Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 595–609
http://www.ssca.org.in/journal

Tests of Contrasts for Mean Vectors with Large Dimensions

Rauf Ahmad
Department of Statistics, Uppsala University, Sweden

Received: 17 April 2024; Revised: 28 September 2024; Accepted: 10 October 2024

Abstract

A test statistic for a fixed contrast comparison of high-dimensional mean vectors is introduced. The statistic can be used when the dimension of the vectors exceeds the sample size, and the data may not necessarily follow a multivariate normal distribution. The components of the test statistics are defined as U-statistics with optimal properties, where the same estimators are given equivalent, computationally highly efficient, formulation for practical applications. The properties of the statistic are studied under a general multivariate model and certain mild assumptions. Through simulations, the statistic is shown to have an accurate size control and high power properties. An extension of a set of fixed orthogonal contrasts is also discussed.

Key words: High-dimensional tests; Multivariate inference; Contrast comparisons; U-statistics.

AMS Subject Classifications: 62H11, 62H30

1. Introduction

Let $\mathbf{X}_{ik} = (X_{ik1}, \ldots, X_{ikp})^{\mathrm{T}} \sim \mathcal{F}_i$, $k = 1, \ldots, n_i$, be a random sample of n_i vectors from ith non-degenerate p-variate distribution, denoted $\mathcal{F}_i$, which need not necessarily be multivariate normal, $i = 1, \ldots, g \geq 2$. Further, the g populations are assumed to be independent, with $\mathrm{E}(\mathbf{X}_{ik}) = \boldsymbol{\mu}_i \in \mathbb{R}^p$ and $\mathrm{Cov}(\mathbf{X}_{ik}) = \boldsymbol{\Sigma}_i \in \mathbb{R}^{p \times p}$, and $\boldsymbol{\Sigma}_i > 0$, $\forall\ i$.

Most of the testing problems in multivariate theory pertain to the two basic parameters, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$; *e.g*, single- and multi-sample hypotheses for $\boldsymbol{\mu}_i$, such as $\boldsymbol{\mu}_i = \mathbf{0}$, $\boldsymbol{\mu}_1 = \ldots = \boldsymbol{\mu}_g$ $(g \geq 2)$. These hypotheses are termed global hypotheses, and their rejection often implies further exploration to sort out potential contributors to the rejection. For example, for $g = 1$, a level profile analysis is carried out to test if all components of $\boldsymbol{\mu}$ are same, *i.e.* if $\mu_1 = \ldots = \mu_p$.

In practice, however, situations exist, mainly in multi-sample cases, where certain specific contrast comparisons among $\boldsymbol{\mu}_i$ are of interest. For example, for $g = 3$, it might be of interest to test if $\boldsymbol{\mu}_1 - 2\boldsymbol{\mu}_2 + \boldsymbol{\mu}_3 = \mathbf{0}$. In general, such a *contrast hypothesis* is formulated

Corresponding Author: Rauf Ahmad
Email: rauf.ahmad@statistik.uu.se

as

$$\mathrm{H}_0 : \sum_{i=1}^{g} c_i \boldsymbol{\mu}_i = \mathbf{0} \quad \text{vs.} \quad \mathrm{H}_1 : \text{Not } \mathrm{H}_0, \tag{1}$$

with $\sum_{i=1}^{g} c_i = 0$, a condition which is an inevitable component of the definition of a contrast. In the aforementioned example, $(c_1, c_2, c_3) = (1, -2, 1)$.

Note that, for $g = 2$, the condition implies $c_2 = -c_1$ which, without loss of generality, can be taken as $c_1 = 1 \Rightarrow c_2 = -1$, so that H_0 reduces to the usual two-sample hypothesis $\mathrm{H}_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Although, it is also a special form of contrast, the main advantage of contrast testing is apparent for the case of more than two populations.

Our objective in this article is to construct tests for H_0 in (1) when the dimension p may be large, and possibly larger than the sample sizes, *i.e.*, $p \gg n_i$, the $\mathcal{F}_i$ may be non-normal, and $\boldsymbol{\Sigma}_i$ may be unequal. For the classical case, *i.e.*, $p < n_i$, with $\mathcal{F}_i$ assumed multivariate normal, and often $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \; \forall \; i$ (homoscedasticity assumption), the multivariate theory offers likelihood-ratio tests leading to Wilks' Λ criterion, which is further related to an F-statistic, and for moderately large sample sizes, follows an approximate χ^2-distribution; see *e.g* Anderson (2003).

As the likelihood-ratio testing framework collapses for high-dimensional data, particularly when $p \gg n_i$, new testing strategies are needed to cope with this issue. In this context, we are interested to introduce tests of (1) for $p \gg n_i$ under multivariate Behrens-Fisher setting, additionally relaxing normality assumption which is replaced with alternative mild assumptions stated below.

The test statistics are composed of estimators defined as U-statistics with optimality properties. The same estimators are alternatively also defined as simple functions of empirical covariance estimators, which makes them computationally very efficient. The U-statistics version, however, helps study their theoretical properties, including limiting distribution, conveniently, where the efficient formulation is useful for practical applications.

Whereas high-dimensional mean testing has generally attracted huge attraction in the recent past (see a list of references in Ahmad, 2019b), problems like contrast comparison have mostly been dealt with under the general rubric of multiple testing theory. For a related work in the classical case, *i.e.*, $n > p$, see Hayter (2014) and the references cited therein. A general, comprehensive reference for multiple testing problems, including for large data, containing abundant further references, is Dickhaus (2014).

Section 2 introduces test statistic for a single contrast hypothesis in (1), with an extension to a set of orthogonal contrasts in Section 3. Evaluation of the proposed tests through simulations is given in Section 4. Some technical results are deferred to the Appendix.

2. Test of a single contrast

Given the data set up in Sec. 1, let $\mathbf{X}_i = (\mathbf{X}_{i1}^{\mathrm{T}}, \ldots, \mathbf{X}_{in_i}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{n_i \times p}$ be the data matrix corresponding to the ith sample, so that the unbiased estimators of $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are defined as

$$\overline{\mathbf{X}}_i = \frac{1}{n_i} \mathbf{X}_i^{\mathrm{T}} \mathbf{1}_{n_i}, \quad \widehat{\boldsymbol{\Sigma}}_i = \frac{1}{n_i - 1} \mathbf{X}_i^{\mathrm{T}} \mathbf{C}_{n_i} \mathbf{X}_i, \tag{2}$$

respectively, where $\mathbf{C}_{n_i} = \mathbf{I}_{n_i} - \mathbf{J}_{n_i}/n_i$ is the centering matrix with $\mathbf{I}_{n_i}$ as identity matrix and $\mathbf{J}_{n_i} = \mathbf{1}_{n_i}\mathbf{1}_{n_i}^{\mathrm{T}}$ with $\mathbf{1}_{n_i}$ a vector of 1s. All vectors are column vectors by default.

Further, we denote vector inner product of $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ as $\langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^{\mathrm{T}}\mathbf{b} : \mathbb{R}^p \mapsto \mathbb{R}$, so that $\|\mathbf{a}\|^2 = \mathbf{a}^{\mathrm{T}}\mathbf{a}$ is the (squared) norm of $\mathbf{a}$, and $\|\mathbf{A}\|^2 = \mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) : \mathbb{R}^{q\times p} \mapsto \mathbb{R}$ is the Frobenius norm of $\mathbf{A} \in \mathbb{R}^{q\times p}$. Moreover, $\otimes$ and $\oplus$ are Kronecker product and sum, respectively.

To consider a test statistic for H_0 in (1), we can logically begin with the point estimator, $\sum_{i=1}^{g} c_i\overline{\mathbf{X}}_i = \overline{\mathbf{X}}_0$, and note that, under independence,

$$\mathrm{E}(\overline{\mathbf{X}}_0) = \boldsymbol{\mu}_0 = \sum_{i=1}^{g} c_i\boldsymbol{\mu}_i \quad \text{and} \quad \mathrm{Cov}(\overline{\mathbf{X}}_0) = \boldsymbol{\Sigma}_0 = \sum_{i=1}^{g} c_i^2\frac{\boldsymbol{\Sigma}_i}{n_i}. \tag{3}$$

In the classical setting, assuming normality and homoscedasticity, a test for (1) can be defined as $\mathrm{T}^2 = c_0^{-1}\overline{\mathbf{X}}_0^{\mathrm{T}}\mathbf{S}_0^{-1}\overline{\mathbf{X}}_0$ with $c_0 = \sum_{i=1}^{g} c_i^2/n_i^2$, where $\mathbf{S}_0 = \sum_{i=1}^{g}(n_i-1)\widehat{\boldsymbol{\Sigma}}_i/(n-g)$ is the pooled estimator of $\boldsymbol{\Sigma}_0$ and $n = \sum_{i=1}^{g} n_i$. The T^2 statistic has optimality properties under the aforementioned assumptions, but its validity rests on the invertibility of $\mathbf{S}_0$ which, in turn, holds if and only if $n - g > p$. As this condition is not satisfied for high-dimensional data, and definitely not when $p \gg n_i$, T^2 collapses in this case and needs a modification.

