CIRJE-F-818

Asymptotic Expansion and Estimation of EPMC for Linear Classification Rules in High Dimension

Tatsuya Kubokawa
University of Tokyo

Masashi Hyodo
Graduate School of Economics, University of Tokyo

Muni S. Srivastava
University of Toronto

September 2011

Asymptotic Expansion and Estimation of EPMC for Linear Classification Rules in High Dimension

Tatsuya Kubokawa[*], Masashi Hyodo[†] and Muni S. Srivastava[‡]

University of Tokyo and University of Toronto

September 29, 2011

Abstract

The problem of classifying a new observation vector into one of the two known groups distributed as multivariate normal with common covariance matrix is considered. In this paper, we handle the situation that the dimension, p, of the observation vectors is less than the total number, N, of observation vectors from the two groups, but both p and N tend to infinity with the same order. Since the inverse of the sample covariance matrix is close to an ill condition in this situation, it may be better to replace it with the inverse of the ridge-type estimator of the covariance matrix in the linear discriminant analysis (LDA). The resulting rule is called the ridge-type linear discriminant analysis (RLDA). The second-order expansion of the expected probability of misclassification (EPMC) for RLDA is derived, and the second-order unbiased estimator of EMPC is given. These results not only provide the corresponding conclusions for LDA, but also clarify the condition that RLDA improves on LDA in terms of EPMC. Finally, the performances of the second-order approximation and the unbiased estimator are investigated by simulation.

Key words and phrases: High dimension, inverted Wishart distribution, linear discriminant analysis, misclassification error, multivariate normal, ridge-type estimation, second-order approximation, Wishart identity.

1 Introduction

In this paper, we consider the classical problem of classifying a $p \times 1$ observation vector $\boldsymbol{x}$ into one of the two population groups Π_1 and Π_2. For each $i = 1, 2$, Π_i denotes a population from a multivariate normal distribution $\mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, and it is supposed that

[*]Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: tatsuya@e.u-tokyo.ac.jp

[†]Faculty of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, E-Mail: hyodohh@yahoo.co.jp

[‡]Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, CANADA M5S 3G3, E-Mail: srivasta@utstat.toronto.edu

$\boldsymbol{x}_{ij}$, $j = 1, \ldots, N_i$, are observed from the population Π_i. Here, $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma}$ are unknown parameters. These unknown parameters are estimated by $\overline{\boldsymbol{x}}_i = N_i^{-1} \sum_{j=1}^{N_i} \boldsymbol{x}_{ij}$, $i = 1, 2$, and $\widehat{\boldsymbol{\Sigma}}_0 = n^{-1}\boldsymbol{S}$, where

$$\boldsymbol{S} = \sum_{i=1}^{2}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \overline{\boldsymbol{x}}_i)', \quad N = N_1 + N_2, \quad n = N - 2.$$

A standard classification method is the Linear Discriminant Analysis (LDA), and $\boldsymbol{x}$ is classified into either Π_1 or Π_2 as

$$W_0 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_0^{-1}\{\boldsymbol{x} - \frac{1}{2}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)\} > (\text{resp.} <)0 \quad \Longrightarrow \quad \boldsymbol{x} \in \Pi_1(\text{resp.}\Pi_2). \tag{1.1}$$

The classification problem has been studied many times in the literature since Fisher (1936) and Wald (1944). Of these, asymptotic expansions of expected probability of misclassification (EPMC) for W_0 have been derived by Okamoto (1963), Siotani (1982) and others in the situation that n is large, but p is bounded. This classical problem has been revisited by Saranadasa (1993), Fujikoshi and Seo (1998), Tonda and Wakaki (2003), Fujikoshi (2004), Srivastava (2006), Srivastava and Kubokawa (2007) and Hyodo and Yamada (2010) in high-dimensional situations that both n and p are large. In this paper, we handle the case that p is a large number with the constraints $n > p$ and $\lim_{n\to\infty} p/n = \gamma < 1$. This asymptotic theory has been discussed by Saranadasa (1993), Fujikoshi and Seo (1998), Tonda and Wakaki (2003) and Fujikoshi (2004) for multivariate discriminant analysis, and by Kubokawa and Srivastava (2011) for the Akaike information criterion in a multivariate linear regression model. In this situation, we are faced with the following problems:

(I) The precision matrix $\widehat{\boldsymbol{\Sigma}}_0^{-1}$ is closer to an ill condition as p is larger and closer to n, which results in a larger fluctuation in LDA based on W_0 and thus gives a larger EPMC.

(II) Asymptotic approximations of EPMC for W_0 were derived by Okamoto (1963), Siotani (1982) and others based on large sample theory, but this classical approximation gets worse for larger p.

A simple way for coping with problem (I) is to use the ridge-type estimators for $\boldsymbol{\Sigma}$. Thus, we consider the estimator given by

$$\widehat{\boldsymbol{\Sigma}}_\lambda = n^{-1}(\boldsymbol{S} + \hat{\lambda}\boldsymbol{I}), \quad \hat{\lambda} = c_n \frac{\operatorname{tr} \boldsymbol{S}}{np},$$

for $c_n = O(1)$. The function $\hat{\lambda}$ was used by Srivastava and Kubokawa (2007) and Kubokawa and Srivastava (2011). Then, the ridge-type linear discrimination analysis (RLDA) as given in Srivastava and Kubokawa (2007) is given by

$$W_\lambda = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\{\boldsymbol{x} - \frac{1}{2}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)\} < (\text{resp.} >)0 \quad \Longrightarrow \quad \boldsymbol{x} \in \Pi_1(\text{resp.}\Pi_2). \tag{1.2}$$

It is noted that the use of RLDA in high dimension was suggested in Fujikoshi (2004), and recently RLDA with another type of $\hat{\lambda}$ was treated by Xu, Brock and Parrish (2009) and Hyodo and Yamada (2010).

For problem (II), we consider the asymptotic theory of both n and p going to infinity such that p/n goes to a constant γ, $0 \leq \gamma < 1$. In this asymptotic theory, Tonda and Wakaki (2003) derived the second-order approximation of EPMC of LDA given by W_0 and showed that the high-dimensional approximation is better than the large sample approximation given by Okamoto (1963). They also gave the second-order unbiased estimator of EPMC. Since LDA is scale-invariant, the distribution of LDA does not depend on $\boldsymbol{\Sigma}$, and Fujikoshi (2000) showed that W_0 can be expressed based on independent standard normal and chi-square random variables. Tonda and Wakaki (2003) used this expression to provide the asymptotic expansion. However, their approach cannot be applied to RLDA since W_λ is not scale-invariant.

In this paper, we derive a second-order approximation of EPMC of RLDA given by W_λ. Our approach is based on a direct approximation with respect to $\boldsymbol{S}^{-1}$, and we need the third and fourth moments of the inverted Wishart matrix. Using the Stein-Haff identity, we derive the higher order moments of the inverted Wishart matrix and evaluate second-order terms of EPMC. As a result, we obtain the second-order approximation of EPMC for W_λ. This yields the approximation of EPMC for LDA as a special case, which can be confirmed to be identical to the approximation given by Tonda and Wakaki (2003) in the sense of second-order approximation. From the approximation derived in this paper, it is seen that a difference between RLDA and LDA appears in the second-order term of their EPMC, and we can establish the condition that RLDA improves on LDA in terms of EPMC. This approximation also gives us a second-order unbiased estimator of EPMC for W_λ, which is an extension of Tonda and Wakaki (2003).

The paper is organized as follows: In Section 2, we derive the second-order expansion of EPMC of W_λ in high dimension with some remarks. Especially, it is noted that the effect on the ridge estimator does not appear in the second-order term in the usual large sample theory, but appears in the high dimension. It is also noted that the usual large sample theory can be induced from the high-dimensional asymptotic theory. In Section 3, we give a second-order unbiased estimator of EPMC for W_λ. In Section 4, we give simulation results for EPMC and estimators of EPMC. Through the simulation results, we can confirm that the second-order approximation and its unbiased estimator derived in this paper are not bad in most cases, and that the asymptotic approximation in high dimension is better than the large sample approximation given by Okamoto (1963) for large p. The higher order moments of the inverted Wishart matrix, evaluations of second-order terms and some proofs are given in the appendix.

2 Second-order Approximation of EPMC for RLDA

In this section, we derive the second-order approximation of EPMC for RLDA under appropriate assumptions. Throughout the paper, we use the notations

$$m = n - p, \quad \boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \quad \text{and} \quad \Delta^2 = \boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}.$$

Also, the expected probability of misclassification (EPMC) of RLDA W_λ and LDA W_0 are denoted by

$$e_\lambda(2|1) = P[W_\lambda < 0|\boldsymbol{x} \in \Pi_1] \quad \text{and} \quad e_0(2|1) = P[W_0 < 0|\boldsymbol{x} \in \Pi_1].$$

Assume the following conditions:

(A1) $n > p+7$, $(n,p) \to \infty$ and $p/n \to \gamma$ for $0 < \gamma < 1$.

(A2) $\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$ is bounded for large p.

(A3) There exist limiting values $\lim_{n\to\infty} \operatorname{tr} \boldsymbol{\Sigma}^i/p$ for $i = -1, 1, 2$.

Under the condition (A3), Srivastava (2005) showed that $\hat{\lambda} = O_p(1)$ and $\hat{\lambda} = \lambda + O_p(n^{-1})$, where

$$\lambda = \lim_{n\to\infty} c_n \operatorname{tr} [\boldsymbol{\Sigma}]/p.$$

Suppose that $\boldsymbol{x} \in \Pi_1$. Under this condition, a conditional distribution given $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{S})$ is written as $(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\{\boldsymbol{x} - 2^{-1}(\overline{\boldsymbol{x}}_1 + \overline{\boldsymbol{x}}_2)\}|(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{S}) \sim \mathcal{N}_p(-U, V)$ where

$$\begin{aligned} U =&(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}(\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - (1/2)(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2), \\ V =&(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2). \end{aligned}$$

Then, EPMC of W_λ can be expressed as

$$e_\lambda(2|1) = E[\Phi(U/\sqrt{V})],$$

where $\Phi(\cdot)$ denotes the distribution function of a standard normal random variable.

To expand U and V stochastically, define random variables $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ by

$$\begin{aligned} \boldsymbol{z}_1 =& N^{-1/2}\{N_1(\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) + N_2(\overline{\boldsymbol{x}}_2 - \boldsymbol{\mu}_2)\}, \\ \boldsymbol{z}_2 =& (N_1N_2/N)^{1/2}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2 - \boldsymbol{\delta}). \end{aligned}$$

It is seen that $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are mutually independently and identically distributed as $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Using these variables, we can rewrite U and V as

$$\begin{aligned} U =& -\frac{1}{2}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} - \frac{N_1 - N_2}{2N_1N_2}\boldsymbol{z}_2'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 - \frac{N_1^{1/2}}{(NN_2)^{1/2}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 + N^{-1/2}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_1 \\ &+ \frac{1}{(N_1N_2)^{1/2}}\boldsymbol{z}_1'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2, \\ V =& \boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} + \frac{N}{N_1N_2}\boldsymbol{z}_2'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 + 2\frac{N^{1/2}}{(N_1N_2)^{1/2}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2. \end{aligned}$$

We begin by expanding U stochastically. Let us define U_0, U_1 and U_2 as

$$\begin{aligned} U_0 =& -\frac{1}{2}\frac{n}{m}\Delta^2 - \frac{1}{2}\frac{N_1 - N_2}{N_1N_2}\frac{np}{m}, \\ U_1 =& -\frac{1}{2}\left\{n\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n}{m}\Delta^2\right\} - \frac{1}{2}\frac{N_1 - N_2}{N_1N_2}\left\{\boldsymbol{z}_2'n\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m}\right\} \\ &+ N^{-1/2}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_1 - \frac{N_1^{1/2}}{(NN_2)^{1/2}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 + \frac{1}{(N_1N_2)^{1/2}}\boldsymbol{z}_1'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2, \\ U_2 =& \frac{1}{2}n\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta} + \frac{1}{2}\frac{N_1 - N_2}{N_1N_2}\hat{\lambda}n\boldsymbol{z}_2'\boldsymbol{S}^{-2}\boldsymbol{z}_2. \end{aligned}$$

Then the stochastic expansion of U is given in the following lemma.

