Tests for the parallelism and flatness hypotheses of multi-group profile analysis for high-dimensional elliptical populations

Tests for the parallelism and flatness hypotheses of multi-group profile analysis for high-dimensional elliptical populations

Masashi Hyodo[†]

[†] *Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University, 1-1, Gakuen-cho, Naka-ku, Sakai-shi, Osaka 599-8531, Japan*
E-Mail: `hyodo@ms.osakafu-u.ac.jp`

Abstract

This paper is concerned with tests for the parallelism and flatness hypotheses in multi-group profile analysis for high-dimensional data. We extend to elliptical distributions the procedures developed for normal populations by Harrar and Kong [S.W. Harrar, X. Kong, High-dimensional multivariate repeated measures analysis with unequal covariance matrices, J. Multivariate Anal. 145 (2016) 1–21]. Specifically, we prove that their statistics continue to be asymptotically normal when the underlying population is elliptical, and we obtain new tests by improving their estimator of the asymptotic variance. Using asymptotic normality, we show that the asymptotic size of the proposed tests is equal to the nominal significance level, and we also derive the asymptotic power. Finally, we present simulation results and find that the power of the new tests is superior to that of the original Harrar–Kong procedure.

AMS 2000 subject classification: Primary 62H15; secondary 62F05.

Key words: Elliptical population, high dimension, profile analysis, statistical hypothesis testing.

1. Introduction

We consider a multi-sample testing problem for profile analysis for populations with elliptically contoured distributions. For group $g \in \{1, \ldots, a\}$, let $\boldsymbol{\mu}_g = (\mu_{g1}, \ldots, \mu_{gp})^\top$ be a p-dimensional real vector, Λ_g be a $p \times p$ nonnegative definite matrix, and ξ_g be a nonnegative function. The $p \times 1$ random vector $\boldsymbol{X}_g$ is said to have an elliptically contoured distribution, denoted $\boldsymbol{X}_g \sim C_p(\xi_g, \boldsymbol{\mu}_g, \Lambda_g)$, if the characteristic function of $\boldsymbol{X}_g$ can be written, for any $\boldsymbol{t} \in \mathbb{R}^p$, as

$$\phi_g(\boldsymbol{t}) = e^{i\boldsymbol{t}^\top \boldsymbol{\mu}_g} \xi_g(\boldsymbol{t}^\top \Lambda_g \boldsymbol{t}).$$

As a result, $\mathrm{E}(\boldsymbol{X}_g) = \boldsymbol{\mu}_g$ and $\mathrm{var}(\boldsymbol{X}_g) = -2\xi_g'(0)\Lambda_g \equiv \Sigma_g$, respectively. Well-known examples of elliptical distributions include the multivariate normal, multivariate Student t, and contaminated normal distributions; see, e.g., Muirhead [9].

Let $\boldsymbol{X}_{g1}, \ldots, \boldsymbol{X}_{gn_g}$ be mutually independent copies of $\boldsymbol{X}_g$. We consider a test of the parallelism hypothesis

$$\mathcal{H}_{01} : \forall_{g \in \{1,\ldots,a-1\}} \ \boldsymbol{\mu}_g - \boldsymbol{\mu}_a = \gamma_g \mathbf{1}_p \quad \text{vs.} \quad \mathcal{A}_{01} : \neg\mathcal{H}_{01}. \tag{1}$$

Here, γ_g is an unknown real constant and $\mathbf{1}_p$ is a $p \times 1$ vector of 1's, i.e., $\mathbf{1}_p = (1, \ldots, 1)^\top$. We also consider tests of the flatness hypothesis

$$\mathcal{H}_{02} : \forall_{g \in \{1,\ldots,a\}} \ \mu_{g1} = \cdots = \mu_{gp} \quad \text{vs.} \quad \mathcal{A}_{02} : \neg\mathcal{H}_{02}, \tag{2}$$

and the level hypothesis

$$\mathcal{H}_{03} : \gamma_1 = \cdots = \gamma_{a-1} = 0 \quad \text{vs.} \quad \mathcal{A}_{03} : \neg\mathcal{H}_{03}. \tag{3}$$

Harrar and Kong [4] give expressions that are equivalent to hypotheses (1)–(3). Expression (1) is equivalent to

$$\widetilde{\mathcal{H}}_{01} : \boldsymbol{\mu}^\top K_{01} \boldsymbol{\mu} = 0 \quad \text{vs.} \quad \widetilde{\mathcal{A}}_{01} : \boldsymbol{\mu}^\top K_{01} \boldsymbol{\mu} > 0$$

with $K_{01} = P_a \otimes P_p$, where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_a^\top)^\top$ and $P_k = I_k - k^{-1}\mathbf{1}_k\mathbf{1}_k^\top$ for $k \in \{a, p\}$. Setting $K_{02} = (a^{-1}\mathbf{1}_a\mathbf{1}_a^\top) \otimes P_p$ and $K_{03} = D_a \otimes (p^{-1}\mathbf{1}_p\mathbf{1}_p^\top)$, we can also express hypotheses $\widetilde{\mathcal{H}}_{0x}$ with $x \in \{2, 3\}$ in the form

$$\widetilde{\mathcal{H}}_{0x} : \boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu} = 0 \quad \text{vs.} \quad \widetilde{\mathcal{A}}_{0x} : \boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu} > 0,$$

where $D_a = \mathrm{diag}(n_1, \ldots, n_a) - n_{(a)}^{-1}\boldsymbol{n}\boldsymbol{n}^\top$. Here, $\boldsymbol{n} = (n_1, \ldots, n_a)^\top$ and $n_{(a)} = n_1 + \cdots + n_a$.

Srivastava [12] derived the likelihood ratio test for hypotheses (1)–(3) for two normal populations. However, the likelihood ratio tests for (1) and (2) cannot be applied to situations where $n_{(a)} \ll p$, e.g., microarray data, even for normal populations with covariance homogeneity.

In profile analysis, Takahashi and Shutoh [13] considered approximate tests for hypotheses (1) and (2) for two normal populations with equal covariance matrices. Harrar and Kong [4] extended these tests to multi-group normal populations without assuming equal covariance matrices. They also obtained the approximate test for hypothesis (3) based on matching moments.

In parallel, the effect of non-normality in profile analysis has been investigated. Okamoto et al. [10] used a perturbation method to obtain the asymptotic expansions of the distributions of test statistics for elliptical populations. Maruyama [7] extended the results under more general conditions using a different method introduced by Kano [6]. Note that these results are derived as $n_{(a)} \to \infty$.

In this paper, we propose new approximate tests for hypotheses (1) and (2) for high-dimensional elliptical populations without assuming equal covariance matrices. We note that the rank of K_{03} is at most $a - 1$, i.e., it does not grow with p; accordingly, $(n_{(a)}, p)$ asymptotic considerations are not relevant in pursuing our primary interest, which is to test (1) and (2). To this end, we show that the asymptotic normality of the test statistics proposed by Harrar and Kong [4] holds when the underlying distribution is elliptical. An improved estimator of the asymptotic variance of these test statistics also enables us to propose new approximate tests for (1) and (2) for high-dimensional elliptical populations.

The remainder of this paper is organized as follows. Preliminary asymptotic results for approximate tests are presented in Section 2. Using these results, we construct approximate tests for (1) and (2) and derive the asymptotic power and size of these tests for elliptical populations in Section 3. In Section 4, the numerical accuracy of the proposed tests is investigated, and the results are illustrated with a short numerical example. Section 5 concludes the paper. Technical proofs are given in the Appendix.