The test statistic that we intend to propose for (1) is based on a modification of T^2-type statistics for testing different hypotheses on location parameters (see *e.g* Ahmad, 2014, 2019b). To see how this modification may work for the present case, first assume, tentatively, that $\boldsymbol{\Sigma}_i$ are known and, to avoid singularity issue of their empirical estimators at a later stage, consider the criterion

$$\mathrm{A} = \frac{\mathrm{A}_1}{\mathrm{tr}(\boldsymbol{\Sigma}_0)}, \tag{4}$$

with $\mathrm{A}_1 = \|\overline{\mathbf{X}}_0\|^2$, $\overline{\mathbf{X}}_0$, $\boldsymbol{\Sigma}_0$ as in (3), and $\mathrm{tr}(\cdot)$ is the trace operator. It follows that

$$\|\overline{\mathbf{X}}_0\|^2 = \left(\sum_{i=1}^{g} c_i\overline{\mathbf{X}}_i\right)^{\mathrm{T}}\left(\sum_{i=1}^{g} c_i\overline{\mathbf{X}}_i\right) = \sum_{i=1}^{g} c_i^2\|\overline{\mathbf{X}}_i\|^2 + \sum_{\substack{i=1\\i\neq j}}^{g}\sum_{j=1}^{g} c_ic_j\langle\overline{\mathbf{X}}_i, \overline{\mathbf{X}}_j\rangle.$$

Partitioning $\|\overline{\mathbf{X}}_i\|^2$ as

$$\|\overline{\mathbf{X}}_i\|^2 = \frac{1}{n_i^2}\sum_{k=1}^{n_i}\|\mathbf{X}_{ik}\|^2 + \frac{1}{n_i^2}\sum_{\substack{k=1\\k\neq r}}^{n_i}\sum_{r=1}^{n_i}\langle\mathbf{X}_{ik}, \mathbf{X}_{ir}\rangle = \frac{1}{n_i}\,\mathrm{E}_i + \frac{n_i-1}{n_i}U_i = Q_i + U_i,$$

we can further write A_1 as

$$\mathrm{A}_1 = \|\overline{\mathbf{X}}_0\|^2 = \sum_{i=1}^{g} c_i^2Q_i + \sum_{i=1}^{g} c_i^2U_i + 2\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g} c_ic_jU_{ij} \;=\; \mathrm{A}_{11} + \mathrm{A}_{12}, \tag{5}$$

with $\mathrm{A}_{11} = \sum_{i=1}^{g} c_i^2Q_i$, where $Q_i = (\mathrm{E}_i - U_i)/n_i$, $\mathrm{E}_i = \sum_{k=1}^{n_i}\|\mathbf{X}_{ik}\|^2/n_i$. Moreover

$$U_i = \frac{1}{n_i(n_i-1)}\sum_{\substack{k=1\\k\neq r}}^{n_i}\sum_{r=1}^{n_i}\langle\mathbf{X}_{ik}, \mathbf{X}_{ir}\rangle \quad \text{and} \quad U_{ij} = \langle\overline{\mathbf{X}}_i, \overline{\mathbf{X}}_j\rangle = \frac{1}{n_in_j}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}\langle\mathbf{X}_{ik}, \mathbf{X}_{jl}\rangle \tag{6}$$

are, one- and two-sample U-statistics with symmetric kernels, $h(\overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{ir}) = \langle \overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{ir} \rangle$ and $h(\overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{jl}) = \langle \overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{jl} \rangle$, respectively. The motivation behind this decomposition becomes clear from the moments of the components of A_1 as summarized in the following theorem, proved in Appendix B.1.

Theorem 1: Given the partition of A_1 in (5) with

$$\mathrm{A}_{11} = \sum_{i=1}^{g} c_i^2 Q_i, \quad \mathrm{A}_{12} = \sum_{i=1}^{g} c_i^2 U_i + 2 \sum_{\substack{i=1 \\ i<j}}^{g} \sum_{j=1}^{g} c_i c_j U_{ij}, \tag{7}$$

we have $\mathrm{E}(\mathrm{A}_{11}) = \mathrm{tr}(\boldsymbol{\Sigma}_0)$, $\mathrm{E}(\mathrm{A}_{12}) = \|\boldsymbol{\mu}_0\|^2$ and $\mathrm{Var}(\mathrm{A}_{12}) = 2\|\boldsymbol{\Sigma}_0\|^2 + \mathrm{R}$, where

$$\begin{aligned} \mathrm{R} \;=\; & 4 \sum_{i=1}^{g} (c_i\boldsymbol{\mu}_i)^{\mathrm{T}} \frac{c_i^2 \boldsymbol{\Sigma}_i}{n_i} (c_i\boldsymbol{\mu}_i) + 4 \left(\sum_{\substack{i=1 \\ i<j}}^{g} \sum_{j=1}^{g} (c_i\boldsymbol{\mu}_i)^{\mathrm{T}} \frac{c_j^2 \boldsymbol{\Sigma}_j}{n_j} (c_i\boldsymbol{\mu}_i) + \sum_{\substack{i=1 \\ i<j}}^{g} \sum_{j=1}^{g} (c_j\boldsymbol{\mu}_j)^{\mathrm{T}} \frac{c_i^2 \boldsymbol{\Sigma}_i}{n_i} (c_j\boldsymbol{\mu}_j) \right) \\ & + 8 \left(\sum_{\substack{i=1 \\ i<j}}^{g} \sum_{j=1}^{g} (c_i\boldsymbol{\mu}_i)^{\mathrm{T}} \frac{c_j^2 \boldsymbol{\Sigma}_j}{n_j} (c_j\boldsymbol{\mu}_j) + \sum_{\substack{i=1 \\ i<j}}^{g} \sum_{j=1}^{g} (c_j\boldsymbol{\mu}_j)^{\mathrm{T}} \frac{c_i^2 \boldsymbol{\Sigma}_i}{n_i} (c_i\boldsymbol{\mu}_i) \right) \\ & + 8 \left(\sum_{\substack{i=1 \\ i<j<j'}}^{g} \sum_{j=1}^{g} \sum_{j'=1}^{g} (c_i\boldsymbol{\mu}_i)^{\mathrm{T}} \frac{c_j^2 \boldsymbol{\Sigma}_j}{n_j} (c_{i'}\boldsymbol{\mu}_{i'}) + \sum_{\substack{i=1 \\ i<i'<j}}^{g} \sum_{i'=1}^{g} \sum_{j=1}^{g} (c_j\boldsymbol{\mu}_j)^{\mathrm{T}} \frac{c_i^2 \boldsymbol{\Sigma}_i}{n_i} (c_{j'}\boldsymbol{\mu}_{j'}) \right) \end{aligned}$$

Under H_0, $\mathrm{E}(\mathrm{A}_{12}) = 0$, $\mathrm{Var}(\mathrm{A}_{12}) = 2\|\boldsymbol{\Sigma}_0\|^2$, where $\mathrm{E}(\mathrm{A}_{11})$, $\mathrm{Var}(\mathrm{A}_{11})$ remain same.

We observe that, $\mathrm{E}(\mathrm{A}_{11})$ is independent of $\boldsymbol{\mu}_i$, hence of $\boldsymbol{\mu}_0$, and $\mathrm{E}(\mathrm{A}_{12})$ is independent of $\boldsymbol{\Sigma}_i$, hence of $\boldsymbol{\Sigma}_0$. Further, under H_0, $\mathrm{E}(\mathrm{A}_{12}) = 0$ and $\mathrm{Var}(\mathrm{A}_{12}) = 2\|\boldsymbol{\Sigma}_0\|^2$, so that

$$\mathrm{E}(\mathrm{A}) \;=\; 1 + \frac{\|\boldsymbol{\mu}_0\|^2}{\mathrm{tr}(\boldsymbol{\Sigma}_0)} = 1 \tag{8}$$

$$\mathrm{Var}(\mathrm{A}) \;=\; \frac{2\|\boldsymbol{\Sigma}_0\|^2}{[\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2}. \tag{9}$$

From the proof in Appendix B.1, we note that we use a slight approximation for $\mathrm{Var}(\mathrm{A}_{12})$ since the first term in $2\|\boldsymbol{\Sigma}_0\|^2$ has denominator $n_i(n_i - 1)$, not n_i^2, which, precisely, gives $\mathrm{Var}(\mathrm{A}_{12}) = 2\|\boldsymbol{\Sigma}_0\|^2[1+o(1)]$ and $\mathrm{Var}(\mathrm{A}) = [2\|\boldsymbol{\Sigma}_0\|^2/[\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2][1+o(1)]$. As $(n_i-1)/n_i$ makes no difference for the final limit as $n_i \to \infty$, we skip $o(1)$ term when the context is clear.

