SUT Journal of Mathematics
Vol. 53, No. 1 (2017), 59–72

High-dimensional properties of AIC, BIC and C_p for estimation of dimensionality in canonical correlation analysis

Yasunori Fujikoshi

(Received November 1, 2016; Revised January 29, 2017)

Abstract.
This paper is concerned with consistency properties of the dimensionality estimation criteria AIC, BIC and C_p in CCA (Canonical Correlation Analysis) between p variables and q $(\leq p)$ variables, based on a sample of size $N = n + 1$. The consistency properties of the criteria are studied under a high-dimensional asymptotic framework such that p and n tend to infinity satisfying $p/n \to c \in [0, 1)$, and under two types of assumptions on the order of the population canonical correlations, where q is fixed. We note that there are cases that the criteria based on AIC and C_p are consistent, but the criterion based on BIC is not consistent. Through a Monte Carlo simulation experiment, our results are checked numerically, and the estimation criteria are compared.

AMS 2010 *Mathematics Subject Classification.* 62H12, 62H30.

Key words and phrases. AIC, BIC, Canonical correlation analysis, C_p, Consistency property, Dimensionality, High-dimensional asymptotic framework.

§1. Introduction

In this paper we are concerned with the dimensionality estimation method by use of the model selection criteria AIC (Akaike (1973)), BIC (Schwarz (1978)) and C_p (Mallows (1973)) in CCA (canonical correlation analysis) with two random vectors of p and q $(q \leq p)$ components, based on the sample size $N = n+1$. The criteria based on AIC and C_p were proposed by Fujikoshi and Veitch (1979), and the criterion based on BIC was studied by Gunderson and Muirhead (1997). It is known in a large-sample asymptotic framework that AIC and C_p are not consistent, but BIC is consistent. For these results, see Fujikoshi (1985), Gunderson and Muirhead (1997).

However, recently it is known that there is a situation such that AIC and C_p have a consistency property, but BIC is not consistent, when the number p of response variables and the sample size n are large under a high-dimensional framework such that $p/n \to c \in [0, 1)$. These results can be found in Fujikoshi, Sakurai and Yanagihara (2014) and Yanagihara, Wakaki and Fujikoshi (2015) for selection of variables in multivariate regression model. Further, such properties have been shown in Fujikoshi and Sakurai (2016) for model selection criteria of estimating the dimensionality in multivariate linear model.

In this paper we consider asymptotic properties of AIC, BIC and C_p for estimation of dimensionality in CCA under a high-dimensional asymptotic framework such that

$$q \text{ is fixed}, \quad p \to \infty, \quad n \to \infty, \quad p/n \to c \in [0, 1). \tag{1.1}$$

It is shown that the AIC and C_p for estimation of dimensionality have consistency properties under two types of assumptions on population canonical correlations, but BIC is not consistent under one type of assumptions on population canonical correlations. It may be noted that these properties are different from the ones in a large-sample case, since in general AIC and C_p have a positive probability of selecting each of the overspecified models. Our results are checked numerically by conducting a Monte Carlo simulation experiment. Further, we compare with the selection rates of the three criteria.

The present paper is organized as follows. In Section 2, we summarize the criteria for estimating the dimensionality in CCA. High-dimensional properties of the criteria are given in Section 3. In Section 4 we check our theoretical results by conducting a Monte Carlo simulation experiment, and compare with the selection rates of the three criteria. In Section 5, we discuss our conclusions and future subjects. The proofs of our results are given in Appendix.

§2. The criteria for estimation of dimensionality

Let

$$\begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{y}_1 \end{pmatrix}, \ldots, \begin{pmatrix} \boldsymbol{x}_N \\ \boldsymbol{y}_N \end{pmatrix} \tag{2.1}$$

be a sample of size $N = n + 1$ of $(\boldsymbol{x}', \ \boldsymbol{y}')'$ from $(p + q)$-dimensional normal distribution $\mathrm{N}_{q+p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{x} : p \times 1$ and $\boldsymbol{y} : q \times 1$. Let S be the sample covariance matrix formed from the sample. In this paper we assume that $q \leq p$ without loss of generality. Corresponding to a partition $(\boldsymbol{x}', \boldsymbol{y}')$, we partition $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and S as

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

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

We are interested in the number of nonzero population canonical correlations, which is called the dimensionality in canonical correlation analysis. Related to the estimation of the dimensionality we consider a dimensionality model:

$$\begin{aligned} M_j : \qquad & \rho_j > \rho_{j+1} = \cdots = \rho_q = 0, \\ & \Leftrightarrow \ \mathrm{rank}(\boldsymbol{\Sigma}_{12}) = j. \end{aligned} \tag{2.2}$$

If M_j is true, we can explain the correlation structure between $\boldsymbol{x}$ and $\boldsymbol{y}$ by the first j canonical correlation variables, since the remaining canonical variables have no power of prediction.

