

LIMITING BEHAVIOR OF EIGENVALUES IN HIGH-DIMENSIONAL MANOVA VIA RMT

BY ZHIDONG BAI^{*,1}, KWOK PUI CHOI^{†,2} AND YASUNORI FUJIKOSHI^{‡,3}

Northeast Normal University^{}, National University of Singapore[†]
 and Hiroshima University[‡]*

In this paper, we derive the asymptotic joint distributions of the eigenvalues under the null case and the local alternative cases in the MANOVA model and multiple discriminant analysis when both the dimension and the sample size are large. Our results are obtained by random matrix theory (RMT) without assuming normality in the populations. It is worth pointing out that the null and nonnull distributions of the eigenvalues and invariant test statistics are asymptotically robust against departure from normality in high-dimensional situations. Similar properties are pointed out for the null distributions of the invariant tests in multivariate regression model. Some new formulas in RMT are also presented.

1. Introduction. It is both basic and important to study the distributions of the eigenvalues in a one-way multivariate analysis of variance (MANOVA) model and multiple discriminant analysis. Suppose there are $q + 1$ groups, and $\{\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}\}$ represents a random sample of p -vectors from the i th group, which has mean vector $\boldsymbol{\mu}_i$ and common covariance matrix $\boldsymbol{\Sigma}$. Various inferential procedures are based on the matrices

$$\mathbf{S}_b = \frac{1}{n} \sum_{i=1}^{q+1} n_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})', \quad \mathbf{S}_e = \frac{1}{n} \sum_{i=1}^{q+1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)',$$

where

$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{q+1} n_i \bar{\mathbf{x}}_i, \quad n = \sum_{i=1}^{q+1} n_i.$$

The matrices $n\mathbf{S}_b$ and $n\mathbf{S}_e$ are called the matrices of sums of squares and products due to between-groups and within-groups, respectively. Let $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$, then $n\mathbf{S}_t$

Received May 2017; revised October 2017.

¹Supported by NSFC 11571067 and 11471140.

²Supported by the Singapore Ministry of Education Academic Research Fund R-155-000-141-112.

³Supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), # 25330038, 2013–2015.

MSC2010 subject classifications. Primary 62H10; secondary 62E20.

Key words and phrases. Asymptotic distribution, eigenvalues, discriminant analysis, high-dimensional case, MANOVA, nonnormality, RMT, test statistics.

is called the matrix of sums of squares and products due to total variation. These matrices are also used in canonical discriminant analysis, which is a statistical procedure designed to discriminate between several different groups in terms of a few discriminant functions.

The large-sample asymptotic distributions of the eigenvalues of $\mathbf{S}_b \mathbf{S}_e^{-1}$ or $\mathbf{S}_b \mathbf{S}_t^{-1}$ were derived under normality by Hsu (1941) and Anderson (1951). A gap in the proof of the main theorem in Hsu (1941) was corrected by Bai (1985). Amemiya (1990) characterized the limiting distributions of the roots of a general determinantal equation. Hsu's results were extended to asymptotic expansions by a number of authors; see Sugiura (1976), Fujikoshi (1977), Muirhead (1978, 1982), Glynn and Muirhead (1978), etc. For extension to elliptical case with different covariance matrices, see Seo, Kanda and Fujikoshi (1994).

As the data dimension, p , increases, it is well known that the accuracy of large sample approximations deteriorates [see, for instance, Bai and Saranadasa (1996)]. As an alternative approach to overcome this shortcoming, it has been considered to derive asymptotic distributions of the eigenvalues in a high-dimensional situation where the dimension p , the sample size n and the number of groups $q + 1$ are large. The high-dimensional asymptotic results when $p/n \rightarrow c \in (0, 1)$ were obtained in Fujikoshi, Himeno and Wakaki (2008) by assuming that the population eigenvalues are simple. Johnstone (2008) derived the limiting distribution of the largest eigenvalue when $q/n \rightarrow \tilde{c} \in (0, 1)$ in addition to $p/n \rightarrow c \in (0, 1)$ under the assumption that all the population eigenvalues are zero, that is, the null case. All these results were obtained under normality assumption of the populations.

In this paper, we derive asymptotic joint distributions of the eigenvalues under the null case and the local alternatives in MANOVA model and multiple discriminant analysis when both the dimension and the sample size are large. Our results are obtained without assuming normality using techniques from random matrix theory (RMT). There are several asymptotic results based on RMT; see, for instances, Bai et al. (2009), Bai, Liu and Wong (2011), Zheng (2012), etc. In this paper, we show that by applying RMT, it is possible to derive the asymptotic distributions of the eigenvalues without assuming normality. In the course of derivation, we obtain some new limit theorems which may be of independent interests.

The high-dimensional asymptotic results can be applied to some MANOVA tests based on the likelihood ratio criterion, Lawley–Hotelling criterion or Bartlett–Nanda–Pillai criterion. It is worth pointing out that our asymptotic results coincide with those under normality assumption. Therefore, these MANOVA tests are robust against departure from normality when the dimension and the sample size are large. Similar results are shown for test statistics in multivariate regression model.

We extend our results to a nonnull case, the local alternatives. The high-dimensional asymptotic distributions of the eigenvalues are shown to be the eigenvalues of a Gaussian orthogonal ensemble and a fixed positive definite matrix. Based on these results, we derive the asymptotic powers of the above multivariate tests in closed form, from which various properties of these tests in high-dimension

are delineated. We have conducted simulation studies to examine the accuracy of the asymptotic approximations of the test statistics. Results of some of these studies are reported in Section 6.

Fuelled by myriad of applications such as pattern recognition, genomic studies, earth observation data, biomedical studies, which are of high data dimension, many tests have been proposed in a high-dimensional framework in which both p, n are large including the possibility that $p > n$ in the past decade. When $p > n$, \mathbf{S}_e is singular and so the classical invariant tests can no longer be applied. It is currently under intense research how to handle the effects of high dimension. We will mention some related work below. For examples, [Srivastava and Fujikoshi \(2006\)](#) modified the classical tests using generalized Moore–Penrose inverse; [Srivastava and Kubokawa \(2013\)](#) introduced a test that is invariant under change of measurement units; [Ullah and Jones \(2015\)](#) compared different regularized MANOVA tests; and [Wang, Peng and Li \(2015\)](#) proposed a novel high-dimensional nonparametric multivariate test. A timely review of the development of tests of significance for high-dimensional mean vectors over the past two decades can now be found in [Hu and Bai \(2016\)](#). [Dempster \(1958, 1960\)](#) introduced two nonexact tests for equality of two population mean vectors to circumvent the singularity problem of \mathbf{S}_e in Hotelling's T^2 . [Bai and Saranadasa \(1996\)](#) studied the effects of high-dimension on statistical inference. [Hu and Bai \(2016\)](#) reviewed the ideas and technicalities of some new approaches: the approach of [Chen and Qin \(2010\)](#) in which they modified the Bai–Saranadasa test to an unbiased test; the approach of [Srivastava and Du \(2008\)](#) for an invariant test; and the approach in [Cai and Xia \(2014\)](#) test.

It is interesting to compare these three tests (i.e., the likelihood ratio criterion, Lawley–Hotelling criterion and Bartlett–Nada–Pillai criterion) with the tests proposed in a high-dimensional situation: $p, n \rightarrow \infty$ with $p/n \rightarrow c \in (0, 1)$. We compare these three tests with two modified Dempster tests based on the asymptotic powers under local alternatives. However, a careful study of this important problem will be left as future work.

The organization of the paper is as follows. In Section 2, we state our main results (Theorem 2.4 and Corollary 2.6) on the limiting joint distributions of the nonzero eigenvalues of $\mathbf{S}_b \mathbf{S}_t^{-1}$ and $\mathbf{S}_b \mathbf{S}_e^{-1}$, based on which we establish the null robustness of some multivariate tests in Corollary 2.1. In the same section, we state some limit theorems, Theorems 2.1 to 2.3, which are of independent interests. The proofs of Theorem 2.4 and Corollaries 2.6 and 2.1 are given in Section 3. In Section 4, we apply our basic results to derive the high-dimensional asymptotic distributions of some tests in multivariate regression model. In Section 5, we extend the results in the null case to local alternatives. In Section 6, our asymptotic results are numerically examined. Proof of Theorem 2.1 via CLT martingale decomposition along the rows for the computation of the covariance structure of the Gaussian orthogonal ensemble is given in Section 7. Theorem 2.2 is proved in Section 8.

2. Statements of main results. Throughout this article, $\{y_{jk} : 1 \leq j \leq p, 1 \leq k \leq n\}$ denotes a double array of independent and identically distributed (*i.i.d.*) random variables with mean 0, variance 1 and finite fourth moment. Let $\mathbf{y}_k = (y_{1k}, \dots, y_{pk})'$ and $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$. Write $\bar{\mathbf{y}} = \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k$, and $\mathbf{S} = \frac{1}{n} \mathbf{Y} \mathbf{Y}'$. For $1 \leq j \leq p$, let \mathbf{Z}_j denote the $(p - 1) \times n$ matrix obtained by removing the j th row, denoted by \mathbf{z}'_j , from \mathbf{Y} . Similarly, for $1 \leq k \leq n$, we let \mathbf{Y}_k denote the $p \times (n - 1)$ matrix obtained by removing the k th column, \mathbf{y}_k , from \mathbf{Y} . We define $\mathbf{S}_k = \frac{1}{n} \mathbf{Y}_k \mathbf{Y}'_k$ for $1 \leq k \leq n$. Similarly, for $1 \leq k \neq \ell \leq n$, $\mathbf{Y}_{k\ell}$ denotes the matrix after removing the k th and the ℓ th columns from \mathbf{Y} , and $\mathbf{S}_{k\ell} = \frac{1}{n} \mathbf{Y}_{k\ell} \mathbf{Y}'_{k\ell}$. Let the minimum and the maximum eigenvalues of \mathbf{S} be denoted by $\lambda_{\min}(\mathbf{S})$ and $\lambda_{\max}(\mathbf{S})$, respectively.

We use $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{D} X$ to denote X_n converges to X in probability and in distribution, respectively. Let \mathbf{I}_n denote the $n \times n$ identity matrix, and $\mathbf{1}_n$ the n -vector of all components 1. Let $\mathbf{0}$ and \mathbf{O} respectively represent a zero vector and a zero matrix of appropriate order, often clear in the context of their appearances. Throughout this article, vectors are column vectors equipped with the Euclidean norm. The norm of a matrix \mathbf{B} is taken to be the spectral norm, $\|\mathbf{B}\| = \{\lambda_{\max}(\mathbf{B} \mathbf{B}')\}^{1/2}$.

For $q \geq 1$, we let $\mathbf{W} = [w_{ij}]_{1 \leq i, j \leq q}$ denote a Gaussian orthogonal ensemble (GOE): \mathbf{W} is a $q \times q$ symmetric matrix with $w_{ii} \sim N(0, 1)$, $w_{ij} \sim N(0, 1/2)$, and $w_{ij} = w_{ji}$ are jointly independent for $1 \leq i \leq j \leq q$. We suppress the dependence of \mathbf{W} on q as the order of the Gaussian orthogonal ensemble is often explicitly spelled out or clear from the context.

We consider an $n \times q$ matrix $\mathbf{A} = [a_{ij}]_{1 \leq i \leq n, 1 \leq j \leq q}$ satisfying the following properties:

$$(2.1) \quad \mathbf{A}' \mathbf{A} = \mathbf{I}_q,$$

$$(2.2) \quad \mathbf{A}' \mathbf{1}_n = \mathbf{0},$$

$$(2.3) \quad \max_{1 \leq i \leq n, 1 \leq j \leq q} |a_{ij}| = O(n^{-1/2}).$$

We first state the following theorems, Theorems 2.1 to 2.3, which are of independent interest. Theorem 2.3 is then applied to deduce the limiting behavior of the eigenvalues in the MANOVA setting under the null case.

THEOREM 2.1. *Let $p \times n$ random matrix \mathbf{Y} be as described at the beginning of Section 2. Suppose $n \times q$ matrix \mathbf{A} satisfies (2.1) and (2.3). Suppose q is fixed, $p, n \rightarrow \infty$ satisfying $p/n \rightarrow c \in (0, 1)$, then we have*

$$\sqrt{n} \left\{ \mathbf{A}' \mathbf{Y}' (\mathbf{Y} \mathbf{Y}')^{-1} \mathbf{Y} \mathbf{A} - \frac{p}{n} \mathbf{I}_q \right\} \xrightarrow{D} \sqrt{2c(1-c)} \mathbf{W},$$

where \mathbf{W} is a $q \times q$ Gaussian orthogonal ensemble.

THEOREM 2.2. *Let $p \times n$ random matrix \mathbf{Y} and p -vector $\bar{\mathbf{y}}$ be as described at the beginning of this section. Let $n \times q$ matrix \mathbf{A} satisfy (2.1)–(2.3). Suppose q is fixed, $p, n \rightarrow \infty$ satisfying $p/n \rightarrow c \in (0, 1)$, then we have*

$$\sqrt{n}\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')^{-1}\mathbf{Y}\mathbf{A} - \mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A}\} \xrightarrow{P} \mathbf{O}.$$

Theorem 2.3 below follows immediately from Theorems 2.1 and 2.2. The proof of Theorem 2.1 will be given in Section 7 and that of Theorem 2.2 in Section 8.