2. Preliminary asymptotic results

We define an $a \times a$ non-random matrix

$$(R_a)_{ij} = \begin{cases} d_i & \text{if } i = j, \\ \psi\delta_i\delta_j & \text{if } i \neq j, \end{cases}$$

with $d_i, \delta_i, \psi \in \mathbb{R}$ for $i, j \in \{1, \ldots, a\}$. Then we consider the random variable

$$T = \overline{\boldsymbol{X}}^\top (R_a \otimes P_p)\overline{\boldsymbol{X}} - \sum_{g=1}^{a} \frac{d_g \mathrm{tr}(P_p S_g)}{n_g},$$

where $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_1^\top, \ldots, \overline{\boldsymbol{X}}_a^\top)^\top$ and

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

with $\overline{\boldsymbol{X}}_g = (\boldsymbol{X}_{g1} + \cdots + \boldsymbol{X}_{gn_g})/n_g$.

Remark 1. If $\psi = -1$, $d_i = 1 - 1/a$, and $\delta_i = 1/\sqrt{a}$ for all $i \in \{1, \ldots, a\}$, then T is the test statistic for $\mathcal{H}_{01}$. If $\psi = d_i = a^{-1}$ and $\delta_i = 1$ for all $i \in \{1, \ldots, a\}$, then T is the test statistic for $\mathcal{H}_{02}$.

Here, T is an unbiased estimator of $\boldsymbol{\mu}^\top(R_a \otimes P_p)\boldsymbol{\mu}$, i.e.,

$$\mathrm{E}(T) = \boldsymbol{\mu}^\top(R_a \otimes P_p)\boldsymbol{\mu}. \tag{4}$$

In addition, the variance of T is given by

$$\sigma^2 = \sum_{g=1}^{a} \frac{2d_g^2 \mathrm{tr}\{(P_p\Sigma_g)^2\}}{n_g(n_g-1)} + \sum_{g=2}^{a}\sum_{h=1}^{g-1} \frac{4\psi^2\delta_g^2\delta_h^2 \mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}{n_g n_h} + 4\boldsymbol{\mu}^\top (R_a\otimes P_p)\left\{\sum_{g=1}^{a}\frac{1}{n_g}(\boldsymbol{e}_g\boldsymbol{e}_g^\top)\otimes\Sigma_g\right\}(R_a\otimes P_p)\boldsymbol{\mu}, \quad (5)$$

where $\boldsymbol{e}_i$ denotes the ith basis vector.

Remark 2. If R_a is an idempotent matrix, then $R_a \otimes P_p$ is also an idempotent matrix, and $\boldsymbol{\mu}^\top(R_a\otimes P_p)\boldsymbol{\mu} = 0$ is equivalent to $(R_a\otimes P_p)\boldsymbol{\mu} = \mathbf{0}$. Thus, if R_a is an idempotent matrix and $\boldsymbol{\mu}^\top(R_a\otimes P_p)\boldsymbol{\mu} = 0$, then

$$\mathrm{E}(T) = 0,\ \sigma^2 = \sum_{g=1}^{a} \frac{2d_g^2 \mathrm{tr}\{(P_p\Sigma_g)^2\}}{n_g(n_g-1)} + \sum_{g=2}^{a}\sum_{h=1}^{g-1} \frac{4\psi^2\delta_g^2\delta_h^2 \mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}{n_g n_h}.$$

We investigate the asymptotic distribution of T for elliptical populations. Our primary objective in this section is to derive the asymptotic distribution of T under some assumptions.

For each $g \in \{1,\ldots,a\}$, let n_g be a function of p, i.e., $n_g = n_g(p)$. For any $g,h \in \{1,\ldots,a\}$, $i \in \{1,2\}$, let $\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^i\}$ be a function of p. We assume the following conditions:

(A1) For all $g,h \in \{1,\ldots,a\}$, $\lim_{p\to\infty} n_g(p) = \infty$, $0 < \lim_{p\to\infty} n_g(p)/n_h(p) < \infty$.

(A2) For each $g \in \{1,\ldots,a\}$, the kurtosis is finite, i.e., one has

$$\kappa_g = \frac{\mathrm{E}[\{(\boldsymbol{X}_g-\boldsymbol{\mu}_g)^\top\Sigma_g^{-1}(\boldsymbol{X}_g-\boldsymbol{\mu}_g)\}^2]}{p(p+2)} - 1 < \infty.$$

(A3) For all $g,h \in \{1,\ldots,a\}$, $\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}/\{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)\}^2 = o(1)$.

Examples of multivariate distributions satisfying Assumptions (A2) and (A3) include the following three, whose density is given here as a function of $z = \Sigma_g^{-1/2}(\boldsymbol{x}_g - \boldsymbol{\mu}_g) \in \mathbb{R}^p$:

(a) The multivariate normal distribution with density

$$f(z) = \frac{1}{(2\pi)^{p/2}}\exp(-z^\top z/2).$$

(b) The ϵ-contaminated normal distribution with density given, for any $\varepsilon \in [0,1]$ and $\eta \in (0,\infty)$, by

$$f(z) = (1-\epsilon)\times\frac{1}{(2\pi)^{p/2}}\exp(-z^\top z/2) + \epsilon\times\frac{1}{(2\pi\eta^2)^{p/2}}\exp\{-z^\top z/(2\eta^2)\}.$$

(c) The multivariate Student t distribution with $k \in \mathbb{N}$ degrees of freedom, with density

$$f(z) = \frac{\Gamma\{(k+p)/2\}}{\Gamma(k/2)(k\pi)^{p/2}}(1 + z^\top z/k)^{-(k+p)/2},$$

where Γ denotes Euler's gamma function.

These distributions satisfy Assumption (A2) with $\kappa_g = 0$ in case (a), $\kappa_g = 1 + \epsilon(\eta^4-1)/\{1+\epsilon(\eta^2-1)\}^2 - 1$ in case (b), and $\kappa_g = 2/(k-4)$ in case (c) when $k > 4$. Examples of covariance matrices that satisfy (A3) are those with compound symmetry. Thus if $\Sigma_g = (1-\rho_g)I_p + \rho_g(\mathbf{1}_p\mathbf{1}_p^\top)$, for all $g \in \{1,\ldots,a\}$ and $\rho_g \in (-1/(p-1), 1)$, then $\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}/\{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)\}^2 = 1/(p-1)$.

The following lemma provides the asymptotic normality of T under Assumptions (A1)–(A3). The lemma assures us that the asymptotic normality of the statistic T is maintained for an elliptical population.

Lemma 1. *Under Assumptions* (A1)–(A3), $T - \boldsymbol{\mu}^\top(R_a\otimes P_p)\boldsymbol{\mu}/\sigma \rightsquigarrow \mathcal{N}(0,1)$ *as* $p\to\infty$.

Proof. See Appendix A.2. □

3. Main results

3.1. Proposed test

In this subsection, we propose approximate tests using a normal approximation based on Lemma 1. The test statistics for hypotheses (1) and (2) are respectively given by

$$T_{01} = \overline{\boldsymbol{X}}^\top K_{01} \overline{\boldsymbol{X}} - \sum_{g=1}^{a} \left(1 - \frac{1}{a}\right) \frac{\mathrm{tr}(P_p S_g)}{n_g}, \quad T_{02} = \overline{\boldsymbol{X}}^\top K_{02} \overline{\boldsymbol{X}} - \sum_{g=1}^{a} \frac{\mathrm{tr}(P_p S_g)}{a n_g}.$$

