High-Dimensional Asymptotic Distributions of Characteristic Roots in Multivariate Linear Models and Canonical Correlation Analysis

Yasunori Fujikoshi

Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi Hiroshima, Hiroshima 739-8626, Japan

Abstract

In this paper, we derive the asymptotic distributions of the characteristic roots in multivariate linear models when the dimension p and the sample size n are large. The results are given for the case that the population characteristic roots have multiplicities greater than unity, and their orders are $\mathrm{O}(np)$ or $\mathrm{O}(n)$. Next, similar results are given for the asymptotic distributions of the canonical correlations when one of the dimensions and the sample size are large, assuming that the order of the population canonical correlations is $\mathrm{O}(\sqrt{p})$ or $\mathrm{O}(1)$.

AMS 2000 Subject Classification: primary 62H10; secondary 62E20

Key Words and Phrases: asymptotic distributions, canonical correlations, characteristic roots, high-dimensional approximations, multivariate linear model.

Abbreviated title: High-Dimensional Approximations of Characteristic Roots.

1. Introduction

The large-sample asymptotic distributions of the characteristic roots in discriminant analysis and canonical correlation analysis were derived under normality by Hsu (1941a, b) and Anderson (1951). The results were extended by considering nonnormal cases and by obtaining their asymptotic expansions, and the results for various such cases were presented by many authors; see, for example, Sugiura (1976), Fujikoshi (1977), Muirehead (1978, 1982), Glynn and Muirhead (1978), and Muirhead and Watermaux (1980).

However, it is known that these large-sample approximations become less accurate as the number of the response variables, that is, the dimensionality, becomes larger. To overcome this, the distributions of the characteristic roots have been studied in high-dimensional situations, where the dimension and the sample size are both large. More precisely, for discriminant analysis with $q+1$ groups, based on n samples of p variables, the asymptotic distributions of the characteristic roots were obtained by Fujikoshi et al. (2008), under a high-dimensional asymptotic framework in which $p/n \to c \in [0,1)$ and q is fixed. For canonical correlation analysis of p variables and $q(\leq p)$ variables, Fujikoshi and Sakurai (2009) obtained the asymptotic distributions of the canonical correlations when $p/n \to c_0 \in [0,1)$ and q is fixed.

For these high-dimensional approximations, it was assumed that the population characteristic roots are simple. In this paper, we extend the results to cases in which the population characteristic roots have arbitrary multiplicities. The characteristic roots in discriminant analysis can be treated as a special case of those of a multivariate linear model. We also consider high-dimensional distributions in which the order of the characteristic roots of the noncentrality matrix in the multivariate linear model is $\mathrm{O}(pn)$ or $\mathrm{O}(n)$. For the case of canonical correlations, the populations canonical correlations are assumed to be $\mathrm{O}(p)$ or $\mathrm{O}(1)$.

Our results show that the consistency found in the sample roots in the large-sample case does not hold in the high-dimensional case. We discuss with some applications.

2. Characteristic Roots in the Multivariate Linear Model

We consider a multivariate linear model of p response variables $y_1, \ldots, y_p$ on a subset of k explanatory variables $x_1, \ldots, x_k$. Suppose that there are n observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ on each of $\boldsymbol{y} = (y_1, \ldots, y_p)'$ and $\boldsymbol{x} = (x_1, \ldots, x_k)'$, respectively, and let $\mathbf{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$ and $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ be the $n \times p$ and $n \times k$ observation matrices of $\boldsymbol{y}$ and $\boldsymbol{x}$, respectively. The multivariate normal linear model is written as

$$\mathbf{Y} \sim \mathrm{N}_{n\times p}(\mathbf{X\Theta}, \mathbf{\Sigma} \otimes \mathbf{I}_n), \tag{2.1}$$

where $\mathbf{\Theta}$ is a $k \times p$ unknown matrix of coefficients, $\mathbf{\Sigma}$ is a $p \times p$ unknown covariance matrix, and $\mathbf{I}_n$ is the identity matrix of order n. The notation $\mathrm{N}_{n\times p}(\cdot, \cdot)$ means the matrix normal distribution such that the mean of $\mathbf{Y}$ is $\mathbf{X\Theta}$ and the covariance matrix of $\mathrm{vec}\,(\mathbf{Y})$ is $\mathbf{\Sigma} \otimes \mathbf{I}_n$, where $\mathrm{vec}\,(\mathbf{Y})$ is the $np \times 1$ vector formed by stacking the columns of $\mathbf{Y}$ under each other. We assume that $n - k > p$ and $\mathrm{rank}(\mathbf{X}) = k$.

Let $\mathbf{C}$ be a given $q \times k$ matrix of $\mathrm{rank}(\mathbf{C}) = q(\leq k)$. When testing or estimating the rank of $\mathbf{C\Theta}$, it is important to study the distribution of the nonzero characteristic roots of $\mathbf{S}_h\mathbf{S}_e^{-1}$,

$$\ell_1 > \cdots > \ell_m > 0, \quad m = \min(p, q), \tag{2.2}$$

where

$$\mathbf{S}_e = \mathbf{Y}'(\mathbf{I}_n - \mathbf{P}_\mathbf{X})\mathbf{Y}, \quad \mathbf{S}_h = (\mathbf{C}\hat{\mathbf{\Theta}})'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\mathbf{C}\hat{\mathbf{\Theta}} \tag{2.3}$$

and $\hat{\mathbf{\Theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Here, without loss of generality, we may assume that $\ell_1 > \cdots > \ell_m > 0$, since the probability that the ℓ_i's are equal is 0. It is well known (see, e.g., Anderson, 2003) that $\mathbf{S}_e$ and $\mathbf{S}_h$ are independently distributed as a Wishart distribution $\mathrm{W}_p(n-k, \boldsymbol{\Sigma})$ and a noncentral Wishart distribution $\mathrm{W}_p(q, \boldsymbol{\Sigma}; \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma}^{1/2})$, respectively, where

$$\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{C}\boldsymbol{\Theta})'\{\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'\}^{-1}\mathbf{C}\boldsymbol{\Theta}\boldsymbol{\Sigma}^{-1/2}. \tag{2.4}$$

In a multivariate regression model, we are often interested in the case $\mathbf{C} = \mathbf{I}_k$.

Consider the characteristic roots used in discriminant analysis with $(q+1)$ p-variate normal populations and common covariance matrix $\boldsymbol{\Sigma}$. Let $\boldsymbol{\mu}_i$ be the mean vector of the ith population. Suppose that a sample of size n_i is available from the ith population, and let $\boldsymbol{y}_{ij}$ be the jth observation from the ith population. Let us denote the between-group and within-group sum of squares and product matrices by

$$\mathsf{S}_b = \sum_{i=1}^{q+1} n_i(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})', \quad \mathsf{S}_w = \sum_{i=1}^{q+1}(n_i - 1)\mathsf{S}_i,$$

respectively, where $\bar{\boldsymbol{y}}_i$ and S_i are the mean vector and sample covariance matrix of the ith population, and $\bar{\boldsymbol{y}}$ is the total mean vector defined by $(1/n)\sum_{i=1}^{q+1} n_i\bar{\boldsymbol{y}}_i$, where $n = \sum_{i=1}^{q+1} n_i$. In general, S_w and S_b are independently distributed as a Wishart distribution $\mathrm{W}_p(n-q-1, \boldsymbol{\Sigma})$ and a noncentral Wishart distribution $\mathrm{W}_p(q, \boldsymbol{\Sigma}; \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\Omega}}\boldsymbol{\Sigma}^{1/2})$, respectively, where

$$\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Sigma}^{-1/2}\sum_{i=1}^{q+1} n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1/2}, \quad \bar{\boldsymbol{\mu}} = (1/n)\sum_{i=1}^{q+1} n_i\boldsymbol{\mu}_i. \tag{2.5}$$

The characteristic roots of $\mathsf{S}_b\mathsf{S}_e^{-1}$ are used for testing and estimating the number of non-zero characteristic roots of $\tilde{\boldsymbol{\Omega}}$, which is the dimensionality in discriminant analysis. For further details, see, for example, Fujikoshi, Ulyanov, and Shimizu (2010). These characteristic roots can be regarded as a special case of the multivariate linear model; this is easily seen by taking $k = q+1$ and choosing Y, $\mathbf{C}$, $\mathbf{X}$ and $\boldsymbol{\Theta}$ as follows:

$$\mathsf{Y} = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1n_1}, \ldots, \boldsymbol{y}_{q+1,1}, \ldots, \boldsymbol{y}_{q+1,n_{q+1}})', \quad \mathbf{C} = (\mathbf{I}_q,\ -\mathbf{1}_q),$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{n_{q+1}} \end{pmatrix}, \quad \boldsymbol{\Theta} = \begin{pmatrix} \boldsymbol{\mu}_1' \\ \boldsymbol{\mu}_2' \\ \vdots \\ \boldsymbol{\mu}_{q+1}' \end{pmatrix},$$

where $\mathbf{1}_n$ is an $n \times 1$ vector whose elements are all one. Then, $\mathsf{S}_h = \mathsf{S}_b$ and $\mathsf{S}_e = \mathsf{S}_w$.