Lemma 2.1 *Assume the conditions (A1)-(A3). Then, U is expanded as $U = U_0 + U_1 + U_2 + O_p(n^{-3/2})$, where $U_0 = O(1)$, $U_1 = O_p(n^{-1/2})$ and $U_2 = O_p(n^{-1})$.*

Proof. For the proof, we use the expression

$$\begin{aligned}(\boldsymbol{S} + \hat{\lambda}\boldsymbol{I})^{-1} =&\boldsymbol{S}^{-1} - \hat{\lambda}\boldsymbol{S}^{-1}(\boldsymbol{I} + \hat{\lambda}\boldsymbol{S}^{-1})^{-1}\boldsymbol{S}^{-1}\\ =&\boldsymbol{S}^{-1} - \hat{\lambda}\boldsymbol{S}^{-2} + \hat{\lambda}^2\boldsymbol{S}^{-2}(\boldsymbol{I} + \hat{\lambda}\boldsymbol{S}^{-1})^{-1}\boldsymbol{S}^{-1},\end{aligned}$$

which is from Srivastava and Khatri (1979). Since $p/n \to \gamma$, $0 < \gamma < 1$, it follows from Bai and Yin (1993) that the smallest and largest eigenvalues of $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}/n$ are almost surely bounded by a constant. Since $\hat{\lambda} = O_p(1)$, it is noted that

$$n\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}(\boldsymbol{I} + \hat{\lambda}\boldsymbol{S}^{-1})^{-1}\boldsymbol{S}^{-1}\boldsymbol{\delta} \le n^{-2}\hat{\lambda}\boldsymbol{\delta}'(\boldsymbol{S}/n)^{-3}\boldsymbol{\delta} = O_p(n^{-2}).$$

Hence, for the first term in U, namely,

$$\begin{aligned}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} =&n\boldsymbol{\delta}'\{\boldsymbol{S}^{-1} - \hat{\lambda}\boldsymbol{S}^{-2} + \hat{\lambda}^2\boldsymbol{S}^{-2}(\boldsymbol{I} + \boldsymbol{S}^{-1})^{-1}\boldsymbol{S}^{-1}\}\boldsymbol{\delta}\\ =&n\boldsymbol{\delta}'\{\boldsymbol{S}^{-1} - \hat{\lambda}\boldsymbol{S}^{-2}\}\boldsymbol{\delta} + O_p(n^{-2}).\end{aligned}$$

We note that $n\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta} = O_p(n^{-1})$. Since $E[\boldsymbol{S}^{-1}] = (m-1)^{-1}\boldsymbol{\Sigma}^{-1}$, it is seen that

$$E[\boldsymbol{\delta}'(\boldsymbol{S}/n)^{-1}\boldsymbol{\delta}] = \frac{n}{m-1}\Delta^2 = \frac{n}{m}\Delta^2 + O(n^{-1}), \quad m = n - p.$$

Taking this fact into account, we see that $\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta}$ can be decomposed as

$$\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} = \frac{n}{m}\Delta^2 + \Big\{n\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n}{m}\Delta^2\Big\} - n^{-1}\hat{\lambda}\boldsymbol{\delta}'(\boldsymbol{S}/n)^{-2}\boldsymbol{\delta} + O_p(n^{-2}). \tag{2.1}$$

For the second term in U, note that $n\hat{\lambda}^2\boldsymbol{z}_2'\boldsymbol{S}^{-2}(\boldsymbol{I}+\hat{\lambda}\boldsymbol{S}^{-1})^{-1}\boldsymbol{S}^{-1}\boldsymbol{z}_2 \le n^{-2}\hat{\lambda}^2\boldsymbol{z}_2'(\boldsymbol{S}/n)^{-3}\boldsymbol{z}_2$ and that $n^{-2}\boldsymbol{z}_2'(\boldsymbol{S}/n)^{-3}\boldsymbol{z}_2 = O_p(n^{-1})$. Also note that with $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}$,

$$E[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2] = E[\mathrm{tr}\,\boldsymbol{W}^{-1}] = \frac{p}{m} + O(n^{-1}).$$

Then the second term can be expanded as

$$\frac{1}{n}\boldsymbol{z}_2'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 = \frac{p}{m} + \frac{1}{n}\Big\{\boldsymbol{z}_2'n\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m}\Big\} - \frac{\hat{\lambda}}{n^2}\boldsymbol{z}_2'(\boldsymbol{S}/n)^{-2}\boldsymbol{z}_2 + O_p(n^{-2}). \tag{2.2}$$

From (2.1) and (2.2), it follows that $U = U_0 + U_1 + U_2 + O_p(n^{-3/2})$. It can be seen that $U_0 = O(1)$ and $U_2 = O_p(n^{-1})$. Since $E[U_1^2] = O(n^{-1})$ as proved in the appendix, it can be verified that $U_1 = O_p(n^{-1/2})$, and the proof is complete. ■

We next expand V using similar arguments. Let us define V_0, V_1 and V_2 as

$$\begin{aligned}V_0 =&\frac{n^3}{m^3}\Delta^2 + \frac{N}{N_1N_2}\frac{pn^3}{m^3},\\ V_1 =&\Big\{n^2\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n^3}{m^3}\Delta^2\Big\}\\ &+ \frac{Nn^2}{N_1N_2}\Big\{\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\Big\} + 2\frac{N^{1/2}}{(N_1N_2)^{1/2}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2,\\ V_2 =& - 2n^2\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{\delta} - 2\frac{Nn^2}{N_1N_2}\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{z}_2.\end{aligned}$$

Then the stochastic expansion of V is given in the following lemma.

Lemma 2.2 *Assume the conditions (A1)-(A3). Then, V is expanded as $V = V_0 + V_1 + V_2 + O_p(n^{-3/2})$, where $V_0 = O(1)$, $V_1 = O_p(n^{-1/2})$ and $V_2 = O_p(n^{-1})$.*

Proof. The first term in V is approximated as

$$\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} = n^2\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta} - 2n^2\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{\delta} + O_p(n^{-2}).$$

Here, from Proposition A.1, it follows that

$$n^2E[\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta}] = n^2E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{\xi}\boldsymbol{\xi}'] = \alpha_2 n^2(n-1)\Delta^2 = \frac{n^3}{m^3}\Delta^2 + O(n^{-1}),$$

where $\alpha_2 = [m(m-1)(m-3)]^{-1}$, $\boldsymbol{\xi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$ and $\boldsymbol{W} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{S}\boldsymbol{\Sigma}^{-1/2}$. Thus, we can consider the expansion

$$\begin{aligned}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta} =& \frac{n^3}{m^3}\Delta^2 + \left\{n^2\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n^3}{m^3}\Delta^2\right\}\\ &- 2n^2\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{\delta} + O_p(n^{-2}).\end{aligned} \tag{2.3}$$

For the second term, it can be seen that

$$n^{-1}\boldsymbol{z}_2'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 = n\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - 2n\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{z}_2 + O_p(n^{-2}).$$

Noting that

$$nE[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2] = nE[\mathrm{tr}\,\boldsymbol{W}^{-2}] = \alpha_2 pn(n-1) = \frac{pn^2}{m^3} + O(n^{-1}),$$

we get the approximation

$$\begin{aligned}n^{-1}\boldsymbol{z}_2'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2 =& \frac{pn^2}{m^3} + n\left\{\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\right\}\\ &- 2n\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{z}_2 + O_p(n^{-2}).\end{aligned} \tag{2.4}$$

From (2.3) and (2.4), it follows that $V = V_0 + V_1 + V_2 + O_p(n^{-3/2})$. It can be seen that $V_0 = O(1)$ and $V_2 = O_p(n^{-1})$. Since $E[V_1^2] = O(n^{-1})$ as proved in the appendix, it can be verified that $V_1 = O_p(n^{-1/2})$, and the proof is complete. ■

Using the Taylor series expansions given in Lemmas 2.1 and 2.2, we observe that

$$\frac{U}{V^{1/2}} = \{U_0 + U_1 + U_2\}\frac{1}{V_0^{1/2}}\left\{1 - \frac{V_1 + V_2}{2V_0} + \frac{3(V_1 + V_2)^2}{8V_0^2}\right\} + O_p(n^{-3/2}), \tag{2.5}$$

which gives the expansion

$$U/V^{1/2} = w_0 + w_1 + w_2 + O_p(n^{-3/2}), \tag{2.6}$$

where

$$
\begin{aligned}
w_0 =& V_0^{-1/2} U_0, \\
w_1 =& V_0^{-1/2} \{U_1 - \frac{U_0}{2V_0} V_1\}, \\
w_2 =& V_0^{-1/2} \{U_2 - \frac{U_0}{2V_0} V_2 + \frac{3U_0}{8V_0^2} V_1^2 - \frac{1}{2V_0} U_1 V_1\}.
\end{aligned}
$$

Using the Taylor series expansion again, we can approximate the EPMC $E[\Phi(U/V^{1/2})]$ as

$$
\begin{aligned}
E[\Phi(U/V^{1/2})] =& E[\Phi(w_0 + (w_1 + w_2))] \\
=& \Phi(w_0) + \phi(w_0) E[w_1 + w_2 - \frac{1}{2} w_0 w_1^2] + O(n^{-3/2}),
\end{aligned}
\tag{2.7}
$$

where $\phi(\cdot)$ is a pdf of the standard normal distribution. Let $H = E[w_1 + w_2 - \frac{1}{2} w_0 w_1^2]$. Then, H can be written as

$$
\begin{aligned}
H =& \frac{1}{V_0^{1/2}} \{E[U_1] + E[U_2]\} - \frac{U_0}{2V_0^{3/2}} \{E[V_1] + E[V_2]\} \\
& - \frac{U_0}{2V_0^{3/2}} E[U_1^2] + \frac{U_0}{8V_0^{5/2}} \Big(3 - \frac{U_0^2}{V_0}\Big) E[V_1^2] - \frac{1}{2V_0^{3/2}} \Big(1 - \frac{U_0^2}{V_0}\Big) E[U_1 V_1].
\end{aligned}
\tag{2.8}
$$

The moments given in (2.8) can be approximated in the following proposition which will be shown in the appendix.

