Journal of Multivariate Analysis 101 (2010) 2554–2570

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A new test for sphericity of the covariance matrix for high dimensional data

Thomas J. Fisher [a,*], Xiaoqian Sun [b], Colin M. Gallagher [b]

[a] Department of Mathematics and Statistics, University of Missouri-Kansas City, Kansas City, MO 64110, USA

[b] Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, USA

ARTICLE INFO

Article history:
Received 12 November 2009
Available online 3 August 2010

AMS subject classifications:
62H10
62H15

Keywords:
Covariance matrix
Hypothesis testing
High-dimensional data analysis

ABSTRACT

In this paper we propose a new test procedure for sphericity of the covariance matrix when the dimensionality, p, exceeds that of the sample size, $N = n + 1$. Under the assumptions that (A) $0 < \mathrm{tr}\Sigma^i/p < \infty$ as $p \to \infty$ for $i = 1, \ldots, 16$ and (B) $p/n \to c < \infty$ known as the concentration, a new statistic is developed utilizing the ratio of the fourth and second arithmetic means of the eigenvalues of the sample covariance matrix. The newly defined test has many desirable general asymptotic properties, such as normality and consistency when $(n, p) \to \infty$. Our simulation results show that the new test is comparable to, and in some cases more powerful than, the tests for sphericity in the current literature.

1. Introduction

Many applications of modern multivariate statistics involve a large number of variables, p, and hence a large covariance matrix. In many situations (e.g. DNA microarray data) the dimensionality exceeds that of the number of observations, $N = n+1$. In this article, we discuss much of the previous work in developing statistics for testing if the covariance matrix is proportional to the identity, more commonly called *Sphericity*. We consider $\mathbf{X}_1, \ldots, \mathbf{X}_N$ as a set of independent observations from a multivariate normal distribution $N_p(\boldsymbol{\mu}, \Sigma)$, where both the mean vector $\boldsymbol{\mu} \in \mathcal{R}^p$ and covariance matrix $\Sigma > 0$ are unknown. We are interested in testing $H_0 : \Sigma = \sigma^2 I$ vs. $H_A : \Sigma \neq \sigma^2 I$, where σ^2 is the unknown scalar proportion. The classical hypothesis testing techniques are based on the likelihood ratio and are degenerate when $p > n$. Motivated by the previous work in the literature, we define a new test statistic under the framework known as *general asymptotics* or (n, p)-asymptotics.

Much of the current work rests on the large body of literature regarding asymptotics for eigenvalues of the sample covariance matrix, such as Arharov [2], Bai [3], Narayanaswamy and Raghavarao [12], Serdobolskii [18,17], Silverstein [20], Yin and Krishnaiah [24] and others. We build on the substantial list of work completed on statistical testing involving large random matrices, such as Bai et al. [4], Saranadasa [13] and most recently the work completed by Ledoit and Wolf [11], Srivastava [21–23] and Schott [14–16].

Ledoit and Wolf [11] show the locally best invariant test based on John's U statistic, see [10], to be (n, p)-consistent when $p/n \to c < \infty$, where c is a constant known as the concentration. However the distribution of the test statistic under the alternative hypothesis is not available. Like that in Ledoit and Wolf [11], Srivastava [21] proposes a test based on the first and second arithmetic means of the eigenvalues of the sample covariance but only requires the more general condition $n = O(p^\delta)$, $0 < \delta \leq 1$. He shows that the test is (n, p)-consistent and provides the distribution of the test statistic under both the null and alternative hypotheses. In [22], he proposes a modified version of the Likelihood Ratio Test (LRT) in which only the first n eigenvalues are used. This test is applicable under the assumptions $n/p \to 0$ and n fixed.

* Corresponding author.
E-mail address: fishertho@umkc.edu (T.J. Fisher).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.07.004

Motivated by the previous literature, we propose a test based on the second and fourth arithmetic means of the eigenvalues of the sample covariance matrix. The asymptotic distribution of the test statistic under both the null and alternative hypotheses is provided. The proposed the test is consistent under the *general asymptotics* framework. Furthermore, we provide a brief simulation study validating our theoretical work, demonstrating an improvement over the previous literature under certain alternative hypotheses and apply our test statistic to real microarray data. Lastly, we provide concluding remarks.

2. Description of the new test

Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_N \sim N_p(\boldsymbol{\mu}, \Sigma)$, $N = n + 1$, and we are interested in testing,

$$H_0 : \Sigma = \sigma^2 I \text{ vs. } H_A : \Sigma \neq \sigma^2 I.$$

Like that of the Likelihood Ratio Test in classical multivariate statistics, testing remains invariant under the transformation $x \to Gx$, where G is an orthogonal matrix. The test is also invariant under the scalar transformation $x \to cx$; thus we may assume without loss of generality $\Sigma = \text{diag}(\lambda_1, \ldots, \lambda_p)$. From the Cauchy–Schwarz inequality, it follows that

$$\left(\sum_{i=1}^{p} \lambda_i^r\right)^2 \le p\left(\sum_{i=1}^{p} \lambda_i^{2r}\right),$$

with equality holding if and only if $\lambda_1 = \cdots = \lambda_p = \lambda$, for all $i = 1, \ldots, p$ and some constant λ. Thus, we may consider testing $H_0 : \psi_r = 1$ vs. $H_A : \psi_r > 1$ with

$$\psi_r = \frac{\left(\sum_{i=1}^{p} \lambda_i^{2r}/p\right)}{\left(\sum_{i=1}^{p} \lambda_i^{r}/p\right)^2}. \tag{1}$$

We note this test is based on the ratio of arithmetic means of the sample eigenvalues. Srivastava [21] considers the case where $r = 1$, we look at the case of $r = 2$.

We make the following assumptions

(A) As $p \to \infty$, $a_i \to a_i^0$, $\quad 0 < a_i^0 < \infty$, $i = 1, \ldots, 16$,

(B) As $(n, p) \to \infty$, $\dfrac{p}{n} \to c \quad$ where $0 < c < \infty$,

where

$$a_i = \frac{1}{p}\text{tr}\,\Sigma^i = \frac{1}{p}\sum_{j=1}^{p} \lambda_j^i$$

and the λ_js are the eigenvalues of the covariance matrix, i.e. a_i is the ith arithmetic mean of the eigenvalues of the covariance matrix.

Theorem 1. *An unbiased and (n, p)-consistent estimator of $a_4 = \sum_{i=1}^{p} \lambda_i^4/p$ is given by*

$$\hat{a}_4 = \frac{\tau}{p}[\text{tr}\,S^4 + b \cdot \text{tr}\,S^3 \text{tr}\,S + c^* \cdot (\text{tr}\,S^2)^2 + d \cdot \text{tr}\,S^2 (\text{tr}\,S)^2 + e \cdot (\text{tr}\,S)^4] \tag{2}$$

where

$$b = -\frac{4}{n}, \qquad c^* = -\frac{2n^2 + 3n - 6}{n(n^2 + n + 2)}, \qquad d = \frac{2(5n + 6)}{n(n^2 + n + 2)}, \qquad e = -\frac{5n + 6}{n^2(n^2 + n + 2)},$$

and

$$\tau = \frac{n^5(n^2 + n + 2)}{(n + 1)(n + 2)(n + 4)(n + 6)(n - 1)(n - 2)(n - 3)}.$$

Proof. From Lemma 3 in the Appendix,

$$\begin{aligned} E[\hat{a}_4] &= \tau\left(\frac{n(n + 2)(n + 4)(n + 6)(n - 1)(n - 2)(n - 3)(n + 1)}{pn^6(n^2 + n + 2)} \sum_{i=1}^{p} \lambda_i^4\right) \\ &= \frac{1}{p}\sum_{i=1}^{p} \lambda_i^4 = a_4. \end{aligned}$$

Using the asymptotic behavior of the variance of $\hat{a}_4$ from Appendix A.5 and an application of Chebyshev's inequality completes the result:

$$\begin{aligned} P\left[\left|\hat{a}_4 - a_4\right| > \epsilon\right] &\leq \frac{1}{\epsilon^2}\mathrm{Var}[\hat{a}_4] \\ &\simeq \frac{1}{\epsilon^2}\left(\frac{32}{np}a_8 + \frac{32}{n^2}\left(a_6a_2 + a_4^2\right) + \frac{16}{n^2}(a_4^2) + \frac{64}{n^2}(ca_4a_2^2/2 + ca_3^2a_2) + \frac{8}{n^2}(c^2a_2^4) + 2\frac{32}{n^2}(a_5a_3)\right) \\ &\to 0 \quad \text{as } (n,p) \to \infty. \quad \square \end{aligned}$$

Srivastava [21] provides an unbiased and consistent estimator for a_2 which is

$$\hat{a}_2 = \frac{n^2}{(n-1)(n+2)}\frac{1}{p}\left[\mathrm{tr}\, S^2 - \frac{1}{n}(\mathrm{tr}\, S)^2\right]. \tag{3}$$

Thus an (n, p)-consistent estimator for ψ_2 is provided by

$$\hat{\psi}_2 = \frac{\hat{a}_4}{\hat{a}_2^2}.$$

The derivation and justification for our estimator $\hat{a}_4$ in (2) is provided in the Appendix. The following theorem and corollary provide the asymptotic distribution under the alternative and null hypotheses. We remind the reader that c is the concentration, not to be confused with the constant c^* in Theorem 1.

