TEST
DOI 10.1007/s11749-017-0567-x

ORIGINAL PAPER

A simultaneous testing of the mean vector and the covariance matrix among two populations for high-dimensional data

Masashi Hyodo[1] · **Takahiro Nishiyama**[2]

Received: 6 March 2017 / Accepted: 2 October 2017

Abstract In this article, we propose an L^2-norm-based test for simultaneous testing of the mean vector and the covariance matrix under high-dimensional non-normal populations. To construct this, we derive an asymptotic distribution of a test statistic based on both differences mean vectors and covariance matrices. We also investigate the asymptotic sizes and powers of the proposed test using this result. Finally, we study the finite sample and dimension performance of this test via Monte Carlo simulations.

Keywords Simultaneous test · High-dimensional data analysis · Asymptotic distribution · Multivariate analysis

Mathematics Subject Classification 62H15 · 62F03 · 62F05

1 Introduction

Let $\boldsymbol{X}_{g1}, \boldsymbol{X}_{g2}, \ldots, \boldsymbol{X}_{gn_g}$ be p-dimensional random vectors from the gth population ($g \in \{1, 2\}$). We denote gth population mean vector with $\boldsymbol{\mu}_g$, and gth population covariance matrix with Σ_g. Assume that the vector $\boldsymbol{X}_{gi}$ have the following model:

$$\boldsymbol{X}_{gi} = \Sigma_g^{1/2} \boldsymbol{Z}_{gi} + \boldsymbol{\mu}_g \quad \text{for} \quad i \in \{1, \ldots, n_g\}, \tag{1.1}$$

Electronic supplementary material The online version of this article (doi:10.1007/s11749-017-0567-x) contains supplementary material, which is available to authorized users.

✉ Masashi Hyodo
hyodoh_h@yahoo.co.jp

1 Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai 599-8531, Osaka, Japan

2 Department of Business Administration, Senshu University, 2-1-1, Higashimita, Tama-ku, Kawasaki-shi 214-8580, Kanagawa, Japan

Published online: 22 October 2017

where $\boldsymbol{Z}_{gi}$ represents p-dimensional random vectors such that $E(\boldsymbol{Z}_{gi}) = \boldsymbol{0}$, $\mathrm{var}(\boldsymbol{Z}_{gi}) = I_p$, a $p \times p$ identity matrix, and $\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1n_1}, \boldsymbol{Z}_{21}, \ldots, \boldsymbol{Z}_{2n_2}$ are mutually independent.

In practical problems, finding heterogeneity of distribution is an important issue. For example, in DNA microarray data analysis, the test for the equality of mean vectors and test for the equality of covariance matrices contribute to discover significance difference in the distribution of expression levels. However, if only one of these tests is tested, significant differences may not be detected in some cases. Thus, it is more reasonable to detect a significant difference by testing both of them simultaneously. Also, if we assume normal populations, our test is equivalent to testing for equality of distribution functions. Recently, one-sample simultaneous test procedure for high-dimensional mean vectors and covariance matrices is proposed by Liu et al. (2017). They referred to the extension of two-sample problem as future tasks. From the above discussion, as part of the effort to discover significant differences between two high-dimensional distributions, we develop in this paper two-sample simultaneous test procedure. Our primary interest is to test

$$\mathcal{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \quad \Sigma_1 = \Sigma_2 \text{ vs. } \mathcal{H}_1 : \text{not } \mathcal{H}_0. \tag{1.2}$$

Testing the above hypothesis when $p > \min\{n_1 - 1, n_2 - 1\}$ is a non-trivial statistical problem. When $p \leq \min\{n_1 - 1, n_2 - 1\}$, the likelihood ratio test [see Muirhead (1982)] may be used for the above hypothesis. If we let

$$\overline{\boldsymbol{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \boldsymbol{X}_{gi}, \quad W_g = \sum_{i=1}^{n_g}(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)^\top,$$

then the likelihood ratio for $\mathcal{H}_0$ is

$$\lambda = \frac{\prod_{g=1}^2 |W_g|^{n_g/2} n^{pn/2}}{|(n_1 n_2)/n(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)^\top + \sum_{g=1}^2 W_g|^{n/2} \prod_{g=1}^2 n_g^{pn_g/2}},$$

where $n = n_1 + n_2$. When $p > \min\{n_1 - 1, n_2 - 1\}$, at least one of the matrices W_g is singular. This causes the likelihood ratio statistic $-2\log(\lambda)$ to be either infinite or undefined. Therefore, we need to consider other methods.

Note that test (1.2) is equivalent to the following test

$$\mathcal{H}_0 : \|\boldsymbol{\delta}\|^2 = 0, \quad \|\Delta\|_F^2 = 0 \text{ vs. } \mathcal{H}_1 : \text{not } \mathcal{H}_0,$$

where $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ and $\Delta = \Sigma_1 - \Sigma_2$. Here, $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_F$ denotes the Frobenius norm. The unbiased estimators of $\|\boldsymbol{\delta}\|^2$ and $\|\Delta\|_F^2$ are given as follows:

$$\widehat{\|\boldsymbol{\delta}\|^2} = \|\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2\|^2 - \frac{\mathrm{tr}(S_1)}{n_1} - \frac{\mathrm{tr}(S_2)}{n_2}, \quad \widehat{\|\Delta\|_F^2} = \sum_{g=1}^2 \widehat{\|\Sigma_g\|_F^2} - 2\mathrm{tr}(S_1 S_2),$$

where

$$\widehat{\|\Sigma_g\|_F^2} = \frac{n_g - 1}{n_g(n_g - 2)(n_g - 3)} \left[(n_g - 1)(n_g - 2)\mathrm{tr}(S_g^2) + \{\mathrm{tr}(S_g)\}^2 - n_g K_g \right].$$

Here,

$$S_g = \frac{1}{n_g - 1} W_g, \;\; K_g = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} \left\| \boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g \right\|^4 .$$

The unbiased estimator $\widehat{\|\Sigma_g\|_F^2}$ is proposed in Himeno and Yamada (2014) and Srivastava et al. (2014). These estimators are useful considering the following three points:

(i) These estimators are unbiased without the normality assumption.
(ii) It can be defined even when $p > \min\{n_1 - 1, n_2 - 1\}$.
(iii) Under appropriate assumptions, these estimators have asymptotic normality when p, n_1, and n_2 are large.

In high-dimensional settings, the unbiased estimators $\widehat{\|\boldsymbol{\delta}\|^2}$ and $\widehat{\|\Delta\|_F^2}$ are used in two-sample tests for mean vectors in Chen and Qin (2010) and for covariance matrices in Li and Chen (2012), respectively.

In this paper, we consider the sum of the two test statistics from the aforementioned two papers and derived the covariance between them to establish the limiting distribution under both null hypothesis and local alternatives. Using limiting distribution of the sum of the two test statistics, we propose the approximate simultaneous test for (1.2), and the asymptotic theoretical power function is also explicitly achieved.

The rest of the paper is organized as follows. Section 2 presents the test procedure and its asymptotic size and power after establishing the asymptotic normality of the test statistic. In Sect. 3, the attained significance levels and powers of the suggested test are empirically analyzed. We also provide an example to apply our test. Finally, Sect. 4 concludes this paper. Some technical details are relegated to Appendix.

2 Simultaneous testing of the mean vector and the covariance matrix in high-dimensional data

2.1 Test statistic

Before formulating the test statistic for hypothesis $\mathcal{H}_0$, we make several assumptions to investigate the asymptotic properties of the unbiased estimators $\widehat{\|\boldsymbol{\delta}\|^2}$ and $\widehat{\|\Delta\|_F^2}$.