3. Derivation Method

When we consider the distribution of the characteristic roots of $\mathsf{S}_h\mathsf{S}_e^{-1}$ in (2.3), we may assume that S_e are S_h are independently distributed as

$$\mathsf{S}_e \sim \mathrm{W}_p(n-k, \mathbf{I}_p), \quad \mathsf{S}_h \sim \mathrm{W}_p(q, \mathbf{I}_p; \tilde{\mathbf{D}}_\omega), \tag{3.1}$$

where $\tilde{\mathbf{D}}_\omega = \mathrm{diag}(\omega_1, \ldots, \omega_p)$, and $\omega_1 \geq \cdots \geq \omega_p \geq 0$ are the characteristic roots of $\tilde{\boldsymbol{\Omega}}$. In this paper, we assume that

$$n - k \geq p \geq q. \tag{3.2}$$

Then, the first q characteristic roots $\ell_1 > \cdots > \ell_q$ are positive, and the remaining $p - q$ roots are zero. Similarly, $\omega_{q+1} = \cdots = \omega_p = 0$, since $\mathrm{rank}(\tilde{\boldsymbol{\Omega}}) \leq q$. We can express S_h as

$$\mathsf{S}_h = \mathsf{Z}\mathsf{Z}', \quad \mathsf{Z};\ p \times q, \tag{3.3}$$

where the columns of Z are independently distributed as $\mathrm{N}_p(\cdot, \mathbf{I}_p)$, $\mathrm{E}(\mathsf{Z}) = (\mathbf{D}_\omega^{1/2}\ \mathbf{O})'$, and $\mathbf{D}_\omega = \mathrm{diag}(\omega_1, \ldots, \omega_q)$. Consider a transform from $(\mathsf{S}_h, \mathsf{S}_e)$ to (B, W) given by

$$\mathsf{B} = \mathsf{Z}'\mathsf{Z}, \quad \mathsf{W} = \mathsf{B}^{1/2}(\mathsf{Z}'\mathsf{S}_e^{-1}\mathsf{Z})^{-1}\mathsf{B}^{1/2}. \tag{3.4}$$

Then, it is known (Fujikoshi et al., 2007; Wakaki et al., 2014) that

R0: The nonzero characteristic roots of $\mathsf{S}_h\mathsf{S}_e^{-1}$ are the same as those of $\mathsf{B}\mathsf{W}^{-1}$, or equivalently of $\mathsf{S}_e^{-1/2}\mathsf{S}_h\mathsf{S}_e^{-1/2}$, and

$$\mathsf{W} \sim \mathrm{W}_q(m, \mathbf{I}_q), \quad \mathsf{B} \sim \mathrm{W}_q(p, \mathbf{I}_q; \mathbf{D}_\omega), \tag{3.5}$$

where W and B are independent, and $m = n - k - p + q$.

Note that the characteristic roots $\ell_1 > \cdots > \ell_q$ are defined in terms of the $q \times q$ matrices W and B with a reduced size. When q is fixed and m tends to infinity, Fro R0 we can use the perturbation method, which was developed for large-sample asymptotic theory. In general, consider a sequence $\{\mathsf{S}_m \mid m = 1, 2, \cdots\}$ of $q \times q$ positive definite random matrices. Suppose that we are interested in the asymptotic distribution of the characteristic roots $\ell_1 > \cdots > \ell_q > 0$ of S_m. Assume that there exists a $q \times q$ diagonal matrix $\mathbf{\Lambda}$ such that the random matrix

$$\mathsf{V}_m = \sqrt{m}(\mathsf{S}_m - \mathbf{\Lambda}) \tag{3.6}$$

converges in distribution to that of a random matrix V. Here, let $\lambda_1 > \cdots \lambda_h \geq 0$ be the distinct diagonal elements of $\mathbf{\Lambda}$ and let q_α be the multiplicity of $\lambda_\alpha, \alpha = 1, \ldots, h$, i.e.,

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 \mathbf{I}_{q_1} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \lambda_2 \mathbf{I}_{q_2} & \cdots & \mathbf{O} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \lambda_h \mathbf{I}_{q_h} \end{pmatrix}. \tag{3.7}$$

Our problem is to obtain the limiting distribution of

$$\tilde{\ell}_i = \sqrt{m}(\ell_i - \lambda_\alpha), \quad i \in J_\alpha, \quad \alpha = 1, \ldots, h, \tag{3.8}$$

where J_α is the set of integers $q_1 + \cdots + q_{\alpha-1} + 1, \cdots, q_1 + \cdots + q_\alpha$ with $q_0 = 0$. Let V be partitioned as

$$\mathsf{V} = \begin{pmatrix} \mathsf{V}_{11} & \mathsf{V}_{12} & \cdots & \mathsf{V}_{1h} \\ \mathsf{V}_{21} & \mathsf{V}_{22} & \cdots & \mathsf{V}_{2h} \\ \vdots & \vdots & \cdots & \vdots \\ \mathsf{V}_{h1} & \mathsf{V}_{h2} & \cdots & \mathsf{V}_{hh} \end{pmatrix}, \quad \mathsf{V}_{ij}; q_i \times q_j.$$

Then, it is known that

R1: The limiting distribution of $\tilde{\ell}_i, i \in J_\alpha$, is given by the distribution of the characteristic roots of $\mathsf{V}_{\alpha\alpha},\ \alpha = 1, \ldots, h$.

Methods similar to R1 were used in Hsu (1941a, b), Anderson (1963), Eton and Tayler (1991), and other studies.

On the other hand, there is a case in which $\lambda_h = 0$ and the distribution of $\mathbf{V}_{hh}$ degenerates, depending on the condition assumed for the noncentrality matrix. Such cases were first considered by Hsu (1941a) and Anderson (1951). In order to treat a more general case, consider a case such that $\mathbf{S}_m$ is expanded as

$$\mathbf{S}_m = \mathbf{\Lambda} + \frac{1}{\sqrt{m}}\mathbf{Q}^{(1)} + \frac{1}{m}\mathbf{Q}^{(2)} + \frac{1}{m\sqrt{m}}\mathbf{Q}^{(3)} + \mathrm{O_p}(m^{-2}). \tag{3.9}$$

Put $\mathbf{Q} = \mathbf{Q}^{(1)} + (1/\sqrt{m})\mathbf{Q}^{(2)} + (1/m)\mathbf{Q}^{(3)}$, and let $\mathbf{\Lambda}$, $\mathbf{Q}$ and $\mathbf{Q}^{(i)}$ be partitioned as

$$\mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{O} \\ \mathbf{O} & \lambda_h \mathbf{I}_{q_h} \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{[11]} & \mathbf{Q}_{[12]} \\ \mathbf{Q}_{[21]} & \mathbf{Q}_{[22]} \end{pmatrix}, \quad \mathbf{Q}^{(i)} = \begin{pmatrix} \mathbf{Q}^{(i)}_{[11]} & \mathbf{Q}^{(i)}_{[12]} \\ \mathbf{Q}^{(i)}_{[21]} & \mathbf{Q}^{(i)}_{[22]} \end{pmatrix},$$

where $\mathbf{Q}_{[22]}$ and $\mathbf{Q}^{(i)}_{[22]}$ are $q_h \times q_h$ matrices. The asymptotic distribution of $\tilde{\ell}_i = \sqrt{m}(\ell_i - \lambda_\alpha), i \in J_\alpha, \alpha = 1, \ldots, h-1$ can be obtained by the method R1. For the derivation of asymptotic distribution of $\ell_i, i \in J_h$, we can use the following result:

R2: The larst q_h characteristic roots $\ell_i,\ i \in J_h$ are given by those of

$$\begin{aligned} \mathbf{L}_h = \lambda_h \mathbf{I}_{q_h} &+ \frac{1}{\sqrt{m}}\mathbf{Q}_{[22]} - \frac{1}{m}\mathbf{Q}_{[21]}\mathbf{\Theta}\mathbf{Q}_{[12]} + \frac{1}{m\sqrt{m}}\Big\{\mathbf{Q}_{[21]}\mathbf{\Theta}\mathbf{Q}_{[11]}\mathbf{\Theta}\mathbf{Q}_{[12]} \\ &-\frac{1}{2}\mathbf{Q}_{[21]}\mathbf{\Theta}^2\mathbf{Q}_{[12]}\mathbf{Q}_{[22]} - \frac{1}{2}\mathbf{Q}_{[22]}\mathbf{Q}_{[21]}\mathbf{\Theta}^2\mathbf{Q}_{[12]}\Big\} + \mathrm{O_p}(\frac{1}{m^2}), \end{aligned} \tag{3.10}$$

where

$$\mathbf{\Theta} = \begin{pmatrix} \theta_{1r}\mathbf{I}_{q_1} & \cdots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & \theta_{h-1,h}\mathbf{I}_{q_h} \end{pmatrix}, \ \theta_{ir} = (\lambda_i - \lambda_h)^{-1},\ i = 1, \ldots, h-1.$$

Expansion formulas similar to (3.10) were used in Lawley (1956,1959), Fujikoshi (1977b), etc.

4. High-Dimensional Asymptotic Distributions in Multivariate Linear Models

4.1. High-Dimensional Asymptotic Framework

We are concerned with the distribution of the characteristic roots of $\mathbf{S}_h\mathbf{S}_e^{-1}$, where $\mathbf{S}_h$ and $\mathbf{S}_e$ are given in (2.3), which is the same as those of $\mathbf{B}\mathbf{W}^{-1}$ (see R1 in Section 3), where $\mathbf{B}$ and $\mathbf{W}$ are given in (3.4). Large-sample asymptotic distributions were studied by Hsu (1941a), Anderson (1951), and others, under the assumptions that (i) p, q, and k are fixed; (ii) n tends to infinity; and (iii) the order of $\mathbf{D}_\omega$ is $\mathrm{O}(n)$. For high-dimensional approximations, we assume that n, p, and k tend to infinity, but the ratio p/n tends to $c_0 \in (0,1)$, and k/n tends to zero. The $q \times q$ noncentrality matrix $\mathbf{D}_\omega$ depends on n and p, and it is assumed to be $\mathbf{D}_\omega = \mathrm{O}(n)$ and $\mathbf{D}_\omega = \mathrm{O}(np)$.