Note also that, $\mathrm{Var}(\mathrm{A}_{11})$ is not reported in Theorem 1. It will be a part of main theorem, Theorem 2, where it is shown that A_{11} is a simple plug-in, consistent estimator of $\mathrm{E}(\mathrm{A}_{11}) = \mathrm{tr}(\boldsymbol{\Sigma}_0)$ for $n_i, p \to \infty$, in the sense that $\mathrm{A}_1 \,/\, \mathrm{tr}(\boldsymbol{\Sigma}_0)$ and $[\mathrm{A}_1 \,/\, \mathrm{tr}(\boldsymbol{\Sigma}_0)][\mathrm{tr}(\boldsymbol{\Sigma}_0) / \, \mathrm{A}_{11}]$ have essentially the same limit. We can thus consider the following test statistic for H_0

$$\mathrm{T} = \frac{\mathrm{A}_1}{\mathrm{A}_{11}} = 1 + \frac{\mathrm{A}_{12}}{\mathrm{A}_{11}}. \tag{10}$$

With normality assumption relaxed, we replace it with a general multivariate model. Given $\mathbf{X}_{ik} \in \mathbb{R}^p$, let $\mathbf{Y}_{ik} = \mathbf{X}_{ik} - \boldsymbol{\mu}_i$, and define

$$\mathbf{Y}_{ik} = \boldsymbol{\Gamma}_i \mathbf{Z}_{ik}, \quad k = 1, \ldots, n_i, \; i = 1, \ldots, g, \tag{11}$$

with $\boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}_i^{1/2}$, $\mathbf{Z}_{ik} \in \mathbb{R}^p$, $\mathbf{Z}_{ik} \sim \mathcal{F}_i$, where $\mathrm{E}(\mathbf{Z}_{ik}) = \mathbf{0}_p$ and $\mathrm{Cov}(\mathbf{z}_{ik}) = \mathbf{I}_p \; \forall \; i$. Here, $\mathbf{0}_p$ is a vector of zeros and $\mathbf{I}_p$ denotes the identity matrix. We supplement Model (11) with the following assumptions, where $\nu_{is} = \lambda_{is}/p$ and λ_{is}, $s = 1, \ldots, p$, denote the eigenvalues of $\boldsymbol{\Sigma}_i$.

Assumption 1: $\mathrm{E}(Y_{iks}^4) = \gamma_{is} \leq \gamma < \infty \; \forall \; s = 1, \ldots, p, \; \forall \; i = 1, \ldots, g, \; \gamma \in \mathbb{R}^+$.

Assumption 2: $\lim_{p \to \infty} \sum_{s=1}^p \nu_{is} = \nu_{i0} \leq \nu \in \mathbb{R}^+, \; \forall \; i = 1, \ldots, g$.

Assumption 3: $\lim_{n_i, p \to \infty} p/n_i = \xi_i \leq \xi = O(1), \; \forall \; i = 1, \ldots, g$.

Assumption 4: $\lim_{n_i \to \infty} n_i/n = \rho_i \leq \rho = O(1), \; \forall \; i = 1, \ldots, g, \; n = \sum_{i=1}^g n_i$.

Assumption 5: $\lim_{p \to \infty} \boldsymbol{\mu}_i^{\mathrm{T}} \boldsymbol{\Sigma}_k \boldsymbol{\mu}_j / p = \phi_{ijk} \leq \phi = O(1), \; \forall \; i, j, k = 1, \ldots, g$.

Assumption 1 helps deal with moments of quadratic forms under Model (11). Assumption 2 is often used in high-dimensional inference. Assumptions 3-4 ensure a non-degenerate limit by controlling simultaneous rates of convergence among sample sizes and in relation to dimension. Assumption 5 is only needed under the alternative. Using Theorem 1 and the probability convergence of A_{11} (see Appendix B.2), we write

$$\mathrm{T} - 1 = \frac{\mathrm{A}_{12}}{\mathrm{tr}(\boldsymbol{\Sigma}_0)}[1 + o_P(1)],$$

with A_{12} as in (7), $\mathrm{E}(\mathrm{T} - 1) = \|\boldsymbol{\mu}_0\|^2 / \mathrm{tr}(\boldsymbol{\Sigma}_0)$ and

$$\sigma_1^2 = \frac{2\|\boldsymbol{\Sigma}_0\|^2 + \mathrm{R}}{[\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2},$$

where, under H_0, $\mathrm{E}(\mathrm{T} - 1) = 0$, $\sigma_0^2 = 2\|\boldsymbol{\Sigma}_0\|^2 / [\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2$. Theorem 2 gives the distribution of $\widetilde{\mathrm{T}} = (\mathrm{T} - \mathrm{E}(\mathrm{T}))/\sigma_{\mathrm{T}}$ with $\widetilde{\mathrm{T}}_0$ as its value under H_0. For proof, see Appendix B.2.

Theorem 2: Let $\widetilde{\mathrm{T}}$ be as defined above. Under Model (11) and Assumptions 1-5, $\widetilde{\mathrm{T}} \xrightarrow{\mathcal{D}} N(0, 1)$, as $n_i, p \to \infty$. In particular, under H_0, $\widetilde{\mathrm{T}}_0 \xrightarrow{\mathcal{D}} N(0, 1)$.

For power of $\widetilde{\mathrm{T}}$, let Z_α be the quantile of $Z \sim N(0, 1)$, and $\widetilde{\mathrm{T}}$, $\widetilde{\mathrm{T}}_0$ be as in Theorem 2. For any n_i and p, $P(\widetilde{\mathrm{T}}_0 \geq Z_\alpha) = \alpha$ and $P(\widetilde{\mathrm{T}} \geq -\delta + \tau Z_\alpha) = 1 - \beta$ define the size and power of the test, respectively, where $\delta = \|\boldsymbol{\mu}_0\|^2 / \sqrt{2\|\boldsymbol{\Sigma}_0\|^2 + \mathrm{R}}$ and $\tau = \sigma_0/\sigma_1$, with σ_0^2 and σ_1^2 as $\mathrm{Var}(\mathrm{T})$ under the null and alternative, respectively, as in Theorem 2. It follows, under the assumptions, that $\tau \to [2 + \xi^{-1}]^{-1/2} O(1) = O(1)$ and $\delta = n_i [2 + \xi^{-1}]^{-1/2} O(1) = O(n_i)$, so that $1 - \beta = 1$ - $P[\widetilde{\mathrm{T}} \leq -(n_i + Z_\alpha) O(1)] \Rightarrow 1$, as $n_i, p \to \infty$.

We need to estimate $\mathrm{Var}(\mathrm{T})$. As $\mathrm{A}_{11} \xrightarrow{\mathcal{P}} \sum_{i=1}^g \sum_{s=1}^\infty c_i^2 \xi_i \nu_{i0}$, $\mathrm{Var}(\mathrm{T})$ basically follows from $\mathrm{Var}(\mathrm{A}_{12})$ which is composed of $\|\boldsymbol{\Sigma}_i\|^2$ and $\|\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_j\|^2$, where $\boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}_i^{1/2}$, since, under H_0,

$$\|\boldsymbol{\Sigma}_0\|^2 = \sum_{i=1}^g \frac{c_i^4}{n_i} \|\boldsymbol{\Sigma}_i\|^2 + 2 \sum_{\substack{i=1 \\ i<j}}^g \sum_{i=1}^g \frac{c_i^2 c_j^2}{n_i n_j} \|\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_j\|^2. \tag{12}$$

The estimators are defined as below, where $\widehat{\mathbf{\Gamma}}_i = \widehat{\mathbf{\Sigma}}_i^{1/2}$.