THEOREM 2.3. *Under the conditions stated in Theorem 2.2, we have*

$$\sqrt{n}\left\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')^{-1}\mathbf{Y}\mathbf{A} - \frac{p}{n}\mathbf{I}_q\right\} \xrightarrow{D} \sqrt{2c(1-c)}\mathbf{W},$$

where \mathbf{W} is a $q \times q$ Gaussian orthogonal ensemble.

Let Σ be a $p \times p$ positive definite matrix and μ_i a p -vector for $i = 1, \dots, q + 1$. We assume our data vectors \mathbf{x}_k 's are generated as

$$(2.4) \quad \mathbf{x}_k = \Sigma^{1/2}\mathbf{y}_k + \mu_i \quad \text{for } N_{i-1} < k \leq N_i,$$

where $N_0 = 0, N_{q+1} = n, N_i = n_1 + \dots + n_i, i = 1, \dots, q + 1$. [See the discussion and extension of (2.4) at the end of Section 2.] We regard the observations of the random sample from the i th group are of the form $\mathbf{x}_k, N_{i-1} < k \leq N_i$, which has mean vector μ_i and covariance matrix Σ . Further, the matrices \mathbf{S}_e and \mathbf{S}_b defined in the Introduction section can be expressed in terms of $\mathbf{y}_k, k = 1, 2, \dots, n$ as follows:

$$\begin{aligned} \mathbf{S}_e &= \Sigma^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{q+1} \sum_{k=N_{i-1}+1}^{N_i} (\mathbf{y}_k - \bar{\mathbf{y}}_i)(\mathbf{y}_k - \bar{\mathbf{y}}_i)' \right\} \Sigma^{1/2} \\ &= \Sigma^{1/2} (\mathbf{S} - \alpha_{n1}\bar{\mathbf{y}}_1\bar{\mathbf{y}}_1' - \dots - \alpha_{n,q+1}\bar{\mathbf{y}}_{q+1}\bar{\mathbf{y}}_{q+1}') \Sigma^{1/2}, \\ (2.5) \quad \mathbf{S}_b &= \Sigma^{1/2} \left\{ \frac{1}{n} \sum_{i=1}^{q+1} n_i (\bar{\mathbf{y}}_i - \bar{\mathbf{y}} + \xi_i)(\bar{\mathbf{y}}_i - \bar{\mathbf{y}} + \xi_i)' \right\} \Sigma^{1/2}, \end{aligned}$$

where $\alpha_{ni} := \frac{n_i}{n}, \bar{\mathbf{y}}_i = \frac{1}{n_i} \sum_{k=N_{i-1}+1}^{N_i} \mathbf{y}_k, \bar{\mathbf{y}} = \frac{1}{n} \sum_{k=1}^n \mathbf{y}_k = \frac{1}{n} \sum_{i=1}^{q+1} n_i \bar{\mathbf{y}}_i, \xi_i = \Sigma^{-1/2}(\mu_i - \bar{\mu}), \bar{\mu} = \frac{1}{n} \sum_{i=1}^{q+1} n_i \mu_i$. Note that $\sum_{i=1}^{q+1} n_i \xi_i = \mathbf{0}$. Let $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$. Matrices $n\mathbf{S}_b$ and $n\mathbf{S}_e$ are commonly referred to as the matrices of sums of squares and products due to between-groups, and within-groups respectively. We impose a natural assumption:

$$(2.6) \quad \alpha_{ni} = \frac{n_i}{n} \rightarrow \alpha_i > 0, \quad i = 1, 2, \dots, q + 1.$$

THEOREM 2.4. *Suppose $n - q - 1 \geq p \geq q \geq 1$. Let $n\mathbf{S}_b$ and $n\mathbf{S}_t$ be the matrices of sums of squares and products due to between-groups and total variation based on the \mathbf{x}_k 's observations, respectively. Let $d_1 > \dots > d_q$ be the nonzero eigenvalues of $\mathbf{S}_b\mathbf{S}_t^{-1}$. Suppose that $\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_{q+1}$. Put*

$$(2.7) \quad \tilde{d}_i = \frac{\sqrt{n}}{\sqrt{2c(1-c)}}(d_i - p/n), \quad i = 1, \dots, q.$$

Suppose q is fixed, $p, n \rightarrow \infty$ satisfying $p/n \rightarrow c \in (0, 1)$ and condition (2.6) holds, then the limiting joint density function, with support in $y_q < \dots < y_1$, of the normalized eigenvalues $(\tilde{d}_1, \tilde{d}_2, \dots, \tilde{d}_q)$ is given by $f(y_1, \dots, y_q; q)$ where

$$(2.8) \quad f(y_1, \dots, y_q; q) = \left[2^{q/2} \prod_{i=1}^q \Gamma\left(\frac{i}{2}\right) \right]^{-1} \exp\left(-\frac{1}{2} \sum_{i=1}^q y_i^2\right) \prod_{1 \leq i < j \leq q} (y_i - y_j).$$

THEOREM 2.5. *Under the conditions stated in Theorem 2.3, we have*

$$\sqrt{n} \left\{ \mathbf{A}'\mathbf{Y}'\mathbf{S}_e^{-1}\mathbf{Y}\mathbf{A} - \frac{p}{(n-p)}\mathbf{I}_q \right\} \xrightarrow{D} \sqrt{2c/(1-c)^3}\mathbf{W},$$

where \mathbf{W} is a $q \times q$ Gaussian orthogonal ensemble.

Theorems 2.4 and 2.5 imply the following Theorem 2.6. The proofs of Theorems 2.4 and 2.5 will be given in the next section.

THEOREM 2.6. *Let $\ell_1 > \dots > \ell_q$ be the nonzero eigenvalues of $\mathbf{S}_b\mathbf{S}_e^{-1}$, and put*

$$(2.9) \quad \tilde{\ell}_i = \sqrt{n(1-c)^3/(2c)}\{\ell_i - p/(n-p)\}, \quad i = 1, \dots, q.$$

Then under the same conditions as in Theorem 2.4, the limiting density function of the normalized eigenvalues $(\tilde{\ell}_1, \dots, \tilde{\ell}_q)$ is given by $f(y_1, \dots, y_q; q)$ in (2.8).

When $q + 1 = 2$, ℓ_1 is reduced to the Hotelling's T^2 -statistic which is given by

$$T^2 = \frac{n_1n_2}{n}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'(n\mathbf{S}_e)^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2).$$

Pan and Zhou (2011) derived the asymptotic distribution of the one-sample Hotelling's T_0^2 statistic. Note that these two asymptotic results are the same.

A significant consequence of our results is the support of null robustness for nonnormality of some multivariate tests even when the dimension and the sample size are large. For example, in testing the hypothesis $H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_{q+1}$ in the MANOVA model, we have the following three test statistics:

$$\tilde{T}_1 = -\log \frac{|\mathbf{S}_e|}{|\mathbf{S}_t|} = -\log \prod_{i=1}^q (1 + \ell_i)^{-1} = -\sum_{i=1}^q \log(1 - d_i),$$

$$\begin{aligned} \tilde{T}_2 &= \text{tr} \mathbf{S}_b \mathbf{S}_e^{-1} = \sum_{i=1}^q \ell_i = \sum_{i=1}^q \frac{d_i}{1 - d_i}, \\ \tilde{T}_3 &= \text{tr} \mathbf{S}_b \mathbf{S}_t^{-1} = \sum_{i=1}^q \frac{\ell_i}{1 + \ell_i} = \sum_{i=1}^q d_i. \end{aligned}$$

The test statistic \tilde{T}_1 is based on the likelihood ratio test. The test statistics \tilde{T}_2 and \tilde{T}_3 are based on the Lawley–Hotelling criterion and Bartlett–Nanda–Pillai criterion, respectively [see, e.g., [Anderson \(2003\)](#)].

Here, we consider the high-dimensional asymptotic null distributions of \tilde{T}_1 , \tilde{T}_2 and \tilde{T}_3 under an asymptotic framework in which $p, n \rightarrow \infty$ satisfying $p/n \rightarrow c \in (0, 1)$. Assuming normality in the populations, [Wakaki, Fujikoshi and Ulyanov \(2014\)](#) showed the following asymptotic results:

$$(2.10) \quad T_i \xrightarrow{D} N\left(0, \frac{2q}{1 - c}\right), \quad i = 1, 2, 3,$$

where

$$(2.11) \quad \begin{aligned} T_1 &= \sqrt{p} \left(1 + \frac{m}{p}\right) \left\{ \tilde{T}_1 - q \log \left(1 + \frac{p}{m}\right) \right\}, \\ T_2 &= \sqrt{p} \left(\frac{m}{p} \tilde{T}_2 - q\right), \\ T_3 &= \sqrt{p} \left(1 + \frac{p}{m}\right) \left\{ \left(1 + \frac{m}{p}\right) \tilde{T}_3 - q \right\}. \end{aligned}$$

Here, the asymptotic variance is the limit of $2q(1 + p/m)$ where $m = n - p + q$. This result is generalized as stated in the following corollary.

COROLLARY 2.1. *Assume the same conditions as stated in Theorem 2.4. With the T_i 's as defined in (2.11),*

$$(2.12) \quad T_i \xrightarrow{D} N\left(0, \frac{2q}{1 - c}\right), \quad i = 1, 2, 3.$$

Corollary 2.1 shows that the results of [Wakaki, Fujikoshi and Ulyanov \(2014\)](#) continue to hold without assuming normality, but under a rather natural additional assumption (2.6). Note also that these three tests have diverging variance as $c \rightarrow 1$. The proof is given in Section 3.

Discussion and extension of model (2.4). The assumption that the components of \mathbf{y}_k are *i.i.d.* in (2.4) can be relaxed to the components of $\mathbf{y}_k = (y_{1k}, \dots, y_{pk})'$ are independent, means 0, variances 1 and that $\{y_{ik}^4\}$ is uniformly integrable. The model used in [Bai and Saranadasa \(1996\)](#) assumes the components are independent, mean 0, variance 1 and $E(y_{ik}^4) = 3 + \Delta < \infty$. Since the work of [Bai and Saranadasa \(1996\)](#) in studying the effects of high-dimension, to the best of our

knowledge, almost all papers on high-dimensional statistical analysis do assume the independent component structure. See, for examples, [Chen and Qin \(2010\)](#) and [Li and Chen \(2012\)](#) used the same model of [Bai and Saranadasa \(1996\)](#); and [Pan and Zhou \(2011\)](#) assumed the components to be i.i.d. The model (2.4) or the generalization above admits a rather wide class of distributions that includes multivariate normal distributions.

We remark that the independent component assumption seems to us “essential” in the following sense. [Bai and Zhou \(2008\)](#) dropped the independent component assumption, but they have to impose a rather weak assumption, namely, $E(\mathbf{y}'\mathbf{B}\mathbf{y} - \text{tr}\mathbf{B})^2 = o(n^2)$ for any nonrandom $p \times p$ matrix \mathbf{B} with bounded norm [Theorem 1.1 in [Bai and Zhou \(2008\)](#)]. This condition is weak because $E(\mathbf{y}'\mathbf{B}\mathbf{y} - \text{tr}\mathbf{B})^2 = O(n^2)$ is always true. However, even for ϕ -mixing \mathbf{y} components, it is not known if this assumption holds. On the other hand, [Silverstein \(1995\)](#) assumed independent components and he did not need this additional assumption.

3. Proofs of Theorems 2.4–2.6. Note that the eigenvalues of the matrix $\mathbf{S}_b\mathbf{S}_t^{-1}$ are independent of Σ provided that μ_i 's are changed to ξ_i 's. Hence, without loss of generality, we shall assume that $\Sigma = \mathbf{I}_p$.

PROOF OF THEOREM 2.4. Our proof is to apply Theorem 2.3 for an appropriately chosen \mathbf{A} which satisfies (2.1) to (2.3). Let \mathbf{y}_i 's be the random vectors generating the observations \mathbf{x}_i 's as depicted in (2.4). Write

$$\begin{aligned} \mathbf{Y}_b &= [\sqrt{\alpha_{n1}}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}), \dots, \sqrt{\alpha_{n,q+1}}(\bar{\mathbf{y}}_{q+1} - \bar{\mathbf{y}})], \\ \Xi_b &= [\sqrt{\alpha_{n1}}\xi_1, \dots, \sqrt{\alpha_{n,q+1}}\xi_{q+1}], \end{aligned}$$

thus $\mathbf{S}_b = (\mathbf{Y}_b + \Xi_b)(\mathbf{Y}_b + \Xi_b)'$. Rewrite

$$\mathbf{Y}_b = \mathbf{Y} \left(\frac{1}{\sqrt{n}}\mathbf{G} - \frac{1}{n}\mathbf{1}_n\mathbf{g}' \right),$$

where $\mathbf{g} = (\sqrt{\alpha_{n1}}, \dots, \sqrt{\alpha_{n,q+1}})'$, $\mathbf{G} = [\frac{\mathbf{e}_1}{\sqrt{n_1}}, \dots, \frac{\mathbf{e}_{q+1}}{\sqrt{n_{q+1}}}]$, and for $1 \leq i \leq q + 1$, $\mathbf{e}'_i = (\mathbf{0}'_{N_{i-1}}, \mathbf{1}'_{n_i}, \mathbf{0}'_{n-N_i})$.