These statistics are also used by Harrar and Kong [4]. From Eqs. (4), (5), and Remark 2, their expectation and variance for elliptical populations are summarized by the following equations. For $x \in \{1, 2\}$,

$$\mathrm{E}(T_{0x}) = \begin{cases} 0 & \text{under } \mathcal{H}_{0x}, \\ \boldsymbol{\mu}^\top K_{0x} \boldsymbol{\mu} \ (> 0) & \text{under } \mathcal{A}_{0x}. \end{cases} \qquad \mathrm{var}(T_{0x}) = \begin{cases} \sigma^2_{\mathcal{H}_{0x}} & \text{under } \mathcal{H}_{0x}, \\ \sigma^2_{\mathcal{A}_{0x}} & \text{under } \mathcal{A}_{0x}, \end{cases}$$

where, for $x \in \{1, 2\}$,

$$\sigma^2_{\mathcal{H}_{01}} = \sum_{g=1}^{a} \left(1 - \frac{1}{a}\right)^2 \frac{2\mathrm{tr}\{(P_p \Sigma_g)^2\}}{n_g(n_g - 1)} + \sum_{g=2}^{a} \sum_{h=1}^{g-1} \frac{4\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)}{a^2 n_g n_h}, \ \sigma^2_{\mathcal{H}_{02}} = \sum_{g=1}^{a} \frac{2\mathrm{tr}\{(P_p \Sigma_g)^2\}}{a^2 n_g(n_g - 1)} + \sum_{g=2}^{a} \sum_{h=1}^{g-1} \frac{4\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)}{a^2 n_g n_h}$$

and

$$\sigma^2_{\mathcal{A}_{0x}} = \sigma^2_{\mathcal{H}_{0x}} + 4\boldsymbol{\mu}^\top K_{0x} \left\{ \sum_{g=1}^{a} \frac{1}{n_g} (\boldsymbol{e}_g \boldsymbol{e}_g^\top) \otimes \Sigma_g \right\} K_{0x} \boldsymbol{\mu}.$$

In practice, it is necessary to estimate the asymptotic variance $\sigma^2_{\mathcal{H}_{0x}}$. Harrar and Kong [4] proposed to use

$$\widetilde{\mathrm{tr}\{(P_p \Sigma_g)^2\}} = \frac{(n_g - 1)^2}{(n_g + 1)(n_g - 2)} \left[\mathrm{tr}\{(P_p S_g)^2\} - \frac{\{\mathrm{tr}(P_p S_g)\}^2}{n_g - 1} \right], \tag{6}$$

$$\widehat{\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)} = \mathrm{tr}(P_p S_g P_p S_h). \tag{7}$$

Assuming that the underlying distribution is elliptical, these statistics have the following expectations:

$$\mathrm{E}[\widetilde{\mathrm{tr}\{(P_p \Sigma_g)^2\}}] = \mathrm{tr}\{(P_p \Sigma_g)^2\} + \frac{\kappa_g(n_g - 1)}{n_g(n_g + 1)} [\{\mathrm{tr}(P_p \Sigma_g)\}^2 + 2\mathrm{tr}\{(P_p \Sigma_g)^2\}], \quad \mathrm{E}\{\widehat{\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)}\} = \mathrm{tr}(P_p \Sigma_g P_p \Sigma_h).$$

Thus, the estimator (6) has a bias for elliptical populations except when $\kappa_g = 0$.

We use the same estimator of $\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)$ as Harrar and Kong [4] but a different estimator of $\mathrm{tr}\{(P_p \Sigma_g)^2\}$, which is defined as follows:

$$\widehat{\mathrm{tr}\{(P_p \Sigma_g)^2\}} = \frac{n_g - 1}{n_g(n_g - 2)(n_g - 3)} \left[(n_g - 1)(n_g - 2)\mathrm{tr}\{(P_p S_g)^2\} + \{\mathrm{tr}(P_p S_g)\}^2 - n_g M_g \right], \tag{8}$$

where

$$M_g = \frac{1}{n_g - 1} \sum_{i=1}^{n_g} \{(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)^\top P_p (\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)\}^2.$$

Some properties of the estimators (7) and (8) are summarized in the following lemma.

Lemma 2. *The estimators* $\widehat{\mathrm{tr}\{(P_p \Sigma_g)^2\}}$ *and* $\widehat{\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)}$ *are unbiased, rate consistent estimator, i.e., under Assumptions* (A1)–(A2),

$$\frac{\widehat{\mathrm{tr}\{(P_p \Sigma_g)^2\}}}{\mathrm{tr}\{(P_p \Sigma_g)^2\}} = 1 + o_p(1), \qquad \frac{\widehat{\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)}}{\mathrm{tr}(P_p \Sigma_g P_p \Sigma_h)} = 1 + o_p(1) \text{ as } p \to \infty.$$

Proof. See Appendix A.3. □

Remark 3. *If P_p is replaced by I_p, $\widehat{\mathrm{tr}\{(P_p\Sigma_g)^2\}}$ is the same estimator as that of Himeno and Yamada [5].*

Using Eqs. (7) and (8), one finds

$$\widehat{\sigma}^2_{\mathcal{H}_{01}} = \sum_{g=1}^{a}\left(1-\frac{1}{a}\right)^2 \frac{2\widehat{\mathrm{tr}\{(P_p\Sigma_g)^2\}}}{n_g(n_g-1)} + \sum_{g=2}^{a}\sum_{h=1}^{g-1}\frac{4\widehat{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}}{a^2 n_g n_h},\ \widehat{\sigma}^2_{\mathcal{H}_{02}} = \sum_{g=1}^{a}\frac{2\widehat{\mathrm{tr}\{(P_p\Sigma_g)^2\}}}{a^2 n_g(n_g-1)} + \sum_{g=2}^{a}\sum_{h=1}^{g-1}\frac{4\widehat{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}}{a^2 n_g n_h},$$

as the unbiased estimators of $\sigma^2_{\mathcal{H}_{01}}$ and $\sigma^2_{\mathcal{H}_{02}}$, respectively.

Since $\widehat{\sigma}_{\mathcal{H}_{0x}}/\sigma_{\mathcal{H}_{0x}} = 1 + o_p(1)$ as $p \to \infty$ under Assumptions (A1)–(A2), Lemma 1 and Slutsky's theorem allow us to conclude that the asymptotic null distribution is normal. In fact for $x \in \{1, 2\}$, under Assumptions (A1)–(A3), and $\mathcal{H}_{0x}$, one has, as $p \to \infty$,

$$T_{0x}/\widehat{\sigma}_{\mathcal{H}_{0x}} \rightsquigarrow \mathcal{N}(0, 1). \tag{9}$$

Based on Eq. (9), we propose the following approximate tests:

$$\text{Reject } \mathcal{H}_{01} \quad \Leftrightarrow \quad T_{01}/\widehat{\sigma}_{\mathcal{H}_{01}} \ge z_\alpha, \tag{10}$$

$$\text{Reject } \mathcal{H}_{02} \quad \Leftrightarrow \quad T_{02}/\widehat{\sigma}_{\mathcal{H}_{02}} \ge z_\alpha, \tag{11}$$

where z_α denotes the upper $100 \times \alpha\%$ percentile of the standard normal distribution, $\mathcal{N}(0, 1)$.

3.2. Asymptotic size and power

Under Assumptions (A1)–(A3), one can first deduce from Eq. (9) that as $p \to \infty$, the size of the tests (10) and (11) is given, for $x \in \{1, 2\}$, by

$$\Pr(T_{0x}/\widehat{\sigma}_{\mathcal{H}_{0x}} \ge z_\alpha \mid \mathcal{H}_{0x}) = \alpha + o(1).$$

Next, using Lemmas 1–2 can be used to prove the following theorem concerning the asymptotic power of tests (10) and (11).

Theorem 1. *For $x \in \{1, 2\}$, under Assumptions (A1)–(A3),*

$$\Pr\left(T_{0x}/\widehat{\sigma}_{\mathcal{H}_{0x}} \ge z_\alpha \mid \mathcal{A}_{0x}\right) = \Phi\left(\frac{\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}} - \frac{\sigma_{\mathcal{H}_{0x}}}{\sigma_{\mathcal{A}_{0x}}} z_\alpha\right) + o(1)$$

as $p \to \infty$, where Φ denotes the cumulative distribution function (CDF) of $\mathcal{N}(0, 1)$.