Definition 1: Estimators of $\|\mathbf{\Gamma}_i\mathbf{\Gamma}_j\|^2$, $\|\mathbf{\Sigma}_i\|^2$, under Model (11), are defined as below, where $\nu_i = (n_i-1)/n_i(n_i-2)(n_i-3)$, $\mathrm{Q}_i = \sum_{k=1}^{n_i}\|\widetilde{\mathbf{X}}_{ik}\|^2$, $\widetilde{\mathbf{X}}_{ik} = \mathbf{X}_{ik} - \overline{\mathbf{X}}_i$, $i = 1,\ldots,g$.

$$\mathrm{E}_{ij} = \|\widehat{\mathbf{\Gamma}}_i\widehat{\mathbf{\Gamma}}_j\|^2, \tag{13}$$

$$\mathrm{E}_i = \nu_i\left[(n_i-1)(n_i-2)\|\widehat{\mathbf{\Sigma}}_i\|^2 + [\|\widehat{\mathbf{\Gamma}}_i\|^2]^2 - n\,\mathrm{Q}_i\right], \tag{14}$$

As functions of empirical $\widehat{\mathbf{\Sigma}}_i$, the estimators are computationally very efficient. They are unbiased and high-dimensional consistent. To prove these properties, however, an alternative formulation of the same estimators, in terms of U-statistics, is very helpful.

Given Model (11), let $\mathbf{D}_{ikr} = \mathbf{Y}_{ik} - \mathbf{Y}_{ir}$ with $\mathrm{E}(\mathbf{D}_{ikr}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{D}_{ikr}) = 2\mathbf{\Sigma}_i = 2\|\mathbf{\Gamma}_i\|^2$, and $\widehat{\mathbf{\Sigma}}_i$ can be written as U-statistic with symmetric kernel $h(\mathbf{X}_{ik},\mathbf{X}_{ir}) = \mathbf{D}_{ikr}\mathbf{D}_{ikr}^{\mathrm{T}}/2$, *i.e.*,

$$\widehat{\mathbf{\Sigma}}_i = \frac{1}{Q(n_i)}\sum_{\substack{k=1\\k\neq r}}^{n_i}\sum_{r=1}^{n_i}\frac{1}{2}\mathbf{D}_{ikr}\mathbf{D}_{ikr}^{\mathrm{T}}$$

where $Q(n_i) = n_i(n_i-1)$. Denote further $\mathrm{A}_{ijkrls} = \mathbf{D}_{ikr}^{\mathrm{T}}\mathbf{D}_{jls}$ and $\mathrm{A}_{ikrls} = \mathbf{D}_{ikr}^{\mathrm{T}}\mathbf{D}_{ils}$ with $\mathrm{E}(\mathrm{A}_{ijklrs}^2) = 4\|\mathbf{\Gamma}_i\mathbf{\Gamma}_j\|^2$, $\mathrm{E}(\mathrm{A}_{ikrls}^2) = 4\|\mathbf{\Sigma}_i\|^2$. The U-statistics forms of E_{ij} and E_i follow as

$$\mathrm{E}_{ij} = \frac{1}{Q(n_i)Q(n_j)}\sum_{\substack{k=1\\\pi(k,r)}}^{n_i}\sum_{r=1}^{n_i}\sum_{\substack{l=1\\\pi(l,s)}}^{n_j}\sum_{s=1}^{n_j}\frac{1}{4}\,\mathrm{A}_{ijklrs}^2 \tag{15}$$

$$\mathrm{E}_i = \frac{1}{P(n_i)}\sum_{k=1}^{n_i}\sum_{\substack{r=1\\\pi(k,r,l,s)}}^{n_i}\sum_{l=1}^{n_i}\sum_{s=1}^{n_i}\mathrm{B}_{ikrls}, \tag{16}$$

where $P(n_i) = n_i(n_i-1)(n_i-2)(n_i-3)$, $\mathrm{B}_{ikrls} = \mathrm{A}_{ikrls}^2 + \mathrm{A}_{iksrl}^2 + \mathrm{A}_{ilrsk}^2$ and $\pi(\cdot)$ implies all involved indices pairwise unequal. Note that, E_i and E_{ij} are one- and two-sample U-statistics with symmetric kernels $\mathrm{B}_{ikrls}/4$ and $\mathrm{A}_{ijklrs}^2/4$, respectively; see *e.g* Koroljuk and Borovskich (1994). The following theorem summarizes the properties of estimators. The proof of this theorem is a tedious computational exercise of projection properties of U-statistics and is omitted for simplicity; see *e.g* Ahmad (2017).

Theorem 3: Given Model (11), Assumption 1, and E_{ij}, E_i as in (15)-(16). Then, $\mathrm{E}(\mathrm{E}_{ij}) = \|\mathbf{\Gamma}_i\mathbf{\Gamma}_j\|^2$ and $\mathrm{E}(\mathrm{E}_i) = \|\mathbf{\Sigma}_i\|^2$. Further,

$$\mathrm{Var}(\mathrm{E}_{ij}) = \frac{2}{(n_i-1)(n_j-1)}\left[(n_i+n_j-1)\|\mathbf{\Sigma}_i\mathbf{\Sigma}_j\|^2 + \left\{\|\mathbf{\Gamma}_i\mathbf{\Gamma}_j\|^2\right\}^2 + M_2O(n) + M_3O(1)\right],$$

$$\mathrm{Var}(\mathrm{E}_i) = \frac{4}{P(n_i)}\left[a(n_i)\|\mathbf{\Sigma}_i^2\|^2 + b(n_i)\left\{\|\mathbf{\Sigma}_i\|^2\right\}^2 + M_2O(n_i^3) + M_3O(n_i^2)\right],$$

$$\mathrm{Cov}(\mathrm{E}_{ij},\mathrm{E}_i) = \frac{4}{Q(n_i)}\left[n_i\,\mathrm{tr}(\mathbf{\Sigma}_i^3\mathbf{\Sigma}_j) + M_2O(n_i)\right],$$

where $a(n_i) = 2n_i^3 - 9n_i^2 + 9n_i - 16$, $b(n_i) = n_i^2 - 3n_i + 8$, $P(n_i) = n_i(n_i-1)(n_i-2)(n_i-3)$, $Q(n_i) = n_i(n_i-1)$, and M_2, M_3 are given in Lemma 1.

Note that, less emphasis on terms involving M_2, M_3 *etc.* is due to the fact that they eventually vanish, exactly under normality, and asymptotically under Model (11) and the assumptions. For the rest, Theorem 3 yields $\text{Var}(\text{E}_i\,/\,\text{E}(\text{E}_i)) \leq O(1/n_i)$, $\text{Var}(\text{E}_{ij}\,/\,\text{E}(\text{E}_{ij})) \leq O(1/n_i+1/n_j)$, $\text{Cov}(\text{E}_i\,/\,\text{E}(\text{E}_i), \text{E}_{ij}\,/\,\text{E}(\text{E}_{ij})) \leq O(1/n_i)$, *i.e.*, the ratios are uniformly bounded in p. This, in particular, implies that p does not influence the non-degenerate limit of $\widetilde{\text{T}}$ in Theorem 2. Following corollary can now replace Theorem 2 for practical applications.

Corollary 3.1: Theorem 2 remains valid if $\text{Var}(\widetilde{\text{T}})$ is replaced with $\widehat{\text{Var}}(\widetilde{\text{T}})$ obtained by substituting E_{ij} for $\|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2$ and E_i for $\|\boldsymbol{\Sigma}_i\|^2$ in (12).

3. Test of a set of orthogonal contrasts

Often, the researcher is interested to simultaneously test a set of multiple contrasts. In principal, this set can be of any cardinality, but only a set of orthogonal contrasts makes sense since any contrast beyond orthogonal set will carry redundant information. For g populations, an orthogonal set consists of $m = g - 1$ contrasts. We are thus interested in simultaneous testing of a set of m contrasts, *i.e.*,

$$\text{H}_{0q} : \sum_{i=1}^{g} c_{iq}\boldsymbol{\mu}_{iq} = \mathbf{0} \;\; \text{vs.} \;\; \text{H}_{1q} : \text{Not } \; \text{H}_{0q}, \quad q = 1, \ldots, m, \tag{17}$$

where $\sum_{i=1}^{g} c_{iq} = 0$, as before, with additional orthogonality constraint, $\sum_{i=1}^{g} c_{iq}c_{iq'} = 0$, $q \neq q'$. Extending the notations in Sec. 2, we can re-write the set of hypotheses in (17) as

$$\text{H}_{0s} : \boldsymbol{\Xi}_s = \mathbf{0} \;\; \text{vs.} \;\; \text{H}_{1s} : \text{Not } \; \text{H}_{0s}, \tag{18}$$

where s refers to the *set* of contrasts, with $\boldsymbol{\Xi}_s = (\boldsymbol{\mu}_{01}^{\text{T}}, \ldots, \boldsymbol{\mu}_{0m}^{\text{T}})$, $\boldsymbol{\mu}_{0q} = \sum_{i=1}^{g} c_{iq}\boldsymbol{\mu}_{iq}$. Letting $\overline{\mathbf{X}}_{0q} = \sum_{i=1}^{g} c_{iq}\overline{\mathbf{X}}_{iq}$ estimate $\boldsymbol{\mu}_{0q}$, an estimator of $\boldsymbol{\Xi}_s \in \mathbb{R}^{m\times p}$ follows as

$$\mathbf{M}_s = (\overline{\mathbf{X}}_{01}^{\text{T}}, \ldots, \overline{\mathbf{X}}_{0m}^{\text{T}}) \in \mathbb{R}^{m\times p}.$$