Theorem 2. *Under assumptions* (A) *and* (B), *as* $(n, p) \to \infty$

$$\frac{n}{\sqrt{8(8+12c+c^2)}}\left(\frac{\hat{a}_4}{\hat{a}_2^2} - \psi_2\right) \xrightarrow{D} N(0, \xi_2^2),$$

with

$$\xi_2^2 = \frac{1}{(8+12c+c^2)a_2^6}\left(\frac{4}{c}a_4^3 - \frac{8}{c}a_4a_2a_6 - 4a_4a_2a_3^2 + \frac{4}{c}a_2^2a_8 + 4a_6a_2^3 + 8a_2^2a_5a_3 + 4ca_4a_2^4 + 8ca_3^2a_2^3 + c^2a_2^6\right). \tag{4}$$

Proof. The result follows from Proposition 2 and an application of the delta-method with some additional algebra. □

Corollary 1. *Under the null hypothesis,* $\psi_2 = 1$, *and under the assumptions* (A) *and* (B), *as* $(n, p) \to \infty$

$$T = \left(\frac{n}{\sqrt{8(8+12c+c^2)}}\right)\left(\frac{\hat{a}_4}{\hat{a}_2^2} - 1\right) \xrightarrow{D} N(0, 1). \tag{5}$$

Proof. Under H_0, each $\lambda_i = \lambda$, for $i = 1, \ldots, p$ and some constant λ. Thus $\xi_2^2 = 1$, which completes the proof. □

From the asymptotic distribution under the alternative hypothesis we are able to determine the (n, p)-asymptotic behavior of the power function of our test statistic.

Theorem 3. *Under assumptions* (A) *and* (B), *as* $(n, p) \to \infty$ *the test statistic* T *in* (5) *is* (n, p)*-consistent.*

Proof. For large n and p, the power function of T is

$$\mathrm{Power}_\alpha(T) \simeq \Phi\left(\frac{n\left(\frac{a_4}{a_2^2} - 1\right)}{\xi_2\sqrt{8(8+12c+c^2)}} - \frac{z_\alpha}{\xi_2}\right).$$

Under assumptions (A) and (B), we know ξ_2^2 from (4) is constant. From the properties of $\Phi(\cdot)$, it is clear that $\mathrm{Power}_\alpha(T) \to 1$ as $(n, p) \to \infty$. □

3. Simulation study

A simulation study shows the effectiveness of our test statistic. We first provide a study verifying the normality of our test statistic by simulating the Attained Significance Level (ASL), or size, of our newly defined test statistic. Draw an independent sample of size $N = n + 1$ from a valid p-dimensional normal distribution under the null hypothesis (i.e. each $\lambda_i = 1$). Replicate this 1000 times. Using T from (5) we calculate

$$\mathrm{ASL}(T) = \frac{(\#T > z_\alpha)}{1000},$$

Table 1
ASL for T in (5).

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.036*	0.040	0.050	0.056
$n = 50$	0.050	0.061	0.050	0.058
$n = 100$	0.060	0.052	0.048	0.054
$n = 150$	0.049	0.048	0.049	0.047
$n = 200$	0.047	0.055	0.057	0.051

Table 2
ASL for T_s from [21].

$p = cn$	$c = 1$	$c = 2$	$c = 4$	$c = 5$
$n = 25$	0.050	0.067*	0.057	0.057
$n = 50$	0.055	0.049	0.051	0.053
$n = 100$	0.057	0.053	0.056	0.060
$n = 150$	0.054	0.046	0.050	0.040
$n = 200$	0.041	0.043	0.052	0.042

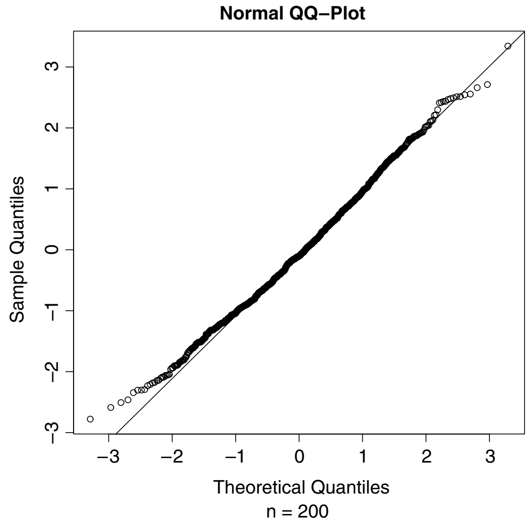

Fig. 1. Normal QQ-Plot for T in (5) under H_0.

denoting the ASL of T where z_α is the upper $100\alpha\%$ critical point of the standard normal distribution. We test with $\alpha = 0.05$. Table 1 provides the results for an assortment of $c = p/n$ values for our newly defined test statistic. Table 2 provides analogous results for the test statistic defined in [22], denoted T_s. Ledoit and Wolf [11] provide similar results for their test statistic based on John's U statistic, denoted U_J. Only in one case in each table do we have a simulated size that is significantly different (see * in Tables 1 and 2) than the predicted size of 0.05. We also look at QQ-Plots for the test statistic T under both the null and alternative hypotheses. Begin by sampling $N = n + 1 = 201$ observations from a $p = 400$ dimensional normal distribution with mean zero vector and an identity covariance matrix, hence $\lambda_i = 1$ for all i. Calculate the test statistic, T, and repeat the process 1000 times. Fig. 1 shows the QQ-Plot of the 1000 observed values of the test statistic under the null hypothesis. Similarly we repeat the simulation under the alternative hypothesis with $\Sigma = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$ with each $\lambda_i \sim U(0.5, 1.5)$ and $n = 200$, $p = 400$. Fig. 2 shows the results for the 1000 observed values of the test statistic. In both cases, the normality result appears to be satisfied by the QQ-Plots for large n and p validating the theoretical result.

Lastly a series of power simulations to confirm the consistency of our test and to demonstrate its improved performance under certain alternative hypotheses is performed. From our simulation studies it appears the newly proposed test statistic performs well when only a few elements of the covariance matrix are different. Define *near* spherical matrices to be of the form,

$$\Sigma = \begin{pmatrix} \Theta & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

where Θ is a $k \times k$ diagonal matrix, $k < p$, with all elements $\theta_i \neq 1$. $\mathbf{I}$ is a $(p - k) \times (p - k)$ identity matrix and $\mathbf{0}$ is a $(p - k)$-vector of zeros. k is chosen to be small, so the *near* spherical matrix will be the identity with the exception of a few elements.

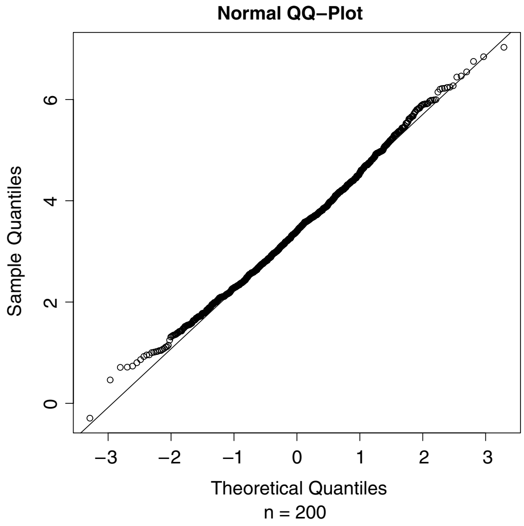

Fig. 2. Normal QQ-Plot for T in (5) under H_A.

Table 3
Simulated power under *near* spherical covariance matrix with $k = 1$.

$p = cn$	$c = 1, \theta = 3$			$c = 2, \theta = 4$		
	Our T	T_s	U_J	Our T	T_s	U_J
$n = 25$	0.505	0.427	0.436	0.580	0.463	0.521
$n = 50$	0.647	0.489	0.633	0.750	0.599	0.773
$n = 100$	0.794	0.529	0.794	0.901	0.641	0.904
$n = 150$	0.858	0.565	0.845	0.938	0.680	0.940
$n = 200$	0.903	0.624	0.912	0.969	0.710	0.975

Table 4
Simulated power with $k = 6$, $\Theta = \text{diag}\,(0.75, 1.25, 1.75, 2.25, 2.75, 3.25)$.

$p = cn$	$c = 1$			$c = 2$		
	Our T	T_s	U_J	Our T	T_s	U_J
$n = 25$	0.609	0.722	0.548	0.416	0.495	0.384
$n = 50$	0.908	0.895	0.895	0.692	0.630	0.695
$n = 100$	0.991	0.974	0.992	0.849	0.722	0.846
$n = 150$	0.999	0.988	0.999	0.899	0.749	0.907
$n = 200$	1.000	0.999	1.000	0.938	0.770	0.938

To make comparisons with the test statistics defined in [21,11], we perform a similar test to that described in [22]. A simulation is used to obtain the critical point of our test statistic (and that from [21,11]). Letting $N = n + 1$ and p increase such that $p/n \to c$, we compute, under $H_0 : \Sigma = I$, 1000 simulated observed values our test statistic T and find T_α such that

$$P(T > T_\alpha) = \alpha.$$

T_α is the estimated critical point at significance level α. The same is repeated for the test statistics described in [21,11]. Then simulate from a p-dimensional normal distribution with zero mean vector and a *near* spherical covariance matrix.