Based on the likelihood of $\mathbf{S}$, it is known (Fujikoshi and Veitch (1979)) that AIC for M_j is given by

$$\begin{aligned} \mathrm{AIC}_j = & -\sum_{i=j+1}^{q} n\log(1-r_i^2) + n(p+q) + (p+q+1)\log|\mathbf{S}| + K \\ & + 2\left\{ j(p+q-j) + \frac{1}{2}p(p+1) + \frac{1}{2}q(q+1) \right\}, \end{aligned} \tag{2.3}$$

where $K = 2\log\left\{\boldsymbol{\Gamma}_{p+q}(\frac{1}{2}/(\frac{1}{2}n)^{(1/2)n(p+q)}\right\}$. Instead of AIC, we may use

$$\begin{aligned} \mathrm{A}_j = & \mathrm{AIC}_j - \mathrm{AIC}_q \\ = & -\sum_{i=j+1}^{q} n\log(1-r_i^2) - 2(p-j)(q-j), \ j = 0, \ldots, q. \end{aligned} \tag{2.4}$$

Here $\mathrm{A}_q = 0$. The BIC and $\mathrm{C}_{p,j}$ corresponding to A_j are given by

$$\mathrm{B}_j = -\sum_{i=j+1}^{q} n\log(1-r_i^2) - (\log n)(p-j)(q-j), \ j = 0, \ldots, q, \tag{2.5}$$

$$\mathrm{C}_{p,j} = n\sum_{i=j+1}^{q} \frac{r_i^2}{1-r_i^2} - 2(p-j)(q-j), \ j = 0, \ldots, q. \tag{2.6}$$

Here $\mathrm{B}_q = 0$ and $\mathrm{C}_{p,q} = 0$. Note that the A_j, B_j and $\mathrm{C}_{p,j}$ based on the likelihood of $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)'$ and $\mathbf{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)'$ can be expressed as the ones in (2.4), (2.5) and (2.6) replaced n by N.

The estimation methods based on A_j, B_j and $\mathrm{C}_{p,j}$ are expressed as

$$\hat{j}_{\mathrm{A}} = \arg\min_{j\in\mathcal{F}} \mathrm{A}_j, \quad \hat{j}_{\mathrm{B}} = \arg\min_{j\in\mathcal{F}} \mathrm{B}_j, \quad \hat{j}_{\mathrm{C}} = \arg\min_{j\in\mathcal{F}} \mathrm{C}_{p,j},$$

respectively.

§3. High-dimensional properties

We denote the model M_j in (2.2) by j simply. Then, the set of all the models is $\mathcal{F} = \{0, 1, \ldots, q\}$. It is assumed that the true covariance matrix is

$$\mathbf{\Sigma}^* = \begin{pmatrix} \mathbf{\Sigma}_{11}^* & \mathbf{\Sigma}_{12}^* \\ \mathbf{\Sigma}_{21}^* & \mathbf{\Sigma}_{22}^* \end{pmatrix}$$

with $\mathrm{rank}(\mathbf{\Sigma}_{12}^*) = j_*$. Then, the true dimension is j_*, where $0 \le j_* \le q$, and the model M_{j_*} means the minimum dimensionality model including M_*. We separate $\mathcal{F}$ into two sets, one is a set of overspecified models, i.e., $\mathcal{F}_+ = \{j_*, j_* + 1, \ldots, q\}$ and the other is a set of underspecified models, i.e., $\mathcal{F}_- = \mathcal{F}_+^c \cap \mathcal{F} = \{0, 1, \ldots j_* - 1\}$. Further, we denote the set of models deleting the true model from $\mathcal{F}_+$ by $\mathcal{F}_+ \backslash \{j_*\}$, i.e., $\mathcal{F}_+ \backslash \{j_*\} = \{j_* + 1, \ldots, q\}$.