Denoting $\boldsymbol{\Sigma}_{0q} = \sum_{i=1}^{g} c_{iq}^2\boldsymbol{\Sigma}_{iq}/n_i$ and using $\text{Cov}(\overline{\mathbf{X}}_{0q}, \overline{\mathbf{X}}_{0q'}) = \mathbf{0}$ for $q \neq q'$,, we get

$$\text{E}(\mathbf{M}_s) = \boldsymbol{\Xi}_s \;\; \text{and} \;\; \text{Cov}(\mathbf{M}_s) = \boldsymbol{\Sigma}_s = \text{diag}(\boldsymbol{\Sigma}_{01}, \ldots, \boldsymbol{\Sigma}_{0m}) = \oplus_{q=1}^{m}\boldsymbol{\Sigma}_{0q},$$

It is obvious then that the theory for m orthogonal contrasts extends straightforwardly from that of one contrast in Sec. 2, where the orthogonality condition particularly simplifies the computations. Thus, partitioning $\|\overline{\mathbf{X}}_{0q}\|^2$ similarly as $\|\overline{\mathbf{X}}_0\|^2$ in Sec. 2, we have

$$\|\overline{\mathbf{X}}_{0q}\|^2 = \sum_{i=1}^{g} c_{iq}^2 Q_{iq} + \sum_{i=1}^{g} c_{iq}^2 U_{iq} + 2\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g} c_{iq}c_{jq}U_{ijq} \;=\; \text{A}_{11q} + \text{A}_{12q},$$

with

$$\text{A}_{11q} = \sum_{i=1}^{g} c_{iq}^2 Q_{iq}, \quad \text{A}_{12q} = \sum_{i=1}^{g} c_{iq}^2 U_{iq} + 2\sum_{\substack{i=1\\i<j}}^{g}\sum_{j=1}^{g} c_{iq}c_{jq}U_{ijq},$$

where $Q_{iq} = (E_{iq} - U_{iq})/n_i$, E_{iq}, U_{iq}, U_{ijq} are defined as for single contrast, except for each q now. The rest of the theory proceeds likewise, so that Theorem 2 and Corollary 3.1 stand

Table 1: Estimated size of $\widetilde{\mathrm{T}}$ for three distributions with three covariance triplets

		Normal			Uniform			Exponential		
n_1, n_2, n_3	p	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
10, 20, 30	50	0.061	0.044	0.060	0.048	0.060	0.052	0.062	0.058	0.065
	100	0.051	0.051	0.057	0.052	0.054	0.056	0.055	0.055	0.058
	300	0.048	0.055	0.054	0.055	0.055	0.053	0.057	0.052	0.054
	500	0.050	0.056	0.049	0.051	0.055	0.053	0.055	0.057	0.057
	1000	0.044	0.052	0.051	0.046	0.048	0.051	0.052	0.054	0.055
20, 30, 50	50	0.050	0.048	0.045	0.049	0.052	0.045	0.049	0.057	0.054
	100	0.052	0.057	0.046	0.056	0.055	0.047	0.047	0.055	0.052
	300	0.054	0.053	0.054	0.052	0.048	0.051	0.055	0.054	0.053
	500	0.047	0.050	0.048	0.055	0.052	0.052	0.058	0.056	0.055
	1000	0.051	0.053	0.055	0.053	0.053	0.048	0.051	0.048	0.051
30, 50, 100	50	0.054	0.053	0.048	0.051	0.054	0.052	0.056	0.049	0.048
	100	0.057	0.049	0.055	0.050	0.051	0.055	0.055	0.054	0.052
	300	0.055	0.048	0.052	0.055	0.054	0.050	0.053	0.052	0.052
	500	0.055	0.047	0.050	0.054	0.052	0.051	0.053	0.050	0.047
	1000	0.049	0.051	0.053	0.049	0.051	0.052	0.049	0.053	0.051

valid for any T_q defined for qth contrast, using corresponding A_{11q} and A_{12q}. We therefore leave the unnecessarily repetitive details, and rather focus on the following important remarks which highlights the essential differences with the single contrast case.

First, the emphasis on making an orthogonal set of contrasts is due to the fact that such a set picks all information from the data without retaining much redundancies. It is further substantiated by the orthogonality condition, $\sum_{i=1}^{g} c_{iq}c_{iq'} = 0$, which, because of $\sum_{i=1}^{g} c_{iq} = 0$, mimics the numerator of a covariance.

Second, the theory of set of orthogonal contrasts pertains to the case of planned comparisons within the ambit of multiple testing. It differs from, *e.g*, Scheffé's method of all possible contrasts (Scheffé, 1959, Ch. 3), originally devised as a post-hoc strategy after global univariate ANOVA hypothesis is rejected. Scheffé's method allows infinitely many contrasts, although practically only a finite set is recommended and practically used in order to keep better error control.

Third, since many hypotheses are tested simultaneously, an error control mechanism is called for. With g relatively small or moderate in practice, a simple Bonferroni adjustment would suffice, which controls the family wise error rate in the strong sense. Otherwise, some researchers recommend a comparison-wise error control. For comprehensive theoretical results on multiple testing and error control procedures, see Dickhaus (2014). For a high-dimensional multiple testing framework, see Ahmad (2019a) and the references therein.

4. Simulations

We assess the accuracy of the proposed test statistics, particularly focusing on its robustness to normality assumption and validity under high-dimensional settings. For simplicity, we consider $\widetilde{\mathrm{T}}$ in Theorem 2 for $g = 3$. We generate p-dimensional random vectors of sizes (n_1, n_2, n_3) from normal, exponential and uniform distributions, where

$$n_1 \in \{10, 20, 30\},\ n_2 = \{20, 30, 50\},\ n_3 = \{30, 50, 100\},\ p = \{50, 100, 300, 500, 1000\}.$$

All (n_1, n_2, n_3, p) quadruplets are used under three triplets of covariance structures, denoted, S_1: (CS, AR, UN), S_2: (CS1, CS2, AR) and S_3: (CS, AR1, AR2), which are composed of compound symmetric (CS), Autoregressive of order 1, AR(1), and unstructured (UN) matrices. They are defined as CS: $\boldsymbol{\Sigma} = (1-\rho)\mathbf{I}_p + \rho\mathbf{J}_p$, AR: $\text{Cov}(X_k, X_l) = \kappa\rho^{|k-l|}$, $\forall\ k, l$, and UN: $\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=1}^p$ with $\sigma_{ij} = 1(1)p$ $(i = j)$ and $\rho_{ij} = (i-1)/p$ $(i > j)$, respectively. Here, $\mathbf{I}$ denotes identity matrix and $\mathbf{J}$ is a matrix of 1s. We use $\kappa = 1$ for all three triplets, where CS1, CS2, and AR1, AR2, refer to CS and AR with $\rho = 0.3$ and $\rho = 0.7$, respectively.

The nominal level is set at $\alpha = 0.05$, for both size and power. Moreover, for power, we set H_1 by letting $\boldsymbol{\mu}_1$ deviate from H_0 in a monotonically increasing fashion, *i.e.*, $\boldsymbol{\mu}_1 = \delta_r\mathbf{p}_1$, $\mathbf{p}_1 = (1/p, \ldots, p/p)$, $\delta_r = 0.2(0.2)1$. Finally, we used two contrast vectors, $(c_1, c_2, c_3) = (1,$ 0, -1) and (1, -2, 1), and due to similarity of results, only the first case is discussed here.

The estimated size and power are obtained by averaging over 2000 simulations. Table 1 reports the size for all distributions and Tables 2-3 report the power, respectively, for normal and uniform distributions. The power results for exponential distributions were very similar to those for uniform, and are therefore not reported.