Let \mathbf{F} be a $(q + 1) \times q$ matrix such that $[\mathbf{F}, \mathbf{g}]$ is a $(q + 1) \times (q + 1)$ orthogonal matrix. It follows that

$$\mathbf{F}'\mathbf{g} = \mathbf{0}, \quad \mathbf{F}'\mathbf{F} = \mathbf{I}_q \quad \text{and} \quad \mathbf{F}\mathbf{F}' + \mathbf{g}\mathbf{g}' = \mathbf{I}_{q+1}.$$

As $(\mathbf{Y}_b + \Xi_b)\mathbf{g} = \mathbf{0}$, we can rewrite \mathbf{S}_b as

$$\begin{aligned} \mathbf{S}_b &= (\mathbf{Y}_b + \Xi_b)[\mathbf{F}, \mathbf{g}]\{(\mathbf{Y}_b + \Xi_b)[\mathbf{F}, \mathbf{g}]\}' \\ &= \frac{1}{n}(\mathbf{Y}\mathbf{A} + \sqrt{n}\Xi_b\mathbf{F})(\mathbf{Y}\mathbf{A} + \sqrt{n}\Xi_b\mathbf{F})', \end{aligned}$$

where

$$(3.1) \quad \mathbf{A} = (\mathbf{G} - n^{-1/2}\mathbf{1}_n\mathbf{g}')\mathbf{F} = \mathbf{G}\mathbf{F}.$$

Under the assumption of Theorem 2.4 (the null case), $\mathbf{E}_b = \mathbf{O}$ together with $\mathbf{\Sigma} = \mathbf{I}_p$,

$$\mathbf{S}_t = \frac{1}{n} \sum_{k=1}^n (\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})' = \frac{1}{n}(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')$$

and

$$\mathbf{S}_b = \mathbf{Y}_b\mathbf{Y}'_b = \frac{1}{n}(\mathbf{Y}\mathbf{A})(\mathbf{Y}\mathbf{A})'.$$

Recalling the fact that $\mathbf{B}_1\mathbf{B}_2$ and $\mathbf{B}_2\mathbf{B}_1$ have the same nonzero eigenvalues leads to the lemma below.

LEMMA 3.1. *The nonzero eigenvalues, $d_1 > \dots > d_q$, of $\mathbf{S}_b\mathbf{S}_t^{-1}$ are the same as that of $\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')^{-1}\mathbf{Y}\mathbf{A}$.*

To complete our proof of Theorem 2.4, it remains to:

- (i) verify that \mathbf{A} satisfies (2.1)–(2.3); and
- (ii) deduce Theorem 2.4 from the Gaussian orthogonal ensemble.

To verify (i), first observe that $\mathbf{A}'\mathbf{A} = (\mathbf{F}'\mathbf{G}')\mathbf{G}\mathbf{F} = \mathbf{F}'\mathbf{I}_{q+1}\mathbf{F} = \mathbf{I}_q$ and $\mathbf{A}'\mathbf{1}_n = \sqrt{n}\mathbf{F}'\mathbf{g} = \mathbf{0}$, so conditions (2.1) and (2.2) hold. Writing $\mathbf{A} = [a_{ij}]$ and $\mathbf{F} = [f_{ij}]$, then we have for $N_{k-1} < i \leq N_k, k = 1, \dots, q + 1$,

$$|a_{ij}| = |f_{kj}/\sqrt{n_k}| \leq 1/\sqrt{n_k}$$

which implies (2.3). Part (ii) follows from Theorem 13.3.1 in Anderson (2003). □

PROOFS OF THEOREMS 2.5 AND 2.6. Under the null case, \mathbf{S}_t and \mathbf{S}_b can be written, respectively, as

$$\mathbf{S}_t = \frac{1}{n}\mathbf{Y}(\mathbf{I}_n - \mathbf{P}_0)\mathbf{Y}' \quad \text{and} \quad \mathbf{S}_b = \frac{1}{n}\mathbf{Y}\mathbf{A}(\mathbf{Y}\mathbf{A})',$$

where $\mathbf{P}_0 = \frac{1}{n}\mathbf{1}_n\mathbf{1}'_n$. The nonzero eigenvalues $d_1 > \dots > d_q > 0$ of $\mathbf{S}_b\mathbf{S}_t^{-1}$ are the same as that of $\mathbf{D} = \frac{1}{n}(\mathbf{Y}\mathbf{A})'\mathbf{S}_t^{-1}\mathbf{Y}\mathbf{A}$. Theorem 2.3 implies that

$$(3.2) \quad \mathbf{D} = \frac{p}{n}\mathbf{I}_q + \frac{1}{\sqrt{n}}\sqrt{2c(1-c)}\mathbf{W} + o_P(n^{-1/2}).$$

The nonzero eigenvalues $\ell_1 > \dots > \ell_q > 0$ of $\mathbf{S}_b\mathbf{S}_e^{-1}$ are the same as that of $\mathbf{L} = \frac{1}{n}(\mathbf{Y}\mathbf{A})'\mathbf{S}_e^{-1}\mathbf{Y}\mathbf{A}$. Since $\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_e$, these eigenvalues are related as follows:

$$(3.3) \quad \ell_i = \frac{d_i}{1 - d_i}, \quad d_i = \frac{\ell_i}{1 + \ell_i}, \quad i = 1, \dots, q.$$

Rewriting \mathbf{S}_e as

$$\mathbf{S}_e = \mathbf{S}_t^{1/2} \left\{ \mathbf{I}_p - \frac{1}{n} \mathbf{S}_t^{-1/2} \mathbf{Y} \mathbf{A} (\mathbf{S}_t^{-1/2} \mathbf{Y} \mathbf{A})' \right\} \mathbf{S}_t^{1/2},$$

and applying the following inverse matrix formula:

$$(3.4) \quad (\mathbf{I}_p + \mathbf{U} \mathbf{V})^{-1} = \mathbf{I}_p - \mathbf{U} (\mathbf{I}_q + \mathbf{V} \mathbf{U})^{-1} \mathbf{V}$$

for any $p \times q$ matrix \mathbf{U} and $q \times p$ matrix \mathbf{V} , we have

$$(3.5) \quad \mathbf{L} = \mathbf{D} + \mathbf{D} (\mathbf{I}_q - \mathbf{D})^{-1} \mathbf{D} = (\mathbf{I}_q - \mathbf{D})^{-1} \mathbf{D}.$$

Substituting (3.2) to (3.5), we have

$$(3.6) \quad \mathbf{L} = \frac{p}{n-p} \mathbf{I}_q + \frac{1}{\sqrt{n}} \left(\frac{n}{n-p} \right)^2 \sqrt{2c(1-c)} \mathbf{W} + o_P(n^{-1/2}).$$

This implies Theorems 2.5 and 2.6. \square

Alternatively, Theorem 2.6 can be obtained by using Theorem 2.4 and relation (3.3).

PROOF OF COROLLARY 2.1. Note that $\tilde{T}_1 = -\log |\mathbf{I}_q - \mathbf{D}|$. Using (3.2), we have

$$\begin{aligned} \tilde{T}_1 &= -\log \left(\frac{n-p}{n} \right)^q \left| \mathbf{I}_q - \frac{n}{(n-p)\sqrt{n}} \sqrt{2c(1-c)} \mathbf{W} \right| + o_P(n^{-1/2}) \\ &= -q \log \left(\frac{n-p}{n} \right) + \frac{n}{(n-p)\sqrt{n}} \sqrt{2c(1-c)} \mathbf{W} + o_P(n^{-1/2}). \end{aligned}$$

This implies that

$$T_1 = \sqrt{2/(1-c)} \operatorname{tr} \mathbf{W} + o_P(1)$$

and hence T_1 converges in distribution to $N(0, 2q/(1-c))$. Similarly, the asymptotic results of T_2 and T_3 are obtained by noting that $\tilde{T}_2 = \operatorname{tr}(\mathbf{D}(\mathbf{I}_q - \mathbf{D})^{-1})$, $\tilde{T}_3 = \operatorname{tr} \mathbf{D}$ and by using (3.2). \square

4. Extension to eigenvalues in multivariate regression model. Suppose that \mathbf{Y} is a $p \times n$ observation matrix whose columns are independently distributed with the same covariance matrix $\mathbf{\Sigma}$ and

$$E(\mathbf{Y}') = \mathbf{1}_n \boldsymbol{\theta}' + \mathbf{A}_1 \boldsymbol{\Theta}_1 + \mathbf{A}_2 \boldsymbol{\Theta}_2,$$

where $\boldsymbol{\theta}$ is an unknown $p \times 1$ vector, $\boldsymbol{\Theta}_i$ are unknown $q_i \times p$ matrices and \mathbf{A}_i are given $n \times q_i$ matrices. Let $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2]$. We assume that $\operatorname{rank}([\mathbf{1}_n \ \mathbf{A}]) = 1 + q_1 + q_2$, and entries of \mathbf{A} satisfies (2.3).

For testing the hypothesis $\Theta_2 = \mathbf{O}$, we may assume by reparametrization if necessary (see Section 2 in the Supplementary Material [Bai et al. (2018)] for details) that

$$(4.1) \quad \mathbf{A}'_i \mathbf{A}_i = \mathbf{I}_{q_i}, \quad \mathbf{A}'_i \mathbf{1}_n = \mathbf{0}, \quad \mathbf{A}'_1 \mathbf{A}_2 = \mathbf{O}, \quad i = 1, 2.$$

Then the matrices of sums of squares and products due to the error (\mathbf{S}_e) and due to the hypothesis (\mathbf{S}_h) are expressed as

$$(4.2) \quad \mathbf{S}_e = \frac{1}{n} \mathbf{Y}(\mathbf{I}_n - \mathbf{P}_0 - \mathbf{A}_1 \mathbf{A}'_1 - \mathbf{A}_2 \mathbf{A}'_2) \mathbf{Y}' \quad \text{and} \quad \mathbf{S}_h = \frac{1}{n} \mathbf{Y} \mathbf{A}_2 \mathbf{A}'_2 \mathbf{Y}',$$

where $\mathbf{P}_0 = (1/n) \mathbf{1}_n \mathbf{1}'_n$. The matrix of sums of squares and products due to the total is

$$(4.3) \quad \mathbf{S}_t = \mathbf{S}_h + \mathbf{S}_e = \frac{1}{n} \mathbf{Y}(\mathbf{I}_n - \mathbf{P}_0 - \mathbf{A}_1 \mathbf{A}'_1) \mathbf{Y}'.$$

We consider the distribution of the eigenvalues of $\mathbf{S}_h \mathbf{S}_t^{-1}$ or $\mathbf{S}_h \mathbf{S}_e^{-1}$ under the null hypothesis, that is, $E(\mathbf{Y}') = \mathbf{1}_n \boldsymbol{\theta}' + \mathbf{A}_1 \boldsymbol{\Theta}_1$. Moreover, without loss of generality, we may assume that $E(\mathbf{Y}) = \mathbf{0}$ since $\mathbf{S}_h, \mathbf{S}_e$ and \mathbf{S}_t remain unchanged if \mathbf{Y} is replaced by $\mathbf{Y} - \boldsymbol{\theta} \mathbf{1}'_n - \boldsymbol{\Theta}'_1 \mathbf{A}'_1$.

We may consider the case $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] = [y_{ij}]$ where y_{ij} are *i.i.d.* random variables with mean 0, variance 1 and finite fourth moments. From our assumptions in (4.1), it is easy to verify that matrix $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2]$ satisfies conditions (2.1) and (2.2); and condition (2.3) by our setup. Let $\mathbf{S}_0 = (1/n) \mathbf{Y}(\mathbf{I}_n - \mathbf{P}_0) \mathbf{Y}'$. By the same way as in (3.2), we have

$$(4.4) \quad \tilde{\mathbf{D}} = \frac{1}{n} (\mathbf{Y} \mathbf{A})' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A} = \frac{p}{n} \mathbf{I}_q + \frac{\sqrt{2c(1-c)}}{\sqrt{n}} \mathbf{W} + o_P(n^{-1/2}),$$

where $q = q_1 + q_2$, \mathbf{W}_{ij} is of order $q_i \times q_j$ for $i, j = 1, 2$, and

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix}.$$

Here, $\mathbf{W} = [w_{ij}]$ is a $q \times q$ symmetric Gaussian orthogonal ensemble. Our purpose is to derive the limiting distribution of the eigenvalues of $\mathbf{D} = \frac{1}{n} (\mathbf{Y} \mathbf{A}_2)' \mathbf{S}_t^{-1} \mathbf{Y} \mathbf{A}_2$, and that of $\mathbf{L} = \frac{1}{n} (\mathbf{Y} \mathbf{A}_2)' \mathbf{S}_e^{-1} \mathbf{Y} \mathbf{A}_2$. Noting that $\mathbf{S}_t = \mathbf{S}_0 - \frac{1}{n} \mathbf{Y} \mathbf{A}_1 \mathbf{A}'_1 \mathbf{Y}'$ and using (3.4), we have

$$\begin{aligned} \mathbf{D} &= \frac{1}{n} \mathbf{A}'_2 \mathbf{Y}' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A}_2 \\ &\quad + \frac{1}{n} \mathbf{A}'_2 \mathbf{Y}' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A}_1 \left(\mathbf{I}_{q_1} - \frac{1}{n} \mathbf{A}'_1 \mathbf{Y}' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A}_1 \right)^{-1} \frac{1}{n} \mathbf{A}'_1 \mathbf{Y}' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A}_2. \end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
 \mathbf{D} &= \frac{1}{n} \mathbf{A}'_2 \mathbf{Y}' \mathbf{S}_0^{-1} \mathbf{Y} \mathbf{A}_2 + o_p(n^{-1/2}) \\
 (4.5) \quad &= \frac{p}{n} \mathbf{I}_{q_2} + \sqrt{2/n} \mathbf{W}_{22} + o_p(n^{-1/2}).
 \end{aligned}$$

Furthermore, using $\mathbf{L} = (\mathbf{I}_{q_2} - \mathbf{D})^{-1} \mathbf{D}$, we have

$$(4.6) \quad \mathbf{L} = \frac{p}{n-p} \mathbf{I}_{q_2} + \sqrt{2/n} \left(\frac{n}{n-p} \right)^2 \mathbf{W}_{22} + o_p(n^{-1/2}).$$

This implies that the eigenvalues of \mathbf{D} (resp., \mathbf{L}) are asymptotically distributed as in Theorem 2.4 (resp., Theorem 2.6) replacing q by q_2 in (4.5) above.