We assume that the jth element of $\boldsymbol{Z}_{gi}$ (denotes Z_{gij}) has a uniformly bounded eighth moment, and for any positive integers r and α_ℓ such that

$$\sum_{\ell=1}^{r} \alpha_\ell \leq 8, \quad E\left(\prod_{\ell=1}^{r} Z_{gij_\ell}^{\alpha_\ell} \right) = \prod_{\ell=1}^{r} E\left(Z_{gij_\ell}^{\alpha_\ell} \right) \tag{2.1}$$

whenever $j_1, j_2, \ldots, j_r$ are distinct indices. Condition (2.1) means that each $\{Z_{gij}\}_{j=1}^{p}$ has a kind of pseudo-independence among its components. Obviously, if Z_{gij}'s have independent components, then (2.1) is trivially true. We also make the following assumptions:

(A1) Let n_1 and n_2 be functions of p. Then, $\min\{n_1, n_2\} \to \infty$ as $p \to \infty$.
(A2) Let $\mathrm{tr}(\Sigma_g \Sigma_h)$ and $\mathrm{tr}\{(\Sigma_g \Sigma_h)^2\}$ be functions of p. Then, for $g, h \in \{1, 2\}$,

$$\lim_{p\to\infty} \frac{\mathrm{tr}\left\{(\Sigma_g \Sigma_h)^2\right\}}{\left\{\mathrm{tr}(\Sigma_g \Sigma_h)\right\}^2} = 0.$$

(A3) Let n_g, $\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2$, $\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2$, and $\|\Sigma_g\|_F^2$ be functions of p. Then, for $g \in \{1, 2\}$,

$$\limsup_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{\|\Sigma_g\|_F^2} < \infty \quad \text{and} \quad \limsup_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{\|\Sigma_g\|_F^4} < \infty.$$

(A4) Let n_g, $\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2$, $\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2$, and $\|\Sigma_g\|_F^2$ be functions of p. Then, for $g \in \{1, 2\}$,

$$\lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{\|\Sigma_g\|_F^2} = \infty \quad \text{or} \quad \lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{\|\Sigma_g\|_F^4} = \infty.$$

We state some remarks on conditions (A2), (A3), and (A4).

(a) The covariance matrix $\Sigma_g = (\sigma_{g\cdot i}\sigma_{g\cdot j}\rho_g^{|i-j|})$ where $\sigma_{g\cdot\ell}^2 = \mathrm{var}(X_{gi\ell})$ are the marginal variance for $\ell \in \{1, \ldots, p\}$. If $\{\sigma_{g,\ell}^2\}$ are uniformly bounded away from infinity and zero, respectively, then $\|\Sigma_g\|_F^2 = O(p)$, and (A2) is satisfied.
(b) For simplicity, we define $\boldsymbol{\delta} = \delta_1 \mathbf{1}_p$ and $\Delta = \delta_2 \mathbf{1}\mathbf{1}^\top$, where $\delta_1, \delta_2 \in (0, \infty)$. For example, if we investigate non-trivial power of Hotelling's test under large-sample case (fixed p case), the local alternative hypothesis has the form of $\delta_1 = O(n^{-1/2})$. In other words, Hotelling's test has non-trivial power under this local alternative hypothesis and has non-power beyond the level of significance under $\delta_2 = o(n^{-1/2})$. On the other hand, the sufficient condition of (A3) is as follows:

$$\delta_1 = O\left(\frac{\|\Sigma_g\|_F^{1/2}}{\sqrt{pn_g}}\right), \quad \delta_2 = O\left(\frac{\|\Sigma_g\|_F}{p\sqrt{n_g}}\right).$$

If $\|\Sigma_g\|_F^2 = O(p)$, then $\delta_1 = O(n_g^{-1/2}p^{-1/4})$ and $\delta_2 = O(n_g^{-1/2}p^{-1/2})$. When $\delta_1 = o(n_g^{-1/2}p^{-1/4})$, $\delta_2 = o(n_g^{-1/2}p^{-1/2})$, our test has non-power beyond the level of significance. Since the high-dimensional data contain more information, it is allowed less differences among each components of two means and of two covariance matrices than fixed p case. Assumption (A3) is obviously satisfied under $\mathcal{H}_0$, and it can be viewed as a high-dimensional version of the local alternative hypotheses.

(c) To understand the performance of our test when (A3) is not valid, we reverse the local alternative condition (A3) to (A4). We consider the same $\boldsymbol{\delta}$ and Δ in (b). If $\|\Sigma_g\|_F^2 = O(p)$ and $\liminf_{p\to\infty} \lambda_{\min}(\Sigma_g) > 0$ hold, then the sufficient condition of (A4) is as $\delta_1 = O(n_g^{\eta_1 - 1/2})$, $\delta_2 = O(n_g^{\eta_2 - 1/2})$ for $\eta_1, \eta_2 \in (0, \infty)$. Here, $\lambda_{\min}(\Sigma_g)$ denotes the smallest eigenvalue of Σ_g. We also note that the asymptotic power of our test has 1 under fixed alternative (A4).

(d) Assumptions (A1) and (A2) are used to derive the asymptotic null distribution. Assumptions (A1), (A2), and (A3) are used to derive the asymptotic distribution under local alternative. Assumptions (A1) and (A4) are a sufficient condition that our proposed test has asymptotic power 1. We also note that (A3) and (A4) are not simultaneously used.

Under assumption (A1), the leading variances of $\widehat{\|\boldsymbol{\delta}\|^2}$ and $\widehat{\|\Delta\|_F^2}$ and the leading covariances of $\widehat{\|\boldsymbol{\delta}\|^2}$ and $\widehat{\|\Delta\|_F^2}$ are

$$
\begin{aligned}
\sigma_1^2 &= \sigma_{10}^2 + \sum_{g=1}^{2} \frac{4\left\|\Sigma_g^{1/2}\boldsymbol{\delta}\right\|^2}{n_g},\\
\sigma_2^2 &= \sigma_{20}^2 + \sum_{g=1}^{2} \frac{8\left\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\right\|_F^2}{n_g} + \sum_{g=1}^{2} \frac{4\kappa_{4g}\mathrm{tr}\left\{\odot^2(\Sigma_g^{1/2}\Delta\Sigma_g^{1/2})\right\}}{n_g},\\
\sigma_{12} &= \sum_{g=1}^{2} \frac{4\kappa_{3g}\sum_{i=1}^{p}\boldsymbol{\delta}^\top\Sigma_g^{1/2}\boldsymbol{e}_i\boldsymbol{e}_i^\top\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\boldsymbol{e}_i}{n_g},
\end{aligned}
$$

where $\kappa_{3g} = E(Z_{gij}^3)$, $\kappa_{4g} = E(Z_{gij}^4) - 3$,

$$
\sigma_{10}^2 = \sum_{g=1}^{2} \frac{2\|\Sigma_g\|_F^2}{n_g^2} + \frac{4\mathrm{tr}(\Sigma_1\Sigma_2)}{n_1 n_2},\quad \sigma_{20}^2 = \sum_{g=1}^{2} \frac{4\|\Sigma_g\|_F^4}{n_g^2} + \frac{8\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2}{n_1 n_2}.
$$

Here, the notation $\odot^i A$ stands for the Hadamard product of i matrices A, i.e., $\odot^i A = A \odot \cdots \odot A$ (i times). Note that $\kappa_{4g} \geq -2$ since $0 \leq \mathrm{var}(Z_{gij}^2) = E(Z_{gij}^4) - 1$.

Remark 1 We assume (A3). Then, $-1 < \lim_{p\to\infty} \sigma_{12}/\sigma_1\sigma_2 < 1$. See Section 2 in "supplemental material" for full details. Also note that if $\Delta = O$ or $\boldsymbol{\delta} = \mathbf{0}$ or $\kappa_{3g} = 0$, then $\sigma_{12} = 0$; if $\boldsymbol{\delta} = \mathbf{0}$, then $\sigma_1^2 = \sigma_{10}^2$; if $\Delta = O$, then $\sigma_2^2 = \sigma_{20}^2$.