When we treat the distributions of the canonical correlations themselves or their function such as $\hat{j}_{\mathrm{A}}$, $\hat{j}_{\mathrm{B}}$ and $\hat{j}_{\mathrm{C}}$, without loss of generality we may assume that

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{I}_p & \mathcal{R}' \\ \mathcal{R} & \mathbf{I}_q \end{pmatrix}, \tag{3.1}$$

$\mathcal{R} = (\mathcal{R}_1,\ \mathbf{O})$, $\mathcal{R}_1 = \mathrm{diag}(\rho_1, \ldots, \rho_q)$. The number of possible nonzero canonical correlations is q. We will consider the transformed population and sample canonical correlations defined by

$$\gamma_j = \frac{\rho_j}{(1 - \rho_j^2)^{1/2}}, \quad d_j = \frac{r_j}{(1 - r_j^2)^{1/2}}, \quad j = 1, \ldots, q. \tag{3.2}$$

Depending on our results, the following assumptions are taken up:

B1 (The true model and dimension j_*): The true model is that the samples in (2.1) are independently and identically distributrd as $\mathrm{N}_{q+p}(\boldsymbol{\mu}, \mathbf{\Sigma})$ with $\boldsymbol{\mu} = \boldsymbol{\mu}_*$ and $\mathbf{\Sigma} = \mathbf{\Sigma}_*$. The true dimensionality is j_*, and the true canonical correlations are

$$\rho_1^* \ge \cdots \ge \rho_{j_*}^* > \rho_{j_*+1}^* = \cdots = \rho_q^* = 0.$$

B2 (The asymptotic framework): q is fixed, $p \to \infty,\ n \to \infty,\ p/n \to c \in [0, 1)$.

B3 (The canonical correlations-1): Let $\gamma_j^* = \rho_j^*/\{1 - (\rho_j^*)^2\}, j = 1, \ldots, q$. For any i $(1 \le i \le j_*)$,
$\rho_i^* = \mathrm{O}(1),\ \gamma_i^* = \mathrm{O}(1)$ and $\lim_{p/n \to c} \rho_i^* = \rho_{i0}^*$.

B4 (The canonical correlations-2): For any $i (1 \le i < j_*)$,
$\gamma_i^* = \sqrt{p}\theta_i^* = \mathrm{O}(\sqrt{p})$ and $\lim_{p/n \to c} \theta_i^* = \theta_{i0}^*$.

In B3 and B4, it is assumed that the multiplicities of the ρ_i's do not depend on p and q. In this paper, the true dimension j_* is fixed, though our results shall be generalized for the case $\lim j_* = j_{*0}$. In the following, we give sufficient conditions for the three criteria to be consistent. Here, the consistency of, e.g., $\hat{j}_{\mathrm{A}}$ means that the probability that $\hat{j}_{\mathrm{A}}$ selects the true dimension tends asymptotically to 1, i.e.,

$$\lim_{p/n \to c} \Pr(\hat{j}_{\mathrm{A}} = j_*) = 1.$$

Here, the notation $\lim_{p/n \to c}$ is used as an abbreviation for the asymptotic framework (1.1).

Theorem 3.1. *Suppose that assumptions* B1 *and* B2 *are satisfied. Further, assume that* $c \in [0, c_{\mathrm{a}})$, *where* c_{a} (≈ 0.797) *is the constant satisfying* $\log(1 - c_{\mathrm{a}}) + 2c_{\mathrm{a}} = 0$.

1. $\hat{j}_{\mathrm{A}}$ *is consistent if* B3 *and the inequality* "$-\log(1-(\rho^*_{j_*})^2) > 2c + \log(1-c)$" *are satisfied.*
2. $\hat{j}_{\mathrm{A}}$ *is consistent if assumption* B4 *is satisfied.*

Theorem 3.2. *Suppose that assumptions* B1 *and* B2 *are satisfied.*

1. BIC *is not consistent if assumptions* B2 *with* $c > 0$ *and* B3 *are satisfied.*
2. BIC *is consistent if assumptions* B2 *and* B4 *are satisfied.*

Theorem 3.3. *Suppose that assumptions* B1 *and* B2 *are satisfied. Further, assume that* $c \in [0, 1/2)$.