We provide examples for two cases of *near* sphericity. Table 3 shows two results for the case where $k = 1$, or Θ is a scalar of element θ. Each element of the covariance matrix is the same, with the exception of one element. Two examples are provided, $\theta = 3$ with $c = 1$ and $\theta = 4$ with $c = 2$. Table 4 provides two results, $c = 1$ and $c = 2$, for the case where $k = 6$ elements differ from the spherical model, i.e. $\Theta = \text{diag}\,(0.75, 1.25, 1.75, 2.25, 2.75, 3.25)$. Tables 3 and 4 show that all three test statistics appear to be consistent as $(n, p) \to \infty$ and that, under the simulated *near* spherical alternative hypothesis, our newly defined test is more powerful than that described in [21] and is comparable to that described in [11]. Simulation studies with other covariance matrices under the alternative hypothesis are available in [8]. They show consistency of the test statistics. The best performing test varies depending on the covariance matrix under the alternative hypothesis.

Lastly we study the effect of θ in the case where $k = 1$. Table 3 indicates our newly proposed statistic is comparable to that of Ledoit and Wolf [11] and tends to perform better than Srivastava [21]. In this study we let $n = 50$, $c = 3$ and the value of θ increases. Fig. 3 provides the simulated power after 1000 runs for our newly proposed test and that of Srivastava [21].

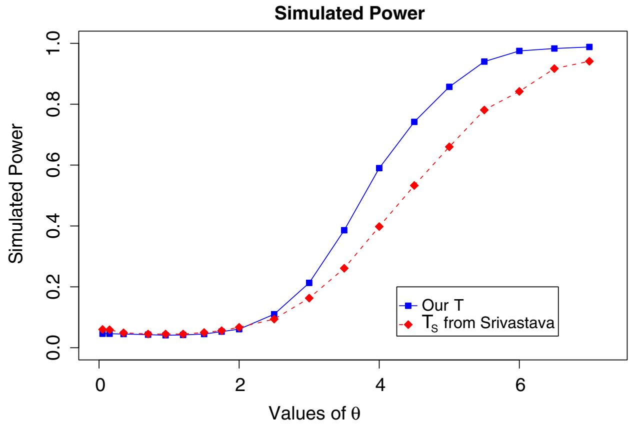

Fig. 3. Simulated power of T and T_s as θ increases.

We see from the plot that neither test performs particularly well when θ is small. As θ increases the newly proposed test appears to dominate that of Srivastava. Both appear to become consistent as θ grows.

4. Data examples

In this section, for a further comparison we test the hypothesis of sphericity against two *classic* data sets. We follow the preprocessing protocol attributed to Dudoit et al. [6] and Dettling and Bühlmann [5] by thresholding, filtering, and a logarithmic transformation but do not follow standardization so as to compare to the results in [22]. Preprocessed data are available at the website of Prof. Tatsuya Kubokawa: http://www.e.u-tokyo.ac.jp/~tatsuya/index.html (last accessed: 27 April 2010).

4.1. Colon dataset

In this dataset, expression levels of 6500 human genes are measured using Affymetrix microarray technology on 40 tumors and 22 normal colon tissues. A selection of 2000 genes with the highest minimal intensity across the samples has been made by Alon et al. [1]. Our dimensionality is $p = 2000$ and the degrees of freedom available to estimate the covariance matrix is only 60. The data is further described and is available at the Princeton Oncology website. Calculate an estimate of the covariance matrix using a pooled covariance matrix with 60 degrees of freedom. We compute test values of $T = 185.8071$ from (5), $T_s = 2771.6538$, and $U_J = 2816.2916$ where T_s is the sphericity test from Srivastava [21] and U_J is that from Ledoit and Wolf [11], respectively. In each case we get a p-value ≈ 0 indicating any assumption of sphericity in the case of these data to be false.

4.2. Leukemia dataset

This dataset contains gene expression levels of 72 patients either suffering from acute lymphoblastic leukemia or acute myeloid leukemia. There are 47 and 25 patients for each respective case and they are obtained on Affymetric oligonucleotide microarrays. The data is attributed to Golub et al. [9]. The data is comprised of $p = 3571$ genes and the degrees of freedom available are only 70. The data is available and described further at the Broad Institute's website. The leukemia data is preprocessed in the same way and we get the observed test statistic values of $T = 242.4386$, $T_s = 2294.9184$, and $U_J = 2326.7520$ for T in (5), T_s from Srivastava [21] and U_J from Ledoit and Wolf [11], respectively. In each case we get a p-value ≈ 0 indicating any assumption of sphericity in the case of these data being false.

5. Concluding remarks

We have proposed a new test for sphericity of the covariance matrix. Like that of Srivastava [21], our test is based on the Cauchy–Schwarz inequality. Unlike Johns U-statistic and Srivastava's T_s test, we look at the second and fourth arithmetic means of the sample eigenvalues. Simulations indicate that the newly defined test statistic, T in (5), appears to perform better in some *near* spherical cases and is comparable to tests in the previous literature.

5.1. Notes on assumptions and limitations

The two underlying assumptions, (A) and (B), are comparable to that of Ledoit and Wolf [11], with the exception that the sixteenth arithmetic mean of the covariance matrix is assumed to be convergent as $p \to \infty$. Both our test and that of Ledoit and Wolf [11] require $p/n \to c$ as $(n, p) \to \infty$. This assumption is more restrictive than that in [21] but does not appear to hinder the application of the test statistic in practice, since c is easily approximated with the ratio of p to n. We further note that the requirement of convergence of the sixteenth arithmetic mean is higher than the eighth in [21] and the fourth in [11].

There is an increase in the variability of our test statistic compared to that of Srivastava [21]. As you look at higher arithmetic means, the variance increases. Although the two tests are asymptotically comparable and the newly defined test appears to be more powerful in *near* spherical cases of Σ, the larger variance of T may be a problem in certain cases.

5.2. Future work and recommendations

Our new test is of the form (1) with $r = 2$. This builds upon the work of Srivastava [21] who defined a test based on $r = 1$. Future work may look at $r = 3, 4, \ldots$. We conjecture that, although more powerful in certain alternative hypotheses, these test will make more restrictive assumptions and the variance of the corresponding test statistic will grow to the point where it may be infeasible to use the statistic. In the case of r being a fraction (e.g. $r = 0.5$) we suspect the test may show an improvement in some cases of Σ and in general will not be hindered by infeasible assumptions and a large variance. However, we suspect the distribution of terms like $\hat{a}_{1/2}$ to be difficult to determine and we leave this question open.

Although each of the tests described is asymptotically comparable, each test seems to perform better under certain alternative hypotheses. We recommend our newly defined test, T in (5), when a *near* spherical covariance matrix is suspected.

Acknowledgment

The first author would like to thank Professor Muni Srivastava for his helpful correspondence and direction when requested.

Appendix

A.1. Expression of estimator for a_4

Obtain expressions for $\operatorname{tr} S$, $\operatorname{tr} S^2$, $(\operatorname{tr} S)^2$, $\operatorname{tr} S^2(\operatorname{tr} S)^2$, $\operatorname{tr} S^3 \operatorname{tr} S$, $\operatorname{tr} S^4$ and $(\operatorname{tr} S)^4$ in terms of chi-squared random variables. We make use of the following well-known result from [17].

Lemma 1. *Consider the sample covariance matrix and recalling $N = n + 1$,*

$$S = \frac{1}{n} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})'.$$

There exists an orthogonal transformation of vectors

$$y_k = \sum_{i=1}^{N} \Omega_{ki} x_i,$$

such that the vectors $y_N = \sqrt{N}\bar{x}$ and $y_k \sim N(0, \Sigma)$, $k = 1, \ldots, n$, are independent, and the sample covariance matrix is equal to

$$S = \frac{1}{n} \sum_{i=1}^{n} y_i y_i'.$$

Let $nS = YY' \sim W_p(\Sigma, n)$, where $Y = (y_1, y_2, \ldots, y_n)$ and each $y_i \sim N_p(0, \Sigma)$ and independent. By orthogonal decomposition, $\Sigma = \Gamma' \Lambda \Gamma$, where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ with λ_i being the ith eigenvalue of Σ and Γ is an orthogonal matrix. Define $U = (u_1, u_2, \ldots, u_n)$, where u_i are i.i.d. $N_p(0, I)$ and we can write $Y = \Sigma^{1/2} U$ where $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$. Define $W' = (w_1, w_2, \ldots, w_p) = U'\Gamma'$ and each w_i are i.i.d. $N_n(0, I)$.

Define $v_{ii} = w_i' w_i$ and it is easy to see that each v_{ii} is an i.i.d. chi-squared random variable with n degrees of freedom. Thus, $n \operatorname{tr} S = \operatorname{tr} W' \Lambda W$.