We have shown that the results in MANOVA can be applied to some tests in multivariate regression model. However, in order to test the significance of the intercept of the linear model, we need to generalize our results in this section which will be left as future work.

5. Eigenvalues in MANOVA under local alternatives. Recall, in Section 3, the matrix of sums of squares and products due to between-groups is represented as

$$n\mathbf{S}_b = (\mathbf{Y}\mathbf{A} + \sqrt{n}\mathbf{\Xi}_b\mathbf{F})(\mathbf{Y}\mathbf{A} + \sqrt{n}\mathbf{\Xi}_b\mathbf{F})',$$

where $\mathbf{\Xi}_b$ and \mathbf{F} are defined there. In this section, we derive the limiting distributions of the nonzero eigenvalues $\ell_1 > \dots > \ell_q$ of $\mathbf{S}_b\mathbf{S}_e^{-1}$ and the nonzero eigenvalues $d_1 > \dots > d_q$ of $\mathbf{S}_b\mathbf{S}_t^{-1}$ under a local alternative case. Specifically, instead of assuming $\mathbf{\Xi}_b = \mathbf{O}$, we assume that as $p, n \rightarrow \infty$ satisfying $p/n \rightarrow c \in (0, 1)$,

$$(5.1) \quad \sqrt{n}(\mathbf{\Xi}_b\mathbf{F})' \mathbf{\Xi}_b\mathbf{F} \rightarrow \mathbf{Q},$$

where \mathbf{Q} is a fixed positive definite matrix.

THEOREM 5.1. *Using the same notation and following the same assumptions as in Theorem 2.4, instead of the null case, we assume the local alternative given in (5.1). Then the limiting distribution of $\sqrt{n}(d_i - p/n)$, $i = 1, \dots, q$ is the same as that of the eigenvalues of $\sqrt{2c(1-c)}\mathbf{W} + (1-c)\mathbf{Q}$, where \mathbf{W} is a $q \times q$ Gaussian orthogonal ensemble.*

Similarly, the limiting distribution of $\sqrt{n}(\ell_i - p/(n-p))$, $i = 1, \dots, q$ is the same as that of the eigenvalues of $\sqrt{2c/(1-c)^3}\mathbf{W} + (1-c)^{-1}\mathbf{Q}$.

By a similar reasoning as in the proof of Theorem 2.4, the limiting distribution of the nonzero eigenvalues $\ell_1 > \dots > \ell_q$ of $\mathbf{S}_b\mathbf{S}_e^{-1}$, is the same as the limiting distribution of the eigenvalues of

$$(5.2) \quad \mathbf{U} := (\mathbf{Y}\mathbf{A} + \sqrt{n}\mathbf{\Xi}_b\mathbf{F})' \{ \mathbf{Y}(\mathbf{I}_n - \mathbf{G}\mathbf{G}')\mathbf{Y}' \}^{-1} (\mathbf{Y}\mathbf{A} + \sqrt{n}\mathbf{\Xi}_b\mathbf{F}).$$

In order to show the result on the eigenvalues $\ell_i, i = 1, \dots, q$ in Theorem 5.1, it suffices to show that the limiting distribution of the nonzero eigenvalues of $\mathbf{S}_b \mathbf{S}_e^{-1}$ normalized by the location $p(n - p)^{-1} \mathbf{I}_q$ and the scale \sqrt{n} is that of the eigenvalues of the sum of $\sqrt{2c/(1 - c)^3} \mathbf{W}$ and $\mathbf{Q}/(1 - c)$. More precisely,

$$(5.3) \quad \sqrt{n} \left(\mathbf{U} - \frac{p}{n-p} \mathbf{I}_q \right) \xrightarrow{D} \tilde{\mathbf{W}} = \frac{\sqrt{2c}}{(1 - c)^{3/2}} \mathbf{W} + \frac{1}{1 - c} \mathbf{Q},$$

which will be proved later. The result of the d_i 's is obtained by using (5.3) and the fact that d_i 's are also the eigenvalues of $\mathbf{U}(\mathbf{I}_q + \mathbf{U})^{-1}$. In fact, using $\mathbf{U} = p(n - p)^{-1} \mathbf{I}_q + n^{-1/2} \tilde{\mathbf{W}} + o_P(n^{-1/2})$, we have

$$\begin{aligned} (\mathbf{I}_q + \mathbf{U})^{-1} &= \frac{n - p}{n} \left(\mathbf{I}_q + \frac{n - p}{n^{3/2}} \tilde{\mathbf{W}} + o_P(n^{-1/2}) \right)^{-1} \\ &= \frac{n - p}{n} \left(\mathbf{I}_q - \frac{n - p}{n^{3/2}} \tilde{\mathbf{W}} + o_P(n^{-1/2}) \right), \end{aligned}$$

which implies

$$\begin{aligned} \mathbf{U}(\mathbf{I}_q + \mathbf{U})^{-1} &= \frac{p}{n} \left(\mathbf{I}_q + \frac{n - p}{p\sqrt{n}} \tilde{\mathbf{W}} \right) \left(\mathbf{I}_q - \frac{n - p}{n^{3/2}} \tilde{\mathbf{W}} \right) + o_P(n^{-1/2}) \\ &= \frac{p}{n} \mathbf{I}_q + \frac{1}{\sqrt{n}} (1 - p/n)^2 \tilde{\mathbf{W}} + o_P(n^{-1/2}) \\ &= \frac{p}{n} \mathbf{I}_q + \frac{1}{\sqrt{n}} [\sqrt{2c(1 - c)} \mathbf{W} + (1 - c) \mathbf{Q}] + o_P(n^{-1/2}). \end{aligned}$$

This proves the assertion on $\sqrt{n}(d_i - p/n), 1 \leq i \leq q$ in Theorem 5.1.

Note that the test statistics, $\tilde{T}_i, i = 1, 2, 3$ can be expressed in terms of \mathbf{U} as

$$\tilde{T}_1 = \log |\mathbf{I}_q + \mathbf{U}|, \quad \tilde{T}_2 = \text{tr} \mathbf{U}, \quad \tilde{T}_3 = \text{tr} [\mathbf{U}(\mathbf{I}_q + \mathbf{U})^{-1}].$$

From (5.3), we have

$$T_i \xrightarrow{D} N(c^{-1/2} \text{tr} \mathbf{Q}, 2q/(1 - c)).$$

Therefore, these three test statistics have the same asymptotic power under the local alternative. Under the null hypothesis, from (2.12), it follows that $T_i/(2q/(1 - c))^{1/2} \sim N(0, 1)$. Therefore, the rejection region is given by $T_i > (2q/(1 - c))^{1/2} z_\alpha$. So the asymptotic powers are given by

$$(5.4) \quad \Phi((2cq)^{-1/2} \sqrt{1 - c} \text{tr} \mathbf{Q} - z_\alpha),$$

where Φ is the distribution function of a standard normal, and z_α is the upper α quantile of a standard normal distribution. Consequently, this shows that the asymptotic power is a decreasing function of c and tends to α as c tends to one.

It is known that accuracy of approximation by the large-sample asymptotic distributions of the invariant test statistics including $T_i, i = 1, 2, 3$ becomes worse as p gets larger and larger. This problem can be solved by using the high-dimensional approximations. Further, from (5.4) we can see that the powers are consistent as p/n tends to one. When p is large but p/n is small, it is expected that the invariant tests $T_i, i = 1, 2, 3$ are useful in comparison with some high-dimensional tests. Among the high-dimensional tests, there are two types of Dempster tests given as generalizations of Dempster (1960) and Bai and Saranadasa (1996). These are seen in Fujikoshi, Himeno and Wakaki (2004), Srivastava and Fujikoshi (2006), Schott (2007), Srivastava and Kubokawa (2013), etc. Here, we consider two test statistics given by

$$(5.5) \quad T_D = \sqrt{p} \left(n \frac{\text{tr} \mathbf{S}_b}{\text{tr} \mathbf{S}_e} - q \right) \quad \text{and} \quad T_B = \sqrt{\frac{q}{p}} \left(\frac{n}{q} \text{tr} \mathbf{S}_b - \text{tr} \mathbf{S}_e \right).$$

Assuming the condition,

$$p^{-1} \text{tr} \boldsymbol{\Sigma}^k = O(1) \quad \text{and} \quad p^{-1} n \text{tr} [(\boldsymbol{\Xi}_b \mathbf{F})' \boldsymbol{\Sigma}^k (\boldsymbol{\Xi}_b \mathbf{F})] = O(1), \quad k = 1, 2.$$

Furthermore, in the local alternatives satisfying (5.4), assume that

$$\mathbf{Q} = \mathbf{M}' \boldsymbol{\Sigma}^{-1} \mathbf{M},$$

where \mathbf{M} is a $p \times q$ constant matrix. Then, under normality, we have from Fujikoshi, Himeno and Wakaki (2004) that the asymptotic power of T_D under the local alternatives can be expressed as

$$(5.6) \quad \Phi \left(\frac{\text{tr} \mathbf{M} \mathbf{M}'}{\sqrt{(2cq/p) \text{tr} \boldsymbol{\Sigma}^2}} - z_\alpha \right).$$

Similarly, we can show that T_B has the same asymptotic local as that of T_D . From (5.4) and (5.6), we can conclude that \tilde{T}_i 's are asymptotically more powerful than T_B and T_D if

$$\frac{\sqrt{1-c}}{\sqrt{2cq}} \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{M} \mathbf{M}') > \frac{\text{tr} \mathbf{M} \mathbf{M}'}{\sqrt{(2cq/p) \text{tr} \boldsymbol{\Sigma}^2}},$$

which is equivalent to

$$R := p^{-1/2} \sqrt{\text{tr} \boldsymbol{\Sigma}^2} \frac{\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{M} \mathbf{M}')}{\text{tr} \mathbf{M} \mathbf{M}'} > (1-c)^{-1/2}.$$

Let $\lambda_1 \geq \dots \geq \lambda_p > 0$ be the eigenvalues of $\boldsymbol{\Sigma}$. Then $\lambda_p \leq p^{-1} \text{tr} \boldsymbol{\Sigma}^2 \leq \lambda_1$. Let $\text{vec}(\mathbf{M})$ be the $pq \times 1$ column vector obtained by stacking the columns of \mathbf{M} on top of one another. Then, noting that

$$\frac{\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{M} \mathbf{M}')}{\text{tr}(\mathbf{M} \mathbf{M}')} = \frac{\text{vec}(\mathbf{M})' \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_q \text{vec}(\mathbf{M})}{\text{vec}(\mathbf{M})' \text{vec}(\mathbf{M})},$$

we have

$$\lambda_1^{-1} = \min_{\mathbf{M}} \frac{\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{M}')}{\text{tr}(\mathbf{M}\mathbf{M}')} \leq \max_{\mathbf{M}} \frac{\text{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{M}\mathbf{M}')}{\text{tr}(\mathbf{M}\mathbf{M}')} = \lambda_p^{-1}.$$

Therefore,

$$(5.7) \quad \frac{\lambda_p}{\lambda_1} \leq R \leq \frac{\lambda_1}{\lambda_p}.$$

In general, if $\frac{\lambda_1}{\lambda_p}$ is large, we can expect that R is large, and the invariant tests \tilde{T}_i 's shall be asymptotically more powerful than T_D and T_B under the condition that c is small.