Our high-dimensional assumptions are summarized as follows.

$$
\begin{aligned}
&\mathrm{A1}: \ q \text{ is fixed}, \ \ k \text{ is fixed or tends to infinity}, \ p \text{ and } n \text{ tend to infinity}, \\
&\qquad\quad c = p/n \to c_0 \in [0,1), \ \ k/n \to 0. \\
&\mathrm{A2}: \ \omega_i = O(n), \ \ i = 1, \ldots, q. \\
&\mathrm{A3}: \ \omega_i = O(pn), \ \ i = 1, \ldots, q.
\end{aligned}
$$

Specifically, we consider two cases: (1) A1 & A2, (2) A1 & A3. Note that under A1, $m = n - k - p$ tends to ∞.

In general, the asymptotic distribution of the characteristic roots $\ell_1 > \cdots > \ell_q$ depends on the multiplicity of the population characteristic roots $\omega_1 \geq \cdots \geq \omega_q$. Under A2, it is assumed that the population characteristic roots have arbitrary multiplicities as follows:

$$
\begin{aligned}
&\omega_1 = \cdots = \omega_{q_1} = n\lambda_1, \\
&\omega_{q_1+1} = \cdots = \omega_{q_1+q_2} = n\lambda_2, \\
&\qquad\quad \vdots \\
&\omega_{q-q_h+1} = \cdots = \omega_q = n\lambda_h,
\end{aligned}
\tag{4.1}
$$

where the λ_i's are O(1) and $\lambda_1 > \lambda_2 > \cdots > \lambda_h \geq 0$. Here, $\sum_{\alpha=1}^{h} q_\alpha = q$. Note that we assume that the multiplicities of the ω_i's do not depend on n and p. The assumption (4.1) can be expressed in matrix notation as

$$\mathbf{D}_\omega = n\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \mathbf{O} \\ \mathbf{O} & \lambda_h \mathbf{I}_{q_h} \end{pmatrix}, \tag{4.2}$$

where $\boldsymbol{\Lambda}_1 = \mathrm{Diag}\left(\lambda_1 \mathbf{I}_{q_1}, \ldots, \lambda_{q_{h-1}} \mathbf{I}_{q_{h-1}}\right)$. Here, Diag means a bock diagonal matrix.

Similarly, under A3, we assume that

$$\begin{aligned} \omega_1 &= \cdots = \omega_{q_1} = np\delta_1, \\ \omega_{q_1+1} &= \cdots = \omega_{q_1+q_2} = np\delta_2, \\ &\vdots \\ \omega_{q-q_h+1} &= \cdots = \omega_q = np\delta_h, \end{aligned} \tag{4.3}$$

where the δ_i's are constants and $\delta_1 > \delta_2 > \cdots > \delta_h \geq 0$. In matrix notation, we have

$$\mathbf{D}_\omega = np\boldsymbol{\Delta} = np \begin{pmatrix} \boldsymbol{\Delta}_1 & \mathbf{O} \\ \mathbf{O} & \delta_h \mathbf{I}_{q_h} \end{pmatrix}, \tag{4.4}$$

where $\boldsymbol{\Delta}_1 = \mathrm{Diag}\left(\delta_1 \mathbf{I}_{q_1}, \ldots, \delta_{q_{h-1}} \mathbf{I}_{q_{h-1}}\right)$.

4.2. Case in which $\mathbf{D}_\omega = n\boldsymbol{\Lambda} = \mathrm{O}(n)$

In this section, we assume that $\mathbf{D}_\omega = n\boldsymbol{\Lambda} = \mathrm{O}(n)$ with λ_α as in (4.2). Let

$$\mathsf{U} = \frac{1}{\sqrt{p}}(\mathsf{B} - p\mathbf{I}_q - n\boldsymbol{\Lambda}), \quad \mathsf{V} = \frac{1}{\sqrt{m}}(\mathsf{W} - m\mathbf{I}_q). \tag{4.5}$$

Then, noting that B and W are Wishart distributions, we have that for a given $q \times q$ symmetric matrix $\mathbf{K}$,

$$\mathrm{E}\left\{\mathrm{etr}(\mathbf{K}\mathsf{V})\right\} = \mathrm{etr}(\mathbf{K}^2)\left\{1 + \mathrm{O}(m^{-1/2})\right\}, \tag{4.6}$$

$$\mathrm{E}\left\{\mathrm{etr}(\mathbf{K}\mathsf{U})\right\} = \mathrm{etr}\left\{\mathbf{K}^2(\mathbf{I}_q + 2(n/p)\boldsymbol{\Lambda})\right\}\left\{1 + \mathrm{O}(p^{-1/2})\right\}. \tag{4.7}$$

Results (4.6) and (4.7) show that the limiting distributions of $\mathbf{U}$ and $\mathbf{V}$ are normal. The random matrices $\mathbf{B}$ and $\mathbf{W}$ are expressed in terms of $\boldsymbol{U}$ and $\boldsymbol{V}$ as

$$\frac{1}{p}\mathbf{B} = \mathbf{I}_q + \frac{n}{p}\mathbf{\Lambda} + \frac{1}{\sqrt{p}}\mathbf{U}, \quad \frac{1}{m}\mathbf{W} = \mathbf{I}_q + \frac{1}{\sqrt{m}}\mathbf{V}.$$

The characteristic roots of $\mathbf{B}\mathbf{W}^{-1}$ are the same as those of

$$\begin{aligned}\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2} &= \frac{p}{m}\left(\frac{1}{m}\mathbf{W}\right)^{-1/2}\left(\frac{1}{p}\mathbf{B}\right)\left(\frac{1}{m}\mathbf{W}\right)^{-1/2} \\ &= \mathbf{D}_\mu + \frac{1}{\sqrt{m}}\mathbf{X} + \mathrm{O_p}(m^{-1}),\end{aligned}$$

where

$$\begin{aligned}\mathbf{D}_\mu &= \mathrm{Diag}(\mu_1\mathbf{I}_{q_1}, \ldots, \mu_r\mathbf{I}_{q_r}), \\ \mathbf{X} &= -\frac{1}{2}\left(\mathbf{V}\mathbf{D}_\mu + \mathbf{D}_\mu\mathbf{V}\right) + \sqrt{p/m}\mathbf{U},\end{aligned}$$

and $\mu_\alpha = p/m + (n/m)\lambda_\alpha$, $\alpha = 1, \ldots, r$. Here, $\mathrm{O_p}$ denotes the order in probability notation. Let $\mathbf{X}$ be partitioned as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1r} \\ \mathbf{X}_{21} & \mathbf{X}_{22} & \cdots & \mathbf{X}_{2r} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{X}_{r1} & \mathbf{X}_{r2} & \cdots & \mathbf{X}_{rr} \end{pmatrix}, \quad \mathbf{X}_{ij}; q_i \times q_j. \tag{4.8}$$

Below, we will show that $\mathbf{X}$ converges in distribution to a random matrix $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}_{ij})$. Therefore, by R1 in Section 2, we have that the limiting distribution of $\sqrt{m}(\ell_i - \mu_\alpha), j \in J_\alpha$, is the same as the distribution of the characteristic roots of $\tilde{\mathbf{X}}_{\alpha\alpha}$. Therefore, we consider the limiting joint distribution of $\{\mathbf{X}_{11}, \ldots, \mathbf{X}_{hh}\}$, based on the characteristic function method. Let $\mathbf{T} = (t_{ij})$ be a $q \times q$ symmetric matrix having $(1 + \delta_{ij})t_{ij}/2$ as its (i, j)th element. Here, δ_{ij} is the Kronecker delta, i.e., $\delta_{ij} = 0 (i \neq j)$ and $\delta_{ii} = 1$. Let $\mathbf{T}$ be partitioned into submatrices as $\mathbf{T} = (\mathbf{T}_{\alpha\beta})$, where $\mathbf{T}_{\alpha\beta}$ is a $q_\alpha \times q_\beta$ submatrix. The joint characteristic function of $\{\mathbf{X}_{11}, \ldots, \mathbf{X}_{hh}\}$ can be expressed as follows:

$$
\begin{aligned}
C(\mathbf{T}_{11},\ldots,\mathbf{T}_{rr}) &= \mathrm{E}\left[\mathrm{etr}\left(i\sum_{\alpha=1}^{r}\mathbf{T}_{\alpha\alpha}\mathbf{X}_{\alpha\alpha}\right)\right] \\
&= \mathrm{E}\left[\mathrm{etr}\left\{i\sum_{\alpha=1}^{r}\mathbf{T}_{\alpha\alpha}\left(-\mu_\alpha\mathbf{V}_{\alpha\alpha}+\sqrt{p/m}\mathbf{U}_{\alpha\alpha}\right)\right\}\right] \\
&= \mathrm{E}\left[\mathrm{etr}\left\{-i\mathbf{D}_{\mu\mathbf{T}}\mathbf{V}+i\sqrt{p/m}\mathbf{D}_{\mathbf{T}}\mathbf{U}\right\}\right],
\end{aligned}
$$

where

$$
\mathbf{D}_{\mathbf{T}} = \mathrm{Diag}(\mathbf{T}_{11},\ldots,\mathbf{T}_{hh}),\ \ \mathbf{D}_{\mu\mathbf{T}} = \mathrm{Diag}(\mu_1\mathbf{T}_{11},\ldots,\mu_h\mathbf{T}_{hh}).
$$

Using (4.6) and (4.7), we have

$$
C(\mathbf{T}_{11},\ldots,\mathbf{T}_{rr}) = \left\{\prod_{\alpha=1}^{h}\mathrm{etr}\left(\frac{1}{2}i^2\sigma_\alpha^2\mathbf{T}_{\alpha\alpha}^2\right)\right\}\left\{1+\mathrm{O}(n^{-1/2})\right\}, \tag{4.9}
$$

where

$$
\begin{aligned}
\sigma_\alpha^2 &= 2\left\{\mu_\alpha^2+\frac{p}{m}+2\frac{n}{m}\lambda_\alpha\right\} \\
&= 2\left\{\frac{p}{m}\left(\frac{p}{m}+1\right)+2\frac{n}{m}\left(\frac{p}{m}+1\right)\lambda_\alpha+\left(\frac{n}{m}\right)^2\lambda_\alpha^2\right\}.
\end{aligned} \tag{4.10}
$$