From [21] we get the following important results

$$n \operatorname{tr} S = \sum_{i=1}^{p} \lambda_i v_{ii},$$

$$n^2 (\operatorname{tr} S)^2 = \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^{p} \lambda_i \lambda_j v_{ii} v_{jj},$$

and

$$n^2 \mathrm{tr}\, S^2 = \sum_{i=1}^{p} \lambda_i^2 v_{ii}^2 + 2 \sum_{i<j}^{p} \lambda_i \lambda_j v_{ij}^2.$$

Using the same approach and the commutative property of the trace operation (i.e. tr $(ABC) = \mathrm{tr}\,(CAB)$), we derive,

$$\begin{aligned} n^4 \mathrm{tr}\, S^4 &= \mathrm{tr}\,(W'\Lambda W)(W'\Lambda W)(W'\Lambda W)(W'\Lambda W) \\ &= \mathrm{tr}\left[\left(\sum_{i=1}^{p} \lambda_i w_i w_i'\right)^4\right] \\ &= \sum_{i=1}^{p} \lambda_i^4 v_{ii}^4 + 4\sum_{i\neq j}^{p} \lambda_i^3 \lambda_j v_{ii}^2 v_{ij}^2 + \sum_{i<j}^{p} \lambda_i^2 \lambda_j^2 (4 v_{ii} v_{jj} v_{ij}^2 + 2 v_{ij}^4) + \sum_{i\neq j<k}^{p} \lambda_i^2 \lambda_j \lambda_k (8 v_{ii} v_{ij} v_{jk} v_{ik} + 4 v_{ij}^2 v_{ik}^2) \\ &\quad + \sum_{i<j<k<l}^{p} \lambda_i \lambda_j \lambda_k \lambda_l (8 v_{ij} v_{jk} v_{kl} v_{il} + 8 v_{ij} v_{jl} v_{kl} v_{ik} + 8 v_{ik} v_{jk} v_{jl} v_{il}). \end{aligned}$$

Likewise we find analogous results for $n^4 \mathrm{tr}\, S^3 \mathrm{tr}\, S$, $n^4 (\mathrm{tr}\, S^2)^2$, $n^4 \mathrm{tr}\, S^2 (\mathrm{tr}\, S)^2$, and $n^4 (\mathrm{tr}\, S)^4$. Consider the constants b, c^*, d, e defined in Theorem 1, then rewrite

$$\frac{\mathrm{tr}\, S^4 + b \mathrm{tr}\, S^3 \mathrm{tr}\, S + c^* (\mathrm{tr}\, S^2)^2 + d \mathrm{tr}\, S^2 (\mathrm{tr}\, S)^2 + e (\mathrm{tr}\, S)^4}{p} = \eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5,$$

where

$$\eta_1 = \frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2+n+2)} \frac{1}{n^4 p} \sum_{i=1}^{p} \lambda_i^4 v_{ii}^4, \tag{6}$$

$$\eta_2 = \frac{4}{n^4 p} \sum_{i\neq j}^{p} \lambda_i^3 \lambda_j \left(\frac{v_{ii}^2 v_{ij}^2 (n^4 - 4n^3 + n^2 + 6n) + v_{ii}^3 v_{jj} (-n^3 + 4n^2 - n - 6)}{n^2(n^2+n+2)} \right), \tag{7}$$

$$\begin{aligned} \eta_3 &= \frac{2}{n^4 p} \sum_{i<j}^{p} \lambda_i^2 \lambda_j^2 \left(\frac{v_{ii} v_{ij}^2 v_{jj} (2n^4 - 10n^3 + 12n^2)}{n^2(n^2+n+2)} \right. \\ &\quad \left. + \frac{v_{ij}^4 (n^4 - 3n^3 - 4n^2 + 12n) + v_{ii}^2 v_{jj}^2 (-2n^3 + 7n^2 + 3n - 18)}{n^2(n^2+n+2)} \right), \end{aligned} \tag{8}$$

$$\begin{aligned} \eta_4 &= \frac{4}{n^4 p} \sum_{i\neq j<k}^{p} \lambda_i^2 \lambda_j \lambda_k \left(\frac{v_{ii}^2 v_{jk}^2 (-2n^3 + 2n^2 + 12n)}{n^2(n^2+n+2)} \right. \\ &\quad + \frac{(v_{ii} v_{ij}^2 v_{kk} + v_{ii} v_{ik}^2 v_{jj})(-3n^3 + 7n^2 + 6n) + v_{ii}^2 v_{jj} v_{kk} (5n^2 - 9n - 18)}{n^2(n^2+n+2)} \\ &\quad \left. + \frac{v_{ii} v_{ij} v_{ik} v_{jk} (2n^4 - 4n^3 - 2n^2 - 12n) + v_{ij}^2 v_{ik}^2 (n^4 - 3n^3 - 4n^2 + 12n)}{n^2(n^2+n+2)} \right), \end{aligned} \tag{9}$$

with the index read as $i \neq j$, $i \neq k$ and $j < k$, and

$$\begin{aligned} \eta_5 &= \frac{8}{n^4 p} \sum_{i<j<k<l}^{p} \lambda_i \lambda_j \lambda_k \lambda_l \left(\frac{n^2(n^2+n+2)}{n^2(n^2+n+2)} (v_{ij} v_{jk} v_{kl} v_{il} + v_{ij} v_{jl} v_{kl} v_{ik} + v_{ik} v_{jk} v_{jl} v_{il}) \right. \\ &\quad - \frac{3n(n^2+n+2)}{n^2(n^2+n+2)} (v_{ii} v_{jk} v_{jl} v_{kl} + v_{jj} v_{ik} v_{il} v_{kl} + v_{kk} v_{ij} v_{il} v_{jl} + v_{ll} v_{ij} v_{ik} v_{jk}) - \frac{n(2n^2+3n-6)}{n^2(n^2+n+2)} (v_{ij}^2 v_{kl}^2 + v_{ik}^2 v_{jl}^2 + v_{il}^2 v_{jk}^2) \\ &\quad \left. + \frac{n(5n+6)}{n^2(n^2+n+2)} (v_{ij}^2 v_{kk} v_{ll} + v_{ik}^2 v_{jj} v_{ll} + v_{il}^2 v_{jj} v_{kk} + v_{jk}^2 v_{ii} v_{ll} + v_{jl}^2 v_{ii} v_{kk} + v_{kl}^2 v_{ii} v_{jj}) - \frac{3(5n+6) v_{ii} v_{jj} v_{kk} v_{ll}}{n^2(n^2+n+2)} \right). \end{aligned} \tag{10}$$

A.2. Calculation of $E[\hat{a}_4]$

We begin by summarizing some results about the random components of our estimator.

Lemma 2. *For $v_{ii} = (w_i' w_i)$ and $v_{ij} = (w_i' w_j)$ for any $i \neq j$,*

$$E[v_{ii} v_{ij}^2] = n(n+2), \qquad E[v_{ii}^2 v_{ij}^2] = n(n+2)(n+4),$$
$$E[v_{ij}^4] = 3n(n+2), \qquad E[v_{ij}^2] = n,$$
$$E[v_{ii} v_{ij}^2 v_{jj}] = n(n+2)^2, \qquad E[v_{ij}^2 v_{ik}^2] = n(n+2),$$
$$E[v_{ij} v_{ik} v_{jk}] = n, \qquad E[v_{ij} v_{il} v_{jk} v_{kl}] = n,$$
$$E[v_{ii} v_{ij} v_{ik} v_{jk}] = n(n+2).$$

Using Lemma 2 we can easily calculate the expected value of $\hat{a}_4$.

Lemma 3. *For $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5$ in (6)–(10) respectively,*

$$E[\eta_1] = \frac{n(n+2)(n+4)(n+6)(n-1)(n-2)(n-3)(n+1)}{pn^6(n^2+n+2)} \sum_{i=1}^{p} \lambda_i^4$$

and

$$E[\eta_2] = E[\eta_3] = E[\eta_4] = E[\eta_5] = 0.$$

Proof. Using the fourth moment of a χ^2 r.v. it is easy to see the first result. Using the results in Lemma 2 it is easy to find

$$E[\eta_2] = \frac{4}{n^4 p} \sum_{i \neq j}^{p} \lambda_i^3 \lambda_j \left(\frac{n(n+2)(n+4)(n^4 - 4n^3 + n^2 + 6n)}{n^2(n^2+n+2)} + \frac{n^2(n+2)(n+4)(-n^3 + 4n^2 - n - 6)}{n^2(n^2+n+2)} \right) = 0.$$

An analogous derivation provides the result for η_3, η_4 and η_5. □

A.3. Calculation of $V[\hat{a}_4]$

To calculate the variance of the estimator in (2) we recall the moments of χ^2 and standard normal random variables when needed. We also need the following lemma.

Lemma 4. *Let Q be an orthogonal matrix such that $A_j = w_j w_j' = Q' D Q$ with $D = Q A_j Q'$ and $D = \text{diag}\,(w_j' w_j, 0, \ldots, 0)$. Given A_j, we can find $x_i = Q w_i \sim N_n(0, I)$ and it follows that x_i is independently distributed of A_j. Now $x_i = (x_{i1}, \ldots, x_{in}) = Q w_i \sim N_n(0, I)$ and thus x_{i1} is independent of $w_j' w_j$ and also x_{ik} for $k = 2, \ldots, n$, hence*

$$E[v_{ij}^2] = E[w_i' w_j w_j' w_i] = E[w_i' A_j w_i] = E[x_{i1}^2 w_j' w_j],$$

and

$$E[v_{ij}^4] = E[(w_i' w_j w_j' w_i)^2] = E[(w_i' A_j w_i)^2] = E[x_{i1}^4 (w_j' w_j)^2].$$

Proof. A_j is a function of the random variable w_j and x_i is a function of the random variable w_i. w_i and w_j are independent by definition, hence x_i and A_j are independent. Furthermore an orthogonal transformation does not alter the distribution of a normal random variable. Matrix algebra provides the remainder of the derivation. □

A.3.1. Variance of η_1

Lemma 5. *The variance of η_1 is given by*

$$V[\eta_1] = \frac{32(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)}{n^{11}(n^2+n+2)^2 p} (n^4 - 5n^3 + 5n^2 + 5n - 6)^2 a_8.$$