1. $\hat{j}_{\mathrm{C}}$ *is consistent if* B3 *and the inequality* "$(\rho^*_{j_*0})^2 \left\{1 - (\rho^*_{j_*0})^2\right\}^{-1} > c(1-2c)$" *are satisfied.*
2. $\hat{j}_{\mathrm{C}}$ *is consistent if assumption* B4 *is satisfied.*

From the proofs of Theorems 3.1 and 3.3 we can see that the AIC and C_p criteria on the dimensionality in canonical correlation analysis satisfy the followings:

$$\text{(i) if } c \in [0, c_{\mathrm{a}}),\ \lim_{p/n \to c} \Pr(\hat{j}_{\mathrm{A}} \in \mathcal{F}_+ \backslash \{j_*\}) = 0, \tag{3.3}$$

$$\text{(ii) if } c \in [0, 0.5),\ \lim_{p/n \to c} \Pr(\hat{j}_{\mathrm{C}} \in \mathcal{F}_+ \backslash \{j_*\}) = 0. \tag{3.4}$$

These results hold without the assumptions on the order of population canonical correlations.

Under a large-sample framework; $n \to \infty$ and B3 it is known that

$$\lim_{n\to\infty} \Pr(\hat{j}_{\mathrm{A}} = j) = \lim_{n\to\infty} \Pr(\hat{j}_{\mathrm{C}} = j) = h(j|j_*), \tag{3.5}$$

where for $j = 0, 1, \ldots, j_* - 1$, $h(j|j_*) = 0$, and for $j = j_*, \ldots, q$, $h(j|j_*)$'s are positive and for their explicit expressions, see Fujikoshi (1985). Gunderson and Muirhead (1997) extended the result to the case of an elliptical distribution.

§4. Numerical study

In this section, we numerically examine the validity of some of our claims, and point some tendencies for the dimensionalities estimated by AIC, BIC and C_p. In our simulation setting, $q = 5$ and the true dimensionality is $j_* = 3$. We consider two types of population canonical correlations. The first type is:

$$\text{(a)}; \quad \rho_1^* = 2\rho, \quad \rho_2^* = \frac{3}{2}\rho, \quad \rho_3^* = \rho, \quad \rho_4^* = \rho_5^* = 0,$$

where

$$\rho = \sqrt{\frac{(4p)/(21)}{p + 1 + (4p)/(21)}}.$$

Then

$$\rho_1^* \to 0.8, \ \rho_2^* \to 0.6, \ \rho_3^* \to 0.4.$$

The second type is defined in terms of $\gamma_i^* = \rho_i^*/\sqrt{1 - (\rho_i^*)^2}$ as follows:

$$\text{(b)}; \quad \gamma_1^* = \frac{\tilde{\rho}}{\sqrt{1 - (\tilde{\rho})^2}}, \quad \gamma_2^* = \frac{3}{4}\gamma_1^*, \quad \gamma_3^* = \frac{1}{2}\gamma_1^*, \quad \gamma_4^* = \gamma_5^* = 0,$$

where

$$\tilde{\rho} = \sqrt{\frac{p}{p+1}} \cdot \sqrt{\frac{(4p)/(21)}{1 + (4p)/(21)}}.$$

In this case

$$\gamma_1^*/\sqrt{p} \to 0.8, \ \gamma_2^*/\sqrt{p} \to 0.6, \ \gamma_3^*/\sqrt{p} \to 0.4,$$

and

$$\text{(b)}; \quad \rho_i^* = \frac{\gamma_i^*}{\sqrt{1 + (\gamma_i^*)^2}}, \ i = 1, 2, 3, \quad \rho_4^* = \rho_5^* = 0.$$

The cases (a) and (b) correspond to the canonical correlations under assumptions B3 and B4. Our experiments were done for $n = 6p$ and $p = 5, 10, 20, 35, 50, 80, 100$. In this case, $p/n = 1/6 \to c = 1/6. \in [0, c_a)$ and

$c \in [0, 1/2)$. Further, the inequalities in Theorem 3.1(1) and Theorem 3.3.(1) have been satisfied for case (a). Our simulation results are given in Tables 1-6. We use the selection probabilities as the relative frequencies.