The sum of the adjusted random variable and leading variance for each estimator is considered as 1:

$$
\frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\sigma_1} + \frac{\widehat{\|\Delta\|_F^2}}{\sigma_2}.
$$

The variances σ_1 and σ_2 are unknown, but we can replace them by the unbiased estimators of σ_1 and σ_2 under $\mathcal{H}_0$:

$$\widehat{\sigma}_{10}^2 = \sum_{g=1}^{2} \frac{2\widehat{\|\Sigma_g\|_F^2}}{n_g^2} + \frac{4\mathrm{tr}(S_1 S_2)}{n_1 n_2}, \quad \widehat{\sigma}_{20}^2 = \sum_{g=1}^{2} \frac{4(\widehat{\|\Sigma_g\|_F^2})^2}{n_g^2} + \frac{8\{\mathrm{tr}(S_1 S_2)\}^2}{n_1 n_2}.$$

Using these estimators, we propose the statistic

$$T = \frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\widehat{\sigma}_{10}} + \frac{\widehat{\|\Delta\|_F^2}}{\widehat{\sigma}_{20}}.$$

From the following theorem, we derive the asymptotic distribution of T.

Theorem 1 *Let*

$$\widetilde{T}(c_1, c_2) = c_1 \frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\sigma_1} + c_2 \frac{\widehat{\|\Delta\|_F^2}}{\sigma_2},$$

where c_1 and c_2 are real constants that do not depend on p, n_1, and n_2 and $(c_1, c_2)^\top \neq$ $\mathbf{0}$. Then, under Assumptions (A1), (A2), *and* (A3),

$$\frac{\widetilde{T}(c_1, c_2) - m(c_1, c_2)}{\sigma(c_1, c_2, \rho^*)} \xrightarrow{d} \mathcal{N}(0, 1) \ \text{ as } \ p \to \infty.$$

Here, $\xrightarrow{d}$ denotes convergence in distribution

$$m(c_1, c_2) = c_1 \frac{\|\boldsymbol{\delta}\|^2}{\sigma_1} + c_2 \frac{\|\Delta\|_F^2}{\sigma_2}, \quad \sigma(c_1, c_2, \rho^*) = \sqrt{c_1^2 + c_2^2 + 2c_1 c_2 \rho^*},$$

where $\rho^* = \lim_{p\to\infty} \sigma_{12}/(\sigma_1\sigma_2)$.

Proof See Appendix B.

Using Lemma A.1, we evaluate

$$\mathrm{var}(\widehat{\|\Sigma_g\|_F^2}) = O\left(\frac{\|\Sigma_g\|_F^4}{n_g}\right), \quad \mathrm{var}\{\mathrm{tr}(S_1 S_2)\} = O\left(\frac{(n_1+n_2)\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2}{n_1 n_2}\right)$$

under (A1). Thus, we obtain

$$\frac{\widehat{\|\Sigma_g\|_F^2}}{\|\Sigma_g\|_F^2} = 1 + o_p(1), \quad \frac{\mathrm{tr}(S_1 S_2)}{\mathrm{tr}(\Sigma_1\Sigma_2)} = 1 + o_p(1)$$

under (A1). Note that

$$\frac{\widehat{\sigma}_{10}^2}{\sigma_{10}^2} = \sum_{g=1,g\neq h}^{2} \frac{\widehat{\|\Sigma_g\|_F^2}/\|\Sigma_g\|_F^2 + n_g^2/\|\Sigma_g\|_F^2\left\{\|\Sigma_h\|_F^2/n_h^2 + 2\mathrm{tr}(\Sigma_1\Sigma_2)/(n_1n_2)\right\}}{1 + n_g^2/\|\Sigma_g\|_F^2\left\{\|\Sigma_h\|_F^2/n_h^2 + 2\mathrm{tr}(\Sigma_1\Sigma_2)/(n_1n_2)\right\}}$$
$$+\frac{\mathrm{tr}(S_1S_2)/\mathrm{tr}(\Sigma_1\Sigma_2) + n_1n_2/\{2\mathrm{tr}(\Sigma_1\Sigma_2)\}\sum_{g=1}^{2}\|\Sigma_g\|_F^2/n_g^2}{1 + n_1n_2/\{2\mathrm{tr}(\Sigma_1\Sigma_2)\}\sum_{g=1}^{2}\|\Sigma_g\|_F^2/n_g^2} - 2,$$
$$\frac{\widehat{\sigma}_{20}^2}{\sigma_{20}^2} = \sum_{g=1,g\neq h}^{2} \frac{(\widehat{\|\Sigma_g\|_F^2}/\|\Sigma_g\|_F^2)^2 + n_g^2/\|\Sigma_g\|_F^4\left[\|\Sigma_h\|_F^4/n_h^2 + 2\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2/(n_1n_2)\right]}{1 + n_g^2/\|\Sigma_g\|_F^4\left[\|\Sigma_h\|_F^4/n_h^2 + 2\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2/(n_1n_2)\right]}$$
$$+\frac{\{\mathrm{tr}(S_1S_2)/\mathrm{tr}(\Sigma_1\Sigma_2)\}^2 + n_1n_2/[2\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2]\sum_{g=1}^{2}\|\Sigma_g\|_F^4/n_g^2}{1 + n_1n_2/[2\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2]\sum_{g=1}^{2}\|\Sigma_g\|_F^4/n_g^2} - 2.$$

Hence, $\widehat{\sigma}_{g0}^2/\sigma_{g0}^2 = 1 + o_p(1)$ under (A1). Moreover, under (A3), $\sigma_g/\sigma_{g0} \to r_g$ for some $r_g \in [1, \infty)$. Therefore,

$$T = \widetilde{T}(r_1, r_2) + o_p(1)$$

under (A1) and (A3). From this fact and Theorem 1, we obtain the following lemma.

Lemma 1 *Under Assumptions* (A1), (A2)*, and* (A3)*,*

$$\frac{T - (\sigma_{10}^{-1}\|\boldsymbol{\delta}\|^2 + \sigma_{20}^{-1}\|\Delta\|_F^2)}{\sigma(r_1, r_2, \rho^*)} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } p \to \infty.$$

Remark 2 $\mathcal{H}_0 \Longrightarrow$ (*A*3).

2.2 Testing procedure and its asymptotic properties

Note that $\sigma(1, 1, 0) = 2$ under $\mathcal{H}_0$. From Lemma 1 and Remark 2, under (A1), (A2), and $\mathcal{H}_0$,

$$\frac{T}{\sqrt{2}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{2.2}$$

Based on the asymptotic normality of T under $\mathcal{H}_0$, we propose the following approximation test:

$$\text{rejecting } \mathcal{H}_0 \Longleftrightarrow T \geq \sqrt{2}z_\alpha. \tag{2.3}$$

Here, z_α is the upper critical value of the standard normal distribution, $\mathcal{N}(0, 1)$.

The size and power of test (2.3) are defined as

$$\text{Size} = \Pr(T \geq \sqrt{2} z_\alpha | \text{sample generated by model (1.1) and } \mathcal{H}_0),$$
$$\text{Power} = \Pr(T \geq \sqrt{2} z_\alpha | \text{sample generated by model (1.1) and } \mathcal{H}_1).$$

Remark 3 From (2.2), under (A1) and (A2), $Size = \alpha + o(1)$ as $p \to \infty$.

First, we investigate the power under (A3). Using Lemma 1, we obtain following proposition:

Proposition 1 *Under* (A1), (A2)*, and* (A3)*,*

$$\text{Power} = \Phi\left(\frac{\sigma_{10}^{-1}\|\boldsymbol{\delta}\|^2 + \sigma_{20}^{-1}\|\Delta\|_F^2}{\sigma(r_1, r_2, \rho^*)} - \frac{\sqrt{2}z_\alpha}{\sigma(r_1, r_2, \rho^*)}\right) + o(1)$$

as $p \to \infty$.

