	17	0.131	0.05	0	0.073	0.05	0
500	36	0.111	0.05	0	0.069	0.05	0
1000	62	0.115	0.05	0	0.065	0.05	0
2000	109	0.112	0.05	0	0.057	0.05	0
3000	151	0.074	0.05	0	0.043	0.05	0

Table 2

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} in Case (ii) with $m = n = \lfloor p^{0.8}/4 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	5	0.164	0.067	0.144	0.088	0.059	0.082
100	9	0.144	0.068	0.152	0.085	0.062	0.110
200	17	0.209	0.094	0.328	0.120	0.079	0.235
500	36	0.295	0.198	0.796	0.177	0.138	0.555
1000	62	0.395	0.303	1.130	0.238	0.209	0.836
2000	109	0.502	0.519	1.693	0.380	0.368	1.307
3000	151	0.693	0.746	2.306	0.543	0.552	1.776

Table 3

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} in Case (iii) with $m = n = \lfloor p^{0.8}/4 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	5	0.094	0.054	0.039	0.045	0.053	0.029
100	9	0.138	0.066	0.139	0.058	0.061	0.099
200	17	0.201	0.089	0.297	0.116	0.075	0.206
500	36	0.226	0.133	0.532	0.139	0.106	0.399
1000	62	0.422	0.336	1.222	0.250	0.192	0.776
2000	109	0.441	0.422	1.448	0.300	0.292	1.098
3000	151	0.532	0.549	1.767	0.408	0.395	1.379

Table 4

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} for Case (i) with $m = n = \lfloor p^{0.8}/2 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	11	0.196	0.05	0	0.102	0.05	0
100	19	0.210	0.05	0	0.105	0.05	0
200	34	0.192	0.05	0	0.097	0.05	0
500	72	0.157	0.05	0	0.077	0.05	0
1000	125	0.165	0.05	0	0.085	0.05	0
2000	218	0.144	0.05	0	0.080	0.05	0
3000	302	0.123	0.05	0	0.070	0.05	0

- Aoshima and Yata [2] classified the asymptotic setting of eigenvalues into a strongly spiked eigenvalue (SSE) model and a non-strongly spiked eigenvalue (NSSE) model, as defined in (1.6) and (1.4) of [2], respectively. Our assumptions on eigenvalues, such as [Assumption 2](#), fall into the latter model. There seems to be room to relax the assumptions used to prove the asymptotic normality of the proposed test statistic into weaker assumptions in the regime of the NSSE model. Moreover, as with the test statistics used for other tests concerning covariance matrices, our test statistic may have another asymptotic behavior under the SSE model.

Table 5

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} for Case (ii) with $m = n = \lfloor p^{0.8}/2 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	11	0.312	0.123	0.484	0.186	0.082	0.254
100	19	0.360	0.199	0.799	0.204	0.120	0.469
200	34	0.471	0.449	1.516	0.300	0.217	0.862
500	72	0.683	0.823	2.570	0.508	0.480	1.594
1000	125	0.901	0.998	4.507	0.783	0.878	2.808
2000	218	0.994	1.000	6.974	0.973	0.998	4.472
3000	302	0.999	1.000	8.492	0.999	1.000	5.691

Table 6

Empirical rejection rates (by numerical simulations) and asymptotic rejection rates of \mathcal{H}_0 for the test statistics T and \tilde{T} for Case (iii) with $m = n = \lfloor p^{0.8}/2 \rfloor$.

p	m, n	T		Δ	\tilde{T}		$\tilde{\Delta}$
		Empirical	Asymptotic		Empirical	Asymptotic	
50	11	0.297	0.103	0.378	0.166	0.075	0.206
100	19	0.353	0.213	0.847	0.217	0.119	0.465
200	34	0.439	0.367	1.305	0.277	0.178	0.720
500	72	0.632	0.785	2.434	0.449	0.415	1.430
1000	125	0.796	0.978	3.664	0.657	0.727	2.248
2000	218	0.979	1.000	6.631	0.943	0.990	3.974
3000	302	0.999	1.000	8.020	0.992	1.000	5.044

CRedit authorship contribution statement

Koji Tsukuda: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Project administration, Funding acquisition. **Shun Matsuura:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – review & editing, Funding acquisition.

Acknowledgments

The authors would like to express their gratitude to the associate editor and the referees who provided valuable comments. The authors also would like to thank the editor who provided many suggestions for improving the presentation and shortening the paper. This study was supported in part by Japan Society for the Promotion of Science KAKENHI Grant Numbers 18K13454, 20K11713, and 21K13836. This study was partly carried out when the first author was a member of Graduate School of Arts and Sciences, the University of Tokyo.

Appendix A. Supplementary data

Some technical details are given in the supplementary material.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jmva.2021.104822>.