Table 1. Selection rates under AIC for case (a)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.00	0.14	0.48	0.31	0.05	0.01
60	10	0.00	0.03	0.47	0.45	0.05	0.00
120	20	0.00	0.00	0.43	0.55	0.02	0.00
210	35	0.00	0.00	0.39	0.61	0.00	0.00
300	50	0.00	0.00	0.36	0.64	0.00	0.00
480	80	0.00	0.00	0.31	0.69	0.00	0.00
600	100	0.00	0.00	0.30	0.70	0.00	0.00

Table 2. Selection rates under BIC for case (a)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.18	0.55	0.24	0.03	0.00	0.00
60	10	0.11	0.74	0.15	0.00	0.00	0.00
120	20	0.06	0.91	0.03	0.00	0.00	0.00
210	35	0.04	0.96	0.00	0.00	0.00	0.00
300	50	0.04	0.96	0.00	0.00	0.00	0.00
480	80	0.06	0.94	0.00	0.00	0.00	0.00
600	100	0.09	0.91	0.00	0.00	0.00	0.00

Table 3. Selection rates under C_p for case (a)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.00	0.06	0.43	0.41	0.08	0.01
60	10	0.00	0.01	0.31	0.58	0.11	0.00
120	20	0.00	0.00	0.21	0.73	0.07	0.00
210	35	0.00	0.00	0.14	0.83	0.03	0.00
300	50	0.00	0.00	0.11	0.88	0.01	0.00
480	80	0.00	0.00	0.05	0.94	0.00	0.00
600	100	0.00	0.00	0.04	0.96	0.00	0.00

Table 4. Selection rates under AIC for case (b)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.02	0.21	0.46	0.26	0.04	0.01
60	10	0.00	0.00	0.15	0.75	0.10	0.00
120	20	0.00	0.00	0.00	0.95	0.05	0.00
210	35	0.00	0.00	0.00	0.99	0.01	0.00
300	50	0.00	0.00	0.00	1.00	0.00	0.00
480	80	0.00	0.00	0.00	1.00	0.00	0.00
600	100	0.00	0.00	0.00	1.00	0.00	0.00

Table 5. Selection rates under BIC for case (b)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.43	0.42	0.13	0.01	0.00	0.00
60	10	0.05	0.45	0.45	0.05	0.00	0.00
120	20	0.00	0.01	0.69	0.31	0.00	0.00
210	35	0.00	0.00	0.10	0.90	0.00	0.00
300	50	0.00	0.00	0.00	1.00	0.00	0.00
480	80	0.00	0.00	0.00	1.00	0.00	0.00
600	100	0.00	0.00	0.00	1.00	0.00	0.00

Table 6. Selection rates under C_p for case (b)

n	p	M_0	M_1	M_2	M_3	M_4	M_5
30	5	0.00	0.10	0.45	0.37	0.07	0.01
60	10	0.00	0.00	0.07	0.74	0.18	0.01
120	20	0.00	0.00	0.00	0.86	0.14	0.00
210	35	0.00	0.00	0.00	0.94	0.06	0.00
300	50	0.00	0.00	0.00	0.98	0.02	0.00
480	80	0.00	0.00	0.00	1.00	0.00	0.00
600	100	0.00	0.00	0.00	1.00	0.00	0.00

From Tables 1, 3, 4 and 6 we can see that AIC and C_p are consistent for the estimation of dimensionality in the high-dimensional settings considered. On the speed of convergence to the true dimension, the case of canonical correlations-2 is faster than the one of canonical correlations-1. Further, in the case of canonical correlations-1 C_p is faster than AIC. These criteria have a tendency of underestimating the dimensionality when (n, p) is not so large.

From Tables 2 and 5 we can see that BIC is not consistent in the case of canonical correlations-1, but is consistent in the case of canonical correlations-2.

§5. Concluding remarks

In general, it is known that under a large sample asymptotic framework; $n \to \infty$ and p and q are fixed, AIC and C_p have no consistency property, in the sense that the probabilities of selecting the true model do not approach to one, but BIC is consistent. However, in this paper, we demonstrated that the AIC and C_p for estimating the dimensionality in canonical correlation analysis have a consistency property, under a high-dimensional framework (1.1). For the consistency, it is required to satisfy some additional assumptions. For AIC, it needs that $c \in [0, c_a)$, where $c_a \approx 0.797$. For C_p, it needs that $c \in [0, 1/2)$. Further, the consistency was considered under two types of assumptions on the population canonical correlations. On the other hand, in a high-dimensional case, we note that BIC is not always consistent.

In this paper we assume that our sample of size $N = n + 1$ comes from $(p+q)$-dimensional normal distribution $\mathrm{N}_{q+p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Further, we assume that for the random vectors of p and q $(\le p)$, q is finite, and p tends to infinity, satisfying $p/n \to c \in [0, 1)$. It is expected to remove the normality assumption, and to consider consistency properties of AIC, BIC and C_p under a general high-dimensional asymptotic framework such that $p/n \to c \in [0, 1)$ and $q/n \to d \in [0, 1)$.