Proof. See Appendix A.4. □

Thus if the difference between $\mathcal{H}_{0x}$ and $\mathcal{A}_{0x}$ is not too small, in that $\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}$ is of the same order as $\sigma_{\mathcal{A}_{0x}}$ or of a higher order, the test will be powerful. Conversely, if the difference between $\mathcal{H}_{0x}$ and $\mathcal{A}_{0x}$ is so small that $\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}$ is of a lower order than $\sigma_{\mathcal{A}_{0x}}$, the test will not be powerful and cannot distinguish $\mathcal{H}_{0x}$ from $\mathcal{A}_{0x}$.

4. Simulation and real example

4.1. Simulation

In this section, we perform Monte Carlo simulations for some selected parameters in order to verify the superiority of our test as compared to Harrar and Kong's tests for (1) and (2) when the kurtosis parameter is not 0.

In our simulation, we compare the empirical size and power of the proposed tests and Harrar and Kong's tests. We generated data from the following model:

$$\forall_{i\in\{1,\dots,n_g\}} \quad \boldsymbol{X}_{gi} = \Sigma_g^{1/2}\boldsymbol{Z}_{gi} + \boldsymbol{\mu}_g, \tag{12}$$

where, for all $g \in \{1, \dots, a\}$,

$$\boldsymbol{\mu}_g = (g-1)\mathbf{1}_p, \quad \Sigma_g = (1-0.1g)I_p + (0.1g)\mathbf{1}_p\mathbf{1}_p^\top.$$

We take $a \in \{2, 3, 4, 5\}$. Under this model, the null hypotheses $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ hold. For the distribution of $\boldsymbol{Z}_{gj}$ in (12), we consider the following options:

(i) Multivariate normal distribution;

(ii) Contaminated normal distribution with $\epsilon = 0.1$ and $\eta = 5$;

(iii) Multivariate Student t distribution with 5 degrees of freedom.

One then has $\kappa_g = 0$ in case (i), $\kappa_g \approx 4.48$ in case (ii), and $\kappa_g = 2$ in case (iii). The sizes calculated with 10,000 replicates are listed in Tables 1 and 2. Here, the nominal significance level is $\alpha = 0.05$.

Please insert Tables 1 and 2 approximately here.

The empirical sizes of our proposed test and Harrar and Kong's test are presented in Tables 1 and 2, respectively. As can be seen in Table 1, our approximate test is within 0.01 from the nominal significance level $\alpha = 0.05$ regardless of the population distribution setting when the dimension p is 200 or 400. From Table 2, we see that this is also true of Harrar and Kong's test but only when the population distribution is a multivariate normal distribution. Indeed, the empirical size of Harrar and Kong's test is significantly less than the nominal significance level when the distribution of $\boldsymbol{Z}_{gj}$ is (ii) or (iii).

For the alternative hypothesis, we choose $\boldsymbol{\mu}_g$ in (12) as follows:

$$\boldsymbol{\mu}_g = \begin{cases} (g-1)\mathbf{1}_p & \text{if } g \in \{1,\ldots,a-1\}, \\ (a-1)(\mathbf{1}^\top_{\lfloor 0.99p \rfloor},\ 0.7 \times \mathbf{1}^\top_{p-\lfloor 0.99p \rfloor})^\top & \text{if } g = a, \end{cases}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The settings of a, p, $\boldsymbol{n}$, the covariance matrix and the distribution of $\boldsymbol{Z}_{gj}$ are the same as for the null hypothesis. Under these models, $\mathcal{H}_{01}$ and $\mathcal{H}_{02}$ do not hold. The estimated power based on 10,000 replicates is displayed in Tables 3 and 4. Here again, the nominal significance level is $\alpha = 0.05$.

The empirical power of our test and Harrar and Kong's original procedure are presented in Tables 3 and 4, respectively. In Table 3, the asymptotic approximation of the power of our proposed test, viz.

$$\text{approx} = \Phi\left(\frac{\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}} - \frac{\sigma_{\mathcal{H}_{0x}}}{\sigma_{\mathcal{A}_{0x}}} z_\alpha\right)$$

is also calculated in each setting. This approximation is based on the result of Theorem 1.

Please insert Tables 3 and 4 approximately here.

It can be seen that these asymptotic approximations are accurate. Therefore, the power of the proposed tests can be roughly estimated as the value obtained by dividing the distance between the null hypothesis and the alternative hypothesis by the variance of T_{0x} under the alternative hypothesis. From Tables 3 and 4, we also see that the powers of the two tests are almost the same when $\boldsymbol{Z}_{gj}$ follows a multivariate normal distribution. In contrast, it can be seen that the proposed test is more powerful than Harrar and Kong's test when the distribution of $\boldsymbol{Z}_{gj}$ is (b) or (c).

From these simulation results, we can see that our tests are more robust against the effects of non-normality as compared to Harrar and Kong's tests. The difference between Harrar and Kong's tests and our test appears in the estimator of $\mathrm{tr}\{(P_g\Sigma_g)^2\}$. To illustrate this point, we compared the ratios

$$\mathrm{tr}\{\widetilde{(P_p\Sigma_g)^2}\}/\mathrm{tr}\{(P_g\Sigma_g)^2\} \quad \text{and} \quad \mathrm{tr}\{\widehat{(P_p\Sigma_g)^2}\}/\mathrm{tr}\{(P_g\Sigma_g)^2\} \tag{13}$$

under model (12) with $\boldsymbol{\mu}_g = \mathbf{0}$ and $\Sigma_g = (1-0.5)I_p + 0.5\,\mathbf{1}_p\mathbf{1}_p^\top$. The biases of these estimators are calculated using 10,000 replicates in each setting. The biases associated with the quantities in (13) are presented in Table 5. The settings of p and n_g can be checked in Table 5.

Please insert Table 5 approximately here.

From Table 5, we can see that $\mathrm{tr}\{\widetilde{(P_p\Sigma_g)^2}\}$ overestimates $\mathrm{tr}\{(P_p\Sigma_g)^2\}$ when the distribution of $\boldsymbol{Z}_{gj}$ is (b) or (c). This overestimation contributes to decreasing the value of the test statistic. Since decreasing the value of the test statistic makes it more difficult to reject the null hypothesis, we expect that the size for Harrar and Kong's test will be lower than the nominal significance level. This observation is consistent with the results of Table 2.

In contrast, we can observe that $\mathrm{tr}\{\widehat{(P_p\Sigma_g)^2}\}$ is almost unaffected by change in distribution. Therefore, changing the estimator of $\mathrm{tr}\{(P_p\Sigma_g)^2\}$ is helpful in all cases considered.

4.2. *Concrete example*

As an illustration of tests (10) and (11), we applied them to a dataset analyzed by Takahashi and Shutoh [13]. The data consist of $a = 2$ groups, with a body weight of $n_g = 10$ rats for each group. The weights of the 20 rats were measured every week for $p = 22$ weeks.

We applied our tests to the parallelism hypothesis $\mathcal{H}_{01}$ and flatness hypothesis $\mathcal{H}_{02}$ at level $\alpha = 0.05$. Given that

$$T_{01}/\widehat{\sigma}_{\mathcal{H}_{01}} \approx -0.657 < z_\alpha \approx 1.6449 \quad \text{and} \quad T_{02}/\widehat{\sigma}_{\mathcal{H}_{02}} \approx 250.264 > z_\alpha \approx 1.6449,$$

the parallelism hypothesis $\mathcal{H}_{01}$ cannot be rejected at the 5% level, but the flatness hypothesis $\mathcal{H}_{02}$ can.