Theorem 2.1 *Assume the conditions (A1)-(A3). The moments in (2.8) are approximated as* $E[U_1] + E[U_2] = H_U(\boldsymbol{\delta}, \boldsymbol{\Sigma}) + O(n^{-2})$, $E[V_1] + E[V_2] = H_V(\boldsymbol{\delta}, \boldsymbol{\Sigma}) + O(n^{-2})$, $E[U_1^2] = H_1(\Delta^2) + O(n^{-2})$, $E[V_1^2] = H_2(\Delta^2) + O(n^{-2})$ *and* $E[U_1 V_1] = H_{12}(\Delta^2) + O(n^{-2})$ *where*

$$
\begin{aligned}
H_U(\boldsymbol{\delta}, \boldsymbol{\Sigma}) =& -\frac{n}{2m^2} \Big\{\Delta^2 + p\frac{N_1 - N_2}{N_1 N_2}\Big\} + \frac{\lambda}{2} \frac{n}{m^2} \Big\{\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-2} \boldsymbol{\delta} + \frac{\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}}{m} \Delta^2 + \frac{N_1 - N_2}{N_1 N_2} \frac{n}{m} \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\Big\}, \\
H_V(\boldsymbol{\delta}, \boldsymbol{\Sigma}) =& \frac{n^2(4n - m)}{m^4} \Big\{\Delta^2 + \frac{pN}{N_1 N_2}\Big\} - \lambda \frac{2n^3}{m^4} \Big\{\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-2} \boldsymbol{\delta} + \frac{2\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}}{m} \Delta^2 + \frac{N(n+p)}{N_1 N_2 m} \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\Big\}, \\
H_1(\Delta^2) =& \frac{n^2}{2m^3} \Big\{\Delta^4 + pn\frac{(N_1 - N_2)^2}{(N_1 N_2)^2} + 2p\frac{N_1 - N_2}{N_1 N_2} \Delta^2 + 2\frac{n}{N_2} (\Delta^2 + \frac{p}{N_1})\Big\}, \\
H_2(\Delta^2) =& 2\frac{n^5}{m^7} \Big\{(4n + p)\Delta^4 + 2\frac{N}{N_1 N_2} \{pn + (n+p)^2\} \Delta^2 + \frac{N^2 p}{(N_1 N_2)^2} \{pn + (n+p)^2\}\Big\}, \\
H_{12}(\Delta^2) =& -\frac{n^4}{m^5} \Big\{2\Delta^4 + 2\frac{n+p}{N_2} \Delta^2 + \frac{N_1^2 - N_2^2}{N_1^2 N_2^2} p(n+p)\Big\}.
\end{aligned}
$$

Based on the approximations given in Theorem 2.1, we can give the second order approximation of EPMC $e_\lambda(2|1)$.

Theorem 2.2 *Assume the conditions (A1)-(A3). The second order approximation of EPMC of RLDA is given by* $e_\lambda(2|1) = \overline{e}_\lambda(2|1) + O(n^{-3/2})$, *where*

$$\overline{e}_\lambda(2|1) = \Phi(U_0 V_0^{-1/2}) + \phi(U_0 V_0^{-1/2}) H(\boldsymbol{\delta}, \boldsymbol{\Sigma}). \tag{2.9}$$

Here, $H(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ *is given by*

$$\begin{aligned} H(\boldsymbol{\delta}, \boldsymbol{\Sigma}) =& \frac{1}{V_0^{1/2}} \Big\{ H_U(\boldsymbol{\delta}, \boldsymbol{\Sigma}) - \frac{U_0}{2V_0} H_V(\boldsymbol{\delta}, \boldsymbol{\Sigma}) - \frac{U_0}{2V_0} H_1(\Delta^2) \\ &+ \frac{U_0}{8V_0^2} \Big(3 - \frac{U_0^2}{V_0}\Big) H_2(\Delta^2) - \frac{1}{2V_0}\Big(1 - \frac{U_0^2}{V_0}\Big) H_{12}(\Delta^2) \Big\}, \end{aligned} \tag{2.10}$$

where $U_0 = -2^{-1}(n/m)\{\Delta^2 + p(N_2^{-1} - N_1^{-1})\}$ *and* $V_0 = (n^3/m^3)\{\Delta^2 + p(N_1^{-1} + N_2^{-1})\}$.

Theorem 2.2 gives the following corollary which was derived in Tonda and Wakaki (2003).

Corollary 2.1 *Let* $H_0(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ *denote the value of* $H(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ *when* $\lambda = 0$. *Then, under the conditions (A1)-(A3), the second order approximation of EPMC of LDA is given by* $e_0(2|1) = \overline{e}_0(2|1) + O(n^{-3/2})$, *where*

$$\overline{e}_0(2|1) = \Phi(U_0 V_0^{-1/2}) + \phi(U_0 V_0^{-1/2}) H_0(\boldsymbol{\delta}, \boldsymbol{\Sigma}). \tag{2.11}$$

In Section 4, we investigate numerically a performance of the approximation given in Theorem 2.2. It is observed that the difference between LDA and RLDA, given in (1.1) and (1.2), appears in $H_U(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ and $H_V(\boldsymbol{\delta}, \boldsymbol{\Sigma})$. Investigating the signs of the coefficients of λ in (2.8), we can see that the sign of the coefficient of λ is not positive if

$$\Delta^4 + 2\frac{N_2 m + N_1 p}{N_1 N_2}\Delta^2 + \frac{N_1^2 - N_2^2}{N_1^2 N_2^2} p^2 \geq 2\frac{mp}{N_1}\frac{\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}}{\operatorname{tr}\boldsymbol{\Sigma}^{-1}}.$$

A sufficient condition for this inequality is given in the following proposition.

Proposition 2.1 *Assume the conditions (A1)-(A3). Then, the second-order approximation* $\overline{e}_\lambda(2|1)$ *of EPMC for RLDA given in (1.2) is smaller than* $\overline{e}_0(2|1)$ *for EPMC of LDA given in (1.1) when the parameters satisfy the inequality*

$$\Delta^4 + \frac{2}{N_1}\Big\{ m + \frac{N_1}{N_2} p - pm \frac{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})}{\operatorname{tr}\boldsymbol{\Sigma}^{-1}} \Big\}\Delta^2 + (N_2^{-2} - N_1^{-2})p^2 \geq 0, \tag{2.12}$$

where $\mathrm{Ch}_{max}(\boldsymbol{A})$ *denotes the largest eigenvalue of matrix* $\boldsymbol{A}$.

Since the condition (2.12) is satified for large Δ^2, Proposition 2.1 means that RLDA improves on LDA in light of minimizing EPMC for large Δ^2. In the case of $N_1 = N_2$, the condition (2.12) can be simplified as

$$\Delta^2 \geq \frac{2}{N_1}\Big\{ pm \frac{\mathrm{Ch}_{max}(\boldsymbol{\Sigma}^{-1})}{\operatorname{tr}\boldsymbol{\Sigma}^{-1}} - n \Big\}. \tag{2.13}$$

This inequality always holds if $pm \le n$, namely, $n \le p^2/(p-1)$.

Although $H(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ consists of many terms, in the case of $N_1 = N_2$, $H(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ can be simplified as

$$
\begin{aligned}
H(\boldsymbol{\delta}, \boldsymbol{\Sigma}) =& -\frac{n}{2m^2}\frac{1}{V_0^{1/2}}\Delta^2 + \frac{n}{2m^3}\frac{\lambda}{V_0^{1/2}}\left[m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta} + (\operatorname{tr}\boldsymbol{\Sigma}^{-1})\Delta^2\right] \\
& - \frac{n^4}{2m^6}\frac{\lambda}{V_0^{3/2}}\Delta^2\left[m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta} + 2\operatorname{tr}\boldsymbol{\Sigma}^{-1}\Delta^2 + 4\frac{n+p}{n}\operatorname{tr}\boldsymbol{\Sigma}^{-1}\right] \\
& + \frac{n^3}{8m^4}\frac{1}{V_0^{3/2}}\Delta^2\Big\{2\frac{4n-m}{m}(\Delta^2 + \frac{4p}{n}) + \Delta^4 + 4\Delta^2 + 8\frac{p}{n}\Big\} \\
& - \frac{n^6}{8m^9}\frac{1}{V_0^{7/2}}(3V_0 - \frac{n^2}{4m^2}\Delta^4)\Big\{(4n+p)\Delta^4 + \frac{8}{n}(pn + (n+p)^2)\Delta^2 \\
& \qquad + \frac{16p}{n^2}(pn + (n+p)^2)\Big\} \\
& + \frac{n^4}{m^5}\frac{1}{V_0^{5/2}}(V_0 - \frac{n^2}{4m^2}\Delta^4)\big\{\Delta^4 + 2\frac{n+p}{n}\big\}. \qquad (2.14)
\end{aligned}
$$

Remark 2.1 In the second order approximation of EPMC in Theorem 2.2, we consider the case that p is a fixed constant and p/n tends to zero. Since

$$
\begin{aligned}
U_0 =& -\frac{\Delta^2}{2} - \frac{p}{2n}\Big\{\Delta^2 + \frac{(N_1 - N_2)n}{N_1N_2}\Big\} + O(n^{-2}), \\
V_0 =& \Delta^2 + \frac{p}{n}\Big\{3\Delta^2 + \frac{Nn}{N_1N_2}\Big\} + O(n^{-2}),
\end{aligned}
$$

it can be shown from Theorem 2.2 that $e_\lambda(2|1) = \overline{e}_{LA}(2|1) + O(n^{-3/2})$, where

$$
\overline{e}_{LA}(2|1) = \Phi(-\frac{\Delta^2}{2}) + \phi(-\frac{\Delta}{2})H_{LA}(\Delta), \qquad (2.15)
$$

where

$$
H_{LA}(\Delta) = \frac{1}{4N_1\Delta}\big\{\frac{\Delta^2}{4} + 3(p-1)\big\} + \frac{1}{4N_2\Delta}\big\{\frac{\Delta^2}{4} - (p-1)\big\} + \frac{1}{4n}(p-1)\Delta.
$$

Since $\overline{e}_{LA}(2|1)$ does not depend on λ, it is seen that $\overline{e}_{LA}(2|1)$ gives the second-order approximation of EPMC for LDA. In fact, this is identical to the expansion derived by Okamoto (1963) who treated the case that $\hat{\lambda} = 0$, and n tends to infinity, but p is bounded. The above expression reveals that the effect of the ridge function in $\widehat{\boldsymbol{\Sigma}}_\lambda$ does not appear in the second order approximation when $n \to \infty$, but p is bounded.

3 Second order Unbiased Estimator of EPMC

We now provide a second order unbiased estimator of EPMC using the second order expansion.

Since the second order expansion given in Theorem 2.2 is a function of Δ^2, $\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}/p$ and $\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}$, we begin by obtaining their consistent estimators. Define $\widehat{\Delta}^2$ and $\widehat{\Delta}^2_{\boldsymbol{\Sigma}}$ by

$$\begin{aligned}\widehat{\Delta}^2 =&\frac{m}{n}\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\widehat{\boldsymbol{\delta}} - \frac{Np}{N_1N_2},\\ \widehat{\Delta}^2_\Sigma =&\frac{m^2}{n^2}\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-2}\widehat{\boldsymbol{\delta}} - \Big(\widehat{\Delta}^2 + \frac{Nn}{N_1N_2}\Big)\frac{\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}}{n},\end{aligned} \tag{3.1}$$

where $\widehat{\boldsymbol{\delta}} = \overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2$. Then, we get the following lemma which will be shown in the appendix.