§6. Appendix: The proofs of Theorems 3.1, 3.2 and 3.3

For the proofs of Theorems 3.1, 3.2 and 3.3, it is basic to establish the limiting behavior of the squares $r_1^2 > \cdots > r_q^2$ of the canonical correlations under a high-dimensional framework. The following results follows from the limiting distributions of $r_1^2 > \cdots > r_q^2$ given by Fujikoshi (2017) (For some special cases, see Fujikoshi and Sakurai (2008)).

Lemma 6.1. *Let $r_1^2 > \cdots > r_q^2$ and $\rho_1^2 \ge \cdots \ge \rho_q^2$ be the squares of the sample and the population canonical correlations between $\boldsymbol{x}; p \times 1$ and $\boldsymbol{y}; q \times 1$ with $p \ge q$, based on a sample of size $N = n + 1$ from $\mathrm{N}_{p+q}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $d_j^2 = r_j^2/(1 - r_j^2)$, $\gamma_j^2 = \rho_j^2/(1 - \rho_j^2)$, $j = 1, \ldots, q$. We assume that the number of nonzero population canonical correlations is a, and hence $\rho_1 \ge \cdots \ge \rho_a > \rho_{a+1} = \cdots = \rho_q = 0$, and the multiplicities of ρ_i's do not depend on p and* *q. Suppose that for the limiting behaviors of* $r_1^2 > \cdots > r_q^2$ *and* $d_1^2 > \cdots > d_q^2$ *under a high-dimensional asymptotic framework*

$$p \to \infty, \quad n \to \infty, \quad p/n \to c \in [0,1),$$

we have the following results:

1. *Suppose that for any* i $(1 \le i \le a)$, $\rho_i^2 = \mathrm{O}(1)$ *and* $\lim_{p/n\to c} \rho_i^2 = \rho_{i0}^2 > 0$. *Let* $\gamma_{i0} = \rho_{i0}/\{1-(\rho_{i0})^2\}^{1/2}, i = 1,\ldots,a$. *Then*

$$r_i^2 \xrightarrow{p} \rho_{i0}^2 + c(1-\rho_{i0}^2), \quad d_i^2 \xrightarrow{p} \frac{c}{1-c} + \frac{1}{1-c}\gamma_{i0}^2; \quad i = 1,\ldots,a,$$
$$r_i^2 \xrightarrow{p} c, \quad d_i^2 \xrightarrow{p} \frac{c}{1-c}; \quad i = a+1,\ldots,q.$$

2. *Suppose that for any* i $(1 \le i \le a)$, $\gamma_i^2 = p\theta_i^2 = \mathrm{O}(p)$ *and* $\lim_{p/n\to c} \theta_i = \theta_{i0} > 0$. *Then*

$$\frac{1}{p} d_i^2 \xrightarrow{p} \frac{1}{1-c}\theta_{i0}^2; \quad i = 1,\ldots,a,$$
$$r_i^2 \xrightarrow{p} c, \quad d_i^2 \xrightarrow{p} \frac{c}{1-c}; \quad i = a+1,\ldots,q.$$

The proofs of Theorems 3.1 and 3.2

We consider a general criterion including A_j and B_j defined by

$$\mathrm{G}_{\nu_n,j} = -\sum_{i=j+1}^{q} n\log(1-r_i^2) - \nu_n(p-j)(q-j), \ j = 0,\ldots,q, \tag{6.1}$$

with $\nu_n > 0$. Note that $\mathrm{G}_{2,j} = \mathrm{A}_j$ and $\mathrm{G}_{\log n,j} = \mathrm{B}_j$. Let $d_j^2 = r_j^2/(1-r_j^2)$. Using $1 - r_j^2 = (1+d_j^2)^{-1}$, the difference between G_j and G_{j_*} is expressed in terms of $d_1^2, \ldots, d_q^2$ as follows. For $j > j_*$,

$$\mathrm{G}_{\nu_n,j} - \mathrm{G}_{\nu_n,j_*} - (j-j_*)\nu_n = -n\log\{(1+d_{j_*+1}^2)\cdots(1+d_j^2)\} + \nu_n(j-j_*)(p+q-j-j_*),$$

and for $j < j_*$,

$$\mathrm{G}_{\nu_n,j} - \mathrm{G}_{\nu_n,j_*} = n\log\{(1+d_{j+1}^2)\cdots(1+d_{j_*}^2)\} + \nu_n(j-j_*)(p+q-j-j_*).$$