5. Discussion and conclusion

In this paper, we proposed new approximate tests for the parallelism and flatness hypotheses in profile analysis for high-dimensional elliptical populations with unequal covariance matrices, and we derived their asymptotic power and size. We showed that the asymptotic size of the proposed tests is close to the nominal value. However, Harrar and Kong's approximate tests do not enjoy this property for all elliptical populations, because the unbiasedness of their estimator of asymptotic variance depends on kurtosis. Furthermore, we found that the asymptotic power depends on the value obtained by dividing the distance between the null and alternative hypotheses by the variance under the alternative hypothesis.

We compared the proposed tests and Harrar and Kong's tests numerically in simulation studies. We found that our tests and Harrar and Kong's tests had approximately the same accuracy when the population is multivariate normal, and we showed that our tests are superior to those of Harrar and Kong for other elliptically contoured distributions. We also confirmed that this superiority is attributable to the asymptotic variance estimator.

In our data illustration, and more generally, it would be desirable to check first whether it is legitimate to assume an elliptical distribution. This could be accomplished, e.g., by extending the method proposed by Batsidis and Zografos [1] to high-dimensional settings. Alternatively, one could try to extend our procedures to a wider class of distributions than the elliptical family. To this end, it would first be necessary to investigate situations in which the symmetry of the distribution is not assumed, such as when a skew elliptical distribution is used. This would make it more difficult to estimate the asymptotic variance, however. This may be the object of future work.

A. Appendix

A.1. *Some moments*

Lemma A.1 Let be $\boldsymbol{X} \sim C_p(\xi, \mathbf{0}, \Lambda)$, and let A and B be $p \times p$ symmetric real matrices. The following expressions are then valid.

(i) $\mathrm{E}(\boldsymbol{X}^\top A\boldsymbol{X}) = \mathrm{tr}(A\Sigma)$.

(ii) $\mathrm{E}(\boldsymbol{X}^\top A\boldsymbol{X}\boldsymbol{X}^\top B\boldsymbol{X}) = (\kappa + 1)\{\mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma) + 2\mathrm{tr}(A\Sigma B\Sigma)\}$.

(iii) $\mathrm{var}(\boldsymbol{X}^\top A\boldsymbol{X}) = \kappa\{\mathrm{tr}(A\Sigma)\}^2 + 2(\kappa + 1)\mathrm{tr}\{(A\Sigma)^2\}$.

(iv) $\mathrm{cov}(\boldsymbol{X}^\top A\boldsymbol{X}, \boldsymbol{X}^\top B\boldsymbol{X}) = \kappa\mathrm{tr}(A\Sigma)\mathrm{tr}(B\Sigma) + 2(\kappa + 1)\mathrm{tr}(A\Sigma B\Sigma)$,

where

$$\Sigma \equiv -2\xi'(0)\Lambda, \quad \kappa \equiv \frac{\xi''(0)}{\{\xi'(0)\}^2} - 1 \quad \left(= \frac{\mathrm{E}\{(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^2\}}{p(p+2)} - 1\right).$$

Proof. See Mathai et al. [8]. □

A.2. Proof of Lemma 1

We define $\boldsymbol{Y}_{gi} = \boldsymbol{X}_{gi} - \boldsymbol{\mu}_g$ for $g \in \{1, \dots, a\}$ and $i \in \{1, \dots, n_g\}$. Note that $\boldsymbol{Y}_{gi}$ is distributed according to an elliptical distribution with $\mathrm{E}(\boldsymbol{Y}_{gi}) = \boldsymbol{0}$ and $\mathrm{var}(\boldsymbol{Y}_{gi}) = \Sigma_g$. Then the statistic T can be rewritten as

$$T = \sum_{g=1}^{a} \frac{d_g}{n_g(n_g-1)} \sum_{\substack{i=1 \\ i\neq j}}^{n_g} \boldsymbol{Y}_{gi}^\top P_p \boldsymbol{Y}_{gj} + \sum_{g\neq h}^{a} \psi\delta_g\delta_h \overline{\boldsymbol{Y}}_g^\top P_p \overline{\boldsymbol{Y}}_h + 2\boldsymbol{\mu}^\top (R_a \otimes P_p)\overline{\boldsymbol{Y}} + \boldsymbol{\mu}^\top (R_a \otimes P_p)\boldsymbol{\mu},$$

where $\overline{\boldsymbol{Y}} = (\overline{\boldsymbol{Y}}_1^\top, \dots, \overline{\boldsymbol{Y}}_a^\top)^\top$. We note that $\boldsymbol{\mu}^\top (R_a \otimes P_p)\overline{\boldsymbol{Y}} = 0$ and $\boldsymbol{\mu}^\top (R_a \otimes P_p)\boldsymbol{\mu} = 0$ if $(R_a \otimes P_p)\boldsymbol{\mu} = \boldsymbol{0}$. That is, the distribution of T does not depend on $\boldsymbol{\mu}$ so long as $(R_a \otimes P_p)\boldsymbol{\mu} = \boldsymbol{0}$.

Let $n_{(0)} = 0$, $n_{(g)} = n_1 + \cdots + n_g$ for $g \in \{1, \dots, a\}$, and set $i' = i - n_{(g-1)}$. We define

$$\varepsilon_i = \frac{2}{\sigma n_g(n_g-1)} \boldsymbol{Y}_{gi'}^\top P_p \boldsymbol{a}_{gi'}$$

for $g \in \{1, \dots, a\}$ and $i \in \{n_{(g-1)} + 1, \dots, n_{(g)}\}$. Here,

$$\boldsymbol{a}_{gi'} = \mathbf{1}(i' \geq 2) d_g \sum_{j=1}^{i'-1} \boldsymbol{Y}_{gj} + \mathbf{1}(g \geq 2)(n_g - 1)\psi\delta_g \sum_{h=1}^{g-1} \delta_h \overline{\boldsymbol{Y}}_h + (n_g-1)\psi\delta_g \sum_{h=1}^{a} \delta_h \boldsymbol{\mu}_h,$$

where $\mathbf{1}$ denotes an indicator function. Then,

$$\frac{1}{\sigma}\{T - \boldsymbol{\mu}^\top (R_a \otimes P_p)\boldsymbol{\mu}\} = \sum_{i=1}^{n_{(a)}} \varepsilon_i.$$

Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and let $\mathcal{F}_i$ for $i \in \mathbb{N}$ be the σ-algebra generated by the random variables $(\varepsilon_1, \dots, \varepsilon_i)$. Then we find that $\mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_\infty$ and $\mathrm{E}(\varepsilon_i \mid \mathcal{F}_{i-1}) = 0$. Thus, (ε_i) is a martingale difference sequence.

We show the asymptotic normality of $\varepsilon_1 + \cdots + \varepsilon_{n_{(a)}}$ by adapting the martingale difference central limit theorem; see, e.g., [3, 11]. It is necessary to check the following two conditions to apply this theorem:

$$\text{(I)} : \sum_{i=1}^{n_{(a)}} \mathrm{E}(\varepsilon_i^2 | \mathcal{F}_{i-1}) = 1 + o_p(1) \text{ as } p \to \infty; \qquad \text{(II)} : \sum_{i=1}^{n_{(a)}} \mathrm{E}(\varepsilon_i^4) = o(1) \text{ as } p \to \infty.$$