Proof. Find the variance of v_{ii}^4 by utilizing the expected values of the eighth and fourth moments of a χ^2 random variable, and the remainder of the algebra is as follows.

$$\begin{aligned} V[\eta_1] &= V\left[\frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^2(n^2+n+2)} \frac{1}{n^4 p} \sum_{i=1}^{p} \lambda_i^4 v_{ii}^4 \right] \\ &= \left(\frac{n^4 - 5n^3 + 5n^2 + 5n - 6}{n^6(n^2+n+2)p} \right)^2 \sum_{i=1}^{p} \lambda_i^8 V[v_{ii}^4] \\ &= \frac{32(n+2)(n+4)(n+6)(n+7)(n^2+14n+60)}{n^{11}(n^2+n+2)^2 p} (n^4 - 5n^3 + 5n^2 + 5n - 6)^2 a_8. \quad \square \end{aligned}$$

A.3.2. Variance of η_2

We provide details on the derivation of the variance of η_2 in (7). From Lemma 3, we know $E[\eta_2] = 0$ and hence $V[\eta_2] = E[\eta_2^2]$. Rewrite η_2 in the form

$$\eta_2 = \frac{C_2}{p}\sum_{i\neq j}^{p}\lambda_i^3\lambda_j V_{ij},$$

where

$$C_2 = \frac{4}{n^6(n^2+n+2)},$$

and

$$V_{ij} = v_{ii}^2 v_{ij}^2 n_1 + v_{ii}^3 v_{jj} n_2, \tag{11}$$

where $n_1 = n^4 - 4n^3 + n^2 + 6n$ and $n_2 = -n^3 + 4n^2 - n - 6$. Then η_2^2 can be expressed as,

$$\frac{p^2}{C_2^2}\eta_2^2 = \sum_{i\neq j}^{p}\lambda_i^6\lambda_j^2 V_{ij}^2 + 2\sum_{i\neq j<k}^{p}\lambda_i^6\lambda_j\lambda_k V_{ij}V_{ik} + 2\sum_{i<j}^{p}\lambda_i^4\lambda_j^4 V_{ij}V_{ji}$$
$$+ 2\sum_{i\neq j\neq k}^{p}\lambda_i^4\lambda_j^3\lambda_k V_{ik}V_{ji} + \sum_{i<j\neq k}^{p}\lambda_i^3\lambda_j^3\lambda_k^2 V_{ik}V_{jk} + 2\sum_{i<j\neq k<l}^{p}\lambda_i^3\lambda_j^3\lambda_k\lambda_l(V_{ik}V_{jl} + V_{il}V_{jk}).$$

To compute the variance of η_2 we simply calculate the expectation of each component above. Much of this derivation follows from the moments of χ^2 and standard normal random variables and by application of Lemma 4. The results for each V_{ij} type component are provided.

Lemma 6. *For V_{ij} defined in (11),*

$$E[V_{ij}^2] = 2n^2(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6)(n+8)(n+10),$$
$$E[V_{ij}V_{ji}] = 2n^2(n+1)^2(n+2)(n+4)^2(n+6)^2(n-1)(n-2)^2(n-3)^2,$$
$$E[V_{ij}V_{ik}] = E[V_{ik}V_{ji}] = E[V_{ik}V_{jk}] = 0,$$
$$E[V_{ik}V_{jl}] = E[V_{ik}]E[V_{jl}] = 0.$$

This leads to the result.

Lemma 7. *The variance of η_2 is provided by*

$$V(\eta_2) = \frac{32(n-1)(n-2)^2(n-3)^2(n+1)^2(n+2)(n+4)(n+6)}{n^{10}(n^2+n+2)^2} \times \left((n^2+18n+80)a_6a_2 + (n^2+10n+24)a_4^2 - \frac{2(n^2+14n+52)}{p}a_8\right).$$

Proof. Using the expected values from Lemma 6 and the following derivations provides the result:

$$\sum_{i\neq j}^{p}\lambda_i^6\lambda_j^2 = \left(\sum_{i=1}^{p}\lambda_i^6\right)\left(\sum_{j=1}^{p}\lambda_j^2\right) - \left(\sum_{i=1}^{p}\lambda_i^8\right) = p^2a_6a_2 - pa_8 = p(pa_6a_2 - a_8),$$

and

$$2\sum_{i<j}^{p}\lambda_i^4\lambda_j^4 = \left(\sum_{i=1}^{p}\lambda_i^4\right)\left(\sum_{j=1}^{p}\lambda_j^4\right) - \left(\sum_{i=1}^{p}\lambda_i^8\right) = p^2a_4^2 - pa_8 = p(pa_4^2 - a_8). \quad \square$$

A.3.3. Variance of η_3, η_4 and η_5

Following the same derivation in the calculation for the variance of the η_2 term in Appendix A.3.2 we can find the variance of η_3, η_4 and η_5. We leave out the tedious algebraic details, available in [8], and provide the results.

Lemma 8. *The variances of η_3, η_4 and η_5 are*

$$V[\eta_3] = \frac{16(n-1)(n-2)^2(n-3)^2(n+1)(n+2)(n+4)(n+6)}{n^{10}(n^2+n+2)^2}(n^3+15n^2+69n+54)\left(a_4^2-\frac{a_8}{p}\right),$$

$$V[\eta_4] = \frac{64(n+2)}{n^{10}(n^2+n+2)^2}\Bigg[\frac{pa_4a_2^2-a_4^2-2a_6a_2+\frac{2}{p}a_8}{2}(n^{10}+12n^9-138n^7-81n^6+3102n^5+200n^4$$
$$-5316n^3-912n^2-144n+3456)+(n+4)\left(pa_3^2a_2-a_6a_2-2a_5a_3+\frac{2}{p}a_8\right)$$
$$\times(n^9+5n^8-44n^7-166n^6+493n^5-79n^4-1554n^3+1380n^2+792n-864)\Bigg],$$

$$V[\eta_5] = \frac{8(n-1)(n+2)}{n^9(n^2+n+2)}(n^5+6n^4+9n^3-56n^2+132n+144)\left(p^2a_2^4-6pa_4a_2^2+8a_6a_2+3a_4^2-\frac{6}{p}a_8\right).$$

Proof. The expectation of the individual components is straightforward and similar to the methodology in Appendix A.3.2. Note the following

$$\begin{aligned} 2\sum_{i\neq j<k}^{p}\lambda_i^4\lambda_j^2\lambda_k^2 &= \left(\sum_{i=1}^{p}\lambda_i^4\right)\left(\sum_{j=1}^{p}\lambda_j^2\right)^2-\left(\sum_{i=1}^{p}\lambda_i^4\right)^2-2\left(\sum_{i=1}^{p}\lambda_i^6\right)\left(\sum_{j=1}^{p}\lambda_j^2\right)+2\left(\sum_{i=1}^{p}\lambda_i^8\right)\\ &= p^3a_4a_2^2-p^2a_4^2-2p^2a_6a_2+2pa_8\\ &= p(p^2a_4a_2^2-pa_4^2-2pa_6a_8+2a_8),\\ 2\sum_{i\neq j<k}^{p}\lambda_i^2\lambda_j^3\lambda_k^3 &= \left(\sum_{i=1}^{p}\lambda_i^3\right)^2\left(\sum_{j=1}^{p}\lambda_j^2\right)-\left(\sum_{i=1}^{p}\lambda_i^6\right)\left(\sum_{i=1}^{p}\lambda_i^2\right)-2\left(\sum_{i=1}^{p}\lambda_i^5\right)\left(\sum_{j=1}^{p}\lambda_j^3\right)+2\left(\sum_{i=1}^{p}\lambda_i^8\right)\\ &= p^3a_3^2a_2-p^2a_6a_2-2p^2a_5a_3+2pa_8\\ &= p(p^2a_3^2a_2-pa_6a_2-2pa_5a_3+2a_8), \end{aligned}$$

and

$$\begin{aligned} 24\sum_{i<j<k<l}^{p}\lambda_i^2\lambda_j^2\lambda_k^2\lambda_l^2 &= p^4a_2^4-6p^3a_4a_2^2+8p^2a_6a_2+3p^2a_4^2-6pa_8\\ &= p(p^3a_2^4-6p^2a_4a_2^2+8pa_6a_2+3pa_4^2-6a_8). \quad \square \end{aligned}$$

A.3.4. Covariance terms of η_1, η_2, η_3, η_4 and η_5

To determine the covariance terms of η_1, with η_2, η_3, η_4, and η_5 we utilize the fact that $E[\eta_i] = 0$ for $i = 2, 3, 4, 5$. Therefore

$$\mathrm{Cov}(\eta_1,\eta_2) = E[\eta_1\eta_2],$$

and due to the independence of many of the random terms in the η_i's, we only have to explore the variables of the form $v_{ii}^4V_{ij}$ and $v_{jj}^4V_{ij}$ (i.e. v_{ii} and V_{jk} are independent). Recall V_{ij} from (11) and see

$$\begin{aligned} v_{ii}^4V_{ij} &= v_{ii}^4\left(v_{ii}^2v_{ij}^2n_1+v_{ii}^3v_{jj}n_2\right)\\ &= v_{ii}^6v_{ij}^2n_1+v_{ii}^7v_{jj}n_2, \end{aligned}$$

where the v_{ii}^4 component from η_1 essentially adds four moments to the random variable. Taking expectations we see,

$$\begin{aligned} E\left[v_{ii}^4V_{ij}\right] &= n(n+2)(n+4)(n+6)(n+8)(n+10)(n+12)(n_1+nn_2)\\ &= 0 \quad \text{with } n_1, n_2 \text{ defined in (11).} \end{aligned}$$

A similar results holds for $v_{jj}^4V_{ij}$ except fourth moments of the v_{jj}s are included. This concept can easily be seen in the results of Lemma 2 in Appendix A.2, specifically with the expected values of $v_{ii}v_{ij}^2$ and $v_{ii}^2v_{ij}^2$. The additional v_{ii} will add an additional moment resulting in the $(n+4)$ in the expected value. In the case of η_1 and η_2, we add a fourth moment of v_{ii} and v_{jj} in the respective calculations to both parts of V_{ij} in (11). Since both expectations are zero, and the other terms are zero by independence, we determine $\mathrm{Cov}(\eta_1,\eta_2) = 0$. Analogous results hold for the covariance terms of η_1 with η_3, η_4 and η_5 respectively.