From Lemma 6.1 we have that for $j_* < i \le q$, $d_i^2 \to c/(1-c)$ in both cases B3 and B4. Therefore we have that for $j > j_*$,

$$\frac{1}{p}\left(\mathrm{G}_{\nu_n,j} - \mathrm{G}_{\nu_n,j_*}\right) - (j-j_*)\nu_n \xrightarrow{p} (j-j_*)c^{-1}\log(1-c).$$

Noting that $\lim_{c\to 0+} c^{-1}\log(1-c) = -1$, we have that in both cases B3 and B4, for any $j > j_*$,

$$\frac{1}{p}\left(\mathrm{A}_j - \mathrm{A}_{j_*}\right) \xrightarrow{p} (j - j_*)\left\{\frac{1}{c}\log(1-c) + 2\right\} > 0, \tag{6.2}$$

if $c \in [0, c_{\mathrm{a}})$, and for any $j > j_*$,

$$\frac{1}{p\log n}\left(\mathrm{B}_j - \mathrm{B}_{j_*}\right) \xrightarrow{p} (j - j_*) > 0. \tag{6.3}$$

Next consider the case $j < j_*$. Noting that under B3

$$1 + d_i^2 \xrightarrow{p} \frac{1}{1-c}\left\{1 - (\rho_{i0}^*)^2\right\}^{-1}, \ for\ i < j_*, \tag{6.4}$$

we have

$$\begin{aligned}
&\frac{1}{p}\left(\mathrm{G}_{\nu_n,j} - \mathrm{G}_{\nu_n,j_*}\right) + (j_* - j)\nu_n \xrightarrow{p} \\
&\quad - (j_* - j)c^{-1}\log(1-c) - c^{-1}\sum_{i=j+1}^{j_*}\log\left\{1 - (\rho_{i0}^*)^2\right\}^2 \\
&\quad \geq (j_* - j)\left\{-\frac{1}{c}\log(1-c) - \frac{1}{c}\log\left\{1 - (\rho_{j_*}^*)^2\right\}\right\}.
\end{aligned}$$

Combining (6.2), (6.3) and the above results, we can get Therem 3.1 (1) and Therem 3.2 (1). Next we assume B4 and $j < j_*$. For $j < j_*$, we have

$$\begin{aligned}
&\frac{1}{n}\left(\mathrm{G}_{\nu_n,j} - \mathrm{G}_{\nu_n,j_*}\right) \\
&\quad \geq (j_* - j)\left\{\log(1 + d_{j_*}^2) - \nu_n(p + q - j - j_*)/n\right\} \equiv D_{\nu_n,j}
\end{aligned}$$

Further, for $i < j_*$, $(1/p)d_i^2 \xrightarrow{p} (1-c)^{-1}(\theta_i^*)^2$, and hence

$$\log(1 + d_i^2) - \log p \xrightarrow{p} \log(1-c)^{-1}(\theta_i^*)^2.$$

Therefor, we have

$$\frac{1}{\log p}D_{2,j} \xrightarrow{p} j - j_*, \ \text{and} \ \frac{1}{\log p}D_{\log n,j} \xrightarrow{p} (j - j_*)(1-c),$$

which implies Theorems 3.1 and 3.2. Here, for the last result, we use

$$\frac{p}{n}\left(\frac{\log n}{\log p}\right) = \frac{p}{n}\left(\frac{\log n}{\log n + \log(p/n)}\right) \to c.$$

The proof of Theorem 3.3

Theorem 3.3 is proved by a similar line as in the proof of Theorem 3.1 as follows. Note that

$$\mathrm{C}_{p,j} - \mathrm{C}_{p,j_*} = -n \sum_{i=j_*+1}^{j} d_i^2 + 2(j - j_*)(p + q - j - j_*), \quad \text{for } j > j_*,$$

$$\mathrm{C}_{p,j} - \mathrm{C}_{p,j_*} = n \sum_{i=j+1}^{j_*} d_i^2 - 2(j_* - j)(p + q - j - j_*), \quad \text{for } j < j_*,$$

where $d_i^2 = r_i^2/(1-r_i^2)$. Under both cases B3 and B4, we can see that by using $d_i^2 \xrightarrow{p} c(1-c)^{-1}$ for $i > j_*$,

$$\frac{1}{p}\left(\mathrm{C}_{p,j} - \mathrm{C}_{p,j_*}\right) \xrightarrow{p} (j - j_*)\left(-\frac{1}{1-c} + 2\right), \text{ for } j > j_*. \tag{6.5}$$