PROOF OF (5.3). We expand \mathbf{U} in (5.2) to obtain

$$(5.8) \quad \mathbf{U} = \mathbf{U}_1 + \mathbf{U}_2 + \mathbf{U}'_2 + \mathbf{U}_3,$$

where

$$\mathbf{U}_1 = \mathbf{A}'\mathbf{Y}'(\mathbf{Y}(\mathbf{I} - \mathbf{G}\mathbf{G}')\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A},$$

$$\mathbf{U}_2 = \sqrt{n}\mathbf{A}'\mathbf{Y}'(\mathbf{Y}(\mathbf{I} - \mathbf{G}\mathbf{G}')\mathbf{Y}')^{-1}\boldsymbol{\Xi}_b\mathbf{F},$$

$$\mathbf{U}_3 = n\mathbf{F}'\boldsymbol{\Xi}'_b(\mathbf{Y}(\mathbf{I} - \mathbf{G}\mathbf{G}')\mathbf{Y}')^{-1}\boldsymbol{\Xi}_b\mathbf{F}.$$

It is easy to see that \mathbf{U}_1 is the same statistic as in the null hypothesis. An easy adaptation of Theorem 2.3 which treats $\mathbf{S}_b\mathbf{S}_t^{-1}$ to $\mathbf{S}_b\mathbf{S}_e^{-1}$ leads to

$$\sqrt{n}\left(\mathbf{U}_1 - \frac{p}{n-p}\mathbf{I}_q\right) \xrightarrow{D} \frac{\sqrt{2c}}{(1-c)^{3/2}}\mathbf{W}.$$

To complete the proof of (5.3), it suffices to derive the limits below

$$(5.9) \quad \sqrt{n}\mathbf{U}_3 \xrightarrow{P} \frac{1}{1-c}\mathbf{Q},$$

$$(5.10) \quad \sqrt{n}\mathbf{U}_2 \xrightarrow{P} \mathbf{0}.$$

Proofs of (5.9) and (5.10) are given in the Supplementary Material. \square

Discussion on Roy's test. Though this paper focuses on the three classical MANOVA tests, Theorems 2.4, 2.6 and 5.1 have important connection with the Roy's test, another classical MANOVA test. Recall Roy's test [see Anderson (2003)] is based on the largest eigenvalue, d_1 , of $\mathbf{S}_b\mathbf{S}_t^{-1}$ (or the largest eigenvalue, ℓ_1 , of $\mathbf{S}_b\mathbf{S}_t^{-1}$). Under normal populations assumption, computing the exact null distribution of d_1 , for given p, q and n , has been studied by many authors; see, for example, Krishnaiah and Chang (1971) and Chiani (2016) and references therein. Dropping normal populations assumption, Theorem 2.4 shows that the asymptotic

null distribution of $\sqrt{n/\{2c(1-c)\}}(d_1 - p/n)$ as $p/n \rightarrow c \in (0, 1)$ corresponds to that of the largest eigenvalue of the GOE, \mathbf{W} . Hence, theoretically speaking, the limiting null distribution of the normalized Roy's test can be found in an integral form, but this integral may not be computationally easy to use. On the other hand, Chiani (2016) has devised an efficient algorithm to compute the distribution of the largest eigenvalue of \mathbf{W} numerically. His method is to compute the determinant of a skew-symmetric matrix whose elements can be recursively computed. It will be an interesting future project to examine the approximation accuracy of the approximation approach based on our limiting result and the approach of Chiani (2016) for finite sample in the nonnormal case.

According to Theorem 5.1, the asymptotic local alternative distribution of $\sqrt{n/\{2c(1-c)\}}(d_1 - p/n)$, corresponds to that of the largest eigenvalue of $\mathbf{W} + \sqrt{(1-c)/(2c)}\mathbf{Q}$. So the limiting local alternative distribution of the normalized Roy's test can also be found in theory. To the best of our knowledge, there is no known computational method to compute it except for \mathbf{Q} is of the form $a\mathbf{I}_q$. For a given local alternative \mathbf{Q} , since in real application, the value of q is not large, we suggest to use the Monte Carlo method to calculate the asymptotic power of Roy's test. Moreover, as the distribution of \mathbf{W} is invariant under the transformation, $\mathbf{U}'\mathbf{W}\mathbf{U}$, for any orthogonal matrix \mathbf{U} , thus we may assume that \mathbf{Q} is diagonal with decreasing diagonal entries in the Monte Carlo simulation.

6. Some simulation studies. Corollary 2.1 shows that the suitably normalized versions of the likelihood ratio statistic, Lawley–Hotelling test statistic and Barlett–Nanda–Pillai test statistic, T_1 , T_2 and T_3 , converge in distribution to $N(0, \frac{2q}{1-c})$. Motivated by relevance with testing problems, our interest is to investigate how well the limiting distribution approximates those of the T_i 's especially in the upper tail. We conduct simulation studies to estimate the actual sizes of these tests given a nominal size, α .

We set $\alpha = 0.05$. Seven population distributions will be considered. They are (i) $N(0, 1)$, (ii) t_3 , (iii) t_4 , (iv) t_5 , (v) χ_3^2 , (vi) exponential with mean 1 and (vii) Poisson with mean 1. Distributions (i), (vi) and (vii) were considered in Pan and Zhou (2011) where they proved the central limit theorem of normalized Hotelling's T^2 statistic. Distributions (i) to (iv) are symmetric, whereas (v)–(vii) are skewed. Moreover, distributions (ii) and (iii) violate the finite fourth moment assumption, an assumption we made throughout this paper due to the methods of proof.

The parameters n , q and c are chosen as follows: $n \in \{40, 80, 160, 320, 480, 640, 800, 960\}$; $q = 1, 3$ and 7 ; and $c \in \{0.1, 0.2, 0.3, 0.5, 0.8, 0.9, 0.95\}$. We let $p = nc$, the data dimension. Here, n is the total sample size, and the sub-population sizes are taken to be equal, that is, $n_i = n/(q + 1)$ for $1 \leq i \leq q + 1$. In total, 1176 cases will be considered.

We describe the simulation procedure for one such case in some details. Given a distribution F (which will be standardized to have mean 0 and variance 1) from

one of the seven distributions, and parameters n, q and c from above, we generate np i.i.d. samples from the standardized F to form a $p \times n$ matrix $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n]$. From \mathbf{Y} , we compute $\mathbf{S}_t = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$. Note that $\mathbf{S}_b = \mathbf{X}\mathbf{X}'$ with

$$\mathbf{X} = \left[\sqrt{\frac{n_1}{n}}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}), \sqrt{\frac{n_2}{n}}(\bar{\mathbf{y}}_2 - \bar{\mathbf{y}}), \dots, \sqrt{\frac{n_{q+1}}{n}}(\bar{\mathbf{y}}_{q+1} - \bar{\mathbf{y}}) \right].$$

We observe that $\mathbf{S}_b\mathbf{S}_t^{-1}$ and $\mathbf{X}'\mathbf{S}_t^{-1}\mathbf{X}$ have the same set of q positive eigenvalues. However, computationally it is much easier to deal with the latter as it is of order $(q + 1) \times (q + 1)$ whereas the former is of order $p \times p$ (note that p is much greater than q and is increasing with n). With these q positive sample eigenvalues, and for each $i = 1, 2, 3$, we compute the value of the test statistic T_i . We repeat this procedure 1000 times generating 1000 test values of these three test statistics. We call the estimate of the actual size of a test empirical size: the proportion of test values exceeding $z_{0.95}\sqrt{2q/(1 - c)}$. If the approximation at the tail is good, we expect the empirical size is close to the nominal size, 0.05.

Figure 1 shows the plots of the empirical sizes of these three tests for various choices of n, c and q . Due to space consideration, we only show the plots for normal distribution and standardized χ_3^2 distribution. These two plots look remarkably similar. See the Supplementary Material for the plots and their numerical values. We summarize some of our observations and comments from the simulation studies.

Surprisingly, the plots for the seven distributions under consideration are almost the same, suggesting that the empirical sizes of the test statistic $T_i, i = 1, 2, 3$, are robust against the underlying distribution of the population. This robustness is also evident in the simulation results of Pan and Zhou (2011) for test T_2 , distributions (i), (vi) and (vii) for $c = 1/2$ and $n = 200, 400$ and 800 .

Size distortions (i.e., empirical size minus nominal size) are consistently (1155 cases out of 1176) positive. Not unexpectedly, size distortion decreases as n increases (across the rows in Figure 1); and increases as q (across the columns) or c (along the x -axis in each plot) increases.

The normalized Barlett–Nanda–Pillai test statistic T_3 has the smallest size distortion, and the normalized Lawley–Hotelling’s T_2 the largest size distortion among the three tests considered in our simulation studies.

If $q = 7$, almost all the empirical sizes of all three statistics exceed 0.08 even when the total sample size reaches 960 (or sub-sample size 120). A few exceptions are only observed for T_3 at $c = 0.1$ and large n such as 800. It will be interesting to investigate if Bartlett-type transforms will improve the approximations.

If one considers the empirical size of a test lying within 0.03 and 0.08 acceptable, and if one would like to use one of these three invariant tests for nominal size 0.05, we recommend the normalized Barlett–Nanda–Pillai test statistic and (a) for $q = 1$, total sample size $n \geq 80$ with $c \leq 0.95$; and (b) for $q = 2, 3$, total sample $n \geq 80$ and $c \leq 0.5$.

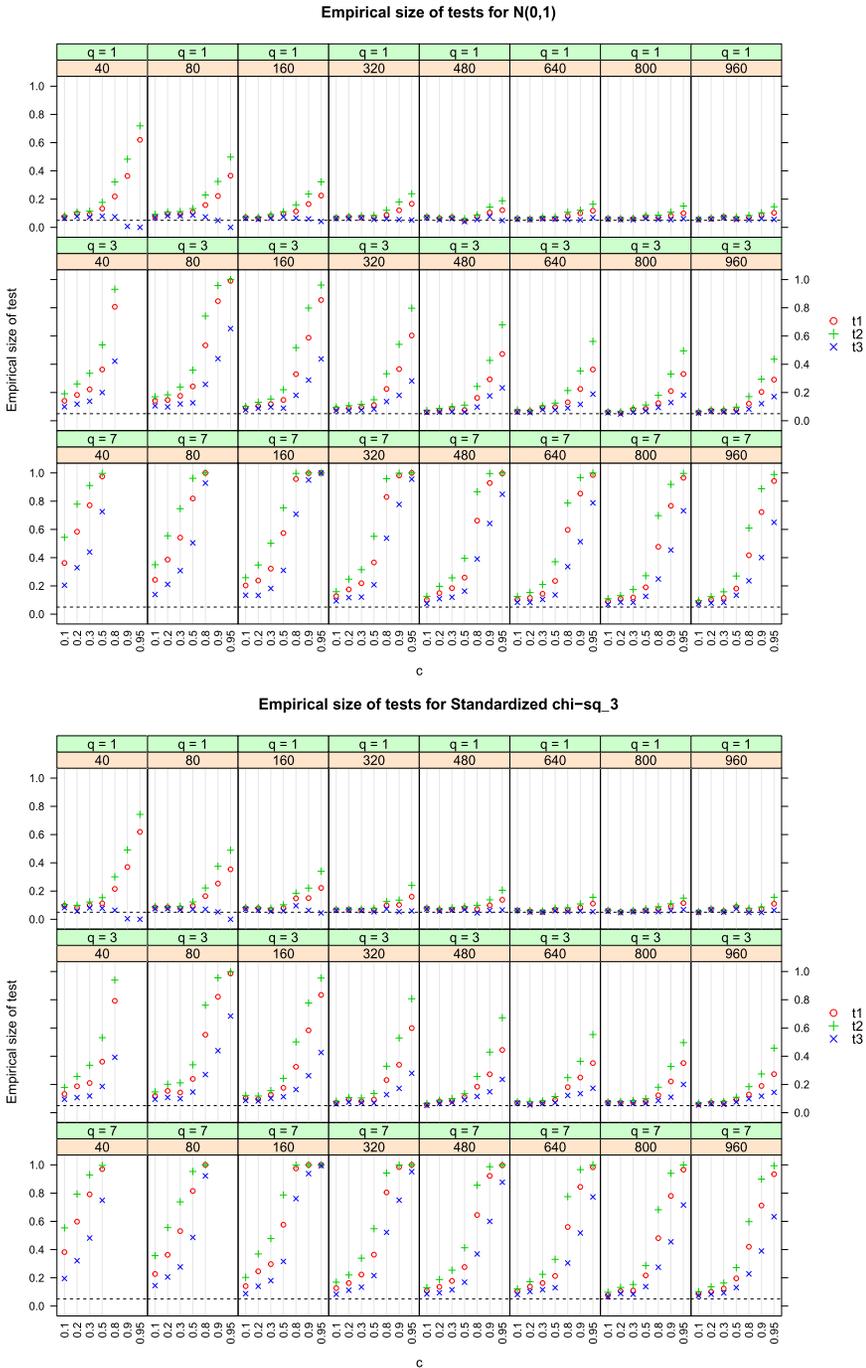


FIG. 1. Plots of the empirical sizes of the normalized versions of the three invariant tests considered in this article. The dotted lines denote the 5% nominal level. The underlying distribution of the population is normal for the top panel; standardized χ^2_3 for the bottom.

7. Proof of Theorem 2.1. By Bai and Silverstein (2004), for any positive constant $\mu < a = (1 - \sqrt{c})^2$ and any given $\ell > 0$, after truncation, we have

$$(7.1) \quad P(\lambda_{\min}(\mathbf{S}) < \mu) = o(n^{-\ell}).$$

Define $\mathcal{B} = 1_{\{\lambda_{\min}(\mathbf{S}) \geq \mu\}}$ which converges to 1 a.s. as $n \rightarrow \infty$. Throughout this paper, all functions under the expectation sign are assumed having the variable \mathcal{B} inserted though not explicitly indicated, for otherwise, the expectation may not exist.

By the same reasoning, the estimation (7.1) holds when \mathbf{S} is replaced by \mathbf{S}_k or \mathbf{S}_{jk} . In the paper, we shall replace \mathcal{B} by \mathcal{B}_k or \mathcal{B}_{jk} as and when needed. Most of the time, we do this implicitly.

We shall present the proofs under the condition that the components of \mathbf{y}_k are *i.i.d.* The proofs for independent components, mean 0, variance 1 and uniformly integrable $\{y_{ik}^4\}$ setting is basically the same except for the truncation and centralization step. For the rest of this paper, we shall assume that there exists a sequence $\eta_n \downarrow 0$ such that the basic random variables, y_{ij} 's, are truncated at $\eta_n \sqrt{n}$ and then renormalized. See Section 1 of the Supplementary Material for details.