Lemma 3.1 *Assume the conditions (A1)-(A3). Then, the estimator $\widehat{\Delta}^2$ is expanded as $\widehat{\Delta}^2 = \Delta^2 + D_1 + D_2 + O_p(n^{-3/2})$ with $D_1 = O_p(n^{-1/2})$ and $D_2 = O_p(n^{-1})$, where*

$$\begin{aligned}D_1 =&(m\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \Delta^2) + \frac{N}{N_1N_2}(m\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - p) + 2\frac{N^{1/2}}{(N_1N_2)^{1/2}}\frac{m}{n}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2,\\ D_2 =&-m\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta} - \frac{Nm}{N_1N_2}\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-2}\boldsymbol{z}_2.\end{aligned}$$

Also, the estimators $(m/n)\mathrm{tr}\,\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}/p$, $\widehat{\Delta}^2_\Sigma$ and $\hat{\lambda}$ are consistent for $\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}/p$, $\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}$ and λ, respectively.

We consider to substitute the consistent estimators given in Lemma 3.1 into the expansion in Theorem 2.2. Let $\widehat{U}_0 = -2^{-1}(n/m)\{\widehat{\Delta}^2 + p(N_2^{-1} - N_1^{-1})\}$ and $\widehat{V}_0 = (n^3/m^3)\{\widehat{\Delta}^2 + p(N_1^{-1} + N_2^{-1})\}$. For the second order term $\phi(U_0V_0^{-1/2})H(\boldsymbol{\delta},\boldsymbol{\Sigma})$, it is sufficient to substitute the consistent estimators into the second order term since $H(\boldsymbol{\delta},\boldsymbol{\Sigma}) = O(n^{-1})$. For the term $\Phi(U_0V_0^{-1/2})$, however, the estimator $\Phi(\widehat{U}_0\widehat{V}_0^{-1/2})$ is not a second order unbiased estimator of $\Phi(U_0V_0^{-1/2})$, since $\Phi(U_0V_0^{-1/2}) = O(1)$.

Since $\widehat{\Delta}^2 = \Delta^2 + D_1 + D_2 + O_p(n^{-3/2})$, it is noted that

$$\begin{aligned}\widehat{U}_0 =&U_0 + c_1(D_1 + D_2) + O_p(n^{-3/2}),\\ \widehat{V}_0 =&V_0 + c_2(D_1 + D_2) + O_p(n^{-3/2}),\end{aligned}$$

for $c_1 = -n/(2m)$ and $c_2 = n^3/m^3$. From (2.5) and (2.6), it follows that

$$\begin{aligned}\widehat{U}_0\widehat{V}_0^{-1/2} =&U_0V_0^{-1/2} + V_0^{-1/2}(c_1D_1 - \frac{U_0}{V_0}c_2D_1)\\ &+ V_0^{-1/2}\{c_1D_2 - \frac{U_0}{2V_0}c_2D_2 + \frac{3U_0}{8V_0^2}c_2^2D_1^2 - \frac{1}{2V_0}c_1c_2D_1^2\} + O_p(n^{-3/2}).\end{aligned}$$

Thus from (2.7), it is observed that

$$E[\Phi(\widehat{U}_0\widehat{V}_0^{-1/2}) = \Phi(U_0V_0^{-1/2}) + \phi(U_0V_0^{-1/2})K(\boldsymbol{\delta},\boldsymbol{\Sigma}) + O(n^{-3/2}), \tag{3.2}$$

where

$$
\begin{aligned}
K(\boldsymbol{\delta},\boldsymbol{\Sigma}) =& \frac{1}{V_0^{1/2}}(c_1 - \frac{U_0}{2V_0}c_2)E[D_1] + \frac{1}{V_0^{1/2}}(c_1 - \frac{U_0}{2V_0}c_2)E[D_2] \\
&+ \frac{1}{2V_0^{3/2}}(\frac{3U_0}{4V_0}c_2^2 - c_1c_2)E[D_1^2] - \frac{U_0}{2V_0^{3/2}}(c_1 - \frac{U_0}{2V_0}c_2)^2E[D_1^2],
\end{aligned}
$$

which can be rewritten as

$$
\begin{aligned}
K(\boldsymbol{\delta},\boldsymbol{\Sigma}) =& \frac{1}{V_0^{1/2}}c_1E[D_1 + D_2] - \frac{U_0}{2V_0^{3/2}}c_2E[D_1 + D_2] \\
&- \frac{U_0}{2V_0^{3/2}}c_1^2E[D_1^2] + \frac{U_0}{8V_0^{5/2}}(3 - \frac{U_0^2}{V_0})c_2^2E[D_1^2] - \frac{1}{2V_0^{3/2}}(1 - \frac{U_0^2}{V_0})c_1c_2E[D_1^2].
\end{aligned}
$$

Combining (2.9) and (3.2), we can see that the approximation of EPMC is expressed as

$$
\begin{aligned}
e_\lambda(2|1) =& \Phi(U_0V_0^{-1/2}) + \phi(U_0V_0^{-1/2})H(\boldsymbol{\delta},\boldsymbol{\Sigma}) + O(n^{-3/2}) \\
=& E[\Phi(\widehat{U}_0\widehat{V}_0^{-1/2})] + \phi(U_0V_0^{-1/2})\{H(\boldsymbol{\delta},\boldsymbol{\Sigma}) - K(\boldsymbol{\delta},\boldsymbol{\Sigma})\} + O(n^{-3/2}), \qquad (3.3)
\end{aligned}
$$

where

$$
\begin{aligned}
&H(\boldsymbol{\delta},\boldsymbol{\Sigma}) - K(\boldsymbol{\delta},\boldsymbol{\Sigma}) \\
=& \frac{1}{V_0^{1/2}}\{H_U(\boldsymbol{\delta},\boldsymbol{\Sigma}) - c_1E[D_1 + D_2]\} - \frac{U_0}{2V_0^{3/2}}\{H_V(\boldsymbol{\delta},\boldsymbol{\Sigma}) - c_2E[D_1 + D_2]\} \\
&- \frac{U_0}{2V_0^{3/2}}\{H_1(\Delta^2) - c_1^2E[D_1^2]\} + \frac{U_0}{8V_0^{5/2}}\Big(3 - \frac{U_0^2}{V_0}\Big)\{H_2(\Delta^2) - c_2^2E[D_1^2]\} \\
&- \frac{1}{2V_0^{3/2}}\Big(1 - \frac{U_0^2}{V_0}\Big)\{H_{12}(\Delta^2) - c_1c_2E[D_1^2]\}. \qquad (3.4)
\end{aligned}
$$

Second order approximations of the moments given above are given in the following lemma which will be shown in the appendix.

Lemma 3.2 *The moments $E[D_1]$, $E[D_2]$ and $E[D_1^2]$ are approximated as*

$$
\begin{aligned}
E[D_1] =& \frac{\Delta^2}{m} + \frac{Np}{N_1N_2m} + O(n^{-2}), \\
E[D_2] =& -\frac{\lambda}{m^2}(m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta} - \operatorname{tr}\boldsymbol{\Sigma}^{-1}\Delta^2) + \lambda\frac{N}{N_1N_2}\frac{n}{m^2}\operatorname{tr}\boldsymbol{\Sigma}^{-1} + O(n^{-2}), \\
E[D_1^2] =& \frac{2}{m}\Delta^4 + 4\frac{N}{N_1N_2}\frac{n+p}{m}\Delta^2 + 2\frac{N^2}{N_1^2N_2^2}\frac{np}{m} + O(n^{-2}).
\end{aligned}
$$

Using this lemma, we can show that

$$\begin{aligned}
H_U(\boldsymbol{\delta}, \boldsymbol{\Sigma}) - c_1 E[D_1 + D_2] =& \frac{n}{N_1 m^2}\Big(p + \lambda \frac{n}{m} \mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\Big), \\
H_V(\boldsymbol{\delta}, \boldsymbol{\Sigma}) - c_2 E[D_1 + D_2] =& \frac{n^2}{m^4}(3n - m)(\Delta^2 + \frac{Np}{N_1 N_2}) - \lambda \frac{n^3}{m^5}\Big\{ m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta} \\
&+ 3(\mathrm{tr}\, \boldsymbol{\Sigma}^{-1})\Delta^2 + \frac{N}{N_1 N_2}(n + 2p)\mathrm{tr}\, \boldsymbol{\Sigma}^{-1}\Big\}, \\
H_1(\Delta^2) - c_1^2 E[D_1^2] =& - \frac{pn^2}{N_1 N_2 m^3}\{2n + (N_1 + 3N_2)\Delta^2\}, \\
H_2(\Delta^2) - c_2^2 E[D_1^2] =& 2\frac{n^5}{m^7}\Big\{(3n + p)\Delta^4 + 2\frac{Np}{N_1 N_2}(2n + p)\Delta^2 + \frac{N^2 p^2}{N_1^2 N_2^2}(3n + p)\Big\}, \\
H_{12}(\Delta^2) - c_1 c_2 E[D_1^2] =& \frac{n^4}{m^5}\Big\{-\Delta^4 + 2\frac{n + p}{N_1}\Delta^2 + \frac{Np}{N_1^2 N_2^2}\{2nN_2 + (N_2 - N_1)p\}\Big\}.
\end{aligned}$$

For notational simplicity, let $\widetilde{H}_1(\Delta^2) = H_1(\Delta^2) - c_1^2 E[D_1^2]$, $\widetilde{H}_2(\Delta^2) = H_2(\Delta^2) - c_2^2 E[D_1^2]$ and $\widetilde{H}_{12}(\Delta^2) = H_{12}(\Delta^2) - c_1 c_2 E[D_1^2]$. Substituting the consistent estimators given in Lemma 3.1 into the above functions, we get the estimators

$$\begin{aligned}
\ell_U =& \frac{n}{N_1 m^2}\{p + \hat{\lambda} \mathrm{tr}\, \widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\}, \\
\ell_V =& \frac{n^2}{m^4}(3n - m)(\widehat{\Delta}^2 + \frac{Np}{N_1 N_2}) - \hat{\lambda}\frac{n^2}{m^4}\Big\{\frac{m^2}{n}\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-2}\widehat{\boldsymbol{\delta}} + 2(\widehat{\Delta}^2 + \frac{pN}{N_1 N_2}\mathrm{tr}\, \widehat{\boldsymbol{\Sigma}}_\lambda^{-1})\Big\},
\end{aligned}$$

$\ell_1 = \widetilde{H}_1(\widehat{\Delta}^2)$, $\ell_2 = \widetilde{H}_2(\widehat{\Delta}^2)$ and $\ell_{12} = \widetilde{H}_{12}(\widehat{\Delta}^2)$. Then a second order unbiased estimator of EPMC is given by

$$\hat{e}_\lambda(2|1) = \Phi(\widehat{U}_0 \widehat{V}_0^{-1/2}) + \phi(\widehat{U}_0 \widehat{V}_0^{-1/2})\ell(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\Sigma}}_\lambda), \tag{3.5}$$

where

$$\ell(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\Sigma}}_\lambda) = \frac{1}{\widehat{V}_0^{1/2}}\Big\{\ell_U - \frac{\widehat{U}_0}{2\widehat{V}_0}\ell_V - \frac{\widehat{U}_0}{2\widehat{V}_0}\ell_1 + \frac{\widehat{U}_0}{8\widehat{V}_0^2}\Big(3 - \frac{\widehat{U}_0^2}{\widehat{V}_0}\Big)\ell_2 - \frac{1}{2\widehat{V}_0}\Big(1 - \frac{\widehat{U}_0^2}{\widehat{V}_0}\Big)\ell_{12}\Big\}, \tag{3.6}$$

for $\widehat{U}_0 = -2^{-1}(n/m)\{\widehat{\Delta}^2 + p(N_2^{-1} - N_1^{-1})\}$ and $\widehat{V}_0 = (n^3/m^3)\{\widehat{\Delta}^2 + p(N_1^{-1} + N_2^{-1})\}$. Combining (3.3) and (3.5), we get the following theorem.