First, we check condition (I). We rewrite

$$\sum_{i=1}^{n_{(a)}} \mathrm{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1}) = 1 + \sum_{j=1}^{7} A_j,$$

where

$$\begin{aligned}
A_1 &= \sum_{g=1}^{a} \frac{4d_g^2}{\sigma^2 n_g^2(n_g-1)^2} \sum_{i=1}^{n_g} (n_g - i)[\boldsymbol{Y}_{gi}^\top P_p \Sigma_g P_p \boldsymbol{Y}_{gi} - \mathrm{tr}\{(P_p\Sigma_g)^2\}], \\
A_2 &= \sum_{g=1}^{a} \frac{8d_g^2}{\sigma^2 n_g^2(n_g-1)^2} \sum_{i=2}^{n_g} \sum_{j=1}^{i-1} (n_g - i)\boldsymbol{Y}_{gi}^\top P_p \Sigma_g P_p \boldsymbol{Y}_{gj}, \\
A_3 &= \sum_{g=2}^{a} \sum_{h=1}^{g-1} \frac{8d_g\psi\delta_g\delta_h}{\sigma^2 n_g^2(n_g-1)} \sum_{i=1}^{n_g} (n_g - i)\boldsymbol{Y}_{gi}^\top P_p \Sigma_g P_p \overline{\boldsymbol{Y}}_h,
\end{aligned}$$

$$A_4 = \sum_{g=1}^{a}\sum_{h=1}^{a} \frac{8d_g\psi\delta_g\delta_h}{\sigma^2 n_g^2(n_g-1)} \sum_{i=1}^{n_g}(n_g-i)\boldsymbol{Y}_{gi}^\top P_p\Sigma_g P_p\boldsymbol{\mu}_h,$$

$$A_5 = \sum_{g=2}^{a}\sum_{h=1}^{g-1} \frac{4\psi^2\delta_g^2\delta_h^2}{\sigma^2 n_g}\left\{\overline{\boldsymbol{Y}}_h^\top P_p\Sigma_g P_p\overline{\boldsymbol{Y}}_h - \frac{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}{n_h}\right\},$$

$$A_6 = \sum_{g=3}^{a}\sum_{h=2}^{g-1}\sum_{\ell=1}^{h-1} \frac{8\psi^2\delta_g^2\delta_h\delta_\ell}{n_g}\overline{\boldsymbol{Y}}_h^\top P_p\Sigma_g P_p\overline{\boldsymbol{Y}}_\ell,$$

$$A_7 = \sum_{g=2}^{a}\sum_{h=1}^{g-1}\sum_{\ell=1}^{a} \frac{8\psi^2\delta_g^2\delta_h\delta_\ell}{\sigma^2 n_g}\overline{\boldsymbol{Y}}_h^\top P_p\Sigma_g P_p\boldsymbol{\mu}_\ell.$$

It is straightforward to show that $\mathrm{E}(A_1+\cdots+A_7)=0$. Hölder's inequality yields

$$\mathrm{var}\left\{\sum_{i=1}^{n_{(a)}}\mathrm{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1})\right\} = \mathrm{E}\left\{\left(\sum_{i=1}^{7}A_i\right)^2\right\} \le 7\sum_{i=1}^{7}\mathrm{E}(A_i^2). \tag{14}$$

The expectations $\mathrm{E}(A_1^2)$ through $\mathrm{E}(A_7^2)$ are evaluated as follows:

$$\mathrm{E}(A_1^2) = O\left(\sum_{g=1}^{a}\frac{1}{n_g}\right) = o(1), \tag{15}$$

$$\mathrm{E}(A_2^2) = O\left(\sum_{g=1}^{a}\frac{\mathrm{tr}\{(P_p\Sigma_g)^4\}}{[\mathrm{tr}\{(P_p\Sigma_g)^2\}]^2}\right) = o(1), \tag{16}$$

$$\mathrm{E}(A_3^2) = O\left(\sum_{g=2}^{a}\sum_{h=1}^{g-1}\sqrt{\frac{\mathrm{tr}\{(P_p\Sigma_g)^4\}}{[\mathrm{tr}\{(P_p\Sigma_g)^2\}]^2}}\sqrt{\frac{\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}}{\{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)\}^2}}\right) = o(1), \tag{17}$$

$$\mathrm{E}(A_4^2) = O\left(\sum_{g=1}^{a}\sqrt{\frac{\mathrm{tr}\{(P_p\Sigma_g)^4\}}{[\mathrm{tr}\{(P_p\Sigma_g)^2\}]^2}}\right) = o(1), \tag{18}$$

$$\mathrm{E}(A_5^2) = O\left(\sum_{g=2}^{a}\sum_{h=1}^{g-1}\left[\frac{1}{n_h} + \frac{\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}}{\{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)\}^2}\right]\right) = o(1), \tag{19}$$

$$\mathrm{E}(A_6^2) = O\left(\sum_{g=3}^{a}\sum_{h=2}^{g-1}\sum_{\ell=1}^{h-1}\frac{\sqrt{\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_\ell)^2\}}}{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)\mathrm{tr}(P_p\Sigma_g P_p\Sigma_\ell)}\right) = o(1), \tag{20}$$

$$\mathrm{E}(A_7^2) = O\left(\sum_{g=2}^{a}\sum_{h=1}^{g-1}\frac{\sqrt{\mathrm{tr}\{(P_p\Sigma_g P_p\Sigma_h)^2\}}}{\mathrm{tr}(P_p\Sigma_g P_p\Sigma_h)}\right) = o(1). \tag{21}$$

Substituting (15)–(21) into (14) yields

$$\mathrm{var}\left\{\sum_{i=1}^{n_{(a)}}\mathrm{E}(\varepsilon_i^2 \mid \mathcal{F}_{i-1})\right\} = o(1).$$

Condition (I) follows.

Next, we check Condition (II). Define

$$\varepsilon_i^{(1)} = \frac{2\mathbf{1}(i'\ge 2)d_g}{\sigma n_g(n_g-1)}\boldsymbol{Y}_{gi'}^\top P_p\sum_{j=1}^{i'-1}\boldsymbol{Y}_{gj}, \quad \varepsilon_i^{(2)} = \frac{2\psi\delta_g\mathbf{1}(g\ge 2)}{\sigma n_g}\boldsymbol{Y}_{gi'}^\top P_p\sum_{h=1}^{\ell-1}\delta_h\overline{\boldsymbol{Y}}_h, \quad \varepsilon_i^{(3)} = \frac{2\psi\delta_g}{\sigma n_g}\boldsymbol{Y}_{gi'}^\top P_p\sum_{h=1}^{a}\delta_h\boldsymbol{\mu}_h.$$

Then, Hölder's inequality yields

$$\sum_{i=1}^{n_{(a)}} \mathrm{E}(\varepsilon_i^4) = \sum_{g=1}^{a} \sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}\left\{\left(\sum_{j=1}^{3} \varepsilon_i^{(j)}\right)^4\right\} \leq \sum_{g=1}^{a} \sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}\left(3^3 \sum_{j=1}^{3} \varepsilon_i^{(j)^4}\right) = 3^3 \sum_{g=1}^{a} \sum_{i=n_{(g-1)}+1}^{n_{(g)}} \sum_{j=1}^{3} \mathrm{E}\left(\varepsilon_i^{(j)^4}\right).$$

Thus, it is sufficient to show that $\sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}(\varepsilon_i^{(j)^4}) = o(1)$ for $j \in \{1,2,3\}$. We have the following expectations:

$$\sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}\left(\varepsilon_i^{(1)^4}\right) = O(n_g^{-1}), \quad \sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}\left(\varepsilon_i^{(2)^4}\right) = O(n_g^{-1}), \quad \sum_{i=n_{(g-1)}+1}^{n_{(g)}} \mathrm{E}\left(\varepsilon_i^{(3)^4}\right) = O(n_g^{-1}).$$