Before proceeding to the proof of Theorem 2.1, we shall introduce further notation and collect some useful preliminary results in the following Section 7.1. The proofs of these preliminary results are given in the Supplementary Material.

7.1. Preliminary lemmas. For $1 \leq k \leq n$ and $1 \leq j \leq p$, recall \mathbf{y}_k and \mathbf{z}'_j denote the k th column and j th row of \mathbf{Y} , respectively; and \mathbf{Y}_k and \mathbf{Z}_j are the matrices after removing the k th column and the j th row of \mathbf{Y} .

Define, for $1 \leq j \leq p$,

$$\begin{aligned} B_j &= \{\mathbf{z}'_j \mathbf{z}_j - \mathbf{z}'_j \mathbf{Z}'_j (\mathbf{Z}_j \mathbf{Z}'_j)^{-1} \mathbf{Z}_j \mathbf{z}_j\}^{-1}, \\ \alpha_j &= n - \text{tr}\{\mathbf{Z}'_j (\mathbf{Z}_j \mathbf{Z}'_j)^{-1} \mathbf{Z}_j\} = n - p + 1, \\ W_j &= \mathbf{z}'_j \{\mathbf{I}_n - \mathbf{Z}'_j (\mathbf{Z}_j \mathbf{Z}'_j)^{-1} \mathbf{Z}_j\} \mathbf{z}_j - \alpha_j \\ \beta_j &= 1/\alpha_j = 1/(n - p + 1); \end{aligned}$$

and for $1 \leq k \leq n$,

$$\begin{aligned} C_k &= \{1 + \mathbf{y}'_k (\mathbf{Y}_k \mathbf{Y}'_k)^{-1} \mathbf{y}_k\}^{-1}, \\ M_k &= \text{tr}(\mathbf{Y}_k \mathbf{Y}'_k)^{-1}, \\ V_k &= \mathbf{y}'_k (\mathbf{Y}_k \mathbf{Y}'_k)^{-1} \mathbf{y}_k - M_k, \\ \bar{C}_k &= 1/(1 + M_k). \end{aligned}$$

With these notation, we have

$$(7.2) \quad B_j = \beta_j - \beta_j B_j W_j = \beta_j - \beta_j^2 W_j + \beta_j^2 B_j W_j^2, \quad 1 \leq j \leq p;$$

$$(7.3) \quad C_k = \bar{C}_k - \bar{C}_k C_k V_k = \bar{C}_k - \bar{C}_k^2 V_k + \bar{C}_k^2 C_k V_k^2, \quad 1 \leq k \leq n.$$

LEMMA 7.1. For any $n \times n$ positive definite matrix \mathbf{B} and n -vector \mathbf{r} , we have

$$(\mathbf{B} + \mathbf{r}\mathbf{r}')^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{r}\mathbf{r}'\mathbf{B}^{-1}}{1 + \mathbf{r}'\mathbf{B}^{-1}\mathbf{r}},$$

$$(\mathbf{B} + \mathbf{r}\mathbf{r}')^{-1}\mathbf{r} = \frac{\mathbf{B}^{-1}\mathbf{r}}{1 + \mathbf{r}'\mathbf{B}^{-1}\mathbf{r}}.$$

LEMMA 7.2. Let $r \geq 1$ be an integer.

(a) We have

$$\lim_{n \rightarrow \infty} \sup_{1 \leq k \leq n} \left| \frac{1}{n} \mathbf{y}'_k \mathbf{S}_k^{-r} \mathbf{y}_k - \frac{1}{n} \text{tr} \mathbf{S}_k^{-r} \right| = 0 \quad \text{a.s.}$$

(b) Moreover, uniformly over $1 \leq k \leq n$ on the set $\bigcup_{k=1}^n \{\lambda_{\min}(\mathbf{S}_k) \geq \mu\}$ where $0 < \mu < a = (1 - \sqrt{c})^2$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \mathbf{S}_k^{-r} \stackrel{\text{a.s.}}{=} c \int_a^b x^{-r} \frac{\sqrt{(b-x)(x-a)}}{2\pi cx} dx := c_r.$$

Here, $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$. In particular, $c_1 = c/(1 - c)$ and $c_2 = c/(1 - c)^3$.

LEMMA 7.3. Let \mathbf{B} and \mathbf{C} be $n \times n$ matrices. Let $\mathbf{y} = (y_1, \dots, y_n)'$ be a random n -vector with y_i 's being independent, mean 0, variance 1 and finite fourth moments. Let $\kappa := \max_{1 \leq i \leq n} E(y_i^4) - 3$.

(a) If, in addition, y_i 's are assumed to be identically distributed, then

$$(7.4) \quad E\{(\mathbf{y}'\mathbf{B}\mathbf{y} - \text{tr}\mathbf{B})(\mathbf{y}'\mathbf{C}\mathbf{y} - \text{tr}\mathbf{C})\} = \text{tr}(\mathbf{B}\mathbf{C}) + \text{tr}(\mathbf{B}\mathbf{C}') + \kappa \sum_{i=1}^n b_{ii} c_{ii}.$$

(b) Suppose \mathbf{B} is symmetric with bounded norm, and that $|y_i| \leq \eta_n \sqrt{n}$ for some sequence $\eta_n \downarrow 0$ in (7.6), then

$$(7.5) \quad E\{(\mathbf{y}'\mathbf{B}\mathbf{y} - \text{tr}\mathbf{B})^2\} \leq (\kappa + 2) \text{tr}(\mathbf{B}^2),$$

$$(7.6) \quad E\{(\mathbf{y}'\mathbf{B}\mathbf{y} - \text{tr}\mathbf{B})^4\} = o(n^3).$$

LEMMA 7.4. As $p, n \rightarrow \infty$ such that $c_{n,p} := p/n \rightarrow c \in (0, 1)$, we have

$$(7.7) \quad \sqrt{n} \max_{1 \leq i \leq n} [E\{\mathbf{y}'_i (\mathbf{Y}\mathbf{Y}')^{-1} \mathbf{y}_i\} - c_{n,p}] = O(n^{-1/2}),$$

$$(7.8) \quad \max_{1 \leq i < j \leq n} E\{\mathbf{y}'_i (\mathbf{Y}\mathbf{Y}')^{-1} \mathbf{y}_j\} = O(n^{-1}).$$

7.2. *Main argument for Theorem 2.1.* Our proof of Theorem 2.1 consists of three steps. In step 1, we compute the mean, $\mathbf{M} = E\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A}\}$, and prove that $\sqrt{n}(\mathbf{M} - \frac{p}{n}\mathbf{I}_q) \rightarrow \mathbf{O}$. Note that \mathbf{M} is asymptotically diagonal. In step 2, we prove the convergence of $\mathbf{V} := \sqrt{n}\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A} - \mathbf{M}\}$ to a symmetric random matrix with entries normally distributed using martingale decomposition. Finally, in step 3, we compute the covariance of any two entries in this limiting random matrix, thus identify the limit as $\sqrt{2c(1-c)}\mathbf{W}$.

Step 1. We shall first compute m_{ij} , the (i, j) th entry of \mathbf{M} and then deduce that

$$(7.9) \quad \sqrt{n}\left(\mathbf{M} - \frac{p}{n}\mathbf{A}'\mathbf{A}\right) = \sqrt{n}(\mathbf{M} - c_{n,p}\mathbf{I}_q) \rightarrow \mathbf{O},$$

where $c_{n,p} = p/n$. Recall \mathbf{a}_k denotes the k th column of \mathbf{A} , we have

$$\begin{aligned} m_{ij} &= \mathbf{a}'_i E\{\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\}\mathbf{a}_j = \sum_{1 \leq r,s \leq n} a_{ri}a_{sj} E\{\mathbf{y}'_r(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_s\} \\ &= \mathbf{a}'_i \mathbf{a}_j E\{\mathbf{y}'_1(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_1\} + \{(\mathbf{1}'_n \mathbf{a}_i)(\mathbf{1}'_n \mathbf{a}_j) - \mathbf{a}'_i \mathbf{a}_j\} E\{\mathbf{y}'_1(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_2\} \\ &= \mathbf{a}'_i \mathbf{a}_j E\{\mathbf{y}'_1(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_1\} + O(n^{-1}) \end{aligned}$$

by using (7.8). Applying (7.7) and (7.8) in Lemma 7.4, we have

$$\sqrt{n}(m_{ii} - c_{n,p}) = \sqrt{n}[E\{\mathbf{y}'_1(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{y}_1\} - c_{n,p} + O(n^{-1})] \rightarrow 0.$$

This proves (7.9), and completes step 1.

Step 2. We shall prove that $\mathbf{V} \xrightarrow{D} \mathbf{W}_c$ via martingale decomposition. Here, \mathbf{W}_c is a symmetric random matrix with normally distributed entries. With E_k, \mathbf{D}_k and $\tilde{\mathbf{D}}_k$ defined below, we outline our approach as follows: Rewrite \mathbf{V} as

$$(7.10) \quad \begin{aligned} \mathbf{V} &= \sqrt{n} \sum_{k=1}^p (E_k - E_{k-1})(\mathbf{D}_k) \\ &= \sqrt{n} \sum_{k=1}^p E_k(\tilde{\mathbf{D}}_k) + o_P(1), \end{aligned}$$

and then apply martingale central limit theorem to show that $\sqrt{n} \sum_{k=1}^n E_k(\tilde{\mathbf{D}}_k)$ converges to a random matrix with entries which are normally distributed.

First, recall \mathbf{Z}_k is the $(p-1) \times n$ matrix obtained from removing the k th row \mathbf{z}'_k from \mathbf{Y} . Let E_k denote the conditional expectation given $\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$, and E_0 is the usual expectation. We apply $(E_k - E_{k-1})\{\mathbf{A}'\mathbf{Z}'_k(\mathbf{Z}_k\mathbf{Z}'_k)^{-1}\mathbf{Z}_k\mathbf{A}\} = \mathbf{0}$ to obtain

$$\begin{aligned} \mathbf{V} &= \sqrt{n} \sum_{k=1}^p (E_k - E_{k-1})\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A} - \mathbf{A}'\mathbf{Z}'_k(\mathbf{Z}_k\mathbf{Z}'_k)^{-1}\mathbf{Z}_k\mathbf{A}\} \\ &= \sqrt{n} \sum_{k=1}^p (E_k - E_{k-1})(\mathbf{D}_k), \end{aligned}$$

where

$$(7.11) \quad \mathbf{D}_k = B_k(\mathbf{A}'\mathbf{P}_k\mathbf{z}_k\mathbf{z}'_k\mathbf{P}_k\mathbf{A})$$

and $\mathbf{P}_k = \mathbf{I}_n - \mathbf{Z}'_k(\mathbf{Z}_k\mathbf{Z}'_k)^{-1}\mathbf{Z}_k$. In the last equality, we used the following result:

$$\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y} = \mathbf{Z}'_k(\mathbf{Z}_k\mathbf{Z}'_k)^{-1}\mathbf{Z}_k + B_k\mathbf{P}_k\mathbf{z}_k\mathbf{z}'_k\mathbf{P}_k.$$

We introduce

$$\mathbf{Q}_k = \mathbf{A}'\mathbf{P}_k(\mathbf{z}_k\mathbf{z}'_k - \mathbf{I}_n)\mathbf{P}_k\mathbf{A}.$$

Note that \mathbf{Q}_k 's are not to be confused with \mathbf{Q} in (5.1).

With B_k, W_k and β_k defined at the beginning of Section 7.1, we define

$$\tilde{\mathbf{D}}_k = \beta_k\mathbf{Q}_k.$$

Note that $\mathbf{D}_k = B_k(\mathbf{Q}_k + \mathbf{A}'\mathbf{P}_k\mathbf{A})$. Applying (7.2), we have

$$\begin{aligned} \mathbf{D}_k - \tilde{\mathbf{D}}_k &= -\beta_k B_k W_k \mathbf{Q}_k + B_k \mathbf{A}'\mathbf{P}_k \mathbf{A} \\ &= -\beta_k B_k W_k \mathbf{Q}_k + \beta_k \mathbf{A}'\mathbf{P}_k \mathbf{A} - \beta_k B_k W_k \mathbf{A}'\mathbf{P}_k \mathbf{A}. \end{aligned}$$

Note also $(E_k - E_{k-1})(\beta_k \mathbf{A}'\mathbf{P}_k \mathbf{A}) = \mathbf{0}$. Since $\beta_k = \frac{1}{n-p+1} \approx \frac{1}{(1-c)n}$, we have

$$\begin{aligned} &E\{\|(E_k - E_{k-1})(\mathbf{D}_k - \tilde{\mathbf{D}}_k)\|^2\} \\ &\leq 2E(\beta_k^2 B_k^2 W_k^2 \|\mathbf{Q}_k\|^2) + 2E(\beta_k^2 B_k^2 W_k^2 \|\mathbf{A}'\mathbf{P}_k \mathbf{A}\|^2) \\ &\leq Kn^{-4}\{E(W_k^2 \|\mathbf{Q}_k\|^2) + E(W_k^2)\} + o(n^{-t}) = o(n^{-2}). \end{aligned}$$

In the second inequality, we used the fact that $\|\mathbf{A}'\mathbf{P}_k \mathbf{A}\| \leq 1$ and considered two complementary events: $|B_k| \leq 1/n$ and $|B_k| > 1/n$ with the latter event giving rise to $o(n^{-t})$. In the last equality, we applied Lemma 7.3 to bound $E(W_k^2) \leq KE(\text{tr}(\mathbf{P}_k^2)) = KE(\text{tr}(\mathbf{P}_k)) \leq Kn$; and Lemma 7.5 below to obtain $E(W_k^2 \|\mathbf{Q}_k\|^2) = o(n^2)$. So

$$E\left\{\left\|\sqrt{n} \sum_{k=1}^p (E_k - E_{k-1})(\mathbf{D}_k - \tilde{\mathbf{D}}_k)\right\|^2\right\} = o(1),$$

proving (7.10).