Theorem 3.1 *Assume the conditions (A1)-(A3). Then,* $E[\hat{e}_\lambda(2|1)] = e_\lambda(2|1) + O(n^{-3/2})$, *namely,* $\hat{e}_\lambda(2|1)$ *is a second order unbiased estimator of* $e_\lambda(2|1)$.

4 Simulation Studies

We now investigate numerical performances of RLDA W_λ, the second-order approximation $\overline{e}_\lambda(2|1)$ and the second-order unbiased estimator $\hat{e}_\lambda(2|1)$ by simulation.

We first investigate the accuracy of asymptotic approximations of EPMC for LDA and RLDA. The EPMC and the approximations are calculated by simulation with 100,000 replications, where in each step, the data sets are generated as $\boldsymbol{x}_{ij} \sim \mathcal{N}_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ for $i = 1, 2$ and $j = 1, \ldots, N_i$, where $\boldsymbol{\Sigma}$ is assumed to be the indentity matrix $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ or to have the serial correlation structure $\boldsymbol{\Sigma} = \left(\rho^{|i-j|}\right)$ for $\rho = 0.5$, and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_1 = p^{-1/2}(\Delta, \ldots, \Delta)'$ for $\Delta^2 = 2, 4$ and $\boldsymbol{\mu}_2 = (0, 0, \ldots, 0)'$.

Let $\Phi_0 = \Phi(U_0 V_0^{-1/2})$ be the limiting value of $\overline{e}_0(2|1)$. We compare the true value $e_0(2|1)$, the limiting value Φ_0, the second-order approximation $\overline{e}_0(2|1)$ given in Theorem 2.2 and the large sample approximation $\overline{e}_{LA}(2|1)$ given in (2.15) which was derived in Okamoto (1963). Concerning RLDA, we compare the true value $e_\lambda(2|1)$ with the limiting value $\Phi_0 = \Phi(U_0 V_0^{-1/2})$ and the second-order approximation $\overline{e}_\lambda(2|1)$ given in (2.9). These values by simulation are reported in Table 1 for $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ and in Table 2 for $\boldsymbol{\Sigma} = (0.5^{|i-j|})$, where the values of $\overline{e}_{LA}(2|1)$ are omitted in Tables 2.

In comparison of the approximations for LDA in Table 1 for $\boldsymbol{\Sigma} = \boldsymbol{I}_p$, it is seen that the second-order approximation $\overline{e}_0(2|1)$ is closer to the true value $e_0(2|1)$ than Φ_0 and $\overline{e}_{LA}(2|1)$ in most cases. As pointed out by Problem (II) in Section 1, the large sample approximation $\overline{e}_{LA}(2|1)$ is not good for large p, but the high-dimensional approximation $\overline{e}_0(2|1)$ improves on $\overline{e}_{LA}(2|1)$ in accuracy of the approximation of EPMC. This gives the similar result as in Tonda and Wakaki (2003) who treated another simulation experiment.

For the approximations of EPMC for RLDA, Tables 1 and 2 show that $\overline{e}_\lambda(2|1)$ gives a superior approximation. In the case of $\boldsymbol{\Sigma} = (0.5^{|i-j|})$ treated in Table 2, the approximations for LDA are the same to the case of $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ since LDA is scale-invariant. Although RLDA is not scale-invariant, Table 2 shows that $\overline{e}_\lambda(2|1)$ gives a good approximation and improves on the first approximation Φ_0.

Table 1. Comparison of approximations of EPMC for LDA and RLDA where $\boldsymbol{\Sigma} = \boldsymbol{I}_p,\ p = 10, 50,\ \Delta^2 = 2, 4$

		LDA				**RLDA**		
Δ^2	(p, N_1, N_2)	$e_0(2\|1)$	Φ_0	$\overline{e}_0(2\|1)$	$\overline{e}_{LA}(2\|1)$	$e_\lambda(2\|1)$	Φ_0	$\overline{e}_\lambda(2\|1)$
	(10,20,20)	0.308	0.310	0.308	0.318	0.304	0.310	0.305
2	(10,30,10)	0.265	0.265	0.267	0.270	0.262	0.265	0.264
	(10,10,30)	0.372	0.377	0.372	0.401	0.370	0.377	0.371
	(10,20,20)	0.219	0.221	0.219	0.217	0.215	0.221	0.216
4	(10,30,10)	0.194	0.193	0.195	0.191	0.191	0.193	0.192
	(10,10,30)	0.264	0.267	0.263	0.264	0.262	0.267	0.260
	(50,40,40)	0.387	0.388	0.387	0.445	0.381	0.388	0.380
2	(50,50,30)	0.353	0.356	0.354	0.382	0.345	0.356	0.345
	(50,30,50)	0.425	0.427	0.425	0.525	0.421	0.427	0.421
	(50,40,40)	0.317	0.319	0.317	0.310	0.308	0.319	0.306
4	(50,50,30)	0.293	0.294	0.292	0.276	0.283	0.294	0.280
	(50,30,50)	0.349	0.349	0.347	0.354	0.341	0.349	0.337

Table 2. Comparison of approximations of EPMC for LDA and RLDA where $\boldsymbol{\Sigma} = (0.5^{|i-j|}),\ p = 10, 50,\ \Delta^2 = 2, 4$

Δ^2	(p, N_1, N_2)	LDA $e_0(2\|1)$	LDA Φ_0	LDA $\overline{e}_0(2\|1)$	RLDA $e_\lambda(2\|1)$	RLDA Φ_0	RLDA $\overline{e}_\lambda(2\|1)$
	(10,20,20)	0.308	0.310	0.308	0.305	0.310	0.305
2	(10,30,10)	0.265	0.265	0.267	0.262	0.265	0.263
	(10,10,30)	0.372	0.377	0.372	0.367	0.377	0.370
	(10,20,20)	0.219	0.221	0.219	0.216	0.221	0.215
4	(10,30,10)	0.194	0.193	0.195	0.191	0.193	0.190
	(10,10,30)	0.264	0.267	0.263	0.263	0.267	0.258
	(50,40,40)	0.387	0.388	0.387	0.378	0.388	0.375
2	(50,50,30)	0.353	0.356	0.354	0.342	0.356	0.339
	(50,30,50)	0.425	0.427	0.425	0.418	0.427	0.418
	(50,40,40)	0.317	0.319	0.317	0.303	0.319	0.299
4	(50,50,30)	0.293	0.294	0.292	0.274	0.294	0.272
	(50,30,50)	0.349	0.349	0.347	0.335	0.349	0.332

Comparing the values of $e_0(2|1)$ and $e_\lambda(2|1)$ in Tables 1 an 2, we see that $e_\lambda(2|1) < e_0(2|1)$, namely RLDA improves on LDA in the sense of minimizing EPMC. Table 3 gives those values in the cases of smaller N_1 and N_2, and shows that the improvement is more significant when n is closer to p for $n = N_1 + N_2 - 2$. In light of Problem (I) raised in Section 1, RLDA is more useful than LDA in high dimension.

Table 3. Comparison of EPMC $e_0(2|1)$ and $e_\lambda(2|1)$ of LDA and RLDA for $\boldsymbol{\Sigma} = \boldsymbol{I}_p,\ p = 10, 50,\ \Delta^2 = 2, 4$

Δ^2	$p=10$: (p, N_1, N_2)	$e_0(2\|1)$	$e_\lambda(2\|1)$	$p=50$: (p, N_1, N_2)	$e_0(2\|1)$	$e_\lambda(2\|1)$
	(10,8,8)	0.388	0.366	(50,28,28)	0.443	0.410
2	(10,9,9)	0.376	0.361	(50,29,29)	0.438	0.410
	(10,10,10)	0.367	0.359	(50,30,30)	0.430	0.407
	(10,8,8)	0.319	0.286	(50,28,28)	0.408	0.352
4	(10,9,9)	0.302	0.278	(50,29,29)	0.396	0.350
	(10,10,10)	0.286	0.268	(50,30,30)	0.384	0.347

We next investigate the bias and mean squared error (MSE) of the second-order unbiased estimator $\hat{e}_\lambda(2|1)$. For comparison, we consider the leave-one-out cross-validation method (CV), which is a popular method for estimating prediction errors for small samples. Set for $j = 1, \ldots, N_1$

$$X_1^{(j)} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_{j+1}, \cdots, \boldsymbol{x}_{N_1}).$$

The set $X_1^{(j)}$ represents the leave-one-out learning set, which is the collection of data with observation $\boldsymbol{x}_i$ removed. It calculates the rate of misclassification when predicting for each specimen using a learning set containing all other observations in the sample. We define the discriminant function using the learning set by

$$W_\lambda^{(j)} = (\overline{\boldsymbol{x}}_1^{(j)} - \overline{\boldsymbol{x}}_2)' \widehat{\boldsymbol{\Sigma}}_\lambda^{(j)^{-1}} \left\{ \boldsymbol{x}_j - \frac{1}{2}(\overline{\boldsymbol{x}}_1^{(j)} + \overline{\boldsymbol{x}}_2) \right\},$$

where $\overline{\boldsymbol{x}}_1^{(j)}$ and $\widehat{\boldsymbol{\Sigma}}_\lambda^{(j)^{-1}}$ are calculated like procedures given around (1.1) based on the learning set $X_1^{(j)}$. Then the CV estimator of $e_\lambda(2|1)$ is given by

$$CV_\lambda(2|1) = \frac{1}{N_1} \sum_{j=1}^{N_1} I_{\{W_\lambda^{(j)} > 0\}}(W_\lambda^{(j)}),$$

where the function $I_A(x)$ is the indicator function defined as

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

The biases and MSEs of the estimators $CV_\lambda(2|1)$ and $\hat{e}_\lambda(2|1)$ are given in Table 4 for $\boldsymbol{\Sigma} = \boldsymbol{I}_p$ and in Table 5 for $\boldsymbol{\Sigma} = (0.5^{|i-j|})$. These tables show that $\hat{e}_\lambda(2|1)$ has smaller MSEs than $CV_\lambda(2|1)$, while $CV_\lambda(2|1)$ has smaller biases than $\hat{e}_\lambda(2|1)$. Since biases and MSEs of both estimators are small, it seems that they are good estimators for EPMC $e_\lambda(2|1)$.