When exploring $\mathrm{Cov}(\eta_2, \eta_3)$ we find that some of random components of η_2 and η_3 interact. Derivation similar to that of Appendix A.3.2 leads to the result,

$$\mathrm{Cov}(\eta_2, \eta_3) = \frac{32(n-1)(n-2)^2(n-3)^2(n+1)^2}{n^{10}(n^2+n+2)^2}(n+2)(n+4)(n+6)^2(n+8)\left(a_5a_3 - \frac{a_8}{p}\right).$$

Similar work reveals no other correlated terms, hence $\mathrm{Cov}(\eta_2, \eta_4) = 0$ and $\mathrm{Cov}(\eta_2, \eta_5) = 0$. We also find $\mathrm{Cov}(\eta_3, \eta_4) = 0$, $\mathrm{Cov}(\eta_3, \eta_5) = 0$ and $\mathrm{Cov}(\eta_4, \eta_5) = 0$.

A.4. Covariance terms of $\hat{a}_4$ and $\hat{a}_2$

Begin by recalling a result for $\hat{a}_2$ in (3) from [21], since $n^2/(n-1)(n+2) \simeq 1$,

$$\hat{a}_2 \simeq \frac{n-1}{n^3p}\sum_{i=1}^{p}\lambda_i^2 v_{ii}^2 + \frac{2}{n^2p}\sum_{i<j}^{p}\lambda_i\lambda_j\left(v_{ij}^2 - \frac{1}{n}v_{ii}v_{jj}\right)$$
$$= q_1 + q_2.$$

The covariance between q_1 of $\hat{a}_2$ and the terms η_2, η_3, η_4, and η_5 is analogous to that of η_1 with the respective terms, resulting in

$$\mathrm{Cov}(\eta_2, q_1) = \mathrm{Cov}(\eta_3, q_1) = \mathrm{Cov}(\eta_4, q_1) = \mathrm{Cov}(\eta_5, q_1) = 0.$$

The covariance of q_1 and η_1 is a straightforward calculation resulting in

$$\mathrm{Cov}(q_1, \eta_1) = \frac{16(n+2)(n+4)(n+5)(n+6)}{n^3(n^2+n+2)p}a_6.$$

Through careful expansion and taking expectations we find there are no correlated terms between q_2 and η_1, η_4 or η_5 resulting in

$$\mathrm{Cov}(\eta_1, q_2) = \mathrm{Cov}(\eta_4, q_2) = \mathrm{Cov}(\eta_5, q_2) = 0.$$

Expansion of $E[\eta_2 q_2]$ and $E[\eta_3 q_2]$ provides the following results,

$$\mathrm{Cov}(\eta_2, q_2) = \frac{16n(n-1)(n-2)(n-3)(n+1)}{n^8(n^2+n+2)p^2}(n+2)(n+4)(n+6)\sum_{i\neq j}^{p}\lambda_i^4\lambda_j^2$$
$$= \frac{16(n-1)(n-2)(n-3)(n+1)}{n^7(n^2+n+2)}(n+2)(n+4)(n+6)\left(a_4a_2 - \frac{a_6}{p}\right),$$

and

$$\mathrm{Cov}(\eta_3, q_2) = \frac{16n(n-1)(n-2)(n-3)(n+1)}{n^8(n^2+n+2)p^2}(n+2)(n+4)(n+6)\sum_{i<j}^{p}\lambda_i^3\lambda_j^3$$
$$= \frac{8(n-1)(n-2)(n-3)(n+1)}{n^7(n^2+n+2)}(n+2)(n+4)(n+6)\left(a_3^2 - \frac{a_6}{p}\right).$$

A.5. Asymptotic variances

We simplify our variance and covariance terms by finding their asymptotic values under assumptions (A) and (B) and as $(n, p) \to \infty$,

$$V(\eta_1) \simeq \frac{32}{np}a_8 = \frac{1}{np}32a_8,$$
$$V(\eta_2) \simeq \frac{32}{n^2}\left(a_6a_2 + a_4^2 - \frac{2}{p}a_8\right) \simeq \frac{32}{n^2}(a_6a_2 + a_4^2) = \frac{1}{np}32c(a_6a_2 + a_4^2),$$
$$V(\eta_3) \simeq \frac{16}{n^2}\left(a_4^2 - \frac{a_8}{p}\right) \simeq \frac{16}{n^2}a_4^2 = \frac{1}{np}16ca_4^2,$$
$$V(\eta_4) \simeq \frac{64}{n^3}\left(\frac{pa_4a_2^2 - (a_4^2 + 2a_6a_2) + \frac{2}{p}a_8}{2} + pa_3^2a_2 - (a_6a_2 + 2a_5a_3) + \frac{2}{p}a_8\right)$$
$$\simeq \frac{64}{n^2}c(a_4a_2^2/2 + a_3^2a_2) = \frac{1}{np}64c^2(a_4a_2^2/2 + a_3^2a_2),$$
$$V(\eta_5) \simeq \frac{8}{n^4}\left(p^2a_2^4 - 6pa_4a_2^2 + (9a_6a_2 + 3a_4^2) - \frac{6}{p}a_8\right) \simeq \frac{8}{n^2}c^2a_2^4 = \frac{1}{np}8c^3a_2^4,$$

and

$$V(q_1) \simeq \frac{8}{np}a_4 = \frac{1}{np}8a_4,$$
$$V(q_2) \simeq \frac{4}{n^2}\left(a_2^2 - \frac{a_4}{p}\right) \simeq \frac{4}{n^2}a_2^2 = \frac{1}{np}4ca_2^2,$$

are provided in [21]. Likewise,

$$\text{Cov}(q_1, \eta_1) \simeq \frac{16}{np}a_6 = \frac{1}{np}16a_6,$$
$$\text{Cov}(\eta_2, \eta_3) \simeq \frac{32}{n^2}\left(a_5a_3 - \frac{a_8}{p}\right) \simeq \frac{32}{n^2}a_5a_3 = \frac{1}{np}32ca_5a_3,$$
$$\text{Cov}(\eta_2, q_2) \simeq \frac{16}{n^2}\left(a_4a_2 - \frac{a_6}{p}\right) \simeq \frac{16}{n^2}a_4a_2 = \frac{1}{np}16ca_4a_2,$$
$$\text{Cov}(\eta_3, q_2) \simeq \frac{8}{n^2}\left(a_3^2 - \frac{a_6}{p}\right) \simeq \frac{8}{n^2}a_3^2 = \frac{1}{np}8ca_3^2,$$

and we note that for τ from Theorem 1, $\tau^2 \simeq 1$ as $n \to \infty$.

A.6. Asymptotic results

To find the asymptotic distribution of our statistic, we utilize the theory of martingale-differences.

Lemma 9. *Let $X_{n,p}$ be a sequence of random variables with $\mathcal{F}_{n,p}$ the σ-field generated by the random variables $(w_1, \ldots, w_p)$, then $\mathcal{F}_{n.0} \subset \mathcal{F}_{n,1} \subset \cdots \subset \mathcal{F}_{n,p}$. If $E\left[X_{n,p}|\mathcal{F}_{n,p-1}\right] = 0$ a.s. then $(X_{n,p}, \mathcal{F}_{n,p})$ is known as a martingale-difference array. If*

(1) $\sum_{j=0}^{p} E\left[(X_{n,j})^2|\mathcal{F}_{n,j-1}\right] \xrightarrow{p} \sigma^2$ *as* $(n,p) \to \infty$.
(2) $\sum_{j=0}^{p} E\left[X_{n,j}^2 I(X_{n,j} > \epsilon)|\mathcal{F}_{n,j-1}\right] \xrightarrow{p} 0$

then $Y_{n,p} = \sum_{j=0}^{p} X_{n,p} \xrightarrow{D} N(0, \sigma^2)$.