Thus, if the difference between $\boldsymbol{\mu}_1$ (Σ_1) and $\boldsymbol{\mu}_2$ (Σ_2) is not so small that $\|\boldsymbol{\delta}\|^2$ $(\|\Delta\|_F^2)$ has the same order as, or a higher order than, that σ_{10} (σ_{20}), the test will be powerful. Conversely, if $\|\boldsymbol{\delta}\|^2$ and $\|\Delta\|_F^2$ are so small that $\|\boldsymbol{\delta}\|^2$ and $\|\Delta\|_F^2$ are of a lower order than σ_{10} and σ_{20}, respectively, the test will not be powerful and cannot distinguish $\mathcal{H}_0$ from $\mathcal{H}_1$.

Next, we investigate the power under (A1) and (A4). We consider

$$\text{(A4-i)}\ \lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{\|\Sigma_g\|_F^2} = \infty, \quad \limsup_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{\|\Sigma_g\|_F^4} < \infty,$$
$$\text{(A4-ii)}\ \limsup_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{\|\Sigma_g\|_F^2} < \infty, \quad \lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{\|\Sigma_g\|_F^4} = \infty,$$
$$\text{(A4-iii)}\ \lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{\|\Sigma_g\|_F^2} = \infty, \quad \lim_{p\to\infty} \frac{n_g\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{\|\Sigma_g\|_F^4} = \infty.$$

Then, under (A1),

$$T - \sqrt{2}z_\alpha = \frac{\|\boldsymbol{\delta}\|^2}{\sigma_{10}} + \frac{\|\Delta\|_F^2}{\sigma_2} + o_p\left(\frac{\|\boldsymbol{\delta}\|^2}{\sigma_{10}}\right) \text{ under (A4} - \text{i)},$$
$$T - \sqrt{2}z_\alpha = \frac{\|\Delta\|_F^2}{\sigma_{20}} + \frac{\|\boldsymbol{\delta}\|^2}{\sigma_1} + o_p\left(\frac{\|\Delta\|_F^2}{\sigma_{20}}\right) \text{ under (A4} - \text{ii)},$$
$$T - \sqrt{2}z_\alpha = \frac{\|\boldsymbol{\delta}\|^2}{\sigma_{10}} + \frac{\|\Delta\|_F^2}{\sigma_{20}} + o_p\left(\frac{\|\boldsymbol{\delta}\|^2}{\sigma_{10}}\right) + o_p\left(\frac{\|\Delta\|_F^2}{\sigma_{20}}\right) \text{ under (A4} - \text{iii)}.$$

See Section 3 in "supplemental material" for full details. Thus, we obtain the following proposition.

Proposition 2 *Under* (A1) *and* (A4)*,* Power $= 1 + o(1)$ *as* $p \to \infty$.

3 Simulation and example

3.1 Simulation

We generate Monte Carlo samples to evaluate the performance of the proposed test procedure, including its size and power. Notice that the data are generated from the model:

$$\boldsymbol{X}_{gi} = \Sigma_g^{1/2} \boldsymbol{Z}_{gi} + \boldsymbol{\mu}_g \quad \text{for} \quad i \in \{1, \ldots, n_g\}. \tag{3.1}$$

We set the null hypothesis $\mathcal{H}_0$ as follows:

$$\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{0},\ \Sigma_1 = \Sigma_2 = B(0.3^{|i-j|})B, \tag{3.2}$$

where

$$B = \text{diag}\left\{\left(0.5 + \frac{1}{p+1}\right)^{1/2}, \left(0.5 + \frac{2}{p+1}\right)^{1/2}, \ldots, \left(0.5 + \frac{p}{p+1}\right)^{1/2}\right\}.$$

In addition, we set the alternative hypothesis $\mathcal{H}_1$ as follows:

$$\begin{aligned} &\boldsymbol{\mu}_1 = \boldsymbol{0},\ \Sigma_1 = B(0.3^{|i-j|})B, \\ &\boldsymbol{\mu}_2 = \|\Sigma_1\|_F^{1/2}(\sqrt{0.1/p}, \ldots, \sqrt{0.1/p})^\top, \quad \Sigma_2 = (1 - \sqrt{0.1})\Sigma_1. \end{aligned} \tag{3.3}$$

Note that $\|\boldsymbol{\delta}\|^2/\|\Sigma_1\|_F = 0.1$ and $\|\Delta\|_F^2/\|\Sigma_1\|_F^2 = 0.1$ under (3.3).

First, we compare the proposed test, denoted as HN, with the modified likelihood ratio test, designated MLR, the proposed likelihood ratio based approximation test, described as HNLR, and simple simultaneous test by using existing approximate tests and Bonferroni correction, described as BF. The MLR and HNLR tests are considered for multivariate normal populations where $p < \min\{n_1 - 1, n_2 - 1\}$. The MLR, proposed by Muirhead (1982), is

$$\text{rejecting } \mathcal{H}_0 \iff -2\rho \log \lambda \geq \chi^2_{p(p+3)/2}(\alpha),$$

where $\chi_a(\alpha)$ denotes the upper 100α percentile of the Chi-squared distribution with a degrees of freedom,

$$\rho = 1 - \frac{(2p^2 + 9p + 11)(n_1^2 + n_2^2 + n_1 n_2)}{6(p+3)n n_1 n_2}.$$

Under a large-sample framework, the size of this approximation test is $\alpha + O(n^{-2})$. Let $\tilde{\lambda} = \prod_{g=1}^2 n_g^{pn_g/2}/(n^{pn/2})\lambda$. By applying the same techniques as in Fujikoshi et al. (2010) for the general moment of $\tilde{\lambda}$ derived by Muirhead (1982), we can express the cumulant-generating function as the following infinite series:

$$K_{\tilde{\lambda}}(t) = \sum_{s=1}^{\infty} \frac{(\sqrt{-1}t)^s \kappa^{(s)}}{s!},$$

where

$$\kappa^{(s)} = (-1)^s \sum_{i=1}^{p} \left\{ \sum_{g=1}^{2} r_g^s \psi^{(s-1)} \left(\frac{n_g - i}{2} \right) - \psi^{(s-1)} \left(\frac{n - i}{2} \right) \right\}.$$

Here, $\psi^{(s)}(\cdot)$ denotes the polygamma function. Using the first- and second-order cumulants, we test the HNLR as follows:

$$\text{rejecting } \mathcal{H}_0 \iff \frac{\tilde{\lambda} - \kappa^{(1)}}{\sqrt{\kappa^{(2)}}} \geq z_\alpha.$$

We test BF as follows:

$$\text{rejecting } \mathcal{H}_0 \iff T_{\text{CQ}} \geq z_{\alpha/2} \text{ or } T_{\text{LC}} \geq z_{\alpha/2}.$$

Here, T_{CQ} is proposed by Chen and Qin (2010) for testing equality of the population mean vectors, and T_{LC} is proposed by Li and Chen (2012) for testing equality of the population covariance matrices. For specific formulas of these statistics, see their papers. The Bonferroni correction sets the significance cutoff at $z_{\alpha/2}$. In general, the Bonferroni correction tends to be a bit too conservative. However, note that since the individual test is an approximate test, this tendency does not necessarily hold with finite samples.

We generated n samples of $\boldsymbol{X}_{gi}$ $(i \in \{1, \ldots, n_g\},\ g \in \{1, 2\})$, from the p-variate normal distribution $\mathcal{N}_p(\boldsymbol{\mu}_g, \Sigma_g)$. We use the simulation settings (3.2) for size and (3.3) for power as $p \in \{10, 20, 40, 80\}$, $(n_1, n_2) \in \{(p+10, p+10), (p+10, p+30), (p+50, p+50), (p+20, p+80)\}$ and conduct 10^5 replications to calculate size of MLR, HNLR, and HN, type I familywise error rate of BF and power. The nominal significance level is $\alpha = 0.05$. Table 1 shows the empirical size of MLR, HNLR, and HN, type I familywise error rate of BF and power. The size of MLR greatly exceeds the significance level of 0.05 as the dimension increases. On the other hand, the size of both HN and HNLR is close to a significance level of 0.05 even in high-dimensional settings. The power of HN exceeds that of HNLR in all cases. In all cases, the size of HN is more close to a significance level of 0.05 than one of BF. In addition, BF is not conservative test despite using the Bonferroni correction. The power of HN exceeds that of BF in most cases. These results indicate that the proposed HN test is superior to the MLR, HNLR, and BF tests in high-dimensional settings.