All simulations are carried out in SAS/IML where data for the three multivariate distributions are generated by using appropriate arguments for corresponding built-in SAS/IML functions. For this, first, by using the square root of the assumed covariance matrix (*e.g*, compound symmetry), a sequence of multivariate normal vectors is generated which follows this structure. Then, by probability integral transform, it is converted into a uniform sequence, say U, which is adjusted for its mean and variance by subtracting a mean vector of length p (with each entry 1/2) and multiplying by the square root of a $p \times p$ diagonal

Table 2: Estimated power of $\widetilde{\text{T}}$ for normal distribution with three covariance triplets

		S_1			S_2			S_3		
n_1, n_2, n_3	p/δ	0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
10, 20, 30	50	0.151	0.963	1.000	0.146	0.962	1.000	0.131	0.923	1.000
	100	0.203	0.998	1.000	0.195	0.985	1.000	0.187	0.998	1.000
	300	0.351	1.000	1.000	0.329	1.000	1.000	0.345	1.000	1.000
	500	0.441	1.000	1.000	0.461	1.000	1.000	0.438	1.000	1.000
	1000	0.528	1.000	1.000	0.545	1.000	1.000	0.582	1.000	1.000
20, 30, 50	50	0.263	1.000	1.000	0.269	1.000	1.000	0.252	1.000	1.000
	100	0.385	1.000	1.000	0.392	1.000	1.000	0.368	1.000	1.000
	300	0.700	1.000	1.000	0.699	1.000	1.000	0.714	1.000	1.000
	500	0.865	1.000	1.000	0.866	1.000	1.000	0.831	1.000	1.000
	1000	0.942	1.000	1.000	0.998	1.000	1.000	0.917	1.000	1.000
30, 50, 100	50	0.426	1.000	1.000	0.435	1.000	1.000	0.423	1.000	1.000
	100	0.637	1.000	1.000	0.647	1.000	1.000	0.649	1.000	1.000
	300	0.956	1.000	1.000	0.948	1.000	1.000	0.947	1.000	1.000
	500	1.000	1.000	1.000	0.994	1.000	1.000	0.999	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

covariance matrix (with each diagonal entry 1/12). The exponential distribution follows by an additional log-transformation of U followed by its corresponding adjustment.

We observe accurate size control under all parameters. In particular, the performance for small or moderate sample sizes and for increasing dimension, for all covariance triplets, is noteworthy. A slight fluctuation of size can be seen for the exponential distribution but it stabilizes itself for even moderate sample sizes. Of particular mention is the power which is not only reasonably high, but also increases for increasing p as well as for increasing n_i.

The performance of the statistic for non-normal cases further implies its robustness under the general model. The overall performance of the statistic supports its use in practice for high-dimensional data with moderate sample sizes and departures from normality.

Table 3: Estimated power of $\widetilde{\text{T}}$ for uniform distribution with three covariance triplets

		S_1			S_2			S_3		
n_1, n_2, n_3	p/δ	0.2	0.6	1.0	0.2	0.6	1.0	0.2	0.6	1.0
10, 20, 30	50	0.143	0.961	1.000	0.134	0.965	1.000	0.142	0.956	1.000
	100	0.198	0.995	1.000	0.190	0.998	1.000	0.181	0.999	1.000
	300	0.315	1.000	1.000	0.329	1.000	1.000	0.351	1.000	1.000
	500	0.463	1.000	1.000	0.472	1.000	1.000	0.446	1.000	1.000
	1000	0.586	1.000	1.000	0.619	1.000	1.000	0.530	1.000	1.000
20, 30, 50	50	0.255	1.000	1.000	0.240	0.998	1.000	0.243	1.000	1.000
	100	0.349	1.000	1.000	0.368	1.000	1.000	0.368	1.000	1.000
	300	0.697	1.000	1.000	0.686	1.000	1.000	0.705	1.000	1.000
	500	0.884	1.000	1.000	0.805	1.000	1.000	0.811	1.000	1.000
	1000	0.883	1.000	1.000	0.936	1.000	1.000	0.960	1.000	1.000
30, 50, 100	50	0.412	1.000	1.000	0.415	1.000	1.000	0.447	1.000	1.000
	100	0.660	1.000	1.000	0.618	1.000	1.000	0.628	1.000	1.000
	300	0.957	1.000	1.000	0.958	1.000	1.000	0.954	1.000	1.000
	500	0.998	1.000	1.000	1.000	1.000	1.000	0.999	1.000	1.000
	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

5. Discussion and remarks

A test statistic for contrast comparison of mean vectors is introduced when the dimension of the vectors is large, even exceeding the number of vectors. Relaxing normality assumptions, properties of the test statistic, including its limit under high-dimensional set up, is provided for a general multivariate model and a few mild assumptions. The statistic is simple and composed of computationally efficient estimators. Simulations are used to demonstrate the theoretical properties of the test statistic. An extension to a set of orthogonal contrasts is also given.

Acknowledgements

The author is thankful to the editors, and particularly to a reviewer whose comments helped improve the section on simulations.

References

Ahmad, R. (2014). A U-statistic approach for a high-dimensional two-sample mean testing problem under non-normality and Behrens-Fisher setting. *Annals of the Institute of Statistical Mathematics*, **66**, 33-61.

Ahmad, R. (2017). Location-invariant multi-sample U-tests for covariance matrices with large dimension. *Scandinavian Journal of Statistics*, **44**, 500-523.

Ahmad, R. (2019a). Multiple comparisons of high-dimensional mean vectors under general conditions. *Journal of Statistical Computations and Simulation*, **89**, 1044-1059.

Ahmad, R. (2019b). A unified approach to testing mean vectors with large dimensions. *AStA: Advances in Statistical Analysis*, **103**, 593-618.

Ahmad, R. (2022). Tests for proportionality of matrices with large dimension. *Journal of Multivariate Analysis*, **189**, 104865.

Anderson, T. W. (2003). *Introduction to Multivariate Statistical Analysis*, 3rd edition, Wiley New York.

Dickhaus T. (2014). *Simultaneous Statistical Inference.* Springer, New York.

Hayter, A. J. (2014). Inferences on linear combinations of normal means with unknown and unequal variances. *Sankhya*, **76A**, 257-279.

Jiang, J. (2010). *Large Sample Techniques for Statistics.* Springer, New York.

Koroljuk, V. S. and Borovskich, Y.V. (1994). *Theory of U-statistics.* Kluwer Press, Dordrecht.

Lee, A. J. (1990). *U-Statistics: Theory & Practice*, Marcel Dekker, New York.

Mardia, K. V, J. T. Kent, and C. C. Taylor (2024). *Multivariate Analysis*, 2nd edition, Wiley, New York.

Scheffé, H. (1959). *The Analysis of Variance*, Wiley, New York.

Searle, S. R. (1971). *Linear Models.* Wiley, New York.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

APPENDIX

A. Some basic results

Lemma 1: For $\mathbf{Z}_{ik} \in \mathbb{R}^p$, $k = 1, \ldots, n_i$, defined in Model (11), let $\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ik}$ and $\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ir}$, $k \neq r$, be quadratic and bilinear form of independent components from sample i, and $\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{jl}$, $k \neq l$, $i \neq j$, be the bilinear form composed of vectors from two independent samples. Also let γ be as defined in Assumption 1 and $\odot$ denotes the Hadamard product. Then $\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ir}) = 0$, $\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{jl}) = 0$, $\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ik}) = \|\boldsymbol{\Gamma}_i\|^2$, $\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ir})^2 = \|\boldsymbol{\Sigma}_i\|^2$, $\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{jl})^2 = \|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2$. Further,

$$\begin{aligned}
\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ik})^2 &= 2\|\boldsymbol{\Sigma}_i\|^2 + [\|\boldsymbol{\Gamma}_i\|^2]^2 + M_1\\
\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{Z}_{ik})^2 &= 2\|\boldsymbol{\Sigma}_i^2\|^2 + [\|\boldsymbol{\Sigma}_i\|^2]^2 + M_2\\
\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{ir})^4 &= 6\|\boldsymbol{\Sigma}_i^2\|^2 + 3[\|\boldsymbol{\Sigma}_i\|^2]^2 + M_3\\
\mathrm{E}(\mathbf{Z}_{ik}^{\mathrm{T}}\mathbf{Z}_{jl})^4 &= 6\|\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\|^2 + 3[\|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2]^2 + M_4,
\end{aligned}$$

with $M_1 = (\gamma - 3)\,\mathrm{tr}(\boldsymbol{\Sigma}_i \odot \boldsymbol{\Sigma}_i)$, $M_2 = (\gamma_i - 3)\,\mathrm{tr}(\boldsymbol{\Sigma}_i^2 \odot \boldsymbol{\Sigma}_i^2)$, $M_3 = 6(\gamma - 3)\,\mathrm{tr}(\boldsymbol{\Sigma}_i^2 \odot \boldsymbol{\Sigma}_i^2) + (\gamma_i - 3)^2\,\mathrm{tr}(\boldsymbol{\Sigma}_i \odot \boldsymbol{\Sigma}_i)^2$, and $M_4 = 6(\gamma - 3)\,\mathrm{tr}(\boldsymbol{\Sigma}_i^2 \odot \boldsymbol{\Sigma}_j^2) + (\gamma - 3)^2\,\mathrm{tr}(\boldsymbol{\Sigma}_i \odot \boldsymbol{\Sigma}_i)\,\mathrm{tr}(\boldsymbol{\Sigma}_j \odot \boldsymbol{\Sigma}_j)$.