References

- [1] M. Aoshima, K. Yata, Two-stage procedures for high-dimensional data, *Sequential Anal.* 30 (2011) 356–399.
- [2] M. Aoshima, K. Yata, Two-sample tests for high-dimension, strongly spiked eigenvalue models, *Statist. Sinica* 28 (2018) 43–62.
- [3] Z. Bai, H. Saranadasa, Effect of high dimension: By an example of a two sample problem, *Statist. Sinica* 6 (1996) 311–329.
- [4] G. Boente, A.M. Pires, I.M. Rodrigues, Robust tests for the common principal components model, *J. Statist. Plann. Inference* 139 (2009) 1332–1347.
- [5] R.J. Boik, Spectral models for covariance matrices, *Biometrika* 89 (2002) 159–182.
- [6] S.X. Chen, L.-X. Zhang, P.-S. Zhong, Tests for high-dimensional covariance matrices, *J. Amer. Statist. Assoc.* 105 (2010) 810–819.
- [7] M.L. Collyer, D.J. Sekora, D.C. Adams, A method for analysis of phenotypic change for phenotypes described by high-dimensional data, *Heredity* 115 (2015) 357–365.
- [8] B.N. Flury, Common principal components in k groups, *J. Amer. Statist. Assoc.* 79 (1984) 892–898.
- [9] B.N. Flury, Asymptotic theory for common principal component analysis, *Ann. Statist.* 14 (1986) 418–430.
- [10] B. Flury, *Common Principal Components and Related Multivariate Models*, John Wiley & Sons, Inc., New York, 1988.
- [11] P. Hall, C.C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, Inc., New York-London, 1980.
- [12] M. Hallin, D. Paindaveine, T. Verdebout, Pseudo-Gaussian inference in heterokurtic elliptical common principal components models, *Ann. I.S.U.P.* 52 (2008) 9–24.
- [13] M. Hallin, D. Paindaveine, T. Verdebout, Testing for common principal components under heterokurticity, *J. Nonparametr. Stat.* 22 (2010) 879–895.
- [14] M. Hallin, D. Paindaveine, T. Verdebout, Optimal rank-based tests for common principal components, *Bernoulli* 19 (2013) 2524–2556.
- [15] D. Houle, J. Mezey, P. Galpern, Interpretation of the results of common principal components analysis, *Evolution* 56 (2002) 433–440.

- [16] M. Hyodo, T. Nishiyama, T. Pavlenko, Testing for independence of high-dimensional variables: ρV -coefficient based approach, *J. Multivariate Anal.* 178 (2020) 104627.
- [17] M. Hyodo, N. Shutoh, T. Nishiyama, T. Pavlenko, Testing block-diagonal covariance structure for high-dimensional data, *Stat. Neerl.* 69 (2015) 460–482.
- [18] A. Ishii, K. Yata, M. Aoshima, Equality tests of high-dimensional covariance matrices under the strongly spiked eigenvalue model, *J. Statist. Plann. Inference* 202 (2019) 99–111.
- [19] A. Ishii, K. Yata, M. Aoshima, Hypothesis tests for high-dimensional covariance structures, *Ann. Inst. Statist. Math.* 73 (2021) 599–622.
- [20] J. Li, S.X. Chen, Two sample tests for high-dimensional covariance matrices, *Ann. Statist.* 40 (2012) 908–940.
- [21] B. Liu, L. Xu, S. Zheng, G.-L. Tian, A new test for the proportionality of two large-dimensional covariance matrices, *J. Multivariate Anal.* 131 (2014) 293–308.
- [22] P.C. Phillips, S.J. Arnold, Hierarchical comparison of genetic variance–covariance matrices. I. Using the Flury hierarchy, *Evolution* 53 (1999) 1506–1515.
- [23] J.R. Schott, A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, *Comput. Statist. Data Anal.* 51 (2007) 6535–6542.
- [24] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, *J. Japan Statist. Soc.* 35 (2005) 251–272.
- [25] M.S. Srivastava, H. Yanagihara, Testing the equality of several covariance matrices with fewer observations than the dimension, *J. Multivariate Anal.* 101 (2010) 1319–1329.
- [26] M.S. Srivastava, H. Yanagihara, T. Kubokawa, Tests for covariance matrices in high dimension with less sample size, *J. Multivariate Anal.* 130 (2014) 289–309.
- [27] T. Sugiyama, M. Hyodo, H. Watanabe, S. Tsukada, T. Seo, Test for equality of generalized variance in high-dimensional and large sample settings, *SUT J. Math.* 55 (2019) 139–154.
- [28] K. Tsukuda, S. Matsuura, High-dimensional testing for proportional covariance matrices, *J. Multivariate Anal.* 171 (2019) 412–420.
- [29] L. Xu, B. Liu, S. Zheng, S. Bao, Testing proportionality of two large-dimensional covariance matrices, *Comput. Statist. Data Anal.* 78 (2014) 43–55.
- [30] Y. Yamada, M. Hyodo, T. Nishiyama, Testing block-diagonal covariance structure for high-dimensional data under non-normality, *J. Multivariate Anal.* 155 (2017) 305–316.
- [31] K. Yata, M. Aoshima, Y. Nakayama, A test of sphericity for high-dimensional data and its application for detection of divergently spiked noise, *Sequential Anal.* 37 (2018) 397–411.