Next consider the case $j < j_*$. Using (6.4), under B3 we have

$$\begin{aligned}\frac{1}{n}(\mathrm{C}_{p,j} - \mathrm{C}_{p,j_*}) &\xrightarrow{p} \sum_{i=j+1}^{j_*} (1-c)^{-1}\left\{c + (\gamma_{i0}^*)^2\right\} - 2c(j_* - j) \\ &\geq (j - j_*)\left[(1-c)^{-1}\left\{c + (\gamma_{j_*0}^*)^2\right\} - 2c\right].\end{aligned}$$

Combining this result with (6.5), we get Theorem 3.3(1). Under B4, $(1/p)d_i^2 \xrightarrow{p} (1-c)^{-1}(\theta_i^*)^2$ for $i < j_*$, and it is easy to see

$$\frac{1}{np}(\mathrm{C}_{p,j} - \mathrm{C}_{p,j_*}) \xrightarrow{p} \frac{1}{1-c}\sum_{i=j+1}^{j_*} (\theta_{i0}^*)^2.$$

Combining this result with (6.5), we get Theorem 3.3(2).

Acknowledgments

The author would like to express his gratitude to the referee and the editor for their many valuable comments and suggestions. The author also would like to express his gratitude to Dr.Tetsuro Sakurai for his help of simulation study. The author's research is partially supported by the Ministry of Education, Science, Sports, and Culture, a Grant-in-Aid for Scientific Research (C), 16K00047, 2016-2018.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd. International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiadó, Budapest.

[2] Anderson, T.W. (2003). *Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, Hoboken, N. J.

[3] Bunea, F., She, Y. and Wegkamp, M. H. (2011). Optimal selection of reduced rank estimators of high-dimensional matrices. *Ann. Statist.*, **39**, 1282–1309.

[4] Bunea, F., She, Y. and Wegkamp, M. H. (2012). Joint variable and rank selection for parsimonious estimation of high-dimensional matrices. *Ann. Statist.*, **40**, 2359–2388.

[5] Chen, L. and Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous and dimension reduction and variable selection. *J. Amer. Statist. Assoc.*, **107**, 1533–1545.

[6] Fujikoshi, Y. and Veitch, L. G. (1979). Estimation of dimensionality in canonical correlation analysis. *Biometrika*, **66**, 345–351.

[7] Fujikoshi, Y. (1985). Two methods for estimation of dimensionality in canonical correlation analysis and the multivariate linear model. In *Statistical Theory and Data Analysis* (K. Matsushita, Ed.), 233–240. Elsevier Science, Amsterdam.

[8] Fujikoshi, Y. and Sakurai, T. (2008). High-dimensional asymptotic expansions for the distributions of canonical correlations. *J. Multivariate Anal.*, **100**, 231–241.

[9] Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010). *Multivariate Statistics: High-Dimensional and Large-Sample Approximations.* Wiley, Hobeken, N.J.

[10] Fujikoshi, Y., Sakurai, T. and Yanagihara, H. (2013). Consistency of high-dimensional AIC-type and C_p-type criteria in multivariate linear regression. *J. Multivariate Anal.*, **149**, 199–212.

[11] Fujikoshi, Y. and Sakurai, T. (2016). High-dimensional consistency of rank estimation criteria in multivariate linear Model. *J. Multivariate Anal.*, **149**, 199-212.

[12] Fujikoshi, Y. (2017). High-dimensional asymptotic distributions of characteristic roots in multivariate linear model and canonical correlation analysis. To appear in Hirosihma Math. J.

[13] Gunderson, B. K. and Muirhead, R. J. (1997). On estimating the dimensionality in canonical correlation analysis. *J. Multivariate Anal.*, **62**, 121–136.

[14] Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques.* Springer, New York.

[15] Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, **15**, 661–675.

[16] Yanagihara, H., Wakaki, H. and Fujikoshi, Y. (2015). A consistency property of the AIC for multivariate linear models when the dimension and the sample size are large. *Electron. J. Stat.*, **9**, 869-897.

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