All moments in Lemma 1 reduce to those under normality for $\gamma = 3$; see Searle (1971).

Lemma 2: (Jiang, 2010, Page 183) Let $Y_1, Y_2, \ldots$ be iid r.vs. with $\mathrm{E}(Y_i) = 0$, $\mathrm{Var}(Y_i) = 1$, and b_{ni} be constants, $1 \leq i \leq n$. Then $\sum_{i=1}^n b_{ni}Y_i \xrightarrow{\mathcal{D}} N(0,1)$ as $n \to \infty$, if $\max_i b_{ni}^2 \to 0$.

B. Main proofs

B.1. Proof of theorem 1

With $\mathrm{E}(Q_i) = \|\boldsymbol{\Gamma}_i\|^2/n_i$, $\mathrm{E}(U_i) = \|\boldsymbol{\mu}_i\|^2$, $\mathrm{E}(U_{ij}) = \langle\boldsymbol{\mu}_i, \boldsymbol{\mu}_j\rangle$, we get, by independence,

$$\mathrm{E}(\mathrm{A}_{11}) = \sum_{i=1}^g c_i^2\|\boldsymbol{\Gamma}_i\|^2/n_i = \mathrm{tr}(\boldsymbol{\Sigma}_0)$$

$$\mathrm{E}(\mathrm{A}_{12}) = \sum_{i=1}^g c_i^2\|\boldsymbol{\mu}_i\|^2 + 2\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_ic_j\langle\boldsymbol{\mu}_i, \boldsymbol{\mu}_j\rangle = \left(\sum_{i=1}^g c_i\boldsymbol{\mu}_i\right)^{\mathrm{T}}\left(\sum_{i=1}^g c_i\boldsymbol{\mu}_i\right) = \|\boldsymbol{\mu}_0\|^2.$$

$$\begin{aligned}
\mathrm{Var}(\mathrm{A}_{12}) &= \mathrm{Var}\left(\sum_{i=1}^g c_i^2U_i + 2\sum_{i<j} c_ic_jU_{ij}\right)\\
&= \mathrm{Var}\left(\sum_{i=1}^g c_i^2U_i\right) + 4\,\mathrm{Var}\left(\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_ic_jU_{ij}\right) + 4\,\mathrm{Cov}\left(\sum_{i=1}^g c_i^2U_i, \sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_ic_jU_{ij}\right)\\
&= \sum_{i=1}^g c_i^4\,\mathrm{Var}(U)_i + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_i^2c_j^2\,\mathrm{Var}(U_{ij}) + 8\sum_{\substack{i=1\\i<j<j'}}^g\sum_{j=1}^g\sum_{j'=1}^g c_i^2c_jc_{j'}\,\mathrm{Cov}(U_{ij}, U_{ij'})\\
&+ 8\sum_{\substack{i=1\\i<i'<j}}^g\sum_{i'=1}^g\sum_{j=1}^g c_ic_{i'}c_j^2\,\mathrm{Cov}(U_{ij}, U_{i'j}) + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_i^3c_j\,\mathrm{Cov}(U_i, U_{ij}) + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_i^3c_j\\
&\mathrm{Cov}(U_j, U_{ij})
\end{aligned}$$

where the remaining covariances vanish when all indices are unequal. Using the second order moments of one- and two-sample U-statistics (see *e.g* Koroljuk and Borovskich, 1994), *i.e.*,

$$\begin{aligned}\mathrm{Var}(U_{n_i}) &= 2\left[2(n_i-1)\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i + \|\boldsymbol{\Sigma}_i\|^2\right]/n_i(n_i-1)\\ \mathrm{Var}(U_{n_in_j}) &= [n_i\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_i + n_j\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_j + \|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2]/n_in_j\end{aligned}$$

with (see also Ahmad, 2019b) $\mathrm{Cov}(U_{n_i}, U_{n_in_j}) = 2\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i/n_i$, $\mathrm{Cov}(U_{n_j}, U_{n_in_j}) = 2\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_j/n_j$, $\mathrm{Cov}(U_{n_in_j}, U_{n_in_{j'}}) = \boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{j'}/n_i$, and $\mathrm{Cov}(U_{n_in_j}, U_{n_{i'}n_j}) = \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_{i'}/n_j$, we get

$$\begin{aligned}\mathrm{Var}(\mathrm{A}_{12}) &= 2\sum_{i=1}^g \frac{c_i^4\|\boldsymbol{\Sigma}_i\|^2}{n_i(n_i-1)} + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g \frac{c_i^2c_j^2\|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2}{n_in_j} + 4\sum_{i=1}^g \frac{c_i^4\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i}{n_i}\\ &\quad + 4\left(\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g \frac{c_i^2c_j^2\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_i}{n_j} + \sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g \frac{c_j^2c_i^2\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_j}{n_i}\right)\\ &\quad + 8\left(\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g \frac{c_ic_j^3\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_j}{n_j} + \sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g \frac{c_i^3c_j\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_i}{n_i}\right)\end{aligned}$$

$$+ 8\left(\sum_{\substack{i=1\\i<j<j'}}^g\sum_{j=1}^g\sum_{j'=1}^g \frac{c_i^2c_jc_{j'}\boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\mu}_{j'}}{n_i} + \sum_{\substack{i=1\\i<i'<j}}^g\sum_{i'=1}^g\sum_{j=1}^g \frac{c_ic_{i'}c_j^2\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_{i'}}{n_j}\right).$$

Slightly re-arranging the terms, we get the required expression as

$$\begin{aligned}\mathrm{Var}(\mathrm{A}_{12}) &= 2\sum_{i=1}^g c_i^4\frac{\|\boldsymbol{\Sigma}_i\|^2}{n_i(n_i-1)} + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_i^2c_j^2\frac{\|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2}{n_in_j} + 4\sum_{i=1}^g (c_i\boldsymbol{\mu}_i)^{\mathrm{T}}\frac{c_i^2\boldsymbol{\Sigma}_i}{n_i}(c_i\boldsymbol{\mu}_i)\\ &\quad + 4\left(\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g (c_i\boldsymbol{\mu}_i)^{\mathrm{T}}\frac{c_j^2\boldsymbol{\Sigma}_j}{n_j}(c_i\boldsymbol{\mu}_i) + \sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g (c_j\boldsymbol{\mu}_j)^{\mathrm{T}}\frac{c_i^2\boldsymbol{\Sigma}_i}{n_i}(c_j\boldsymbol{\mu}_j)\right)\\ &\quad + 8\left(\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g (c_i\boldsymbol{\mu}_i)^{\mathrm{T}}\frac{c_j^2\boldsymbol{\Sigma}_j}{n_j}(c_j\boldsymbol{\mu}_j) + \sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g (c_j\boldsymbol{\mu}_j)^{\mathrm{T}}\frac{c_i^2\boldsymbol{\Sigma}_i}{n_i}(c_i\boldsymbol{\mu}_i)\right)\\ &\quad + 8\left(\sum_{\substack{i=1\\i<i'<j}}^g\sum_{j=1}^g\sum_{j'=1}^g (c_i\boldsymbol{\mu}_i)^{\mathrm{T}}\frac{c_j^2\boldsymbol{\Sigma}_j}{n_j}(c_{i'}\boldsymbol{\mu}_{i'}) + \sum_{\substack{i=1\\i<j<j'}}^g\sum_{i'=1}^g\sum_{j=1}^g (c_j\boldsymbol{\mu}_j)^{\mathrm{T}}\frac{c_i^2\boldsymbol{\Sigma}_i}{n_i}(c_{j'}\boldsymbol{\mu}_{j'})\right)\\ &= 2\sum_{i=1}^g c_i^4\frac{\|\boldsymbol{\Sigma}_i\|^2}{n_i(n_i-1)} + 4\sum_{\substack{i=1\\i<j}}^g\sum_{j=1}^g c_i^2c_j^2\frac{\|\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_j\|^2}{n_in_j}. \qquad (19)\end{aligned}$$