Table 4. Comparison of bias and MSE of $CV_\lambda(2|1)$ and $\hat{e}_\lambda(2|1)$ for $\boldsymbol{\Sigma} = \boldsymbol{I}_p,\ p = 10, 50,\ \Delta^2 = 2, 4$

		$CV_\lambda(2\|1)$		$\hat{e}_\lambda(2\|1)$	
Δ^2	(p, N_1, N_2)	Bias	MSE	Bias	MSE
	(10,20,20)	0.000	0.009	0.003	0.005
2	(10,30,10)	0.001	0.019	0.005	0.005
	(10,10,30)	0.008	0.033	0.017	0.012
	(10,20,20)	0.003	0.007	0.004	0.004
4	(10,30,10)	-0.007	0.011	0.001	0.004
	(10,10,30)	0.007	0.022	0.007	0.007
	(50,40,40)	-0.002	0.006	-0.020	0.003
2	(50,50,30)	-0.005	0.009	0.009	0.004
	(50,30,50)	0.004	0.014	0.016	0.006
	(50,40,40)	-0.004	0.006	-0.018	0.002
4	(50,50,30)	-0.008	0.008	-0.004	0.004
	(50,30,50)	-0.003	0.011	0.014	0.006

Table 5. Comparison of bias and MSE of $CV_\lambda(2|1)$ and $\hat{e}_\lambda(2|1)$ for $\boldsymbol{\Sigma} = (0.5^{|i-j|}),\ p = 10, 50,\ \Delta^2 = 2, 4$

		$CV_\lambda(2\|1)$		$\hat{e}_\lambda(2\|1)$	
Δ^2	(p, N_1, N_2)	Bias	MSE	Bias	MSE
	(10,20,20)	0.003	0.009	0.006	0.005
2	(10,30,10)	0.004	0.019	0.008	0.005
	(10,10,30)	0.014	0.033	0.023	0.012
	(10,20,20)	0.006	0.007	0.007	0.004
4	(10,30,10)	-0.004	0.011	0.004	0.004
	(10,10,30)	0.008	0.022	0.008	0.007
	(50,40,40)	0.007	0.006	-0.011	0.003
2	(50,50,30)	0.006	0.009	0.020	0.004
	(50,30,50)	0.011	0.014	0.023	0.006
	(50,40,40)	0.010	0.006	-0.004	0.002
4	(50,50,30)	0.011	0.008	0.015	0.004
	(50,30,50)	0.011	0.011	0.028	0.006

5 Concluding Remarks

In this paper, we have derived the second-order approximation of EPMC of RLDA given by (1.2) under the high-dimensional situation that $(n, p) \to \infty$ and $\lim_{n\to\infty} p/n = \gamma$ for $0 \le \gamma < 1$. We also have obtained the second-order unbiased estimator of EPMC of RLDA. As their by-products, the second-order approximation of EPMC for LDA and its second-order unbiased estimator of EPMC have been provided. The difference between RLDA and LDA in terms of EPMC appears in the second-order term, and the condition for RLDA improving on LDA has been extracted. It is noted that their difference does not appear in the second-order term in the usual large sample theory.

The goodness of the asymptotic approximation of EPMC for RLDA and LDA and the bias and mean squared error of the second-order unbiased estimator have been investigated by simulation. Through the simulation results, we have confirmed that the second-order approximation and the unbiased estimator are not bad in most cases, and that the asymptotic approximation in high dimension is better than the large sample approximation given by Okamoto (1963) for large p. We also have checked that RLDA improves on LDA in terms of EPMC for larger p. These tell us answers for the problems (I) and (II) raised in Section 1.

Acknowledgments.

Research of the first author was supported in part by Grant-in-Aid for Scientific Research (21540114 and 23243039) from Japan Society for the Promotion of Science. The research of the third author was supported by NSERC.

References

[1] Bai, Z.D., and Yin, Y.Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Prob.*, **21**, 1275-1294.

[2] Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Ann. Eugenics*, **7**, 179-188.

[3] Fujikoshi, Y. (2000). Error bounds for asymptotic approximations of the linear discriminant function when the sample sizes and dimensionality are large. *J. Multivariate Anal.*, **73**, 1-17.

[4] Fujikoshi, Y. (2004). Multivariate analysis for the case when the dimension is large compared to the sample size. *J. Korean Statist. Soc.*, **33**, 1-24.

[5] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant function when the sample size and dimension are large. *Random Oper. Stochastic Equations*, **6**, 260-280.

[6] Haff, L.R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.*, **9**, 531-542.

[7] Haff, L.R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. *Sankhya*, **44**, Series B, 245-258.

[8] Hyodo, M., and Yamada, T. (2010). Asymptotic properties of the EPMC for modified linear discriminant analysis when sample size and dimension are both large. *J. Statist. Plann. Inf.*, **140**, 2739-2748.

[9] Kubokawa, T. and Srivastava, M.S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *J. Multivariate Analysis*, **99**, 1906-1928.

[10] Kubokawa, T., and Srivastava, M.S. (2011). Selection of variables in multivariate regression models for large dimensions. *Commun. Statist.-Theory and Methods*, to appear

[11] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist.*, **34**, 1286-1301.

[12] Letac, G., and Massam, H. (2004). All invariant moments of the Wishart distribution. *Scand. J. Statist.*, **31**, 295-318.

[13] Saranadasa, H. (1993). Asymptotic expansion of the misclassification probabilities of D- and A-criteria for discrimination from the two high dimensional populations using the theory of large dimensional metrices. *J. Multivariate Anal.*, **46**, 154-174.

[14] Siotani, M. (1982). Large sample approximations and asymptotic expansions of classification statistics. *Classification, pattern recognition and reduction of dimensionality*, Handbook of Statistics, vol.2, North-Holland, Amsterdam, 61-100.

[15] Srivastava, M.S. (2005). Some test concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.*, **35**, 251-272.

[16] Srivastava, M.S. (2006). Minimum distance classification rules for high dimensional data. *J. Multivariate Analysis*, **97**, 2057-2070.

[17] Srivastava, M.S. (2007). Multivariate theory for analyzing high dimensional data. *J. Japan Statist. Soc.*, **37**, 53-86.

[18] Srivastava, M.S., and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics.* North-Holland, New York.

[19] Srivastava, M.S. and Kubokawa, T. (2007). Comparison of discrimination methods for high dimensional data. *J. Japan Statist. Soc.*, **37**, 123-134.

[20] Srivastava, M.S. and Kubokawa, T. (2008). Akaike information criterion for selecting components of the mean vector in high dimensional data with fewer observations. *J. Japan Statist. Soc.*, **38**, 259-283.

[21] Stein, C. (1977). Lectures on multivariate estimation theory. (In Russian.) In *Investigation on Statistical Estimation Theory* I, 4-65. Zapiski Nauchych Seminarov LOMI im. V.A. Steklova AN SSSR vol. 74, Leningrad.(In Russian)

[22] Tonda, T., and Wakaki, H. (2003). EPMC estimation in discriminant analysis when the dimension and sample sizes are large. TR No. 03-08, Statistical Research Group, Hiroshima University.

[23] von Rosen, D. (1988). Moments for the inverted Wishart distribution. *Scand. J. Statist.*, **15**, 97-109.

[24] Wald, A. (1944). On the statistical problem arising in the classification of an individual into one of two groups. *Ann. Math. Statist.*, **15**, 145-162.

[25] Xu, P., Brock, G.N., and Parrish, R.S. (2009). Modified linear discrimanant analysis approaches for classification of high-dimensional microarray data. *Comput. Statist. Data Analy.*, **53**, 1674-1687.

A Appendix

A.1 Higher order moments of the inverted Wishart matrix

Let $\boldsymbol{W} = (w_{ij})$ be a $p \times p$ random matrix having the Wishart distribution $\mathcal{W}_p(\boldsymbol{I}, n)$. Let $\boldsymbol{A} = (A_{ij})$ and $\boldsymbol{B} = (B_{ij})$ be $p \times p$ symmetric matrices. To evaluate the second order terms in the expansion of EPMC, we need the moments $E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}]$, $E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})]$, $E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$, $E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})]$, $E[\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$ and $E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})]$. The higher order moments of the inverted Wishart matrix have been derived by Haff (1982), von Rosen (1988), Letac and Massam (2004) and others. However, we cannot find any results for some of the above moments in the literature. For example, it is hard to calculate the expectations $E[\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$ and $E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})]$. We here use the Stein-Haff identity to calculate those moments systematically. Let $\boldsymbol{G}(\boldsymbol{W})$ be a $p \times p$ matrix such that the (i,j) element $g_{ij}(\boldsymbol{W})$ is a function of $\boldsymbol{W}$ and denote

$$\{\mathcal{D}\boldsymbol{G}(\boldsymbol{W})\}_{ij} = \sum_a d_{ia} g_{aj}(\boldsymbol{W}),$$

where for $\boldsymbol{W} = (w_{ij})$,

$$d_{ia} = \frac{1}{2}(1+\delta_{ia})\frac{\partial}{\partial w_{ia}},$$

with $\delta_{ia} = 1$ for $i = a$ and $\delta_{ia} = 0$ for $i \neq a$. Then Stein (1977) and Haff (1979) derived the following identity:

$$E\left[\mathrm{tr}\, \boldsymbol{G}(\boldsymbol{W})\right] = E\left[(m-1)\mathrm{tr}\, \boldsymbol{G}(\boldsymbol{W})\boldsymbol{W}^{-1} + 2\mathrm{tr}\, \mathcal{D}\boldsymbol{G}(\boldsymbol{W})\right], \quad \text{for} \quad m = n - p. \tag{A.1}$$

For differentiating $\boldsymbol{W}^{-1}$, we use the equation

$$\frac{\partial}{\partial w_{ij}}\boldsymbol{W}^{-1} = -\boldsymbol{W}^{-1}(\boldsymbol{e}_i\boldsymbol{e}_j' + \boldsymbol{e}_j\boldsymbol{e}_i')\boldsymbol{W}^{-1}/(1+\delta_{ij}),$$

where $\boldsymbol{e}_i$ is a $p \times 1$ vector such that the i-th element is one and the others are zero. Denote the (i,j) element of $\boldsymbol{W}^{-1}$ by w^{ij}. Then,

$$\begin{aligned} d_{ij}w^{ab} =& \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial w_{ij}}\boldsymbol{e}_a'\boldsymbol{W}^{-1}\boldsymbol{e}_b = -\frac{1}{2}\boldsymbol{e}_a'\boldsymbol{W}^{-1}(\boldsymbol{e}_i\boldsymbol{e}_j' + \boldsymbol{e}_j\boldsymbol{e}_i')\boldsymbol{W}^{-1}\boldsymbol{e}_b \\ =& -\frac{1}{2}(w^{ai}w^{jb} + w^{aj}w^{ib}). \end{aligned} \tag{A.2}$$

Also,

$$\begin{aligned} d_{ij}\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A} =& \frac{1}{2}(1+\delta_{ij})\frac{\partial}{\partial w_{ij}}\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A} = -\frac{1}{2}\mathrm{tr}\, \boldsymbol{W}^{-1}(\boldsymbol{e}_i\boldsymbol{e}_j' + \boldsymbol{e}_j\boldsymbol{e}_i')\boldsymbol{W}^{-1}\boldsymbol{A} \\ =& -\frac{1}{2}\{(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})_{ji} + (\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})_{ij}\}. \end{aligned} \tag{A.3}$$