Result (5.27) implies that $\mathbf{X}_{11},\ldots,\mathbf{X}_{rr}$ are asymptotically independent, and $\sigma_{\alpha\alpha}^{-1}\mathbf{X}_{\alpha\alpha}$ converges to a $q_\alpha\times q_\alpha$ symmetric Gaussian Wigner matrix $\mathbf{F}$ in which the elements are independent, and its diagonal and off-diagonal elements are distributed as $\mathrm{N}(0,1)$ and $\mathrm{N}(0,1/2)$, respectively. Let

$$
z_i = \frac{\sqrt{m}}{\sigma_\alpha}(\ell_i-\mu_\alpha), \quad i\in J_\alpha, \quad \alpha=1,\ldots,h, \tag{4.11}
$$

and

$$
\begin{aligned}
\boldsymbol{z} &= (\boldsymbol{z}_1',\ldots,\boldsymbol{z}_h')', \\
\boldsymbol{z}_\alpha &= (z_{q_1+\cdots+q_{\alpha-1}+1},\ldots,z_{q_1+\cdots+q_\alpha})', \quad \alpha=1,\ldots,h,
\end{aligned} \tag{4.12}
$$

where $q_0 = 0$. The limiting distribution of $\boldsymbol{z}_\alpha$ is the distribution of the characteristic roots of the $q_\alpha \times q_\alpha$ symmetric Gaussian Wigner matrix F_α, whose density is given by

$$f_\alpha(\boldsymbol{z}_\alpha) = \frac{\pi^{q_\alpha(q_\alpha-1)/4}}{2^{q_\alpha/2}\Gamma_{q_\alpha}\left(\frac{1}{2}q_\alpha\right)} \exp\left(-\frac{1}{2}\sum_{i\in J_\alpha} z_i^2\right) \prod_{i<j;i,j\in J_\alpha} (z_i - z_j). \tag{4.13}$$

Summarizing the above results, we have the following theorem.

Theorem 4.1. *Let S_h and S_e be the random matrices in* (2.3)*, and let $\ell_1 > \cdots > \ell_q$ be the nonzero characteristic roots of $\mathsf{S}_h\mathsf{S}_e^{-1}$ under $n-k \geq p \geq q$. Suppose that the characteristic roots of the noncentrality matrix $\tilde{\boldsymbol{\Omega}}$ in* (2.4) *have arbitrary multiplicities as in* (4.1)*, but the multiplicities do not depend on n and p. Further, assume* A1 *and* A2 *except for the case that $c_0 = 0$ and $\lambda_h = 0$. Then, the standardized roots $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_h$ defined by* (4.11) *and* (4.12) *are asymptotically independent, and the limiting density of $\boldsymbol{z}_\alpha$ is given by* (4.13).

The result when all the nonzero roots $\omega_1, \ldots, \omega_q$ are simple was derived by Fujikoshi, Himeno, and Wakaki (2008). Recently Bai, Choi, and Fujikoshi (2015) attempted to extend the result to the nonnormal case when $\Omega = O$, while Johnstone (2008) studied the distribution of the largest root ℓ_1 when $p/n \to c_0 \in (0,1)$ and $q/n \to c_1 \in (0,1)$.

The characteristic roots $d_1 > \cdots > d_q > 0$ of $\mathsf{S}_h(\mathsf{S}_e + \mathsf{S}_h)^{-1}$ are also used instead of those of $\mathsf{S}_h\mathsf{S}_e^{-1}$. The correspondence between those characteristic roots is as follows:

$$d_i = \frac{\ell_i}{1+\ell_i}, \quad i = 1, \ldots, q.$$

Noting that $\{\ell/(1+\ell)\}' = (1+\ell)^{-2}$, we consider the standardized characteristic roots of d_i defined by

$$y_i = \frac{\sqrt{m}}{\sigma_\alpha}(1+\mu_\alpha)^2\left(d_i - \frac{\mu_\alpha}{1+\mu_\alpha}\right), \quad i \in J_\alpha, \quad \alpha = 1, \ldots, h, \tag{4.14}$$

and set

$$
\begin{aligned}
&\boldsymbol{y} = (\boldsymbol{y}_1', \dots, \boldsymbol{y}_h')', \\
&\boldsymbol{y}_\alpha = (y_{q_1+\cdots+q_{\alpha-1}+1}, \dots, y_{q_1+\cdots+q_\alpha})', \quad \alpha = 1, \dots, h.
\end{aligned}
\tag{4.15}
$$

Then, from Theorem 4.1, we have the following asymptotic result.

Corollary 4.1. *Under the same assumptions as in Theorem* 4.1*, the normalized roots* $\boldsymbol{y}_1, \dots, \boldsymbol{y}_h$ *defined by* (4.14) *and* (4.15) *are asymptotically independent, and the limiting density of* $\boldsymbol{y}_\alpha$ *is given by* $f_\alpha(\boldsymbol{y}_\alpha)$ *in* (4.13)*.*

4.3. Case in which $\mathbf{D}_\omega = np\boldsymbol{\Delta} = \mathrm{O}(np)$

In this section, we assume that $\mathbf{D}_\omega = np\boldsymbol{\Delta} = \mathrm{O}(np)$ with the δ_i's as in (4.3). Let

$$
\tilde{\mathbf{U}} = \frac{1}{\sqrt{np}}(\mathbf{B} - p\mathbf{I}_q - np\boldsymbol{\Delta}), \quad \mathbf{V} = \frac{1}{\sqrt{m}}(\mathbf{W} - m\mathbf{I}_q). \tag{4.16}
$$

Here, note that the usual standardization $\mathbf{U} = \sqrt{n}\tilde{\mathbf{U}}$ as in (4.5) diverges and has no limiting distribution, but $\tilde{\mathbf{U}}$ has the limiting distribution. In fact, the characteristic function of $\tilde{\mathbf{U}}$ can be expressed as

$$
\begin{aligned}
C_{\tilde{\mathbf{U}}}(\mathbf{T}) &= \mathrm{E}\left\{\mathrm{etr}\left(i\mathbf{T}\tilde{\mathbf{U}}\right)\right\} \\
&= \mathrm{etr}\left\{-\sqrt{p/n}\,\mathbf{T}(\mathbf{I}_q + n\boldsymbol{\Delta})\right\}\left|\mathbf{I}_q - \frac{2i}{\sqrt{np}}\mathbf{T}\right|^{-p/2} \\
&\times \mathrm{etr}\left\{np\boldsymbol{\Delta}\frac{i}{\sqrt{np}}\mathbf{T}\left(\mathbf{I}_q - \frac{2i}{\sqrt{np}}\mathbf{T}\right)^{-1}\right\} \\
&= \mathrm{etr}(2i^2\boldsymbol{\Delta}\mathbf{T}^2) + \mathrm{O}(n^{-1/2}).
\end{aligned}
$$

The above result implies that the limiting distribution of $\tilde{\mathbf{U}}$ is normal if δ_h is positive. In order to see the limiting distribution of $\tilde{\mathbf{U}}$ in the case $\delta_h = 0$, let $\tilde{\mathbf{U}}$ be partitioned as

$$
\tilde{\mathbf{U}} = \begin{pmatrix} \tilde{\mathbf{U}}_{[11]} & \tilde{\mathbf{U}}_{[12]} \\ \tilde{\mathbf{U}}_{[21]} & \tilde{\mathbf{U}}_{[22]} \end{pmatrix}, \quad \tilde{\mathbf{U}}_{[22]};\ q_h \times q_h.
$$

Similar notation is used for the matrices partitioned from $\mathbf{T}$. Then, if $\delta_h = 0$,

$$C_{\tilde{\mathbf{U}}}(\mathbf{T}) = \mathrm{etr}(2i^2 \mathbf{\Delta}_1 \mathbf{T}_{[11]}^2) + \mathrm{O}(n^{-1/2}).$$

The result implies that the limiting distribution of $\tilde{\mathbf{U}}_{[11]}$ is normal, and the terms of $\tilde{\mathbf{U}}_{[12]}$, $\tilde{\mathbf{U}}_{[21]}$, and $\tilde{\mathbf{U}}_{[22]}$ are $\mathrm{O_p}(n^{-1/2})$. In order to see the $\mathrm{O_p}(n^{-1/2})$ term in $\tilde{\mathbf{U}}_{[22]}$, consider the characteristic function of $\sqrt{n}\tilde{\mathbf{U}}_{[22]}$, which is asymptotically approximated as

$$\begin{aligned} C_{\sqrt{n}\tilde{\mathbf{U}}_{[22]}}(\mathbf{T}_{[22]}) &= \mathrm{E}\left[\mathrm{etr}(i\sqrt{n}\mathbf{T}_{[22]}\tilde{\mathbf{U}}_{[22]})\right] \\ &= \mathrm{etr}(-i\sqrt{p}\mathbf{T}_{[22]}) \left| \mathbf{I}_{q_r} - \frac{2i}{\sqrt{p}}\mathbf{T}_{[22]} \right|^{-p/2} \\ &= \mathrm{etr}(i^2 \mathbf{T}_{[22]}^2) \left\{ 1 + \mathrm{O}(p^{-1/2}) \right\}. \end{aligned}$$

Therefore, the limiting distribution of $\sqrt{n}\tilde{\mathbf{U}}_{[22]}$ is normal.