Next, we investigate the size and power of the proposed test under some distributions. For $j = 1, \ldots, n_g,\ g = 1, 2$, $\boldsymbol{Z}_{gj} = (Z_{gij})$ emerges in (3.1) for the following distributions:

Table 1 Comparison of modified likelihood ratio test

$p = 10$	(n_1, n_2)	(20, 20)	(20, 40)	(60, 60)	(30, 90)
Size	MLR	0.068	0.066	0.048	0.055
	HNLR	0.056	0.057	0.054	0.056
	HN	0.074	0.072	0.068	0.070
	BF	0.098	0.093	0.087	0.092
Power	MLR	0.146	0.232	0.594	0.490
	HNLR	0.120	0.209	0.620	0.495
	HN	0.302	0.350	0.796	0.572
	BF	0.314	0.363	0.763	0.554
$p = 20$	(n_1, n_2)	(30, 30)	(30, 50)	(70, 70)	(40, 100)
Size	MLR	0.156	0.130	0.055	0.081
	HNLR	0.057	0.054	0.056	0.052
	HN	0.066	0.065	0.063	0.064
	BF	0.081	0.083	0.076	0.079
Power	MLR	0.307	0.381	0.653	0.607
	HNLR	0.128	0.210	0.656	0.528
	HN	0.458	0.532	0.903	0.753
	BF	0.427	0.479	0.872	0.703
$p = 40$	(n_1, n_2)	(50, 50)	(50, 70)	(90, 90)	(60, 120)
Size	MLR	0.665	0.484	0.093	0.233
	HNLR	0.055	0.054	0.053	0.055
	HN	0.060	0.060	0.061	0.059
	BF	0.073	0.074	0.074	0.070
Power	MLR	0.862	0.822	0.785	0.863
	HNLR	0.155	0.236	0.682	0.570
	HN	0.765	0.833	0.985	0.946
	BF	0.708	0.764	0.963	0.910
$p = 80$	(n_1, n_2)	(90, 90)	(90, 110)	(130, 130)	(100, 160)
Size	MLR	1.000	0.998	0.451	0.897
	HNLR	0.051	0.050	0.052	0.051
	HN	0.057	0.057	0.056	0.057
	BF	0.065	0.065	0.065	0.070
Power	MLR	1.000	1.000	1.000	1.000
	HNLR	0.226	0.326	0.756	0.668
	HN	0.990	0.995	0.987	0.999
	BF	0.975	0.985	0.988	0.999

Table 2 Third- and fourth-order cumulants

	(D1)	(D2)	(D3)	(D4)
κ_3	0.00	0.89	0.00	– 0.67
κ_4	0.00	1.20	1.00	0.51

Table 3 (D1) Normal distribution

$(n_1, n_2)\backslash p$	32	64	128	256
(20, 20)				
Size	0.065	0.062	0.060	0.059
Power	0.302	0.301	0.297	0.297
(30, 10)				
Size	0.062	0.058	0.056	0.054
Power	0.278	0.271	0.260	0.257
(50, 50)				
Size	0.061	0.058	0.057	0.056
Power	0.757	0.782	0.799	0.814
(70, 30)				
Size	0.061	0.058	0.055	0.054
Power	0.721	0.745	0.760	0.775
(80, 80)				
Size	0.061	0.058	0.056	0.055
Power	0.958	0.974	0.982	0.987
(120, 40)				
Size	0.060	0.057	0.055	0.054
Power	0.910	0.933	0.948	0.959
(100, 100)				
Size	0.061	0.057	0.055	0.054
Power	0.991	0.996	0.998	0.999
(150, 50)				
Size	0.060	0.056	0.054	0.053
Power	0.972	0.984	0.990	0.993

(D1) $Z_{gij} \sim \mathcal{N}(0, 1)$,

(D2) $Z_{gij} = (U_{gij} - 10)/\sqrt{20}$ for $U_{gij} \sim \chi^2_{10}$,

(D3) $Z_{gij} = U_{gij}/\sqrt{5/4}$ for $U_{gij} \sim \mathcal{T}_{10}$,

(D4) $Z_{gij} = \left(1 - \dfrac{9}{5\pi}\right)^{-1/2} \left(U_{gij} + \dfrac{3}{\sqrt{5\pi}}\right)$ for $U_{gij} \sim \mathcal{SN}(-3)$.

Table 2 presents the third-order cumulant κ_3 and the fourth-order cumulant κ_4 for each distribution. We use the simulation settings (3.2) for size and (3.3) for power as $p \in \{32, 64, 128, 256\}$, $(n_1, n_2) \in \{(20, 20), (30, 10), (50, 50), (70, 30), (80, 80), (120, 40), (100, 100), (150, 50)\}$. The nominal significance level is $\alpha = 0.05$. Tables 3, 4, 5, and 6 show the empirical size and power of HN for each distribu-

Table 4 (D2) Standardized Chi-squared distribution

$(n_1, n_2)\backslash p$	32	64	128	256
(20, 20)				
Size	0.075	0.068	0.064	0.062
Power	0.178	0.173	0.167	0.165
(30, 10)				
Size	0.071	0.064	0.060	0.057
Power	0.186	0.173	0.165	0.159
(50, 50)				
Size	0.071	0.065	0.061	0.058
Power	0.443	0.447	0.450	0.455
(70, 30)				
Size	0.070	0.064	0.059	0.057
Power	0.437	0.435	0.436	0.435
(80, 80)				
Size	0.070	0.064	0.060	0.057
Power	0.716	0.739	0.758	0.770
(120, 40)				
Size	0.070	0.063	0.059	0.057
Power	0.647	0.665	0.681	0.692
(100, 100)				
Size	0.070	0.064	0.059	0.057
Power	0.849	0.873	0.891	0.904
(150, 50)				
Size	0.069	0.063	0.058	0.056
Power	0.776	0.806	0.825	0.840

tion. From Tables 3, 4, 5 and 6, we find that the power is monotonically increasing for n and p, except for $n = 40$. The proposed test has the highest power in case (D4). We note that case (D4) has a third-order cumulant, the lowest among (D1)–(D4). Conversely, the proposed test has the lowest power in case (D2). We also note that case (D2) has third- and fourth-order cumulants, the highest among (D1)–(D4). In any case, the third- and fourth-order cumulants affect the power of the test in these simulations.

3.2 An example

In this data analysis, we investigate whether the difference of distribution has been made by the difference of mean vector $\boldsymbol{\delta}$ or by the difference of covariance matrix Δ when differences of distribution can be found by using proposed test (2.3). So, we construct a multiple testing procedure based on (2.3) and we adopt this procedure to Leukemia dataset. We also proposed multiple pairwise comparison procedure based on (2.3) and we adopt this procedure to Khan dataset.