To explain instructively how to calculate higher order terms from lower order terms, we begin with calculating the second order moments $E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}]$ and $E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})]$. Letting $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{W}) = \boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}$, from (A.2), we can see that

$$\begin{aligned} (\mathcal{D}\boldsymbol{G})_{ii} =& \sum_{a,b,c} d_{ia}A_{ab}w^{bc}B_{ci} = -\frac{1}{2}\sum_{a,b,c} A_{ab}(w^{bi}w^{ac} + w^{ba}w^{ic})B_{ci} \\ =& -\frac{1}{2}\{(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B})_{ii} + (\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\boldsymbol{W}^{1}\boldsymbol{B})_{ii}\}, \end{aligned} \tag{A.4}$$

which implies that

$$\mathrm{tr}\, \mathcal{D}\boldsymbol{G} = -\frac{1}{2}\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B} - \frac{1}{2}(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B}).$$

Then, the Stein-Haff identity (A.1) gives that

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] =& E[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}\boldsymbol{W}^{-1} - \mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B} - (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] \\
=&(m-2)E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] - E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})]. \qquad \text{(A.5)}
\end{aligned}
$$

Letting $\boldsymbol{G} = (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})\boldsymbol{B}$, from (A.3), we observe that

$$
\begin{aligned}
(\mathcal{D}\boldsymbol{G})_{ii} =& \sum_a d_{ia}B_{ai}\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A} = -\frac{1}{2}\sum_a B_{ai}\{(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})_{ai} + (\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})_{ia}\} \\
=& -(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B})_{ii}, \qquad \text{(A.6)}
\end{aligned}
$$

which implies that

$$\mathrm{tr}\,\mathcal{D}\boldsymbol{G} = -\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}.$$

Then, the Stein-Haff identity (A.1) gives that

$$E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})] = (m-1)E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] - 2E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] \quad \text{(A.7)}$$

Since $E[\boldsymbol{W}^{-1}] = (m-1)^{-1}\boldsymbol{I}$, the equations (A.5) and (A.7) provide the simultaneous equations

$$
\begin{cases}
(m-2)E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] - E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] = (m-1)^{-1}\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B}, \\
-2E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] + (m-1)E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] = \dfrac{(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})}{m-1}.
\end{cases} \quad \text{(A.8)}
$$

The solutions of the simultaneous equations can be easily obtained, and we get the following proposition.

Proposition A.1 *Assume that $m = n - p > 3$. Let $\alpha_2 = [m(m-1)(m-3)]^{-1}$. Then,*

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}] =& \alpha_2[(m-1)\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})], \qquad \text{(A.9)} \\
E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})] =& \alpha_2[2\mathrm{tr}\,\boldsymbol{A}\boldsymbol{B} + (m-2)(\mathrm{tr}\,\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{B})]. \qquad \text{(A.10)}
\end{aligned}
$$

Of course, (A.9) can be derived directly from $E[\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}] = [m(m-3)]^{-1}\boldsymbol{A} + [m(m-1)(m-3)]^{-1}\mathrm{tr}\,[\boldsymbol{A}]\boldsymbol{I}$, which appeared in the literature. For example, see Srivastava and Khatri (1979). From Proposition A.1, it follows that

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{W}^{-2}] =& \alpha_2 p(n-1), \\
E[\{\mathrm{tr}\,\boldsymbol{W}^{-1}\}^2] =& \alpha_2 p\{n-1+(p-1)(m-3)\}, \qquad \text{(A.11)} \\
E[(\mathrm{tr}\,\boldsymbol{W}^{-1})\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}] =& \alpha_2\{(m-2)p+2\}\Delta^2,
\end{aligned}
$$

for $\boldsymbol{\xi} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}$ and $\Delta^2 = \boldsymbol{\xi}'\boldsymbol{\xi}$.

The arguments as used above can be employed to evaluate the third order moments $E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$ and $E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})]$.

Proposition A.2 *Assume that $m = n - p > 5$. Let $\alpha_3 = \alpha_2[m(m-1)(m-3)]^{-1} = [(m+1)m(m-1)(m-3)(m-5)]^{-1}$. Then,*

$$E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] = \alpha_3(n-1)[(m-1)\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})], \tag{A.12}$$

$$E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})] = \alpha_3(n-1)[4\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})]. \tag{A.13}$$

Proof. For $\boldsymbol{G} = (\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})\boldsymbol{A}$, the Stein-Haff identity gives that

$$E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{A})] = (m-1)E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})] - 4E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}],$$

where we used the equality $\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B} = \mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B}\boldsymbol{W}^{-2}\boldsymbol{A}$. Exchanging $\boldsymbol{A}$ and $\boldsymbol{B}$ yields

$$E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})] = (m-1)E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})] - 4E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}],$$

so that it is seen that $E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})] = E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})]$, since from Proposition A.1, $E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{A})] = E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})] = \alpha_2(n-1)(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})$. Hence, we get one of the equations given by

$$\begin{aligned} -4E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] + (m-1)E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})] \\ = \alpha_2(n-1)(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B}). \end{aligned} \tag{A.14}$$

For $\boldsymbol{G} = \boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}$, the Stein-Haff identity gives

$$E[\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{B}] = (m-3)E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] - 2E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})],$$

or

$$\begin{aligned} (m-3)E[\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] - 2E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})] \\ = \alpha_2(n-1)(\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B}). \end{aligned} \tag{A.15}$$

Solving the simultaneous equations (A.14) and (A.15) gives the solutions given in (A.12) and (A.13), and the proof is complete. ■

We now derive the fourth order moments $E[\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$ and $E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})]$, which are much harder to evaluate.

Proposition A.3 *Assume that $m = n - p > 7$. Let $\alpha_4 = \alpha_3[(m+2)(m-2)(m-7)]^{-1} = [(m+2)(m+1)m(m-1)(m-2)(m-3)(m-5)(m-7)]^{-1}$. Then,*

$$\begin{aligned} &E[\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] \\ &= \alpha_4(n-1)\Big\{\{(m-1)(n-2)-6\}[(m-1)\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})] \\ &\qquad + (2m+3p-2)[4\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})]\Big\}, \end{aligned} \tag{A.16}$$

$$\begin{aligned} &E[(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{B})] \\ &= \alpha_4(n-1)\Big\{2(2m+3p-2)[(m-1)\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + 2(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})] \\ &\qquad + \{(m-4)(n-1)-6\}[4\mathrm{tr}\, \boldsymbol{A}\boldsymbol{B} + (m-3)(\mathrm{tr}\, \boldsymbol{A})(\mathrm{tr}\, \boldsymbol{B})]\Big\}. \end{aligned} \tag{A.17}$$

Proof. We first note that $E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{B})] = E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{A})]$ and $E[(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] = E[(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{B})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})]$, which can be verified by the same arguments as in the proof of Proposition A.2.

Letting $\boldsymbol{G} = \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}$, $\boldsymbol{G} = \boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1}\boldsymbol{B}\boldsymbol{W}^{-1}$ and $\boldsymbol{G} = (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B})\boldsymbol{I}$, from the Stein-Haff identity, we can get the equations

$$
\begin{aligned}
E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] =& E[(m-3)\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B} - (\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})\\
&- \mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-3}\boldsymbol{B} - (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{B})\\
&- (\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B})],\\
E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] =& E[(m-3)\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-3}\boldsymbol{B} - \mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}\\
&- (\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B}) - 2(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B})],\\
E[p\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}] =& E[(m-1)(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B})\\
&- 2\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B} - 4\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-3}\boldsymbol{B}],
\end{aligned}
$$

respectively. Letting $\boldsymbol{G} = (\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})\boldsymbol{W}^{-1}\boldsymbol{A}$, $\boldsymbol{G} = (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})\boldsymbol{W}^{-1}\boldsymbol{B}\boldsymbol{W}^{-1}$ and $\boldsymbol{G} = (\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})\boldsymbol{I}$, we can also get the equations

$$
\begin{aligned}
E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] =& E[(m-2)(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B}) - 2\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}\\
&- 2\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-3}\boldsymbol{B} - (\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})],\\
E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] =& E[(m-3)(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{B}) - 2\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}\\
&- 2(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})],\\
E[p(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})] =& E[(m-1)(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})]\\
&- 2(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B}) - 4(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{B})],
\end{aligned}
$$

respectively. For simplicity, let $C = E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}]$ and $D = E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})]$. Also let

$$
\begin{aligned}
&a_1 = E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B}],\ a_2 = E[(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})],\ a_3 = E[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-3}\boldsymbol{B}],\\
&a_4 = E[(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-3}\boldsymbol{B})],\ a_5 = E[(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-2}\boldsymbol{B})],\\
&a_6 = E[(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{A})(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{B})].
\end{aligned}
$$

Then, we get the simultaneous equations

$$
\begin{cases}
(m-3)a_1 - a_2 - a_3 - a_4 - a_5 = C,\\
-a_1 - a_2 + (m-3)a_3 - 2a_5 = C,\\
-2a_1 - 4a_3 + (m-1)a_5 = pC,\\
-2a_1 + (m-2)a_2 - 2a_3 - a_6 = D,\\
-2a_1 + (m-3)a_4 - 2a_6 = D,\\
-2a_2 - 4a_4 + (m-1)a_6 = pD.
\end{cases}
\tag{A.18}
$$

Eliminating a_5 and a_6 from the equations (A.18) gives

$$\begin{cases} (2m-5)a_1 - a_2 - (m-1)a_3 - 2a_4 = C, \\ (m^2-4m+1)a_1 - (m-1)a_2 - (m+3)a_3 - (m-1)a_4 = (n-1)C, \\ -2a_1 + 2(m-2)a_2 - 4a_3 - (m-3)a_4 = D, \\ -2(m-1)a_1 + m(m-3)a_2 - 2(m-1)a_3 - 4a_4 = (n-1)D. \end{cases} \tag{A.19}$$

Eliminating a_3 and a_4 from (A.19), we can get

$$\begin{cases} (m-2)(m-4)a_1 - 3(m-2)a_2 = (n-2)C - D, \\ -6(m-2)a_1 + (m-1)(m-2)a_2 = (n-1)D - 2C. \end{cases} \tag{A.20}$$

The values of C and D are given in Proposition A.2. Thus, the values of a_1 and a_2 can be derived from (A.20), and the proof of Proposition A.3 is complete. ■

A.2 Evaluation of the second order term

Using the higher order moments of the inverted Wishart matrix given in the previous section, we can evaluate the expectations $E[U_1]$, $E[U_2]$, $E[V_1]$, $E[V_2]$, $E[U_1^2]$, $E[U_1V_1]$ and $E[V_1^2]$ up to $O(n^{-1})$. This gives a proof of Theorem 2.1. Also, from the fact that $E[U_1^2] = O(n^{-1})$ and $E[V_1^2] = O(n^{-1})$, it follows that $U_1 = O_p(n^{-1/2})$ and $V_1 = O_p(n^{-1/2})$ in Lemmas 2.1 and 2.2.