Let $E^{(k)}$ denote the expectation with respect to \mathbf{z}_k . Lemma 7.6 below shows that $E^{(k)}(\|\mathbf{Q}_k\|^4) = o(n)$. Consequently,

$$\sum_{k=1}^p E(\|\sqrt{n}\tilde{\mathbf{D}}_k\|^4) = n^2 \sum_{k=1}^p \beta_k^4 E(\|\mathbf{Q}_k\|^4) = o(1).$$

In other words, the sequence of martingale differences $\{\sqrt{n}E_k(\tilde{\mathbf{D}}_k)\}$ satisfies the Lyapunov condition, and thus $\mathbf{V} \xrightarrow{D} \mathbf{W}_c$. The matrix norm used for the above proof

is, for convenience purposes, the Euclidean norm which differs from the spectral norm by a fixed factor since the order of the matrix under consideration is fixed.

Step 3. This step concerns the computation of the covariance of a pair of entries in \mathbf{W}_c . Let unit n -vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ denote 4 columns of \mathbf{A} , either identical or orthogonal. We need to compute

$$\begin{aligned} & \frac{1}{n(1-c)^2} \sum_{k=1}^p E_{k-1} [E_k \{ \mathbf{a}' \mathbf{P}_k (\mathbf{z}_k \mathbf{z}'_k - \mathbf{I}_n) \mathbf{P}_k \mathbf{b} \} E_k \{ \mathbf{c}' \mathbf{P}_k (\mathbf{z}_k \mathbf{z}'_k - \mathbf{I}_n) \mathbf{P}_k \mathbf{d} \}] \\ &= \frac{1}{n(1-c)^2} \sum_{k=1}^p E_{k-1} [\{ \mathbf{z}'_k E_k (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k) \mathbf{z}_k - \text{tr} E_k (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k) \} \\ & \quad \times \{ \mathbf{z}'_k E_k (\mathbf{P}_k \mathbf{d} \mathbf{c}' \mathbf{P}_k) \mathbf{z}_k - \text{tr} (E_k \mathbf{P}_k \mathbf{d} \mathbf{c}' \mathbf{P}_k) \}] \\ &= \frac{1}{n(1-c)^2} \sum_{k=1}^p E_{k-1} [\text{tr} \{ E_k (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k) E_k (\mathbf{P}_k \mathbf{d} \mathbf{c}' \mathbf{P}_k) \} \\ & \quad + \text{tr} \{ E_k (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k) E_k (\mathbf{P}_k \mathbf{c} \mathbf{d}' \mathbf{P}_k) \}] + R_n \\ &= \frac{1}{n(1-c)^2} \sum_{k=1}^p E_k \{ \text{tr} (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{d} \mathbf{c}' \check{\mathbf{P}}_k) + \text{tr} (\mathbf{P}_k \mathbf{b} \mathbf{a}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{c} \mathbf{d}' \check{\mathbf{P}}_k) \} + R_n \\ &= \frac{1}{n(1-c)^2} \sum_{k=1}^p E_k \{ (\mathbf{a}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{d}) (\mathbf{b}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{c}) + (\mathbf{a}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{c}) (\mathbf{b}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{d}) \} + R_n \\ &= \frac{1}{(1-c)^2} \{ I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) + I(\mathbf{a}, \mathbf{b}, \mathbf{d}, \mathbf{c}) \} + R_n, \end{aligned}$$

where we applied Lemma 7.3(a) in the second equality. Here,

$$(7.12) \quad I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = \frac{1}{n} \sum_{k=1}^p E_k \{ (\mathbf{a}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{d}) (\mathbf{b}' \mathbf{P}_k \check{\mathbf{P}}_k \mathbf{c}) \},$$

$$R_n = \frac{E(y_{11}^4)}{n(1-c)^2} \sum_{k=1}^p E_{k-1} \left(\sum_{i=1}^n \mathbf{a}' \mathbf{P}_k \mathbf{e}_i \mathbf{b}' \mathbf{P}_k \mathbf{e}_i \mathbf{c}' \check{\mathbf{P}}_k \mathbf{e}_i \mathbf{d}' \check{\mathbf{P}}_k \mathbf{e}_i \right),$$

and $\check{\mathbf{P}}_k$ is the projection matrix onto the space spanned by the vectors $\mathbf{z}_1, \dots, \mathbf{z}_{k-1}, \check{\mathbf{z}}_{k+1}, \dots, \check{\mathbf{z}}_p$. Here, $\check{\mathbf{z}}_j$'s are *i.i.d.* copies of \mathbf{z}_j 's.

By Lemma 7.7 below, we have $R_n = O_P(n^{-1})$. Therefore, we only need to compute the limit for $I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$. We proceed to remove successively row by row starting from the last row until we reach the k th row. We may assume $k < p$. For $k < j \leq p$, let \mathbf{Z}_{kj} denote the $(j-2) \times n$ matrix consisting of the rows $\mathbf{z}'_1, \dots, \mathbf{z}'_{k-1}, \mathbf{z}'_{k+1}, \dots, \mathbf{z}'_{j-1}$ and $\mathbf{P}_{kj} = \mathbf{I}_n - \mathbf{Z}'_{kj} (\mathbf{Z}_{kj} \mathbf{Z}'_{kj})^{-1} \mathbf{Z}_{kj}$. Define

$$B_{kj} = 1/(\mathbf{z}'_j \mathbf{P}_{kj} \mathbf{z}_j),$$

$$\beta_{kj} = 1/\text{tr } \mathbf{P}_{kj} = 1/(n - j + 2),$$

$$W_{kj} = \mathbf{z}'_j \mathbf{P}_{kj} \mathbf{z}_j - \text{tr } \mathbf{P}_{kj}.$$

Note that \mathbf{P}_k is a projection matrix onto the orthogonal complement of the space spanned by the rows of \mathbf{Z}_k , we have

$$\mathbf{P}_k = \mathbf{P}_{kp} - B_{kp} \mathbf{P}_{kp} \mathbf{z}_p \mathbf{z}'_p \mathbf{P}_{kp}$$

$$= (1 - \beta_{kp}) \mathbf{P}_{kp} - \beta_{kp} \mathbf{R}_{k1} + \beta_{kp} \mathbf{R}_{k2},$$

where

$$\mathbf{R}_{k1} = \mathbf{P}_{kp} (\mathbf{z}_p \mathbf{z}'_p - \mathbf{I}_n) \mathbf{P}_{kp},$$

$$\mathbf{R}_{k2} = B_{kp} W_{kp} \mathbf{P}_{kp} \mathbf{z}_p \mathbf{z}'_p \mathbf{P}_{kp}.$$

Substituting this into (7.12) and noting that \mathbf{z}_p is independent of $\check{\mathbf{P}}_k$, we obtain

$I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$

$$(7.13) \quad = \frac{1}{n} \sum_{k=1}^p \left(\frac{n-p+1}{n-p+2} \right)^2 E_k \mathbf{a}' \mathbf{P}_{kp} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{P}_{kp} \check{\mathbf{P}}_k \mathbf{c} + \frac{1}{n} \sum_{k=1}^p \sum_{j=1}^6 \Delta_{kj},$$

where

$$\Delta_{k1} = (n - p + 2)^{-2} E_k (\mathbf{a}' \mathbf{R}_{k1} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{R}_{k1} \check{\mathbf{P}}_k \mathbf{c}),$$

$$\Delta_{k2} = (n - p + 1)/(n - p + 2)^2 E_k (\mathbf{a}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{P}_{kp} \check{\mathbf{P}}_k \mathbf{c}),$$

$$\Delta_{k3} = (n - p + 1)/(n - p + 2)^2 E_k (\mathbf{a}' \mathbf{P}_{kp} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{c}),$$

$$\Delta_{k4} = -(n - p + 2)^{-2} E_k (\mathbf{a}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{R}_{k1} \check{\mathbf{P}}_k \mathbf{c}),$$

$$\Delta_{k5} = -(n - p + 2)^{-2} E_k (\mathbf{a}' \mathbf{R}_{k1} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{c}),$$

$$\Delta_{k6} = (n - p + 2)^{-2} E_k (\mathbf{a}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{d} \mathbf{b}' \mathbf{R}_{k2} \check{\mathbf{P}}_k \mathbf{c}).$$

By Lemma 7.7, we have $\Delta_{kj} = O_P(n^{-2})$, for $j = 1, 2, 3, 4, 5, 6$. Expanding $\check{\mathbf{P}}_k$ in a similar way, we obtain

$I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$

$$(7.14) \quad = \frac{1}{n} \sum_{k=1}^p \left(\frac{n-p+1}{n-p+2} \right)^4 E_k (\mathbf{a}' \mathbf{P}_{kp} \check{\mathbf{P}}_{kp} \mathbf{d} \mathbf{b}' \mathbf{P}_{kp} \check{\mathbf{P}}_{kp} \mathbf{c}) + O_P(n^{-2}).$$

Repeating this step successively for $j = p - 1, \dots, k + 1$, we obtain

$I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$

$$= \frac{1}{n} \sum_{k=1}^p \left(\frac{n-p+1}{n-j+2} \right)^4 E_k (\mathbf{a}' \mathbf{P}_{kj} \check{\mathbf{P}}_{kj} \mathbf{d} \mathbf{b}' \mathbf{P}_{kj} \check{\mathbf{P}}_{kj} \mathbf{c})$$

$$\begin{aligned}
 (7.15) \quad & + O_P\left(\frac{p-j+1}{n^2}\right) \\
 & = \frac{1}{n} \sum_{k=1}^p \left(\frac{n-p+1}{n-k+2}\right)^4 E_k(\mathbf{a}'\mathbf{P}_{kk}\check{\mathbf{P}}_{kk}\mathbf{d}\mathbf{b}'\mathbf{P}_{kk}\check{\mathbf{P}}_{kk}\mathbf{c}) + O_P(n^{-1}) \\
 & = \frac{1}{n} \sum_{k=1}^p \left(\frac{n-p+1}{n-k+2}\right)^4 \mathbf{a}'\mathbf{P}_{kk}\mathbf{d}\mathbf{b}'\mathbf{P}_{kk}\mathbf{c} + O_P(n^{-1}),
 \end{aligned}$$

where $\mathbf{P}_{kk} = \mathbf{I}_n - \mathbf{Z}'_{kk}(\mathbf{Z}_{kk}\mathbf{Z}'_{kk})^{-1}\mathbf{Z}_{kk}$ and \mathbf{Z}_{kk} is $(k-1) \times n$ consisting of rows $\mathbf{z}'_1, \dots, \mathbf{z}'_{k-1}$. The last step in (7.15) follows from the fact that $\check{\mathbf{P}}_{kk} = \mathbf{P}_{k,k}$ that is a projection matrix.

By Lemma 7.8, we have

$$\begin{aligned}
 (7.16) \quad I(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) & = \frac{\mathbf{a}'\mathbf{d}\mathbf{b}'\mathbf{c}}{n} \sum_{k=1}^p \frac{(n-k+1)^2}{n^2} \left(\frac{n-p+1}{n-k+2}\right)^4 + o_p(1) \\
 & \rightarrow \mathbf{a}'\mathbf{d}\mathbf{b}'\mathbf{c} \int_0^c \frac{(1-t)^4}{(1-t)^2} dt = c(1-c)^3 \mathbf{a}'\mathbf{d}\mathbf{b}'\mathbf{c}.
 \end{aligned}$$

So the limiting variance of a diagonal element of \mathbf{V} is $2I(\mathbf{a}, \mathbf{a}, \mathbf{a}, \mathbf{a})/(1-c)^2 = 2c(1-c)$; and that of an off-diagonal element is $\{I(\mathbf{a}, \mathbf{b}, \mathbf{a}, \mathbf{b}) + I(\mathbf{a}, \mathbf{b}, \mathbf{b}, \mathbf{a})\}/(1-c)^2 = c(1-c)$ as $\mathbf{a} \perp \mathbf{b}$. Similarly, the limiting covariance of any two diagonal elements corresponds to $2I(\mathbf{a}, \mathbf{a}, \mathbf{c}, \mathbf{c})/(1-c)^2 = 0$ as $\mathbf{a} \perp \mathbf{c}$; and that of a diagonal and an off-diagonal elements is $\{I(\mathbf{a}, \mathbf{a}, \mathbf{c}, \mathbf{d}) + I(\mathbf{a}, \mathbf{a}, \mathbf{d}, \mathbf{c})\}/(1-c)^2 = 0$ as \mathbf{a}, \mathbf{c} and \mathbf{d} are mutually orthogonal. Similarly, the limiting covariance of any two off-diagonal elements is also 0. This completes the proof of Theorem 2.1. \square

7.3. Lemmas 7.5–7.8. We shall state Lemmas 7.5–7.8 which were used in the proof of Theorem 2.1 in Section 7.2. The proofs of these lemmas will be given in the Supplementary Material.