In general, using

$$\frac{1}{np}\mathbf{B} = \mathbf{\Delta} + \frac{1}{\sqrt{np}}\tilde{\mathbf{U}} + \frac{1}{n}\mathbf{I}_q, \text{ and } \quad \frac{1}{m}\mathbf{W} = \mathbf{I}_q + \frac{1}{\sqrt{m}}\mathbf{V},$$

we have

$$\begin{aligned} \frac{m}{np}\mathbf{W}^{-1/2}\mathbf{B}\mathbf{W}^{-1/2} &= \left(\frac{1}{m}\mathbf{W}\right)^{-1/2} \left(\frac{1}{p}\mathbf{B}\right) \left(\frac{1}{m}\mathbf{W}\right)^{-1/2} \\ &= \mathbf{\Delta} + \frac{1}{\sqrt{m}}\mathbf{Q}^{(1)} + \frac{1}{m}\mathbf{Q}^{(2)} + \frac{1}{m\sqrt{m}}\mathbf{Q}^{(3)} + \mathrm{O_p}(m^{-2}), \end{aligned} \tag{4.17}$$

where

$$\begin{aligned} \mathbf{Q}^{(1)} &= -\frac{1}{2}\mathbf{V}\mathbf{\Delta} - \frac{1}{2}\mathbf{\Delta}\mathbf{V}, \\ \mathbf{Q}^{(2)} &= \frac{3}{8}\mathbf{V}^2\mathbf{\Delta} + \frac{3}{8}\mathbf{\Delta}\mathbf{V} + \frac{1}{4}\mathbf{V}\mathbf{\Delta}\mathbf{V} + \frac{m}{n}\mathbf{I}_q, \\ \mathbf{Q}^{(3)} &= -\frac{5}{16}\mathbf{V}^3\mathbf{\Delta} - \frac{5}{16}\mathbf{\Delta}\mathbf{V}^3 - \frac{3}{16}\mathbf{V}^2\mathbf{\Delta}\mathbf{V} - \frac{3}{16}\mathbf{V}\mathbf{\Delta}\mathbf{V}^2 \\ &\quad + \frac{m\sqrt{m}}{\sqrt{np}}\tilde{\mathbf{U}} - \frac{m}{\sqrt{np}}\mathbf{V}. \end{aligned}$$

First we consider the case $\lambda_h > 0$. By a similar argument as in the case $\mathbf{D}_\omega = \mathrm{O}(n)$, we can see that for $\alpha = 1, \ldots, h$, the limiting distribution of

$$\sqrt{m}\left(\frac{m}{np}\ell_i - \delta_\alpha\right), \ i \in J_\alpha$$

is the same as that of the characteristic roots of the limiting distribution of $\mathbf{Q}^{(1)}_{\alpha\alpha} = (-\delta_\alpha)\mathbf{V}_{\alpha\alpha}$. Let

$$z_i = \frac{\sqrt{m}}{\tau_\alpha}\left(\frac{1}{p}\ell_i - \nu_\alpha\right), \quad i \in J_\alpha, \quad \alpha = 1, \ldots, h, \tag{4.18}$$

and

$$\nu_\alpha = \frac{n}{m}\delta_\alpha, \quad \tau_\alpha = \left(\frac{2n}{m}\right)\delta_\alpha, \quad \alpha = 1, \ldots, h. \tag{4.19}$$

Then, the limiting distribution of $z_i, i \in J_\alpha$ is the same as the limiting distribution of the characteristic roots of $(1/\sqrt{2})\mathbf{V}_{\alpha\alpha}$, i.e., a $q_\alpha \times q_\alpha$ symmetric Gaussian Wigner matrix.

Next we consider the case $\lambda_h = 0$. It is easy to see that the limiting distribution of $z_i, i \in J_\alpha, \alpha = 1, \ldots, h-1$ are the same as in the case $\lambda_h > 0$. So, we consider asymptotic distribution of $\ell_i, i \in J_h$. By using R2 in Section 2, it can be seen that the distribution of $\{m/(np)\}\ell_i, i \in J_h$ is asymptotically expressed as the characteristic roots of

$$\mathbf{L}_h = \frac{1}{m}\frac{m}{n}\mathbf{I}_{q_h} + \frac{1}{m\sqrt{m}}\left\{\frac{m\sqrt{m}}{\sqrt{np}}\tilde{\mathbf{U}}_{[22]} - \frac{m}{\sqrt{np}}\mathbf{V}_{[22]}\right\} + \mathrm{O_p}(m^{-2}) \tag{4.20}$$

This implies that the distribution of $\sqrt{m}(\ell_i - p/m), i \in J_h$, is asymptotically expressed as the characteristic roots of

$$\frac{\sqrt{p}}{\sqrt{m}}\sqrt{n}\tilde{\mathbf{U}}_{[22]} - \frac{\sqrt{np}}{m}\mathbf{V}_{[22]}$$

whose characteristic function is expanded as

$$\mathrm{etr}\left\{\frac{1}{2}i^2\frac{2p}{m}\left(1 + \frac{n}{m}\right)\mathbf{T}^2_{[22]}\right\}\left(1 + \mathrm{O}(m^{-1/2})\right).$$

Now, we define

$$\tilde{z}_i = \frac{\sqrt{m}}{\tilde{\tau}_h}\left(\ell_i - \tilde{\nu}_h\right), \quad i \in J_h \tag{4.21}$$

where

$$\tilde{\nu}_h = \frac{p}{m}, \quad \tilde{\tau}_h = \left\{ \frac{2p}{m} \left(1 + \frac{n}{m} \right) \right\}^{1/2}. \tag{4.22}$$

For these $z_i, i \in J_\alpha, \alpha = 1, \ldots, h-1$ and $\tilde{z}_i \in J_h$, using the same vector notation as in (4.12), we have the following theorem and corollary.

Theorem 4.2. *Let* $\mathbf{S}_h$ *and* $\mathbf{S}_e$ *be the random matrices in* (2.3). *Let* $\ell_1 > \ldots > \ell_q$ *be the nonzero characteristic roots of* $\mathbf{S}_h \mathbf{S}_e^{-1}$ *under* $n-k \geq p \geq q$. *We make the same assumption as in Theorem* 4.1, *except that we assume* A3 *instead of* A2. *When* $\delta_h > 0$,*the standardized roots* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_h$ *defined by* (5.29) *and* (4.21) *are asymptotically independent, and the limiting density of* $\boldsymbol{z}_\alpha$ *is given by* (4.13). *When* $\delta_h = 0$, *the limiting distribution of* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{h-1}$ *is the same as in the case* $\delta > 0$. *The limiting distribution of* $\tilde{\boldsymbol{z}}_i, i \in J_h$ *defined by* (4.21) *is given by the one with the density* $f_\alpha(\tilde{\boldsymbol{z}}_h)$ *in* (4.13). *Further,* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{h-1}$ *and* $\tilde{\boldsymbol{z}}_h$ *are asymptotically independent.*

Corollary 4.2. *Suppose the same assumption as in Theorem* 4.2 *with* $\delta_h = 0$. *Let*

$$\tilde{y}_i = \frac{\sqrt{m}}{\tilde{\tau}_h} (1 + \tilde{\nu}_h)^2 \left(d_i - \frac{\tilde{\nu}_h}{1 + \tilde{\nu}_h} \right), \quad i \in J_h. \tag{4.23}$$

Then, the limiting density fuction of $\tilde{y}_i, \in J_h$ *is given by* f_h *in* (4.13).

5. High-Dimensional Asymptotic Distributions of Canonical Correlations

5.1. Preliminaries

In this section, we consider asymptotic distributions of the canonical correlations between the two vectors $\boldsymbol{x}$, which is $p \times 1$, and $\boldsymbol{y}$, which is $q \times 1$. Let $\mathbf{S}$ be the sample covariance matrix of $(\boldsymbol{x}', \ \boldsymbol{y}')'$ based on a sample of size $N = n + 1$ from a $(p+q)$-dimensional normal distribution $\mathrm{N}_{q+p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Without loss of generality, we may assume that $q \le p$. Corresponding to a partition $(\boldsymbol{x}', \boldsymbol{y}')$, we partition $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, and $\mathbf{S}$ as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}. \tag{5.1}$$

Let $\rho_1 \ge \cdots \ge \rho_q \ge 0$ and $r_1 > \cdots > r_q > 0$ be the population and the sample canonical correlations between $\boldsymbol{x}$ and $\boldsymbol{y}$. Then, $\rho_1^2 \ge \cdots \ge \rho_q^2 \ge 0$ and $r_1^2 > \cdots > r_q^2 > 0$ are the characteristic roots of $\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ and $\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}$, respectively. We shall obtain the distribution of the canonical correlations by deriving the transformed canonical correlations

$$d_i = r_i/\sqrt{1 - r_i^2}, \quad i = 1, \dots, q, \tag{5.2}$$

whose population transformed canonical correlations are expressed as

$$\gamma_i = \rho_i/\sqrt{1 - \rho_i^2}, \quad i = 1, \dots, q. \tag{5.3}$$

We use the following notation for diagonal matrices:

$$\mathbf{D}_d = \mathrm{diag}(d_1, \dots, d_q), \quad \mathbf{D}_\gamma = \mathrm{diag}(\gamma_1, \dots, \gamma_q).$$

Hsu (1941b) derived the asymptotic distributions of the canonical correlations under a large sample framework;

$$\mathrm{B0}: \ p \text{ and } q \text{ are fixed}, \quad n \to \infty.$$

However, the results do not work well as the dimension q or p becomes large, and so some high-dimensional approximations were considered under

$$\begin{aligned} \mathrm{B1}: q; \text{fixed}, \quad & p \to \infty, \quad n \to \infty, \quad m = n - p \to \infty, \\ & c = p/n \to c_0 \in [0, 1). \end{aligned}$$

For the population roots, the following two cases are considered:

$$\begin{aligned} \mathrm{B2}: \quad & \gamma_i = \mathrm{O}(1), \ i = 1, \dots, q, \\ \mathrm{B3}: \quad & \gamma_i = \mathrm{O}(\sqrt{p}), \ i = 1, \dots, q. \end{aligned}$$