The above results complete the proof of (II). □

A.3. Proof of Lemma 2

First, we show the unbiasedness and consistency of $\widehat{\mathrm{tr}(P_p\Sigma_g)^2}$. From Lemma A.1, it follows that

$$\begin{aligned}
\mathrm{E}[\mathrm{tr}\{(P_pS_g)^2\}] &= \frac{\kappa_g+1}{n_g}\left[\{\mathrm{tr}(P_p\Sigma_g)\}^2 + 2\mathrm{tr}\{(P_p\Sigma_g)^2\}\right] + \frac{n_g^2-2n_g+2}{n_g(n_g-1)}\mathrm{tr}\{(P_p\Sigma_g)^2\} + \frac{1}{n_g(n_g-1)}\{\mathrm{tr}(P_p\Sigma_g)\}^2,\\
\mathrm{E}[\{\mathrm{tr}(P_pS_g)\}^2] &= \frac{\kappa_g+1}{n_g}\left[\{\mathrm{tr}(P_p\Sigma_g)\}^2 + 2\mathrm{tr}\{(P_p\Sigma_g)^2\}\right] + \frac{2}{n_g(n_g-1)}\mathrm{tr}\{(P_p\Sigma_g)^2\} + \frac{n_g-1}{n_g}\{\mathrm{tr}(P_p\Sigma_g)\}^2,\\
\mathrm{E}(M_g) &= \frac{(\kappa_g+1)(n_g^2-3n_g+3)}{n_g^2}\left[\{\mathrm{tr}(P_p\Sigma_g)\}^2 + 2\mathrm{tr}\{(P_p\Sigma_g)^2\}\right] + \frac{4n_g-6}{n_g^2}\mathrm{tr}\{(P_p\Sigma_g)^2\} + \frac{2n_g-3}{n_g^2}\{\mathrm{tr}(P_p\Sigma_g)\}^2.
\end{aligned}$$

Then we solve simultaneously for $\mathrm{tr}\{(P_p\Sigma_g)^2\}$, $\{\mathrm{tr}(P_p\Sigma_g)\}^2$ and κ_g. The solutions of the simultaneous equations can be obtained easily. We find

$$\mathrm{tr}\{(P_p\Sigma_g)^2\} = \frac{n_g-1}{n_g(n_g-2)(n_g-3)}\left\{(n_g-1)(n_g-2)\mathrm{E}[\mathrm{tr}\{(P_pS_g)^2\}] + \mathrm{E}[\{\mathrm{tr}(P_pS_g)\}^2] - n_g\mathrm{E}(M_g)\right\}.$$

Thus, the unbiased estimator of $\mathrm{tr}\{(P_p\Sigma_g)^2\}$ is $\mathrm{tr}\{\widehat{(P_p\Sigma_g)}^2\}$. Furthermore, the variance of $\mathrm{tr}\{\widehat{(P_p\Sigma_g)}^2\}$ is

$$\mathrm{var}[\mathrm{tr}\{\widehat{(P_p\Sigma_g)}^2\}] = O\left(\frac{1}{n_g}\left[\mathrm{tr}\{(P_p\Sigma_g)^2\}\right]^2\right).$$

Thus under Assumptions (A1)–(A2), as $p \to \infty$,

$$\mathrm{tr}\{\widehat{(P_p\Sigma_g)}^2\}/\mathrm{tr}\{(P_p\Sigma_g)^2\} = 1 + o_p(1).$$

Next, we show the unbiasedness and consistency of $\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)$. We can rewrite the estimator $\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)$ as

$$\begin{aligned}
\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h) &= \frac{1}{n_gn_h}\sum_{i=1}^{n_g}\sum_{j=1}^{n_h}(\boldsymbol{Y}_{gi}^\top P_p\boldsymbol{Y}_{hj})^2 - \frac{1}{n_gn_h(n_h-1)}\sum_{i=1}^{n_g}\sum_{\substack{j,k=1\\ j\neq k}}^{n_h}\boldsymbol{Y}_{gi}^\top P_p\boldsymbol{Y}_{hj}\boldsymbol{Y}_{gi}^\top P_p\boldsymbol{Y}_{hk}\\
&\quad -\frac{1}{n_g(n_h-1)n_h}\sum_{\substack{i,j=1\\ i\neq j}}^{n_g}\sum_{k=1}^{n_h}\boldsymbol{Y}_{gi}^\top P_p\boldsymbol{Y}_{hk}\boldsymbol{Y}_{gj}^\top P_p\boldsymbol{Y}_{hk} + \frac{1}{n_gn_h(n_g-1)(n_h-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{n_g}\sum_{\substack{k,\ell=1\\ k\neq\ell}}^{n_h}\boldsymbol{Y}_{gi}^\top P_p\boldsymbol{Y}_{hk}\boldsymbol{Y}_{gj}^\top P_p\boldsymbol{Y}_{h\ell}.
\end{aligned}$$

Using Lemma A.1, we get $\mathrm{E}\{\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)\} = \mathrm{tr}(P_p\Sigma_gP_p\Sigma_h)$. Furthermore, the variance of $\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)$ is

$$\mathrm{var}[\{\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)\}^2] = O\left(\frac{n_g+n_h}{n_gn_h}\{\mathrm{tr}(P_p\Sigma_gP_p\Sigma_h)\}^2\right).$$

Thus under Assumptions (A1)–(A2), as $p \to \infty$,

$$\frac{\mathrm{tr}(\widehat{P_p\Sigma_gP_p}\Sigma_h)}{\mathrm{tr}(P_p\Sigma_gP_p\Sigma_h)} = 1 + o_p(1).$$

This completes the proof of Lemma 2. □

A.4. Proof of Theorem 1

We assume $\mathcal{A}_{0x}$. From Lemma 3.1 and $\sigma^2_{\mathcal{H}_{0x}}/\sigma^2_{\mathcal{A}_{0x}} < 1$, under Assumptions (A1)–(A2), $\widehat{\sigma}^2_{\mathcal{H}_{0x}}/\sigma^2_{\mathcal{A}_{0x}} = \sigma^2_{\mathcal{H}_{0x}}/\sigma^2_{\mathcal{A}_{0x}} + o_p(1)$ as $p \to \infty$. Thus, under (A1) and (A2),

$$\Pr(T_{0x}/\widehat{\sigma}_{\mathcal{H}_{0x}} \geq z_\alpha) = \Pr\left(\frac{T_{0x} - \boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}} \geq \frac{\sigma_{\mathcal{H}_{0x}}}{\sigma_{\mathcal{A}_{0x}}} z_\alpha - \frac{\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}}\right) + o(1)$$

as $p \to \infty$. Furthermore, from Lemma 1, under Assumptions (A1)–(A3),

$$\Pr\left(\frac{T_{0x} - \boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}} \geq \frac{\sigma_{\mathcal{H}_{0x}}}{\sigma_{\mathcal{A}_{0x}}} z_\alpha - \frac{\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}}\right) = \Phi\left(\frac{\boldsymbol{\mu}^\top K_{0x}\boldsymbol{\mu}}{\sigma_{\mathcal{A}_{0x}}} - \frac{\sigma_{\mathcal{H}_{0x}}}{\sigma_{\mathcal{A}_{0x}}} z_\alpha\right) + o(1)$$

as $p \to \infty$. Combining these two equations yields the theorem. □

Acknowledgments. Thanks are due to Mr. Ogawa for the Monte Carlo simulation used to obtain the numerical results. The author is also grateful to 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 author's research was supported in part by a Grant-in-Aid for Young Scientists (B) (17K14238) from the Japan Society for the Promotion of Science.

References

References

[1] A. Batsidis, K. Zografos, A necessary test of fit of specific elliptical distributions based on an estimator of Song's measure, J. Multivariate Anal. 113 (2013) 91–105.

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

[3] P. Hall, C.C. Heyde, Martingale Limit Theory and its Application, Academic Press, New York, 1980.

[4] S.W. Harrar, X. Kong, High-dimensional multivariate repeated measures analysis with unequal covariance matrices, J. Multivariate Anal. 145 (2016) 1–21.

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

[6] Y. Kano, An asymptotic expansion of the distribution of Hotelling's T^2-statistic under general distributions, Amer. J. Math. Manag. Sci., 15 (1995) 317–341.

[7] Y. Maruyama, Asymptotic expansions of the null distributions of some test statistics for profile analysis in general conditions, J. Statist. Plann. Inference 137 (2007) 506–526.