B.2. Proof of theorem 2

The strategy, as explained around Theorem 2, is to combine the consistency of A_{11} and weak limit of A_{12}. First, $\mathrm{E}(\mathrm{E}_i) = \mathrm{tr}(\boldsymbol{\Sigma}_i)/n_i + \|\boldsymbol{\mu}_i\|^2$, $\mathrm{E}(U_i) = \|\boldsymbol{\mu}_i\|^2$, give $\mathrm{E}(Q_i) = \mathrm{tr}(\boldsymbol{\Sigma}_i)/n_i$, independent of $\boldsymbol{\mu}_i$. As c_i are known constants, we get, for $\mathrm{A}_{11} = \sum_{i=1}^{g} c_i^2 Q_i$, by independence,

$$\mathrm{Var}(\mathrm{A}_{11}) = \sum_{i=1}^{g} c_i^4 \,\mathrm{Var}(Q_i),$$

where $Q_i = (\mathrm{E}_i - U_i)/n_i$. It thus suffices to focus on Q_i. From Lemma 1 and Sec. B.1,

$$\begin{aligned}
\mathrm{Var}(Q_i) &\leq \frac{1}{n_i^2}\{\mathrm{Var}(\mathrm{E}_i) + \mathrm{Var}(U_i)\} = \frac{1}{n_i^2}\left\{\frac{1}{n_i}\mathrm{Var}(\|\mathbf{X}_{ik}\|^2) + \frac{2\|\boldsymbol{\Sigma}_i\|^2}{n_i(n_i-1)}\right\} \\
&\leq \frac{1}{n_i^2}\left\{\frac{(\gamma_i - 1)\|\boldsymbol{\Sigma}_i\|^2}{n_i} + \frac{2\|\boldsymbol{\Sigma}_i\|^2}{n_i(n_i-1)}\right\} = \frac{\gamma_i + 1}{n_i^3}\|\boldsymbol{\Sigma}_i\|^2 \\
&\leq (\gamma_i + 1)c_i^2 O\left(\frac{1}{n_i}\right),
\end{aligned}$$

under the assumptions. It proves the consistency of Q_i, hence of A_{11}, as $n_i, p \to \infty$. Now consider T in (10) which, using the consistency of A_{11}, can be written as

$$\mathrm{T} - 1 = \frac{\mathrm{A}_{12}}{\mathrm{tr}(\boldsymbol{\Sigma}_0)} \cdot \frac{\mathrm{tr}(\boldsymbol{\Sigma}_0)}{\mathrm{A}_{11}} = \frac{\mathrm{A}_{12}}{\mathrm{tr}(\boldsymbol{\Sigma}_0)}[1 + o_P(1)].$$

Using moments in Sec. B.1 and, for convenience, ignoring the $o_P(1)$ factor, we have

$$\mathrm{E}(\mathrm{T} - 1) = \frac{\|\boldsymbol{\mu}_0\|^2}{\mathrm{tr}(\boldsymbol{\Sigma}_0)}, \quad \sigma_1^2 = \frac{2\|\boldsymbol{\Sigma}_0\|^2 + \mathrm{R}}{[\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2},$$

which, under H_0, reduce, respectively, to $\mathrm{E}(\mathrm{T}-1) = 0$, $\sigma_0^2 = 2\|\boldsymbol{\Sigma}_0\|^2/[\mathrm{tr}(\boldsymbol{\Sigma}_0)]^2$, where R is given in Theorem 1. Denote $\mathbf{U} = (\mathbf{U}_1^{\mathrm{T}},\ \mathbf{U}_2^{\mathrm{T}})^{\mathrm{T}}$, where the sub-vectors,

$$\mathbf{U}_1 = (c_1^2 U_1, \ldots, c_g^2 U_g)^{\mathrm{T}}, \quad \mathbf{U}_2 = (c_1 c_2 U_{12}, \ldots, c_1 c_g U_{1g}, c_2 c_1 U_{21}, c_2 c_3 U_{23}, \ldots, c_{g-1} c_g U_{g-1,g})^{\mathrm{T}}$$

are composed of one- and two-sample U-statistics of all distinct pairs, respectively. We can write $\mathrm{A}_{12} = \mathbf{1}_G^{\mathrm{T}}\mathbf{U}$, with $\mathbf{1}_G$ a vector of all 1s of dimension $G = g + g(g-1) = g^2$. Note that, elements in $\mathbf{U}_2$ such as U_{12} and U_{21} are same, by symmetry of the kernel, but are repeated to count all possible cases, so that A_{12} can be represented as a linear combination of the entire vector $\mathbf{U}$. We note that $\mathrm{E}(\mathrm{A}_{12}) = \mathbf{1}^{\mathrm{T}}\,\mathrm{E}(\mathbf{U}) = \|\boldsymbol{\mu}_0\|^2$ and $\mathrm{Var}(\mathrm{A}_{12}) = \mathbf{1}^{\mathrm{T}}\,\mathrm{Cov}(\mathbf{U})\mathbf{1} = 2\|\boldsymbol{\Sigma}_0\|^2 + \mathrm{R}$, as in Theorem 1, where

$$\mathrm{Cov}(\mathbf{U}) = \begin{pmatrix} \mathrm{Cov}(\mathbf{U}_1) & \mathrm{Cov}(\mathbf{U}_1, \mathbf{U}_2) \\ \mathrm{Cov}(\mathbf{U}_2, \mathbf{U}_1) & \mathrm{Cov}(\mathbf{U}_2) \end{pmatrix}.$$

It follows that $\mathrm{Cov}(\mathbf{U}_1)$ and $\mathrm{Cov}(\mathbf{U}_2)$, on the diagonal of $\mathrm{Cov}(\mathbf{U})$, lead to $2\|\boldsymbol{\Sigma}_0\|^2$ in $\mathrm{Var}(\mathrm{A}_{12})$, where $\mathrm{Cov}(\mathbf{U}_1, \mathbf{U}_2)$ leads to R. Further, under independence, $\mathrm{Cov}(\mathbf{U}_1)$ is a diagonal matrix, and off-diagonal elements of $\mathrm{Cov}(\mathbf{U}_2)$, *i.e.*, $\mathrm{Cov}(U_{ij}, U_{i'j'})$, are also zero when $i \neq i'$, $j \neq j'$. The rest of the terms in $\mathrm{Cov}(\mathbf{U})$ are of the form, *e.g*, $\mathrm{Cov}(U_{ij}, U_{i'j}) = \boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\mu}_{i'}/n_i$, which constitute R and, under the assumptions, are uniformly bounded in the limit, and the same holds for the elements of off-diagonal blocks, $\text{Cov}(\mathbf{U}_1, \mathbf{U}_2)$. However, $\text{R} = 0$ under H_0, and also $\text{R}/[\text{tr}(\boldsymbol{\Sigma}_0)]^2 \to 0$ asymptotically under H_1, so that $\sigma_1^2/\sigma_0^2 \to 1$ in the limit. Hence, $\text{Cov}(\mathbf{U})/[\text{tr}(\boldsymbol{\Sigma}_0)]^2$ can be considered as a diagonal matrix for the limit.

Further, $\text{E}(\text{T} - 1)$ is uniformly bounded, and so is $2\|\boldsymbol{\Sigma}_0\|^2/[\text{tr}(\boldsymbol{\Sigma}_0)]^2 \leq 2$, under the assumptions, where these bounds remain intact for any p, so that we can use a sequential limit. Writing $\mathbf{1}^{\text{T}}(\mathbf{U} - \text{E}(\mathbf{U})) = \text{A}_{12} - \text{E}(\text{A}_{12})$, with corresponding elements $U_i - \text{E}(U_i)$ and $U_{ij} - \text{E}(U_{ij})$, and associated kernels, $\langle \overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{ir} \rangle - \|\boldsymbol{\mu}_i\|^2$, and $\langle \overline{\mathbf{X}}_{ik}, \overline{\mathbf{X}}_{jl} \rangle - \langle \boldsymbol{\mu}_i, \boldsymbol{\mu}_j \rangle$, it follows, from the asymptotic theory of U-statistics (Koroljuk and Borovskich, 1994), that, for any p,

$$n_i c_i^2 U_i \xrightarrow{\mathcal{D}} \sum_{s=1}^{p} \lambda_{is}(z_{is}^2 - 1) \quad \text{and} \quad \sqrt{n_i n_j} U_{n_i n_j} \xrightarrow{\mathcal{D}} \sum_{s=1}^{p} \lambda_{is}\lambda_{js} z_{is} z_{js},$$

as $n_i \to \infty$, where z_{is}, z_{js} are iid $N(0, 1)$ variables, and independent of each other, and λ_{is} are the eigenvalues of $\boldsymbol{\Sigma}_i$. Now, taking p and the denominator into account, and applying Lemma 2 for $p \to \infty$, the required limit follows by a simple application of the Cramér-Wold device and Slutsky's theorem (van der Vaart, 1998), as was similarly done in Ahmad (2019b).