Under the assumption that all the population roots are simple, in addition to B1 and B2, Fujikoshi and Sakurai (2009) obtained the following result:

$$\sqrt{n}(r_\alpha^2 - \tilde{\rho}_\alpha^2) \ \stackrel{d}{\to}\ \mathrm{N}(0, \sigma_\alpha^2), \tag{5.4}$$
$$\sqrt{n}(r_\alpha - \tilde{\rho}_\alpha) \ \stackrel{d}{\to}\ \mathrm{N}(0, \frac{1}{4}\sigma_\alpha^2\tilde{\rho}_\alpha^{-2}), \tag{5.5}$$

where

$$\tilde{\rho}_\alpha = \{\rho_\alpha^2 + c(1-\rho_\alpha^2)\}^{1/2},\ \sigma_\alpha^2 = 2(1-c)(1-\rho_\alpha^2)^2\{2\rho_\alpha^2 + c(1-2\rho_\alpha^2)\}.$$

In particular, letting $c = 0$ in (5.4) and (5.5), we have the large sample results:

$$\sqrt{n}(r_\alpha^2 - \rho_\alpha^2) \ \stackrel{d}{\to}\ \mathrm{N}(0, 4\rho_\alpha^2(1-\rho_\alpha^2)^2), \tag{5.6}$$
$$\sqrt{n}(r_\alpha - \rho_\alpha) \ \stackrel{d}{\to}\ \mathrm{N}(0, (1-\rho_\alpha^2)^2). \tag{5.7}$$

Here, we note that the high-dimensional asymptotic results (5.4) and (5.5) depend on p through $c = p/n$, but the large-sample results (5.6) and (5.7) do not depend on p and thus are the same for all p. In this paper, we extend the high-dimensional results to the case in which the population roots have multiplicity greater than unity. Further, we consider the case in which the population roots satisfy B3.

Let $\mathbf{A} = n\mathbf{S}$, which is distributed as a Wishart distribution $\mathrm{W}_{q+p}(n, \mathbf{\Sigma})$, and partition $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

corresponding to a partition of $\mathbf{S}$. Then, $d_1^2 > \cdots > d_q^2$ are the characteristic roots of $\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1}$, where $\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$. When we consider the distribution of the characteristic roots $d_1^2 > \cdots > d_q^2$ or $d_1 > \cdots > d_q$, without loss of generality, we may assume (see Fujikoshi and Sakurai, 2010) that

(i) $\mathbf{A}_{22\cdot 1} \sim \mathrm{W}_q(m, \mathbf{I}_q)$;

(ii) $\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \sim \mathrm{W}_q(p, \mathbf{I}_p; \mathbf{D}_\gamma\mathbf{G}\mathbf{D}_\gamma)$, when the first $q \times q$ matrix $\mathbf{G}$ of $\mathbf{A}_{11}$ is given; here, $\mathbf{G} \sim \mathrm{W}_q(n, \mathbf{I}_q)$;

(iii) $\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ and $\mathbf{A}_{22\cdot 1}$ are independent;

where $m = n - p$.

5.2. The Case $\mathbf{D}_\gamma = \mathrm{O}(1)$

Under B2, we assume that the population characteristic roots γ_i's have arbitrary multiplicities as follows:

$$\begin{aligned}
\gamma_1 &= \cdots = \gamma_{q_1} = \lambda_1, \\
\gamma_{q_1+1} &= \cdots = \gamma_{q_1+q_2} = \lambda_2, \\
&\vdots \\
\gamma_{q-q_h+1} &= \cdots = \gamma_q = \lambda_h \geq 0,
\end{aligned} \tag{5.8}$$

where λ_i's are fixed constants and $\lambda_1 > \lambda_2 > \cdots > \lambda_h \geq 0$. In a matrix notation,

$$\mathbf{D}_\gamma = \mathbf{\Lambda} = \begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{O} \\ \mathbf{O} & \lambda_h \mathbf{I}_{q_h} \end{pmatrix}, \tag{5.9}$$

where $\mathbf{\Lambda}_1 = \mathrm{Diag}\left(\lambda_1 \mathbf{I}_{q_1}, \ldots, \lambda_{q_{h-1}} \mathbf{I}_{q_{h-1}}\right)$.

Let

$$\mathsf{H} = \sqrt{n}(n^{-1}\mathsf{G} - \mathbf{I}_q), \tag{5.10}$$

whose limiting distribution is normal, since for any symmetric matrix $\mathbf{K}$,

$$\mathrm{E}\left\{\mathrm{etr}(\mathbf{K}\mathsf{H})\right\} = \mathrm{etr}(\mathbf{K}^2)\left\{1 + \mathrm{O}(n^{-1/2})\right\}. \tag{5.11}$$

The conditional noncentrality matrix of $\mathsf{A}_{21}\mathsf{A}_{11}^{-1}\mathsf{A}_{12}$ is expressed as

$$\mathbf{\Omega} \equiv \mathbf{D}_\gamma \mathsf{G} \mathbf{D}_\gamma = n\mathbf{\Lambda}^2 + \sqrt{n}\mathbf{\Lambda}\mathsf{H}\mathbf{\Lambda} = \mathrm{O}_\mathrm{p}(n).$$

Let

$$\mathsf{U} = \frac{1}{\sqrt{p}}(\mathsf{A}_{21}\mathsf{A}_{11}^{-1}\mathsf{A}_{12} - p\mathbf{I}_q - n\mathbf{\Lambda}^2 - \sqrt{n}\mathbf{\Lambda}\mathsf{H}\mathbf{\Lambda}), \tag{5.12}$$

$$\mathsf{V} = \frac{1}{\sqrt{m}}(\mathsf{A}_{22\cdot 1} - m\mathbf{I}_q). \tag{5.13}$$

Each the limiting distributions of $\mathbf{U}$ and $\mathbf{V}$ are normal. The results can be seen that for any symmetric matrix $\mathbf{K}$,

$$\mathrm{E}\left\{\mathrm{etr}(\mathbf{K}\mathbf{V})\right\} = \mathrm{etr}(\mathbf{K}^2)\left\{1 + \mathrm{O}(m^{-1/2}\right\}, \tag{5.14}$$

$$\mathrm{E}\left\{\mathrm{etr}(\mathbf{K}\mathbf{U})\right\} = \mathrm{etr}\left\{\mathbf{K}^2(\mathbf{I}_q + 2(n/p)\boldsymbol{\Lambda})\right\}\left\{1 + \mathrm{O}(p^{-1/2})\right\}, \tag{5.15}$$

Noting that $\mathbf{D}_\gamma = \boldsymbol{\Lambda}$, we have

$$\begin{aligned}
\mathbf{A}_{22\cdot 1}^{-1/2}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1/2} &= \frac{p}{m}\left(\mathbf{I}_q + \frac{1}{\sqrt{m}}\mathbf{V}\right)^{-1/2} \\
&\times \left(\mathbf{I}_q + \frac{n}{p}\boldsymbol{\Gamma}^2 + \frac{\sqrt{n}}{p}\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Lambda} + \frac{1}{\sqrt{p}}\mathbf{U}\right)\left(\mathbf{I}_q + \frac{1}{\sqrt{m}}\mathbf{V}\right)^{-1/2} \\
&= \mathbf{D}_\mu + \frac{1}{\sqrt{m}}\mathbf{X} + \mathrm{O_p}(m^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{D}_\mu &= \mathrm{Diag}(\mu_1\mathbf{I}_{q_1}, \ldots, \mu_r\mathbf{I}_{q_r}), \\
\mathbf{X} &= -\frac{1}{2}\left(\mathbf{V}\mathbf{D}_\mu + \mathbf{D}_\mu\mathbf{V}\right) + \sqrt{\frac{p}{m}}\mathbf{U} + \sqrt{\frac{n}{m}}\mathbf{D}_\gamma\mathbf{H}\mathbf{D}_\gamma,
\end{aligned}$$

and

$$\mu_\alpha = p/m + (n/m)\lambda_\alpha^2, \quad \alpha = 1, \ldots, h. \tag{5.16}$$

The joint characteristic function of $\{\mathbf{X}_{11}, \ldots, \mathbf{X}_{hh}\}$ can be expressed as follows:

$$\begin{aligned}
C(\mathbf{T}_{11}, \ldots, \mathbf{T}_{rr}) &= \mathrm{E}\left[\mathrm{etr}\left(i\sum_{\alpha=1}^{r}\mathbf{T}_{\alpha\alpha}\mathbf{X}_{\alpha\alpha}\right)\right] \\
&= \mathrm{E}\left[\mathrm{etr}\left\{-i\mathbf{D}_{\mu\mathbf{T}}\mathbf{V} + i\sqrt{p/m}\mathbf{D}_{\mathbf{T}}\mathbf{U} + \sqrt{n/m}\mathbf{D}_{\lambda^2\mathbf{T}}\mathbf{H}\right\}\right],
\end{aligned}$$

where

$$\begin{aligned}
&\mathbf{D}_{\mathbf{T}} = \mathrm{Diag}(\mathbf{T}_{11}, \ldots, \mathbf{T}_{hh}),\ \mathbf{D}_{\mu\mathbf{T}} = \mathrm{Diag}(\mu_1\mathbf{T}_{11}, \ldots, \mu_h\mathbf{T}_{hh}), \\
&\mathbf{D}_{\lambda^2\mathbf{T}} = \mathrm{Diag}(\lambda_1^2\mathbf{T}_{11}, \ldots, \lambda_h^2\mathbf{T}_{hh}).
\end{aligned}$$

Using (5.11), (5.14) and (5.15), we have

$$C(\mathbf{T}_{11},\ldots,\mathbf{T}_{rr}) = \left\{\prod_{\alpha=1}^{h} \operatorname{etr}\left(\frac{1}{2}i^2\sigma_\alpha^2\mathbf{T}_{\alpha\alpha}^2\right)\right\}(1+\mathrm{O}(n^{-1/2})), \tag{5.17}$$

where

$$\begin{aligned}\sigma_\alpha^2 &= 2\left\{\mu_\alpha^2 + \frac{p}{m}\left(1+2\frac{n}{p}\lambda_\alpha^2\right)+\frac{n}{m}\lambda_\alpha^4\right\}\\ &= \frac{2n}{m^2}(n+m)\left(1+\lambda_\alpha^2\right)\left(\frac{p}{m+n}+\lambda_\alpha^2\right). \end{aligned}\tag{5.18}$$

Put

$$z_i = \frac{\sqrt{m}}{\sigma_\alpha}(d_i^2-\mu_\alpha), \quad i\in J_\alpha, \quad \alpha=1,\ldots,h, \tag{5.19}$$

and $\boldsymbol{z}=(\boldsymbol{z}_1',\ldots,\boldsymbol{z}_h')'$ with $\boldsymbol{z}_\alpha, \alpha=1,\ldots,h$ as in (4.12). Then, by arguments as in multivariate linear model we have the following theorem.