The second condition is known as the Lindeberg condition. The result can be found in numerous texts, see [7] or [19]. The second condition can be satisfied with the stronger Lyapounov type condition

$$\sum_{j=0}^{p} E[X_{n,j}^4|\mathcal{F}_{n,j-1}] \xrightarrow{p} 0.$$

Proposition 1. *Under assumptions* (A) *and* (B), *as* $(n, p) \to \infty$

$$\sqrt{np}\begin{pmatrix} q_1 - a_2 \\ \eta_1 - a_4 \\ q_2 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{q_1}^2 & \sigma_{q_1\eta_1} & 0 & 0 & 0 & 0 & 0 \\ \sigma_{q_1\eta_1} & \sigma_{\eta_1}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{q_2}^2 & \sigma_{q_2\eta_2} & \sigma_{q_2\eta_3} & 0 & 0 \\ 0 & 0 & \sigma_{q_2\eta_2} & \sigma_{\eta_2}^2 & \sigma_{\eta_2\eta_3} & 0 & 0 \\ 0 & 0 & \sigma_{q_2\eta_3} & \sigma_{\eta_2\eta_3} & \sigma_{\eta_3}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{\eta_4}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\eta_5}^2 \end{pmatrix}\right),$$

where $\sigma_{q_1}^2, \sigma_{\eta_1}^2, \sigma_{q_2}^2, \sigma_{\eta_2}^2, \sigma_{\eta_3}^2, \sigma_{\eta_4}^2$, and $\sigma_{\eta_5}^2$ are the asymptotic variances of $q_1, \eta_1, q_2, \eta_2, \eta_3, \eta_4$ and η_5 respectively with the convergence rate of $\sqrt{np}$. $\sigma_{q_1\eta_1}, \sigma_{q_2\eta_2}, \sigma_{q_2\eta_3}$ and $\sigma_{\eta_2\eta_3}$ are the asymptotic covariance terms of q_1 and η_1, q_2 and η_2, and η_3 defined in Appendices A.4 *and* A.5.

Proof. Consider a set of arbitrary non-zero constants k_is such that

$$\sqrt{np}K = \sqrt{np}\left(k_1(q_1 - a_2) + k_2(\eta_1 - a_4) + k_3q_2 + k_4\eta_2 + k_5\eta_3 + k_6\eta_4 + k_7\eta_5\right)$$

and without loss of generality, $k_1 + \cdots + k_7 = 1$. With respect to the increasing set of σ-fields, $F_{n,l} = \sigma\{w_1, \ldots, w_l\}$ we note that K will satisfy the conditions of Lemma 9 if each term also satisfies the requirements. Condition (1) is satisfied by noting

$$\begin{aligned} V[K] &= V[k_1(q_1 - a_2)] + \cdots + V[k_2\eta_5] + \text{Cov}[k_1(q_1 - a_2), k_2(\eta_1 - a_4)] + \cdots + \text{Cov}[k_6, \eta_4, k_7\eta_5] \\ &\le V[k_1(q_1 - a_2)] + \cdots + V[k_2\eta_5] + V[k_1(q_1 - a_2)]^{1/2}V[k_2(\eta_1 - a_4)]^{1/2} + \cdots + V[k_6\eta_4]^{1/2}V[k_7\eta_5]^{1/2}. \end{aligned}$$

To satisfy condition (2) of Lemma 9 we will use the well known inequality for any random variables $Y_1, \ldots, Y_n$

$$E\left|\sum_{i=1}^{n} Y_i\right|^p \leq n^{p-1}\sum_{i=1}^{n} E[|Y_i|^p]$$

and the Lyapounov condition. That is,

$$E[K^4] \leq 7^3\left(k_1^4 E[(q_1 - a_2)^4] + \cdots + k_7^4 E[\eta_5^4]\right)$$

and if each component goes to zero, then the fourth moment of K will also go to zero resulting in asymptotic joint-normality.

The centralized forms of q_1 and η_1 have i.i.d. components and satisfy the requirements for the typical central limit theorem, so they clearly satisfy the requirements for martingale-difference. The other terms require additional work. Here we provide the details for η_2, the remainder of the terms follow the same procedure and details can be found in [8].

Define the σ-field generated by the random variables $\mathcal{F}_{n,j} = \sigma\{w_1, w_2, \ldots, w_j\}$. We then rewrite η_2 as

$$\begin{aligned} n\eta_2 &= \frac{K_2}{p}\sum_{i\neq j}^{p}\lambda_i^3\lambda_j V_{ij} = \frac{K_2}{p}\sum_{i<j}^{p}\lambda_i^3\lambda_j V_{ij} + \lambda_i\lambda_j^3 V_{ji} \\ &= \frac{K_2}{p}\sum_{j=2}^{p}\sum_{i=1}^{j-1}\lambda_i^3\lambda_j V_{ij} + \lambda_i\lambda_j^3 V_{ji} \end{aligned}$$

using the notation from Appendix A.3.2 and

$$K_2 = \frac{4}{n^5(n^2+n+2)} = nC_2.$$

Define

$$X_{n,j} = \sum_{i=1}^{j-1}\frac{K_2}{p}\left(\lambda_i^3\lambda_j V_{ij} + \lambda_i\lambda_j^3 V_{ji}\right) = \sum_{i=1}^{j-1} Y_{ij}$$

and note the following conditional expectations

$$\begin{aligned} &E[v_{ii}^2 v_{ij}^2|\mathcal{F}_{n,j-1}] = v_{ii}^3, \qquad E[v_{ii}^3 v_{jj}|\mathcal{F}_{n,j-1}] = n v_{ii}^3, \\ &E[v_{jj}^2 v_{ij}^2|\mathcal{F}_{n,j-1}] = (n+2)(n+4)v_{ii}, \qquad E[v_{jj}^3 v_{ii}|\mathcal{F}_{n,j-1}] = n(n+2)(n+4)v_{ii}, \end{aligned}$$

hence $E[V_{ij}|\mathcal{F}_{n,j-1}] = 0$ and $E[V_{ji}|\mathcal{F}_{n,j-1}] = 0$, so $E[X_{n,j}|\mathcal{F}_{n,j-1}] = 0$. Following the methodology from Appendix A.3.2 we can calculate and show that condition (1) from Lemma 9 holds. Begin by noting

$$X_{n,j}^2 = \sum_{i=1}^{j-1} Y_{ij}^2 + 2\sum_{i<k}^{j-1} Y_{ij}Y_{kj}. \tag{12}$$

We use the well-known result about expectations

$$E[E[X_{n,j}^2|\mathcal{F}_{n,j-1}]] = E[X_{n,j}^2]$$

and

$$\begin{aligned} E[X_{n,j}^2] &= E\left[\left(\sum_{i=1}^{j-1} Y_{ij}\right)^2\right] \\ &= \sum_{i=1}^{j-1} E[Y_{ij}^2] + 2\sum_{i<k}^{j-1} E[Y_{ij}Y_{kj}]. \end{aligned}$$

By the methodology from Appendix A.3.2,

$$E[Y_{ij}Y_{kj}] = 0$$

and for large n

$$\sum_{i=1}^{j-1} E[Y_{ij}^2] = \sum_{i=1}^{j-1}\frac{K_2^2}{p^2}\left(\lambda_i^6\lambda_j^2 E[V_{ij}^2] + 2\lambda_i^4\lambda_j^4 E[V_{ij}V_{ji}] + \lambda_i^2\lambda_j^6 E[V_{ji}^2]\right)$$

$$= \sum_{i=1}^{j-1} \frac{16}{O\left(n^{14}\right)} \frac{1}{p^2} \left(\lambda_i^6 \lambda_j^2 O\left(n^{14}\right) + 2\lambda_i^4 \lambda_j^4 O\left(n^{14}\right) + \lambda_i^2 \lambda_j^6 O\left(n^{14}\right)\right)$$
$$= \frac{16}{p} \left(\lambda_j^2 a_6 + 2\lambda_j^4 a_4 + \lambda_j^6 a_2\right).$$

Hence

$$E\left[\sum_{j=2}^{p} E\left[X_{n,j}^2 | F_{n,j-1}\right]\right] = \sum_{j=2}^{p} E[X_{n,j}^2]$$
$$= \sum_{j=2}^{p} \frac{16}{p} \left(\lambda_j^2 a_6 + 2\lambda_j^4 a_4 + \lambda_j^6 a_2\right)$$
$$= 32(a_6 a_2 + a_4^2).$$

If we can show that

$$V\left[\sum_{j=2}^{p} E\left[X_{n,j}^2 | F_{n,j-1}\right]\right] \to 0 \quad \text{as } (n,p) \to \infty,$$

then by the law of large numbers, condition (1) of Lemma 9 will be satisfied.

Using (12) we can find the conditional expectation given the σ-field $\mathcal{F}_{n,j-1}$. It is fairly straightforward to show

$$E[Y_{ij} Y_{kj} | \mathcal{F}_{n,j-1}] = 0.$$

Hence

$$E[X_{n,j}^2 | \mathcal{F}_{n,j-1}] = \sum_{i=1}^{j-1} E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]$$

and for large n

$$E[Y_{ij}^2 | \mathcal{F}_{n,j-1}] = \frac{K_2^2}{p^2} \left(\lambda_i^6 \lambda_j^2 O(n^8) v_{ii}^6 + 2\lambda_i^4 \lambda_j^4 O(n^{10}) v_{ii}^4 + \lambda_i^2 \lambda_j^6 O(n^{12}) v_{ii}^2\right). \tag{13}$$

Utilizing the triangular sequencing allows us to compute

$$V\left[\sum_{j=2}^{p} E\left[X_{n,j}^2 | F_{n,j-1}\right]\right] = \sum_{i=1}^{p-1} V\left[\sum_{j=i+1}^{p} E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]\right]$$

and from (13)

$$\sum_{i=1}^{p-1} V\left[\sum_{j=i+1}^{p} E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]\right] \le \frac{K_2^4}{p^2} \sum_{i=1}^{p} \Big(\lambda_i^{12} a_2^2 O(n^{16}) V[v_{ii}^6] + \lambda_i^8 a_4^2 O(n^{20}) V[v_{ii}^4]$$
$$+ \lambda_i^4 a_6^2 O(n^{24}) V[v_{ii}^2] + \lambda_i^{10} a_2 a_4 O(n^{18}) \mathrm{Cov}(v_{ii}^6, v_{ii}^4)$$
$$+ \lambda_i^8 a_2 a_6 O(n^{20}) \mathrm{Cov}(v_{ii}^6, v_{ii}^2) + \lambda_i^6 a_6 a_4 O(n^{22}) \mathrm{Cov}(v_{ii}^4, v_{ii}^2)\Big)$$

and its straightforward to calculate $V[v_{ii}^6] = O(n^{11})$ and the other variance and covariance terms. When including the $K_2^4 = O(n^{-28})$ it is easy to see that $V\left[E[Y_{ij}^2 | \mathcal{F}_{n,j-1}]\right] = O(n^{-1}p^{-2})$. From here it is clear that $V\left[\sum_{j=2}^{p} E\left[X_{n,j}^2 | F_{n,j-1}\right]\right] = O((np)^{-1}) \to 0$ as $(n,p) \to \infty$.