LEMMA 7.5. We have $E(W_k^2 \|\mathbf{Q}_k\|^2) = o(n^2)$.

LEMMA 7.6. We have

$$(7.17) \quad \sum_{i=1}^n E(\|\mathbf{A}'\mathbf{P}_k \mathbf{e}_i\|^4) = O(n^{-1})$$

and

$$(7.18) \quad E^{(k)}(\|\mathbf{Q}_k\|^4) \leq K \left\{ \|\mathbf{A}'\mathbf{P}_k \mathbf{A}\| + \sum_{i=1}^n \|\mathbf{A}'\mathbf{P}_k \mathbf{e}_i\|^8 E(y_{11}^8) \right\} = o(n).$$

LEMMA 7.7. We have $R_n = O_P(n^{-1})$ and $\Delta_{kj} = O_P(n^{-2})$ for $1 \leq k \leq p$, $1 \leq j \leq 6$.

LEMMA 7.8. Recall \mathbf{a} and \mathbf{d} are any two columns of \mathbf{A} , either identical or orthogonal. Recall also that $\mathbf{P}_{kk} = \mathbf{I}_n - \mathbf{Z}'_{kk}(\mathbf{Z}_{kk}\mathbf{Z}'_{kk})^{-1}\mathbf{Z}_{kk}$ where \mathbf{Z}_{kk} is the $(k - 1) \times n$ matrix consisting of the rows $\mathbf{z}'_1, \dots, \mathbf{z}'_{k-1}$. Then

$$\mathbf{a}'\mathbf{P}_{kk}\mathbf{d} = \frac{n - k + 1}{n}\mathbf{a}'\mathbf{d} + o_P(1).$$

REMARK. Pan and Zhou (2011) deduced the central limit theorem of the one-sample Hotelling’s T_0^2 statistic from a more general result (Theorem 2 in their paper). They used Stieltjes’ transform approach in RMT whereas we used the CLT of martingale decomposition. Pan and Zhou (2011) basically considered the CLT of the renormalized T_2 , which is an F statistic, whereas we consider the CLT of the normalized F matrix. In fact, what is considered by Pan and Zhou (2011) is equivalent to the normalized diagonal entry. Consequently, they do not need to consider the off-diagonal elements and the correlations between them. Our martingale decomposition is made along the rows of the data matrix.

8. Proof of Theorem 2.2. We outline the proof of Theorem 2.2 here leaving the technical details to Lemmas 8.1 and 8.2 with their proofs given in the Supplementary Material. By Lemma 7.1,

$$\begin{aligned} & \sqrt{n}\{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')^{-1}\mathbf{Y}\mathbf{A} - \mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A}\} \\ &= \sqrt{n}\mathbf{A}'\mathbf{Y}'\{(\mathbf{Y}\mathbf{Y}' - n\bar{\mathbf{y}}\bar{\mathbf{y}}')^{-1} - (\mathbf{Y}\mathbf{Y}')^{-1}\}\mathbf{Y}\mathbf{A} \\ &= n^{3/2}\frac{\mathbf{A}'\mathbf{Y}'(\mathbf{Y}\mathbf{Y}')^{-1}\bar{\mathbf{y}}\bar{\mathbf{y}}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A}}{1 - n\bar{\mathbf{y}}'(\mathbf{Y}\mathbf{Y}')^{-1}\bar{\mathbf{y}}}. \end{aligned}$$

Theorem 2.2 follows immediately from (8.1) and (8.2) below

$$(8.1) \quad n\bar{\mathbf{y}}'(\mathbf{Y}\mathbf{Y}')^{-1}\bar{\mathbf{y}} = \bar{\mathbf{y}}'\mathbf{S}^{-1}\bar{\mathbf{y}} \xrightarrow{P} c < 1,$$

$$(8.2) \quad n^{3/4}\bar{\mathbf{y}}'(\mathbf{Y}\mathbf{Y}')^{-1}\mathbf{Y}\mathbf{A} = n^{-1/4}\bar{\mathbf{y}}'\mathbf{S}^{-1}\mathbf{Y}\mathbf{A} \xrightarrow{P} \mathbf{0}.$$

Since

$$\bar{\mathbf{y}}'\mathbf{S}^{-1}\bar{\mathbf{y}} = \frac{1}{n^2} \sum_{k=1}^n \mathbf{y}'_k \mathbf{S}^{-1} \mathbf{y}_k + \frac{2}{n^2} \sum_{1 \leq j < k \leq n} \mathbf{y}'_j \mathbf{S}^{-1} \mathbf{y}_k,$$

we apply Lemmas 7.1 and 7.2 to obtain

$$\frac{1}{n^2} \sum_{k=1}^n \mathbf{y}'_k \mathbf{S}^{-1} \mathbf{y}_k = \frac{1}{n} \sum_{k=1}^n \frac{\frac{1}{n} \mathbf{y}'_k \mathbf{S}^{-1} \mathbf{y}_k}{1 + \frac{1}{n} \mathbf{y}'_k \mathbf{S}^{-1} \mathbf{y}_k} \xrightarrow{\text{a.s.}} \frac{c/(1 - c)}{1 + c/(1 - c)} = c;$$

and Lemma 8.1 to show that $\frac{2}{n^2} \sum_{1 \leq j < k \leq n} \mathbf{y}'_j \mathbf{S}^{-1} \mathbf{y}_k \xrightarrow{P} 0$. This proves (8.1).

Since q is fixed, and each of the q columns of \mathbf{A} is orthogonal to $\mathbf{1}_n$, therefore, (8.2) follows from Lemma 8.2 below.

LEMMA 8.1. Let \mathbf{a} and \mathbf{d} denote any two columns of \mathbf{A} , either identical or orthogonal. Let a_k and d_k denote their k th entries, respectively. We have

$$(8.3) \quad \sum_{k=1}^n a_k d_k \left(\frac{1}{n} \mathbf{y}'_k \mathbf{S}^{-1} \mathbf{y}_k - c_{n,p} \right) = o_P(1),$$

$$(8.4) \quad \sum_{1 \leq j < k \leq n} a_j d_k \left(\frac{1}{n} \mathbf{y}'_j \mathbf{S}^{-1} \mathbf{y}_k \right) = O_P(n^{-1/2}).$$

LEMMA 8.2. Let \mathbf{a} be a unit n -vector satisfying $\mathbf{1}'_n \mathbf{a} = 0$. Then

$$n^{-1/4} \bar{\mathbf{y}}' \mathbf{S}^{-1} \mathbf{Y} \mathbf{a} \xrightarrow{P} 0.$$

Proofs of Lemmas 8.1 and 8.2 are given in the Supplementary Material.

Acknowledgements. We are grateful to the Editor, an Associate Editor and the reviewers for their careful reading of our manuscript and many constructive comments and suggestions. We also thank Mr Gregory Ang's assistance in writing the computer codes for the simulation study.

SUPPLEMENTARY MATERIAL

Supplement to “Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT” (DOI: [10.1214/17-AOS1646SUPP](https://doi.org/10.1214/17-AOS1646SUPP); .pdf). In the supplementary material, we presented (i) the truncation and normalization techniques as mentioned at the beginning of Section 7 of this paper; (ii) details of reparametrization for assumptions in (4.1) hold; (iii) proofs of (5.9) and (5.10); (iv) more plots of the empirical sizes of the three invariant tests considered in this paper; (v) the proofs of Lemmas 7.1 to 7.8; and (vi) the proofs of Lemmas 8.1 and 8.2.

REFERENCES

- AMEMIYA, Y. (1990). A note on the limiting distribution of certain characteristic roots. *Statist. Probab. Lett.* **9** 465–470. [MR1060089](https://doi.org/10.1080/00036819008839589)
- ANDERSON, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1950 103–130. Univ. California Press, Berkeley, CA. [MR0044075](https://doi.org/10.1214/aos/1176344075)
- ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, NJ. [MR1990662](https://doi.org/10.1002/9781118130203)
- BAI, Z. D. (1985). A note on the limiting distribution of the eigenvalues of a class of random matrices. *J. Math. Res. Exposition* **5** 113–118. [MR0842111](https://doi.org/10.1080/08837508508839589)
- BAI, Z., CHOI, K. P. and FUJIKOSHI, Y. (2018). Supplement to “Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT.” DOI:[10.1214/17-AOS1646SUPP](https://doi.org/10.1214/17-AOS1646SUPP).
- BAI, Z. D., LIU, H. X. and WONG, W. K. (2011). Asymptotic properties of eigenmatrices of a large sample covariance matrix. *Ann. Appl. Probab.* **21** 1994–2015. [MR2884057](https://doi.org/10.1214/10-AOS1646)
- BAI, Z. D. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329.

- BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 535–605.
- BAI, Z. D. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **5** 425–442.
- BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. [MR2572444](#)
- CAI, T. T. and XIA, Y. (2014). High-dimensional sparse MANOVA. *J. Multivariate Anal.* **131** 174–196.
- CHEN, S. X. and QIN, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. [MR2604697](#)
- CHIANI, M. (2016). Distribution of the largest root of a matrix for Roy’s test in multivariate analysis of variance. *J. Multivariate Anal.* **143** 467–471.
- DEMPSTER, A. P. (1958). A high dimensional two sample significance test. *Ann. Math. Statist.* **29** 995–1010. [MR0112207](#)
- DEMPSTER, A. P. (1960). A significance test for the separation of two highly multivariate small samples. *Biometrics* **16** 41–50.
- FUJIKOSHI, Y. (1977). Asymptotic expansions for the distributions of the latent roots in MANOVA and the canonical correlations. *J. Multivariate Anal.* **7** 386–396.
- FUJIKOSHI, Y., HIMENO, T. and WAKAKI, H. (2004). Asymptotic results of a high-dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *J. Japan Statist. Soc.* **34** 19–26.
- FUJIKOSHI, Y., HIMENO, T. and WAKAKI, H. (2008). Asymptotic results in canonical discriminant analysis when the dimension is large compared to the sample. *J. Statist. Plann. Inference* **138** 3457–3466.
- GLYNN, W. J. and MUIRHEAD, R. J. (1978). Inference in canonical correlation analysis. *J. Multivariate Anal.* **8** 468–478. [MR0512614](#)
- HSU, P. L. (1941). On the limiting distribution of roots of a determinantal equation. *J. Lond. Math. Soc.* **16** 183–194.
- HU, J. and BAI, Z. D. (2016). A review of 20 years of naive tests of significance for high-dimensional mean vectors and covariance matrices. *Sci. China Math.* **59** 2281–2300.
- JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rate of convergence. *Ann. Statist.* **36** 2638–2716. [MR2485010](#)
- KRISHNAIAH, P. R. and CHANG, T. C. (1971). On the exact distributions of the extreme roots of the Wishart and MANOVA matrices. *J. Multivariate Anal.* **1** 108–117. [MR0301860](#)
- LI, J. and CHEN, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. [MR2985938](#)
- MUIRHEAD, R. J. (1978). Latent roots and matrix variates. *Ann. Statist.* **6** 5–33.
- MUIRHEAD, R. J. (1982). *Aspect of Multivariate Statistical Theory*. Wiley, New York.
- PAN, G. M. and ZHOU, W. (2011). Central limit theorem for Hotelling’s T^2 statistic under large dimension. *Ann. Appl. Probab.* **21** 1860–1910. [MR2884053](#)
- SCHOTT, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *J. Multivariate Anal.* **98** 1825–1839.
- SEO, T., KANDA, T. and FUJIKOSHI, Y. (1994). Asymptotic distributions of the sample roots in MANOVA models under nonnormality. *J. Japan Statist. Soc.* **24** 133–140.
- SILVERSTEIN, J. (1995). Strong convergence of empirical distribution of eigenvalues of large dimensional random matrices. *J. Multivariate Anal.* **55** 331–339.
- SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension under non-normality. *J. Multivariate Anal.* **99** 386–402.
- SRIVASTAVA, M. S. and FUJIKOSHI, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *J. Multivariate Anal.* **97** 1927–1940.

- SRIVASTAVA, M. S. and KUBOKAWA, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *J. Multivariate Anal.* **115** 204–216.
- SUGIURA, N. (1976). Asymptotic expansions of the distributions of the latent roots and latent vectors of the Wishart and multivariate F matrices. *J. Multivariate Anal.* **6** 500–525.
- ULLAH, I. and JONES, B. (2015). Regularised MANOVA for high-dimensional data. *Aust. N. Z. J. Stat.* **57** 377–389. [MR3405507](#)
- WAKAKI, H., FUJIKOSHI, Y. and ULYANOV, V. (2014). Asymptotic expansions of the distributions of MANOVA test statistics when the dimension is large. *Hiroshima Math. J.* **44** 247–259.
- WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669.
- ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large-dimensional F -matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 444–476. [MR2954263](#)

Z. BAI
KLAS MOE AND
SCHOOL OF MATH & STAT
NORTHEAST NORMAL UNIVERSITY
CHANGCHUN 130024
CHINA
E-MAIL: baizd@nenu.edu.cn

K. P. CHOI
DEPARTMENT OF STATISTICS AND
APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE 117546
REPUBLIC OF SINGAPORE
E-MAIL: stackp@nus.edu.sg

Y. FUJIKOSHI
DEPARTMENT OF MATHEMATICS
GRADUATE SCHOOL OF SCIENCE
HIROSHIMA UNIVERSITY
HIGASHI-HIROSHIMA
HIROSHIMA, 739-8526
JAPAN
E-MAIL: fujikoshi_y@yahoo.co.jp