Theorem 5.1. *Let* S *be the sample covariance matrix which is decomposed as in* (5.1)*. Let* $d_i^2 = r_i^2/(1-r_i^2), i=1,\ldots,q,$ *where* $1>r_1^2>\cdots>r_q^2>0$ *are the squares of the canonical correlations,* i.e., *the characteristic roots of* $\mathsf{S}_{21}\mathsf{S}_{11}^{-1}\mathsf{S}_{12}\mathsf{S}_{22}^{-1}$. *Suppose that the population canonical correlations have arbitrary multiplicities as in* (5.22)*, but the multiplicities do not depend on* n *and* p*. Further, assume* B1 *and* B2 *except for the case that* $c_0=0$ *and* $\lambda_h=0$*. Then, the standardized roots* $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_h$ *defined by* (5.19) *are asymptotically independent, and the limiting density of* $\boldsymbol{z}_\alpha$ *is given by* (4.13)*.*

Corollary 5.1. *Under the same assumptions as in Theorem* 5.1*, consider the normalized variables of the squares of the canonical correlations and the canonical correlations themselves defined by*

$$y_i = \sqrt{m}\eta_\alpha^{-1}\left(r_i^2-\xi_\alpha^2\right), \quad i\in J_\alpha,\ \alpha=1,\ldots,h, \tag{5.20}$$

$$\tilde{y}_i = \sqrt{m}(2\xi_\alpha)\eta_\alpha^{-1}\left(r_i-\xi_\alpha\right), \quad i\in J_\alpha,\ \alpha=1,\ldots,h, \tag{5.21}$$

where

$$\begin{aligned}\xi_\alpha &= (p/n+\lambda_\alpha^2)^{1/2}(1+\lambda_\alpha^2)^{-1/2},\\ \eta_\alpha &= \sqrt{2}(m/n)(1+\lambda_\alpha^2)^{-3/2}\left\{(p/n)(1-\lambda_\alpha^2)+2\lambda_\alpha^2\right\}^{1/2}.\end{aligned}$$

Then $\boldsymbol{y} = (y_1, \ldots, y_q)'$ and $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_q)'$ have the same limiting distribution as the one of $\boldsymbol{z}$.

Here, note that

$$\xi_\alpha = \left(\frac{\mu_\alpha}{1+\mu_\alpha}\right)^{1/2}, \quad \eta_\alpha = \frac{\sigma_\alpha}{(1+\mu_\alpha)^2}, \quad \alpha = 1, \ldots, h.$$

5.3. The Case $\mathbf{D}_\gamma = \mathrm{O}(\sqrt{p})$

Under B3, we assume that the population transformed canonical correlations γ_i's have arbitrary multiplicities as follows:

$$\begin{aligned}
\gamma_1 &= \cdots = \gamma_{q_1} = \sqrt{p}\delta_1, \\
\gamma_{q_1+1} &= \cdots = \gamma_{q_1+q_2} = \sqrt{p}\delta_2, \\
&\vdots \\
\gamma_{q-q_h+1} &= \cdots = \gamma_q = \sqrt{p}\delta_h \geq 0,
\end{aligned} \tag{5.22}$$

where δ_i's are fixed constants and $\delta_1 > \delta_2 > \cdots > \delta_h \geq 0$. Under (5.22), $\mathbf{D}_\gamma$ is expressed as

$$\mathbf{D}_\gamma = \sqrt{p}\boldsymbol{\Delta} = \sqrt{p}\begin{pmatrix} \boldsymbol{\Delta}_1 & \mathbf{O} \\ \mathbf{O} & \delta_h \mathbf{I}_{q_h} \end{pmatrix}, \tag{5.23}$$

where $\boldsymbol{\Delta}_1 = \mathrm{Diag}\left(\delta_1 \mathbf{I}_{q_1}, \ldots, \delta_{q_{h-1}} \mathbf{I}_{q_{h-1}}\right)$. In this case, let

$$\tilde{\mathbf{U}} = \frac{1}{\sqrt{pn}}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - p\mathbf{I}_q - np\boldsymbol{\Delta}^2 - \sqrt{n}p\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta}), \tag{5.24}$$

and $\mathbf{V} = \frac{1}{\sqrt{m}}(\mathbf{A}_{22\cdot 1} - m\mathbf{I}_q)$ which is the same one as in (5.13). Then, we have

$$\begin{aligned}
\frac{1}{np}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} &= \boldsymbol{\Delta}^2 + \frac{1}{\sqrt{n}}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta} + \frac{1}{n}\mathbf{I}_q + \frac{1}{\sqrt{np}}\tilde{\mathbf{U}}, \\
\left(\frac{1}{m}\mathbf{A}_{22\cdot 1}\right)^{-1/2} &= \mathbf{I}_q - \frac{1}{2\sqrt{m}}\mathbf{V} + \frac{3}{8m}\mathbf{V}^2 - \frac{5}{16m\sqrt{m}}\mathbf{V}^3 + \mathrm{O}_\mathrm{p}(m^{-3/2}).
\end{aligned}$$

Therefore

$$\frac{m}{np}\mathbf{A}_{22\cdot 1}^{-1/2}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1/2} = \boldsymbol{\Delta}^2 + \frac{1}{\sqrt{m}}\tilde{\mathbf{Q}}^{(1)} + \frac{1}{m}\tilde{\mathbf{Q}}^{(2)} + \frac{1}{m\sqrt{m}}\tilde{\mathbf{Q}}^{(3)} + \mathrm{O_p}(m^{-2}), \tag{5.25}$$

where

$$\begin{aligned}
\tilde{\mathbf{Q}}^{(1)} &= \mathbf{Q}^{(1)}(\boldsymbol{\Delta}^2) + \sqrt{m/n}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta}, \\
\tilde{\mathbf{Q}}^{(2)} &= \mathbf{Q}^{(2)}(\boldsymbol{\Delta}^2) - \frac{1}{2}\sqrt{m/n}\left(\mathbf{V}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta} + \boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta}\mathbf{V}\right), \\
\tilde{\mathbf{Q}}^{(3)} &= \mathbf{Q}^{(2)}(\boldsymbol{\Delta}^2) + \frac{3}{8}\sqrt{m/n}\left(\mathbf{V}^2\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta} + \boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta}\mathbf{V}^2\right).
\end{aligned}$$

Here $\mathbf{Q}^{(i)}(\boldsymbol{\Delta}^2), i = 1, 2, 3$ are the ones obtained from $\mathbf{Q}^{(i)}, i = 1, 2, 3$ in (4.17) by substituting $\boldsymbol{\Delta}^2$ to $\boldsymbol{\Delta}$. In general, we have

$$\sqrt{m}\left(\frac{m}{np}\mathbf{A}_{22\cdot 1}^{-1/2}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1/2} - \boldsymbol{\Delta}^2\right) = \tilde{\mathbf{Q}}^{(1)} + \frac{1}{\sqrt{m}}\tilde{\mathbf{Q}}^{(2)} + \frac{1}{m}\tilde{\mathbf{Q}}^{(3)} + \mathrm{O_p}(m^{-3/2}).$$

Let

$$\mathbf{X} \equiv \tilde{\mathbf{Q}}^{(1)} = -\frac{1}{2}\mathbf{V}\boldsymbol{\Delta}^2 - \frac{1}{2}\boldsymbol{\Delta}^2\mathbf{V} + \sqrt{m/n}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta}, \tag{5.26}$$

which is partitioned as in (4.8). Now, we consider the limiting joint distribution of $\{\mathbf{X}_{11}, \dots, \mathbf{X}_{hh}\}$, The joint characteristic function of $\{\mathbf{X}_{11}, \dots, \mathbf{X}_{hh}\}$ can be expressed as follows:

$$\begin{aligned}
C(\mathbf{T}_{11}, \dots, \mathbf{T}_{rr}) &= \mathrm{E}\left[\mathrm{etr}\left(i\sum_{\alpha=1}^{r}\mathbf{T}_{\alpha\alpha}\mathbf{X}_{\alpha\alpha}\right)\right] \\
&= \mathrm{E}\left[\mathrm{etr}\left\{i\sum_{\alpha=1}^{r}\mathbf{T}_{\alpha\alpha}\left(-\delta_\alpha^2\mathbf{V}_{\alpha\alpha} + \sqrt{m/n}\delta_\alpha^2\mathbf{H}_{\alpha\alpha}\right)\right\}\right] \\
&= \mathrm{E}\left[\mathrm{etr}\left\{i\boldsymbol{\Delta}^2\mathbf{D_T}(-\mathbf{V} + \sqrt{p/m}\mathbf{H})\right\}\right],
\end{aligned}$$

where $\mathbf{D_T} = \mathrm{Diag}(\mathbf{T}_{11}, \dots, \mathbf{T}_{hh})$. Using (5.11) and (5.14), we have

$$C(\mathbf{T}_{11}, \dots, \mathbf{T}_{rr}) = \left\{\prod_{\alpha=1}^{h}\mathrm{etr}\left(\frac{1}{2}i^2\tau_\alpha^2\mathbf{T}_{\alpha\alpha}^2\right)\right\}\left\{1 + \mathrm{O}(n^{-1/2})\right\}, \tag{5.27}$$

where

$$\tau_\alpha = \sqrt{2(1+m/n)}\delta_\alpha^2, \quad \alpha = 1, \ldots, h. \tag{5.28}$$