To show the Lyapounov type condition consider

$$X_{n,j}^4 = \sum_{i=1}^{j-1} Y_{ij}^4 + 4\sum_{i \ne k}^{j-1} Y_{ij}^3 Y_{kj} + 6\sum_{i<k}^{j-1} Y_{ij}^2 Y_{kj}^2 + 12\sum_{i \ne k < l}^{j-1} Y_{ij}^2 Y_{kj} Y_{lj} + 24\sum_{i<k<l<m}^{j-1} Y_{ij} Y_{kj} Y_{lj} Y_{mj}$$

and only the Y_{ij}^4 and $Y_{ij}^2 Y_{kj}^2$ have non-zero expectation. For large n

$$\sum_{i=1}^{j-1} E\left[Y_{ij}^4\right] = \frac{256}{p^3} \left(\lambda_j^4 a_{12} + 4\lambda_j^6 a_{10} + 6\lambda_j^8 a_8 + 4\lambda_j^{10} a_6 + \lambda_j^{12} a_4\right)$$

and

$$\sum_{i<k}^{j-1} E\left[Y_{ij}^2 Y_{kj}^2\right] = \frac{256}{p^2}\left(\lambda_j^4 a_6^2 + 4\lambda_j^6 a_4 a_6 + 4\lambda_j^8 a_4^2 + 2\lambda_j^8 a_2 a_6 + 4\lambda_j^{10} a_2 a_4 + \lambda_j^{12} a_2^2\right).$$

Hence

$$\sum_{j=2}^{p} E\left[X_{n,j}^4\right] = \frac{256}{p^2} O(1) + \frac{256}{p} O(1) \to 0 \quad \text{as } (n,p) \to \infty$$

and we have satisfied the second condition of Lemma 9. This completes the proof for the asymptotic normality of $n\eta_2$. We utilize assumption (B) to rewrite our result with the same convergent rate of $\sqrt{np}$.

As previously stated, we leave out the details for the remaining terms, available in [8], but we do make the following points of interest about η_4 and η_5:η_4 can be rewritten as

$$\eta_4 = \sum_{k=3}^{p}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1} \lambda_i^2\lambda_j\lambda_k V_{ijk} + \lambda_i\lambda_j^2\lambda_k V_{jik} + \lambda_i\lambda_j\lambda_k^2 V_{kij}$$

with V_{ijk}, V_{jik} and V_{kij} being the random components. Calculate the conditional expectation based on the σ-field with index based on k. We note the index is assigned to match that of the the η_4 term but is the same increasing set of σ-fields as before. In the V_{kij} term, the following conditional expectation is tricky and must be found through expansion of summations,

$$E[v_{ik}^2 v_{jk}^2 | \mathcal{F}_{n,k-1}] = 2v_{ij}^2 + v_{ii}v_{jj}.$$

For η_5, the σ-field is defined with an index based on l. In the arguments for η_4 and η_5, assumption (B) must be utilized to handle the double and triple summations in the defined $X_{n,k}$ and $X_{n,l}$ respectively. □

A simple linear transformation provides the following important result,

Proposition 2.

$$\sqrt{np}\begin{pmatrix}\hat{a}_2\\ \hat{a}_4\end{pmatrix} \xrightarrow{D} N_2\left(\begin{pmatrix}a_2\\ a_4\end{pmatrix}, \begin{pmatrix}\sigma_2^2 & \sigma_{24}\\ \sigma_{24} & \sigma_4^2\end{pmatrix}\right),$$

where σ_2^2, σ_{24} and σ_4^2 are the (n, p)-asymptotic variance and covariance of $\hat{a}_2$ and $\hat{a}_4$ with respect to the $\sqrt{np}$ convergent rate defined as

$$\sigma_2^2 \simeq 8a_4 + 4ca_2^2,$$
$$\sigma_{24} \simeq 16a_6 + 16ca_4a_2 + 8ca_3^2,$$

and

$$\sigma_4^2 \simeq 32a_8 + \frac{32}{n}(pa_6a_2 + pa_4^2) + \frac{16}{n}pa_4^2 + \frac{64pa_5a_3}{n} + \frac{32pca_4a_2^2}{n} + \frac{64pca_3^2a_2}{n} + \frac{8pc^2a_2^4}{n}$$
$$= 32a_8 + 48ca_4^2 + 32ca_6a_2 + 64ca_5a_3 + \frac{3}{2}c^2a_4a_2^2 + 64c^2a_3^2a_2 + 8c^3a_2^4.$$

References

[1] U. Alon, N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci. USA 96 (12) (1999) 6745–6750.

[2] L.V. Arharov, Limit theorems for the characteristic roots of a sample covariance matrix, Dokl. Akad. Nauk SSSR 199 (1971) 994–997.

[3] Z.D. Bai, Convergence rate of expected spectral distributions of large random matrices. II. Sample covariance matrices, Ann. Probab. 21 (2) (1993) 649–672.

[4] Z.D. Bai, P.R. Krishnaiah, L.C. Zhao, On rates of convergence of efficient detection criteria in signal processing with white noise, IEEE Trans. Inform. Theory 35 (2) (1989) 380–388.

[5] M. Dettling, P. Bühlmann, Boosting for tumor classification with gene expression data, Bioinformatics 19 (9) (2003) 1061–1069.

[6] S. Dudoit, J. Fridlyand, T.P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data, J. Amer. Statist. Assoc. 97 (457) (2002) 77–87.

[7] R. Durrett, Probability: Theory and Examples, 2nd ed., Duxbury Press, Belmont, CA, 1996.

[8] T.J. Fisher, On the testing and estimation of high-dimensional covariance matrices, Ph.D. Dissertation, Clemson University, Department of Mathematical Sciences, 2009.

[9] T. Golub, D. Slonim, P. Tamayo, C. Huard, M Gaasenbeek, J. Mesirov, H. Coller, M. Loh, J. Downing, M. Caligiuri, C. Bloomfield, E. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (5439) (1999) 531–537.

[10] S. John, Some optimal multivariate tests, Biometrika 58 (1971) 123–127.

[11] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (4) (2002) 1081–1102.

[12] C.R. Narayanaswamy, D. Raghavarao, Principal component analysis of large dispersion matrices, J. R. Stat. Soc. Ser. C 40 (2) (1991) 309–316.

[13] H. Saranadasa, Asymptotic expansion of the misclassification probabilities of *D*- and *A*-criteria for discrimination from two high-dimensional populations using the theory of large-dimensional random matrices, J. Multivariate Anal. 46 (1) (1993) 154–174.
[14] J.R. Schott, Testing for complete independence in high dimensions, Biometrika 92 (4) (2005) 951–956.
[15] J.R. Schott, A high-dimensional test for the equality of the smallest eigenvalues of a covariance matrix, J. Multivariate Anal. 97 (4) (2006) 827–843.
[16] J.R. Schott, A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, Comput. Statist. Data Anal. 51 (12) (2007) 6535–6542.
[17] V. Serdobolskii, Multivariate statistical analysis, in: Theory and Decision Library. Series B: Mathematical and Statistical Methods, vol. 41, Kluwer Academic Publishers, Dordrecht, 2000, a high-dimensional approach.
[18] V.I. Serdobol'skiĭ, Theory of essentially multivariate statistical analysis, Uspekhi Mat. Nauk 54 (2(326)) (1999) 85–112.
[19] A.N. Shiryaev, Probability, 2nd ed., in: Graduate Texts in Mathematics, vol. 95, Springer-Verlag, New York, 1996, Translated from the first (1980) Russian edition by R.P. Boas.
[20] J.W. Silverstein, Eigenvalues and eigenvectors of large-dimensional sample covariance matrices, in: Random Matrices and their Applications, Amer. Math. Soc., Providence, RI, 1986, pp. 153–159.
[21] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2) (2005) 251–272.
[22] M.S. Srivastava, Some tests criteria for the covariance matrix with fewer observations than the dimension, Acta Comment. Univ. Tartu. Math. 10 (2006) 77–93.
[23] M.S. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc. 37 (1) (2007) 53–86.
[24] Y.Q. Yin, P.R. Krishnaiah, A limit theorem for the eigenvalues of product of two random matrices, J. Multivariate Anal. 13 (4) (1983) 489–507.