[8] A.M. Mathai, S.B. Provost, T. Hayakawa, Bilinear Forms and Zonal Polynomials, Springer, New York, 1995.

[9] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.

[10] N. Okamoto, N. Miura, T. Seo, On the distributions of some test statistics for profile analysis in elliptical populations, Amer. J. Math. Manag. Sci. 26 (2006) 1–31.

[11] A.N. Shiryaev, Probability, 2nd Ed., Springer, Berlin, 1984.

[12] M.S. Srivastava, Profile analysis of several groups, Commun. Stat. Theory Meth. 16 (1987) 909–926.

[13] S. Takahashi, N. Shutoh, Tests for parallelism and flatness hypotheses of two mean vectors in high-dimensional settings, J. Stat. Comput. Simul. 86 (2016) 1150–1165.

Table 1: The empirical size of proposed test

p	$\boldsymbol{n}^\top$	Parallelism			Flatness		
		(i)	(ii)	(iii)	(i)	(ii)	(iii)
	$(25, 25)$	0.056	0.059	0.053	0.054	0.062	0.056
100	$(10, 20, 20)$	0.075	0.062	0.068	0.060	0.062	0.062
	$(10, 10, 10, 20)$	0.076	0.072	0.069	0.062	0.060	0.054
	$(10, 10, 10, 10, 10)$	0.064	0.065	0.064	0.053	0.055	0.074
	$(50, 50)$	0.057	0.058	0.057	0.053	0.058	0.058
200	$(20, 40, 40)$	0.057	0.055	0.059	0.057	0.060	0.056
	$(20, 20, 20, 40)$	0.057	0.059	0.056	0.058	0.059	0.055
	$(40, 40, 40, 40, 40)$	0.058	0.053	0.059	0.056	0.057	0.054
	$(100, 100)$	0.053	0.051	0.057	0.052	0.052	0.057
400	$(40, 80, 80)$	0.056	0.053	0.052	0.055	0.056	0.054
	$(40, 40, 40, 80)$	0.054	0.054	0.057	0.054	0.053	0.054
	$(40, 40, 40, 40, 40)$	0.052	0.056	0.053	0.056	0.057	0.055

Table 2: The empirical size of Harrar and Kong's test

p	$\boldsymbol{n}^\top$	Parallelism			Flatness		
		(i)	(ii)	(iii)	(i)	(ii)	(iii)
	$(25, 25)$	0.057	0.000	0.006	0.054	0.000	0.017
100	$(10, 20, 20)$	0.071	0.000	0.005	0.062	0.000	0.010
	$(10, 10, 10, 20)$	0.075	0.000	0.003	0.056	0.000	0.007
	$(10, 10, 10, 10, 10)$	0.064	0.000	0.001	0.055	0.000	0.011
	$(50, 50)$	0.057	0.000	0.005	0.053	0.000	0.008
200	$(20, 40, 40)$	0.057	0.000	0.001	0.057	0.000	0.005
	$(20, 20, 20, 40)$	0.057	0.000	0.000	0.058	0.000	0.005
	$(40, 40, 40, 40, 40)$	0.058	0.000	0.000	0.056	0.000	0.006
	$(100, 100)$	0.053	0.000	0.004	0.052	0.000	0.005
400	$(40, 80, 80)$	0.056	0.000	0.007	0.055	0.000	0.002
	$(40, 40, 40, 80)$	0.054	0.000	0.000	0.054	0.000	0.002
	$(40, 40, 40, 40, 40)$	0.052	0.000	0.000	0.057	0.000	0.005

Table 3: The empirical power of proposed test

p	$\boldsymbol{n}^\top$	Parallelism (i)	Parallelism (ii)	Parallelism (iii)	Parallelism approx	Flatness (i)	Flatness (ii)	Flatness (iii)	Flatness approx
	$(25, 25)$	0.069	0.076	0.073	0.063	0.071	0.076	0.072	0.063
100	$(10, 20, 20)$	0.093	0.104	0.094	0.078	0.079	0.087	0.084	0.072
	$(10, 10, 10, 20)$	0.117	0.140	0.126	0.102	0.093	0.103	0.097	0.082
	$(10, 10, 10, 10, 10)$	0.166	0.201	0.167	0.149	0.107	0.111	0.107	0.096
	$(50, 50)$	0.095	0.104	0.094	0.089	0.098	0.097	0.088	0.089
200	$(20, 40, 40)$	0.153	0.191	0.161	0.152	0.131	0.145	0.138	0.123
	$(20, 20, 20, 40)$	0.266	0.320	0.275	0.265	0.162	0.187	0.170	0.161
	$(40, 40, 40, 40, 40)$	0.482	0.532	0.500	0.486	0.212	0.244	0.222	0.221
	$(100, 100)$	0.200	0.204	0.192	0.195	0.197	0.208	0.198	0.195
400	$(40, 80, 80)$	0.491	0.535	0.507	0.501	0.348	0.379	0.345	0.349
	$(40, 40, 40, 80)$	0.858	0.861	0.863	0.862	0.505	0.523	0.513	0.513
	$(40, 40, 40, 40, 40)$	0.995	0.991	0.994	0.995	0.713	0.727	0.716	0.717

Table 4: The empirical power of Harrar and Kong's method

p	$\boldsymbol{n}^\top$	Parallelism (i)	Parallelism (ii)	Parallelism (iii)	Flatness (i)	Flatness (ii)	Flatness (iii)
	$(25, 25)$	0.069	0.001	0.014	0.071	0.001	0.014
100	$(10, 20, 20)$	0.093	0.001	0.007	0.079	0.001	0.016
	$(10, 10, 10, 20)$	0.117	0.001	0.006	0.093	0.002	0.018
	$(10, 10, 10, 10, 10)$	0.167	0.003	0.005	0.107	0.001	0.029
	$(50, 50)$	0.095	0.000	0.014	0.098	0.000	0.011
200	$(20, 40, 40)$	0.152	0.000	0.007	0.131	0.000	0.018
	$(20, 20, 20, 40)$	0.266	0.000	0.007	0.162	0.001	0.028
	$(40, 40, 40, 40, 40)$	0.482	0.000	0.016	0.212	0.000	0.049
	$(100, 100)$	0.200	0.000	0.027	0.197	0.000	0.028
400	$(40, 80, 80)$	0.491	0.000	0.034	0.344	0.000	0.054
	$(40, 40, 40, 80)$	0.858	0.000	0.098	0.505	0.001	0.124
	$(40, 40, 40, 40, 40)$	0.995	0.000	0.394	0.713	0.011	0.307

Table 5: Bias of $\mathrm{tr}\{\widetilde{(P_p\Sigma_1)^2}\}/\mathrm{tr}\{(P_p\Sigma_1)^2\}$ and of $\mathrm{tr}\{\widehat{(P_p\Sigma_1)^2}\}/\mathrm{tr}\{(P_p\Sigma_1)^2\}$

		$\mathrm{tr}\{\widetilde{(P_p\Sigma_1)^2}\}/\mathrm{tr}\{(P_p\Sigma_1)^2\}$			$\mathrm{tr}\{\widehat{(P_p\Sigma_1)^2}\}/\mathrm{tr}\{(P_p\Sigma_1)^2\}$		
p	n_1	(i)	(ii)	(iii)	(i)	(ii)	(iii)
100	10	0.00	37.80	14.92	0.00	0.05	0.00
	25	0.00	16.91	7.67	0.00	0.00	0.00
200	20	0.00	40.57	14.76	0.00	−0.01	−0.01
	50	0.00	17.33	8.12	0.00	0.00	0.00
400	40	0.00	43.03	16.31	0.00	0.02	−0.01
	100	0.00	17.54	6.75	0.00	−0.01	0.00