Table 5 (D3) Standardized t distribution

$(n_1, n_2)\backslash p$	32	64	128	256
(20, 20)				
Size	0.068	0.064	0.061	0.059
Power	0.285	0.282	0.280	0.275
(30, 10)				
Size	0.064	0.059	0.057	0.055
Power	0.272	0.257	0.250	0.243
(50, 50)				
Size	0.064	0.060	0.057	0.056
Power	0.720	0.743	0.762	0.777
(70, 30)				
Size	0.063	0.059	0.056	0.055
Power	0.686	0.710	0.722	0.734
(80, 80)				
Size	0.063	0.059	0.057	0.055
Power	0.942	0.962	0.971	0.978
(120, 40)				
Size	0.062	0.058	0.056	0.054
Power	0.883	0.911	0.931	0.943
(100, 100)				
Size	0.063	0.059	0.056	0.054
Power	0.985	0.990	0.996	0.997
(150, 50)				
Size	0.062	0.058	0.055	0.053
Power	0.957	0.973	0.983	0.989

3.2.1 Leukemia dataset

We apply our test to a microarray dataset analyzed by Dudoit et al. (2002). The dataset, obtained from Affymetrix oligonucleotide microarrays, contains 72 cases of either acute lymphoblastic leukemia (ALL, $n_1 = 47$) or acute myeloid leukemia (AML, $n_2 = 25$). The dataset is publicly available at http://portals.broadinstitute.org/gpp/public/. We preprocess the dataset using the protocol written in Dudoit et al. (2002). The preprocessed dataset comprises $p = 3571$ variables. Using this dataset, we calculate

$$\widehat{\|\boldsymbol{\delta}\|^2} \approx 50.2, \quad \widehat{\|\Delta\|_F^2} \approx 23758.9, \quad \widehat{\sigma}_{10} \approx 11.7, \quad \widehat{\sigma}_{20} \approx 2340.4.$$

From these values, we construct a closed testing procedure to simultaneously test the mean vector and the covariance matrix. Let $\mathcal{C} = \{\mathcal{H}_{\delta,\Delta}, \mathcal{H}_\delta, \mathcal{H}_\Delta\}$. Here, $\mathcal{H}_{\delta,\Delta} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2,\ \Sigma_1 = \Sigma_2$, $\mathcal{H}_\delta : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, $\mathcal{H}_\Delta : \Sigma_1 = \Sigma_2$. Note that the family $\mathcal{C}$ is closed.

Table 6 (D4) Standardized skew normal distribution

$(n_1, n_2)\backslash p$	32	64	128	256
(20, 20)				
Size	0.070	0.066	0.063	0.061
Power	0.397	0.403	0.404	0.402
(30, 10)				
Size	0.067	0.061	0.057	0.056
Power	0.360	0.350	0.345	0.341
(50, 50)				
Size	0.066	0.061	0.059	0.057
Power	0.877	0.902	0.919	0.934
(70, 30)				
Size	0.066	0.061	0.058	0.055
Power	0.839	0.870	0.890	0.906
(80, 80)				
Size	0.065	0.061	0.058	0.056
Power	0.992	0.996	0.998	0.999
(120, 40)				
Size	0.065	0.060	0.056	0.055
Power	0.969	0.982	0.989	0.994
(100, 100)				
Size	0.065	0.060	0.057	0.055
Power	0.999	1.000	1.000	1.000
(150, 50)				
Size	0.065	0.060	0.057	0.055
Power	0.994	0.998	0.999	1.000

Then, we test the hypotheses in $\mathcal{C}$ by using the following procedure: Then, we test the hypotheses

Step 1 We test hypothesis $\mathcal{H}_{\delta,\Delta}$.

Case 1 If $T > \sqrt{2}z_\alpha$, we reject $\mathcal{H}_{\delta,\Delta}$ and go to Step 2.

Case 2 If $T \leq \sqrt{2}z_\alpha$, we retain all hypotheses, $\mathcal{H}_{\delta,\Delta}, \mathcal{H}_\delta,$ and $\mathcal{H}_\Delta$.

Step 2 We test hypotheses $\mathcal{H}_\delta$ and $\mathcal{H}_\Delta$. If $\widehat{\|\boldsymbol{\delta}\|^2}/\widehat{\sigma}_{10} > z_\alpha$, we reject $\mathcal{H}_\delta$. If $\widehat{\|\boldsymbol{\delta}\|^2}/\widehat{\sigma}_{10} \leq z_\alpha$, we retain $\mathcal{H}_\delta$. If $\widehat{\|\Delta\|_F^2}/\widehat{\sigma}_{20} > z_\alpha$, we reject $\mathcal{H}_\Delta$. If $\widehat{\|\Delta\|_F^2}/\widehat{\sigma}_{20} \leq z_\alpha$, we retain $\mathcal{H}_\Delta$.

We assume (A1) and (A2). Then, the following statements hold:

- If $\mathcal{H}_{\delta,\Delta}$, $\mathcal{H}_\delta$ and $\mathcal{H}_\Delta$ are true, $\Pr\left(T > \sqrt{2}z_\alpha,\ \frac{\widehat{\|\Delta\|_F^2}}{\widehat{\sigma}_{20}} > z_\alpha,\ \frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\widehat{\sigma}_{10}} > z_\alpha\right) \leq \Pr\left(T > \sqrt{2}z_\alpha\right) = \alpha + o(1)$.
- If $\mathcal{H}_\delta$ is true, $\Pr\left(T > \sqrt{2}z_\alpha,\ \frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\widehat{\sigma}_{10}} > z_\alpha\right) \leq \Pr\left(\frac{\widehat{\|\boldsymbol{\delta}\|^2}}{\widehat{\sigma}_{10}} > z_\alpha\right) = \alpha + o(1)$.

- If $\mathcal{H}_\Delta$ is true, $\Pr\left(T > \sqrt{2}z_\alpha,\ \frac{\widehat{\|\Delta\|_F^2}}{\widehat{\sigma}_{20}} > z_\alpha\right) \le \Pr\left(\frac{\widehat{\|\Delta\|_F^2}}{\widehat{\sigma}_{20}} > z_\alpha\right) = \alpha + o(1)$.

From the above, the type I familywise error rate of this procedure is not greater than α.

We set a $\alpha = 0.05$ level of significance. In Step 1, the test statistic T is calculated as $T \approx 48.6$, and the approximate critical values of T, at a $\alpha = 0.05$ level of significance, are obtained as $\sqrt{2}z_{0.05} \approx 2.326$. We reject $\mathcal{H}_{\delta,\Delta}$ and go to Step 2. In Step 2, we calculate $\widehat{\|\boldsymbol{\delta}\|^2}/\widehat{\sigma}_{10} \approx 4.3$ and $\widehat{\|\Delta\|_F^2}/\widehat{\sigma}_{20} \approx 10.2$. Both values are greater than $z_{0.05} \approx 1.645$. Thus, we reject $\mathcal{H}_\delta$ and $\mathcal{H}_\Delta$.

3.2.2 Khan dataset

We compare the population mean of four types of gene expression using training data. Khan et al. (2001) study the expression of genes in four types of small round blue cell tumors of childhood (SRBCT). These were the Ewing family of tumors (EWS, 23 cases), Burkitt lymphoma, a subset of non-Hodgkin lymphoma (BL, 8 cases), neuroblastoma (NB, 12 cases), and rhabdomyosarcoma (RMS, 21 cases). The data include the gene expression profiles obtained from both tumor biopsy and cell line samples. These were downloaded from a Web site containing a filtered dataset of 2308 gene expression profiles as described by Khan et al. (2001). This dataset is available from http://bioinf.ucd.ie/people/aedin/R/. First, we construct a hypothesis test for all differences in the population mean vectors $\boldsymbol{\mu}_{EWS}$, $\boldsymbol{\mu}_{\mathrm{BL}}$, $\boldsymbol{\mu}_{\mathrm{NB}}$, and $\boldsymbol{\mu}_{\mathrm{RMS}}$. By combining Bonferroni correction with the proposed test (2.1), we obtain pairwise comparison procedures for all possible pairs.