First we consider the case $\delta_h > 0$. By a similar argument as in the case $\mathbf{D}_\gamma = \mathrm{O}(1)$, we can see that for $\alpha = 1, \ldots, h$, the limiting distribution of

$$\sqrt{m}\left(\frac{m}{np}d_i^2 - \delta_\alpha^2\right), \ i \in J_\alpha$$

is the same as that of the characteristic roots of the limiting distribution of $\tilde{\mathbf{Q}}_{\alpha\alpha}^{(1)} = -\delta_\alpha^2 \mathbf{V}_{\alpha\alpha} + \sqrt{m/n}\delta_\alpha^2 \mathbf{H}_{\alpha\alpha}$. Let

$$z_i = \frac{\sqrt{m}}{\tau_\alpha}\left(\frac{1}{p}d_i^2 - \nu_\alpha\right), \quad i \in J_\alpha, \quad \alpha = 1, \ldots, h, \tag{5.29}$$

and

$$\nu_\alpha = \frac{n}{m}\delta_\alpha^2, \quad \tau_\alpha = \sqrt{2n(n+m)}m^{-1}\delta_\alpha^2, \quad \alpha = 1, \ldots, h. \tag{5.30}$$

Then, the limiting distribution of $z_i, i \in J_\alpha$ is the same as the limiting distribution of the characteristic roots of a $q_\alpha \times q_\alpha$ symmetric Gaussian Wigner matrix $\mathbf{F}_\alpha$.

Next we consider the case $\lambda_h = 0$. It is easy to see that the limiting distribution of $z_i, i \in J_\alpha, \alpha = 1, \ldots, h-1$ are the same as in the case $\lambda_h > 0$. So, we consider asymptotic distribution of $d_i^2, i \in J_h$. By using R2 in Section 2, it can be seen that the distribution of $\{m/(np)\}d_i^2, i \in J_h$ is asymptotically expressed as the characteristic roots of

$$\tilde{\mathbf{L}}_h = \frac{1}{m}\frac{m}{n}\mathbf{I}_{q_h} + \frac{1}{m\sqrt{m}}\left\{\frac{m\sqrt{m}}{\sqrt{np}}\tilde{\mathbf{U}}_{[22]} - \frac{m}{\sqrt{np}}\mathbf{V}_{[22]}\right\} + \mathrm{O_p}(m^{-2}). \tag{5.31}$$

Here, note that the expansion formula (5.31) is the same as that in (4.20) up to the order $\mathrm{O_p}(m^{-3/2})$. The difinition of $\tilde{\mathbf{U}}$ and $\mathbf{V}$ in (5.31) is different from that in (4.20), but the distributions of $\mathbf{V}$ in both cases are the same, and the limiting distributions of $\tilde{\mathbf{U}}$ in both cases are also the same. Therefore, from the distributional results on $\mathbf{L}_h$ in (4.20)it follows that the asymptotic distribution of

$$\tilde{z}_i = \frac{\sqrt{m}}{\tilde{\tau}_h}\left(d_i^2 - \tilde{\nu}_h\right), \quad i \in J_h \tag{5.32}$$

is the same as that in (4.21), where $\tilde{\nu}_h$ and $\tilde{\tau}_h$ are the same as those in (4.22).

For these $z_i, i \in J_\alpha, \alpha = 1, \dots, h-1$ and $\tilde{z}_i \in J_h$, using the same vector notation as in (4.12), we have the following theorem and corollary.

Theorem 5.2. *Let* S *be the sample covariance matrix which is decomposed as in* (5.1). *Let* $d_i^2 = r_i^2/(1-r_i^2), i = 1, \dots, q,$ *where* $1 > r_1^2 > \cdots > r_q^2 > 0$ *are the squares of the canonical correlations,* i.e., *the characteristic roots of* $\mathsf{S}_{21}\mathsf{S}_{11}^{-1}\mathsf{S}_{12}\mathsf{S}_{22}^{-1}$. *We make the same assumption as in Theorem* 5.1, *except that we assume* B3 *instead of* B2. *When* $\delta_h > 0$,*the standardized roots* $\boldsymbol{z}_1, \dots, \boldsymbol{z}_h$ *defined by* (5.29) *and* (4.21) *are asymptotically independent, and the limiting density of* $\boldsymbol{z}_\alpha$ *is given by* (4.13). *When* $\delta_h = 0$, *the limiting distribution of* $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{h-1}$ *is the same as in the case* $\delta > 0$. *The limiting distribution of* $\tilde{\boldsymbol{z}}_i, i \in J_h$ *defined by* (4.21) *is given by the one with the density* $f_\alpha(\tilde{\boldsymbol{z}}_h)$ *in* (4.13). *Further,* $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{h-1}$ *and* $\tilde{\boldsymbol{z}}_h$ *are asymptotically independent.*

Corollary 5.2. *Under the same assumption as in Theorem* 5.2 *with* $\delta_h = 0$, *let*

$$\tilde{y}_i = \sqrt{m}\tilde{\tau}_h^{-1}(1+\tilde{\nu}_h)^2\left\{r_i^2 - \tilde{\nu}_h(1+\tilde{\nu}_h)^{-1}\right\}, \quad i \in J_h. \tag{5.33}$$

Then, the limiting density fuction of $\tilde{y}_i, \in J_h$ *is given by* f_h *in* (4.13). *Similarly, let*

$$\tilde{r}_i = \sqrt{m}2\sqrt{\tilde{\nu}_h}\tilde{\tau}_h^{-1}(1+\tilde{\nu}_h)^{3/2}\left\{r_i - \sqrt{\tilde{\nu}_h(1+\tilde{\nu}_h)^{-1}}\right\}, \quad i \in J_h. \tag{5.34}$$

Then, the limiting density fuction of $\tilde{r}_i, \in J_h$ *is given by* f_h *in* (4.13).

6. Concluding Remarks

In general, it is known that under the large-sample asymptotic framework the characteristic roots in a multivariate linear model are consistent, and the canonical correlations are also consistent. However, from our high-dimensional asymptotic results we have shown that the characteristic roots and the canonical correlation coefficients are not consistent. It will be possible to construct high-dimensional consistent estimators of the population characteristic roots and the population canonical correlations.

The present results have been also used to show that some model selection criteria for estimating the dimensionality in a multivariate linear model and canonical correlation analysis are high-dimensional consistent. For the results, see Fujikoshi and Sakurai (2016) and Fujikoshi (2016). It is expected that our results are basic in studying high-dimensional properties for multivariate inferential methods based on characteristic roots.

References

[1] Anderson, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. In *Proc. Second Berkerley Symp. on Math. Statist. and Prob.*, J. Neyman, ed., Univ. of California Press, Berkeley, CA, 103-130.

[2] Anderson, T.W. (1963). Asymptotic theory for principal components. *Ann. Math. Statist.*, **34**, 122–148. Eaton, M. L. and Tyler, D. E. (1991), On Wiedlandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.*, **19**, 260–271.

[3] Fujikoshi, Y. (1977a). Asymptotic expansion for the distributions of the latent roots in MANOVA and the canonical correlations. *J. Multivariate Anal.*, **7**, 386–396.

[4] Fujikoshi, Y. (1977b). An asymptotic expansion for the distributions of the latent roots of the Wishart matrix with multiple population roots. *Ann. Inst. Statist. Math.*, **29**, 379–387.

[5] 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.

[6] FUJIKOSHI, Y. and SAKURAI, T. (2009). High-dimensional asymptotic expansions for the distributions of canonical correlations. *J. Multivariate Anal.*, **100**, 231–241.

[7] FUJIKOSHI, Y., ULYANOV, V. and SHIMIZU, R. (2010). *Multivariate Analysis: High-Dimensional and Large-Sample Approximations*. Wiley, Hoboken, New Jersey.

[8] FUJIKOSHI, Y. and SAKURAI, T. (2016). High-dimensional consistency of rank estimation criteria in multivariate linear model. To appear in *J. Multivariate Anal.*.

[9] FUJIKOSHI, Y. (2016). High-dimensional properties of AIC, BIC and Cp for estimation of dimensionality in CCA. Submitted for publication.

[10] HSU, P. L. (1941a). On the limiting distribution of roots of a determinantal equation. *J. London Math. Soc.*, **16**, 183–194.

[11] HSU, P. L. (1941b). On the limiting distribution of roots of the canonical correlations. *Biometrika*, **32**, 38–45.

[12] GLYNN, W. J. (1980). Asymptotic representations of the densities of canonical correlations and latent roots in MANOVA when the populatio parameters have arbitrary multiplicities. *Ann. Statist.*, **8**, 468–477.

[13] LAWLEY, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–136.

[14] LAWLEY, D. N. (1959). Tests of significance in canonical analysis. *Biometrika*, **46**, 59–66.

[15] MUIRHEAD, R. J. (1978), Latent roots and matrix variates. *Ann. Statist.*, **6**, 5–33.

[16] Muirhead, R. J. (1982), *Aspect of Multivariate Statistical Theory*, Wiley, New York.

[17] Muirhead, R. J. and Watermaux, C. M. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate nonnormal populations. *Biometrika*, **67**, 31–43.

[18] Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rate of convergence. *Ann. Statist.*, **36**, 2638–2716.

[19] 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.

[20] Sugiura, N. and Fujikoshi, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Statist.*, **40**, 942–952.