Step 1 For $g, h \in \{1, \ldots, 4\},\ g \neq h$, we define $\mathcal{H}_0^{(g,h)}$ as $\boldsymbol{\mu}_g = \boldsymbol{\mu}_h,\ \Sigma_g = \Sigma_h$. We set the family of hypotheses:

$$\mathcal{F} = \left\{\mathcal{H}_0^{(g,h)}\ ;\ g < h,\ g, h = 1, \ldots, 4\right\}.$$

Step 2 We set the significance level at $\alpha \in (0, 1)$.
Step 3 We choose $\mathcal{H}_0^{(g,h)}$ in $\mathcal{F}$.
Step 4 By using (g, h) determined in Step 3, we calculate the statistic $t_{gh} = \widehat{\|\boldsymbol{\delta}_{gh}\|^2}/\widehat{\sigma}_{gh,10} + \widehat{\|\Delta_{gh}\|_F^2}/\widehat{\sigma}_{gh,20}$.
Step 5 We reject $\mathcal{H}_0^{(g,h)}$ if $t_{gh} > \sqrt{2}z_{\alpha/6}$. We perform Steps 4 and 5 for all contrasts taken up in Step 3. The realized values of the test statistic t_{gh} are summarized in Table 7.

Table 7 The values of t_{gh}

	EWS	BL	NB	RMS
EWS	–	25.51*	28.40*	15.90*
BL	–	–	17.67*	30.20*
NB	–	–	–	24.35*

The mark "$*$" represents 5% significance

4 Conclusion and discussion

This paper treated two-sample simultaneous test for high-dimensional mean vectors and covariance matrices under non-normal populations. When $p \geq \min\{n_1 - 1, n_2 - 1\}$, the likelihood ratio test is not applicable for this problem. Therefore, we considered an L^2-norm-based test to simultaneously test the mean vector and the covariance matrix. The L^2-norm-based test for mean vectors ($\mathcal{H}_0 : \boldsymbol{\delta} = \mathbf{0}$ vs. $\mathcal{H}_1 : \boldsymbol{\delta} \neq \mathbf{0}$) was developed by Chen and Qin (2010) and the L^2-norm-based test for covariance matrices ($\mathcal{H}_0 : \Delta = O$ vs. $\mathcal{H}_1 : \Delta \neq O$) by Li and Chen (2012). Our test statistic is based on the sum of standardized estimators used in different tests, and we investigated the asymptotic size and power of the proposed test under a few asymptotic frameworks. We therefore infer that the power of this test depends on the distance between Σ_1 ($\boldsymbol{\mu}_1$) and Σ_2 ($\boldsymbol{\mu}_2$), a trace of the power of Σ_g, and the third- and fourth-order cumulants.

Further, we studied the finite sample and dimension performance of this test under a few non-normal settings via Monte Carlo simulations. Through the simulation results, we confirmed that the proposed test is satisfactory in most cases. The size of the proposed test is only slightly larger than the nominal significance level when the total sample size n and the dimension p are relatively large. The power monotonically increases with the total sample size n and the dimension p. Even when the relatively dimension p is relatively small and $p < \min\{n_1 - 1, n_2 - 1\}$, our test demonstrated better accuracy than the likelihood ratio test.

In conclusion, our test can be recommended for two-sample simultaneous testing of mean vectors and covariance matrices when the total sample size n and the dimension p are both large. However, our non-normal assumption does not include the family of elliptical distributions. A discussion on this distribution family is a task for the future.

Acknowledgements The authors would like to thank Mr. Hayate Ogawa for his help. We also would like to thank the Editor-in-Chief, an Associate Editor, and two reviewers for many valuable comments and helpful suggestions which led to an improved version of this paper. The research of the author was supported in part by a Grant-in-Aid for Young Scientists (B) (17K14238) from the Japan Society for the Promotion of Science. The second author was supported in part by a Grant-in-Aid for Young Scientists (B) (26730020) from the Japan Society for the Promotion of Science.

Appendix A: Some results for moment

Lemma A. 1 *Let A, B be $p \times p$ non-random symmetric matrices and $\boldsymbol{b}$ be p-dimensional non-random vector. Assume that the ith element of p-dimensional random vectors $\mathbf{Z}$ (denotes Z_i) and $\mathbf{W}$ (denotes W_i) has uniformly bounded 8th moment, and there exist finite constants κ_3, κ_4 such that $E(Z_i^3) = E(W_i^3) = \kappa_3$, $E(Z_i^4) = E(W_i^4) = \kappa_4 + 3$ ($\kappa_4 \geq -2$) and for any positive integers r and α_ℓ such that $\sum_{\ell=1}^r \alpha_\ell \leq 8$, $E(\prod_{\ell=1}^r Z_{i_\ell}^{\alpha_\ell}) = \prod_{\ell=1}^r E(Z_{i_\ell}^{\alpha_\ell})$ and $E(\prod_{\ell=1}^r W_{i_\ell}^{\alpha_\ell}) = \prod_{\ell=1}^r E(W_{i_\ell}^{\alpha_\ell})$ whenever $i_1, i_2, \ldots, i_r$ are distinct indices. Also, $\mathbf{Z}$ and $\mathbf{W}$ are independent. Then, it holds that*

(i) $E(\boldsymbol{Z}^\top A\boldsymbol{Z}\boldsymbol{b}^\top \boldsymbol{Z}) = \kappa_3 \sum_{i=1}^{p} \boldsymbol{e}_i^\top A\boldsymbol{e}_i\boldsymbol{b}^\top \boldsymbol{e}_i$,

(ii) $E(\boldsymbol{Z}^\top A\boldsymbol{Z}\boldsymbol{Z}^\top B\boldsymbol{Z}) = \kappa_4 \mathrm{tr}(A \odot B) + \mathrm{tr}A\mathrm{tr}B + 2\mathrm{tr}(AB)$,

(iii) $E\{(\boldsymbol{Z}^\top A\boldsymbol{W})^4\} = \kappa_4^2 \mathrm{tr}\{(\odot^2 A)^2\} + 6\kappa_4 \mathrm{tr}(\odot^2 A^2) + 3\|A\|_F^4 + 6\|A^2\|_F^2$,

(iv) $E\{(\boldsymbol{Z}^\top A\boldsymbol{Z} - \mathrm{tr}A)^4\} \leq const.\|A\|_F^4$.

Proof See Section 1 in "supplemental material."

Appendix B: Proof of Theorem 1

Under assumptions (A1) and (A2),

$$\widetilde{T}(c_1, c_2) = m(c_1, c_2) + \sum_{k=1}^{n} \varepsilon_k + o_p(1),$$

where for $1 \leq k \leq n_1$,

$$\varepsilon_k = \frac{2c_1 \boldsymbol{b}_{1k}^\top \boldsymbol{Y}_{1k}}{\sigma_1 n_1(n_1 - 1)} + \frac{2c_2\{\boldsymbol{Y}_{1k}^\top A_{1k}\boldsymbol{Y}_{1k} - \mathrm{tr}(A_{1k}\Sigma_1)\}}{\sigma_2 n_1(n_1 - 1)},$$

for $n_1 + 1 \leq k \leq n$,

$$\varepsilon_k = \frac{2c_1 \boldsymbol{b}_{2k}^\top \boldsymbol{Y}_{2\ k-n_1}}{\sigma_1 n_2(n_2 - 1)} + \frac{2c_2\{\boldsymbol{Y}_{2\ k-n_1}^\top A_{2k}\boldsymbol{Y}_{2\ k-n_1} - \mathrm{tr}(A_{2k}\Sigma_2)\}}{\sigma_2 n_2(n_2 - 1)}.$$

Here,

$$A_{1k} = \sum_{j=1}^{k-1} \mathcal{E}_j^{(1)} + (n_1 - 1)\Delta, \quad A_{2k} = \sum_{j=1}^{k-n_1-1} \mathcal{E}_j^{(2)} - \frac{n_2 - 1}{n_1} \sum_{j=1}^{n_1} \mathcal{E}_j^{(1)} - (n_2 - 1)\Delta,$$

$$\boldsymbol{b}_{1k} = \sum_{j=1}^{k-1} \boldsymbol{Y}_{1j} + (n_1 - 1)\boldsymbol{\delta}, \quad \boldsymbol{b}_{2k} = \sum_{j=1}^{k-n_1-1} \boldsymbol{Y}_{2j} - \frac{n_2 - 1}{n_1} \sum_{j=1}^{n_1} \boldsymbol{Y}_{1j} - (n_2 - 1)\boldsymbol{\delta},$$

where $\mathcal{E}_j^{(1)} = \boldsymbol{Y}_{1j}\boldsymbol{Y}_{1j}^\top - \Sigma_1$, $\mathcal{E}_j^{(2)} = \boldsymbol{Y}_{2j}\boldsymbol{Y}_{2j}^\top - \Sigma_2$.

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_k$ $(1 \leq k)$ be the σ-algebra by the random variables $(\varepsilon_1, \ldots, \varepsilon_k)$. Then, we find that

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_\infty$$

and $E(\varepsilon_k|\mathcal{F}_{k-1}) = 0$. We show asymptotic normality of $\sum_{k=1}^{n} \varepsilon_k$ by adapting martingale difference central limit theorem, see Hall and Heyde (1980). It is necessary to check the following two conditions to apply this theorem:

$$\text{(I)} \sum_{k=1}^{n} E(\varepsilon_k^2|\mathcal{F}_{k-1}) = \sigma(c_1, c_2, \rho^*)^2 + o_p(1) \text{ as } p \to \infty,$$
$$\text{(II)} \sum_{k=1}^{n} E(\varepsilon_k^4) = o(1) \text{ as } p \to \infty.$$

For any $(c_1, c_2)^\top \in \mathcal{R}^2,\ (c_1, c_2)^\top \neq \mathbf{0}, \sigma(c_1, c_2, \rho^*)^2 > 0$ under (A3).

First we check (I). We define the following four random variables:

$$C_k^{(g)} = \frac{\boldsymbol{b}_{gk}^\top \Sigma_g \boldsymbol{b}_{gk}}{n_g^2(n_g - 1)^2},$$
$$D_k^{(g)} = \frac{\mathrm{tr}\{(\Sigma_g A_{gk})^2\}}{n_g^2(n_g - 1)^2},$$
$$E_k^{(g)} = \frac{\mathrm{tr}\{\odot^2(\Sigma_g^{1/2} A_{gk} \Sigma_g^{1/2})\}}{n_g^2(n_g - 1)^2},$$
$$F_k^{(g)} = \frac{\sum_{i=1}^{p} (\boldsymbol{b}_{gk}^\top \Sigma_g^{1/2} \boldsymbol{e}_i \boldsymbol{e}_i^\top \Sigma_g^{1/2} A_{gk} \Sigma_g^{1/2} \boldsymbol{e}_i)}{n_g^2(n_g - 1)^2}.$$

Then, the sum of conditional mean $\sum_{k=1}^{n} E(\varepsilon_k^2|\mathcal{F}_{k-1})$ partially decomposes as

$$4\sum_{k=1}^{n_1} \left\{ \frac{c_1^2}{\sigma_1^2} C_k^{(1)} + \frac{c_2^2}{\sigma_2^2}(2D_k^{(1)} + \kappa_{41} E_k^{(1)}) + \frac{2c_1c_2\kappa_{31}}{\sigma_1\sigma_2} F_k^{(1)} \right\}$$
$$+4\sum_{k=n_1+1}^{n} \left\{ \frac{c_1^2}{\sigma_1^2} C_k^{(2)} + \frac{c_2^2}{\sigma_2^2}(2D_k^{(2)} + \kappa_{42} E_k^{(2)}) + \frac{2c_1c_2\kappa_{32}}{\sigma_1\sigma_2} F_k^{(2)} \right\}.$$

Under Assumptions (A1) and (A2), it holds the following four statements:

$$\text{(A)}\ \ \sigma_1^{-2} \left(\sum_{k=1}^{n_1} C_k^{(1)} + \sum_{k=n_1+1}^{n} C_k^{(2)} \right) - \frac{c_0}{\sigma_1^2} = o_p(1),$$
$$\text{(B)}\ \ \sigma_2^{-2} \left(\sum_{k=1}^{n_1} D_k^{(1)} + \sum_{k=n_1+1}^{n} D_k^{(2)} \right) - \frac{d_0}{\sigma_2^2} = o_p(1),$$
$$\text{(C)}\ \ \sigma_2^{-2} \left(\sum_{k=1}^{n_1} E_k^{(1)} + \sum_{k=n_1+1}^{n} E_k^{(2)} \right) - \frac{e_0}{\sigma_2^2} = o_p(1),$$
$$\text{(D)}\ \ (\sigma_1\sigma_2)^{-1} \left(\sum_{k=1}^{n_1} F_k^{(1)} + \sum_{k=n_1+1}^{n} F_k^{(2)} \right) - \frac{f_0}{\sigma_1\sigma_2} = o_p(1),$$

where

$$
\begin{aligned}
c_0 &= \sum_{g=1}^{2} \frac{\|\Sigma_g\|_F^2}{2n_g^2} + \frac{\mathrm{tr}(\Sigma_1\Sigma_2)}{n_1 n_2} + \sum_{g=1}^{2} \frac{\|\Sigma_g^{1/2}\boldsymbol{\delta}\|^2}{n_g},\\
d_0 &= \sum_{g=1}^{2} \frac{\|\Sigma_g\|_F^4}{2n_g^2} + \frac{\{\mathrm{tr}(\Sigma_1\Sigma_2)\}^2}{n_1 n_2} + \sum_{g=1}^{2} \frac{\|\Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\|_F^2}{n_g},\\
e_0 &= \sum_{g=1}^{2} \frac{\mathrm{tr}\{\odot^2(\Sigma_g^{1/2}\Delta\Sigma_g^{1/2})\}}{n_g},\\
f_0 &= \frac{\sum_{i=1}^{p} \boldsymbol{\delta}^\top \Sigma_g^{1/2}\boldsymbol{e}_i\boldsymbol{e}_i^\top \Sigma_g^{1/2}\Delta\Sigma_g^{1/2}\boldsymbol{e}_i}{n_g}.
\end{aligned}
$$

These results are obtained by evaluating the moment of four random variables. See Sections 4–7 in "supplemental material" for full details. Since (A)–(D) is satisfied, the statement (I) is true.

Next we check (II). Under assumptions (A1) and (A2), it holds the following two statements:

$$
\text{(E)} \sum_{k=1}^{n_1} E(\varepsilon_k^4) = o(1) \quad \text{as} \quad p \to \infty,
$$

$$
\text{(F)} \sum_{k=n_1+1}^{n} E(\varepsilon_k^4) = o(1) \quad \text{as} \quad p \to \infty.
$$

See Sections 8 and 9 in "supplemental material" for full details. Thus, the asymptotic normality of $T(c_1, c_2)$ is established.

References

Chen SX, Qin YL (2010) A two-sample test for high dimensional data with applications to gene-set testing. Ann Stat 38:808–835

Dudoit S, Fridlyand J, Speed PT (2002) Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc 97:77–87

Fujikoshi Y, Ulyanov VV, Shimizu R (2010) Multivariate statistics: high-dimensional and large-sample approximations. Wiley, Hobeken

Hall P, Heyde CC (1980) Martingale limit theory and its application. Academic Press, New York

Himeno T, Yamada T (2014) Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. J Multivar Anal 130:27–44

Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 7:673–679

Li J, Chen SX (2012) Two sample tests for high-dimensional covariance matrices. Ann Stat 40:908–940

Liu Z, Liu B, Zheng S, Shi NZ (2017) Simultaneous testing of mean vector and covariance matrix for high-dimensional data. J Stat Plan Inference 188:82–93

Muirhead RJ (1982) Aspects of multivariate statistical theory. Wiley, New York

Srivastava MS, Yanagihara H, Kubokawa T (2014) Tests for covariance matrices in high dimension with less sample size. J Multivar Anal 130:289–309