Before giving the proof, we note the following equality for $p \times p$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, and a random vector $\boldsymbol{y}$ having $\mathcal{N}_p(\boldsymbol{0}, \boldsymbol{I})$:

$$E[\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}] = 2\mathrm{tr}\,[\boldsymbol{A}\boldsymbol{B}] + \mathrm{tr}\,[\boldsymbol{A}]\mathrm{tr}\,[\boldsymbol{B}]. \tag{A.21}$$

[1] Evaluation of $E[U_1]$ and $E[V_1]$. For the expectations $E[U_1]$ and $E[V_1]$, from the results in the previous subsection, it is noted that

$$\begin{aligned} nE[\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta}] =& nE[\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'] = \frac{n}{m}\Delta^2 + \frac{n}{m^2}\Delta^2 + O(n^{-2}), \\ E[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2] =& E[\mathrm{tr}\,\boldsymbol{W}^{-1}] = \frac{p}{m} + \frac{p}{m^2} + O(n^{-2}), \\ n^2E[\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta}] =& n^2E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{\xi}\boldsymbol{\xi}] = \frac{n^3}{m^3}\Delta^2 + \frac{n^2(4n-m)}{m^4}\Delta^2 + O(n^{-2}), \\ nE[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2] =& nE[\mathrm{tr}\,\boldsymbol{W}^{-2}] = \frac{pn^2}{m^3} + \frac{pn(4n-m)}{m^4} + O(n^{-2}), \end{aligned} \tag{A.22}$$

which can be used to get that

$$\begin{aligned} E[U_1] =& -\frac{n}{2m^2}\left\{\Delta^2 + p\frac{N_1-N_2}{N_1N_2}\right\} + O(n^{-3/2}), \\ E[V_1] =& \frac{n^2(4n-m)}{m^4}\left\{\Delta^2 + p\frac{N}{N_1N_2}\right\} + O(n^{-3/2}). \end{aligned} \tag{A.23}$$

[2] Evaluation of $E[U_2]$. For evaluation of $E[U_2]$, it is noted that $\hat{\lambda}-\lambda=O_p(n^{-1})$ for $\lambda=c_n\mathrm{tr}\,\boldsymbol{\Sigma}/p$. Since $n\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}=n\lambda\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}+n(\hat{\lambda}-\lambda)\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}=n\lambda\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}+O_p(n^{-2})$, it is sufficient to get the moment $n\lambda E[\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}]$. For $\boldsymbol{\xi}=\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\delta}$, it is observed that

$$nE[\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}]=nE[\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^{-1}\boldsymbol{\xi}]=n\alpha_2[(m-1)\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})\Delta^2],$$

so that

$$E[n\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta}]=\lambda\Big\{\frac{n}{m^2}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+\frac{n}{m^3}(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}\Big\}+O(n^{-2}). \tag{A.24}$$

Similarly, $E[\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-2}\boldsymbol{z}_2]$ can be approximated as

$$E[\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-2}\boldsymbol{z}_2]=\lambda E[\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{\Sigma}^{-1}]+O(n^{-2})=\frac{n}{m^3}\lambda(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})+O(n^{-2}), \tag{A.25}$$

so that $E[U_2]$ is evaluated as

$$\begin{aligned}E[U_2]=&\frac{\lambda}{2}\Big\{\frac{n}{m^2}\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+\frac{n}{m^3}(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})\Delta^2\Big\}\\&+\frac{\lambda}{2}\frac{N_1-N_2}{N_1N_2}\frac{n^2}{m^3}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}+O(n^{-3/2}).\end{aligned} \tag{A.26}$$

[3] Evaluation of $E[V_2]$. For $E[V_2]$, the same arguments as in the evaluation of $E[U_2]$ can be used, but we need to calculate the third moments of $\boldsymbol{S}^{-1}$. It is seen from (A.12) that

$$\begin{aligned}n^2E[\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{\delta}]=&n^2\lambda E[\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^{-1}\boldsymbol{\xi}]+O(n^{-2})\\=&\lambda\alpha_3n^2(n-1)[(m-1)\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+2\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\Delta^2]+O(n^{-2})\\=&\lambda\frac{n^3}{m^5}(m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+2\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\Delta^2)+O(n^{-2}).\end{aligned}$$

Also,

$$\begin{aligned}nE[\hat{\lambda}\boldsymbol{z}_2\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-2}\boldsymbol{z}_2]=&\lambda nE[\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{W}^{-3}]+O(n^{-2})\\=&\lambda\alpha_3n(n-1)(n+p-1)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}+O(n^{-2})\\=&\lambda\frac{n^2}{m^5}(n+p)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}+O(n^{-2}),\end{aligned}$$

so that $E[V_2]$ is estimated as

$$E[V_2]=-2\lambda\frac{n^3}{m^5}\Big\{m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta}+2(\mathrm{tr}\,\boldsymbol{\Sigma}^{-1})\Delta^2)+\frac{N}{N_1N_2}(n+p)\mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\Big\}+O(n^{-2}). \tag{A.27}$$

[4] Evaluation of $E[U_1^2]$. For the moment $E[U_1^2]$, it can be approximated as

$$\begin{aligned}E[U_1^2]=&\frac{1}{4}E[\{n\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}-\frac{n}{m}\Delta^2\}^2]+\frac{1}{4}\frac{(N_1-N_2)^2n^2}{(N_1N_2)^2}E[\big(\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2-\frac{p}{m}\big)^2]\\&+\frac{N_1-N_2}{2N_1N_2}n(n\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}-(n/m)\Delta^2)(\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2-p/m)\\&+\frac{n}{N_2}\big\{nE[\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi}]+\frac{n}{N_1}E[\mathrm{tr}\,\boldsymbol{W}^{-2}]\big\}+O(n^{-3/2})\\=&\frac{1}{4}I_1+\frac{1}{4}\frac{(N_1-N_2)^2n^2}{(N_1N_2)^2}I_2+\frac{N_1-N_2}{2N_1N_2}nI_3+\frac{n}{N_2}I_4+O(n^{-3/2}),\quad\text{(say)}.\end{aligned}$$

First, I_1 is rewritten as

$$
\begin{aligned}
I_1 =& E[\{n\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} - \frac{n}{m}\Delta^2\}^2] \\
=& n^2 E[\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}] - 2\frac{n^2}{m}\Delta^2 E[\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}] + \frac{n^2}{m^2}\Delta^4 \\
=& n^2\{\alpha_2 m - \frac{2}{m(m-1)} + \frac{1}{m^2}\}\Delta^4 \\
=& 2\frac{n^2}{m^3}\Delta^4 + O(n^{-2}).
\end{aligned}
\tag{A.28}
$$

We next evaluate I_2. Note that

$$
E[(\mathrm{tr}\,\boldsymbol{W}^{-1})^2] = \alpha_2(2 + (m-2)p)p = \frac{p^2}{m^2} + 2\frac{p^2}{m^3} + O(n^{-2}).
$$

Then, from (A.21),

$$
\begin{aligned}
I_2 =& E[(\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - p/m)^2] \\
=& E[(\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2)^2] - 2\frac{p}{m}E[\boldsymbol{z}_2\boldsymbol{S}^{-1}\boldsymbol{z}_2] + \frac{p^2}{m^2} \\
=& E[2\mathrm{tr}\,\boldsymbol{W}^{-2} + (\mathrm{tr}\,\boldsymbol{W}^{-1})^2] - 2\frac{p}{m}E[\mathrm{tr}\,\boldsymbol{W}^{-1}] + \frac{p^2}{m^2} \\
=& \frac{2pn}{m^3} + O(n^{-2}).
\end{aligned}
\tag{A.29}
$$

For I_3, note that

$$
nE[(\mathrm{tr}\,\boldsymbol{W}^{-1})\mathrm{tr}\,\boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'] = \alpha_2 n(2 + p(m-2))\Delta^2 = \frac{np}{m^2}\Delta^2 + 2\frac{np}{m^3}\Delta^2 + O(n^{-2}),
$$

which is used to evaluate I_3 as

$$
\begin{aligned}
I_3 =& nE[(\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} - m^{-1}\Delta^2)(\mathrm{tr}\,\boldsymbol{W}^{-1} - p/m)] \\
=& nE[\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}\mathrm{tr}\,\boldsymbol{W}^{-1}] - 2\frac{np}{m(m-1)}\Delta^2 + \frac{np}{m^2}\Delta^2 \\
=& O(n^{-2}).
\end{aligned}
\tag{A.30}
$$

Using (A.22), we can approximate I_4 as

$$
I_4 = nE[\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi}] + \frac{n}{N_1}E[\mathrm{tr}\,\boldsymbol{W}^{-2}] = \frac{n^2}{m^3}(\Delta^2 + \frac{p}{N_1}) + O(n^{-2}).
$$

Hence, $E[U_1^2]$ is evaluated as

$$
E[U_1^2] = \frac{1}{2}\frac{n^2}{m^3}\Delta^4 + \frac{1}{2}\frac{(N_1 - N_2)^2}{(N_1 N_2)^2}\frac{pn^3}{m^3} + \frac{n^3}{N_2 m^3}(\Delta^2 + \frac{p}{N_1}) + O(n^{-2}). \tag{A.31}
$$

The fact that $E[U_1^2] = O(n^{-1})$ implies that $U_1 = O_p(n^{-1/2})$.

[5] Evaluation of $E[V_1^2]$. We next evaluate the moment $E[V_1^2]$, which is expressed as

$$
\begin{aligned}
E[V_1^2] =& E[\{n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\}^2] + \frac{N^2n^2}{(N_1N_2)^2}E[\{n\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn^2}{m^3}\}^2] \\
&+ 2\frac{Nn}{N_1N_2}E[\{n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\}\{n\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn^2}{m^3}\}] \\
&+ 4\frac{N}{N_1N_2}E[\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta}] \\
=& J_1 + \frac{N^2n^2}{(N_1N_2)^2}J_2 + 2\frac{Nn}{N_1N_2}J_3 + 4\frac{Nn}{N_1N_2}J_4 + O(n^{-3/2}), \quad \text{(say)}.
\end{aligned}
$$

For J_1, it is noted from Propositions A.1 and A.3 that

$$
n^4E[(\mathrm{tr}\,\boldsymbol{W}^{-2}\boldsymbol{\xi}\boldsymbol{\xi}')^2] = \frac{n^6}{m^6}\Delta^4 + 2\frac{n^5(7n+2p)}{m^7}\Delta^4 + O(n^{-2}).
$$

Noting this equality and (A.22), we can approximate J_1 as

$$
\begin{aligned}
J_1 =& E[\{n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\}^2] \\
=& E[n^4(\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi})^2] - 2\frac{n^3}{m^3}\Delta^2E[\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi}] + \frac{n^6}{m^6}\Delta^4 \\
=& 2\frac{n^5(4n+p)}{m^7}\Delta^4 + O(n^{-2}).
\end{aligned}
$$

For J_2, from Proposition A.3, it is noted that

$$
\begin{aligned}
n^2E[\mathrm{tr}\,\boldsymbol{W}^{-4}] =& \frac{pn^3\{np + (n+p)^2\}}{m^7} + O(n^{-2}), \\
n^2E[(\mathrm{tr}\,\boldsymbol{W}^{-2})^2] =& \frac{p^2n^4}{m^6} + 2\frac{p^2n^3(4n-m)}{m^7} + O(n^{-2}), \\
\frac{pn^3}{m^3}E[\mathrm{tr}\,\boldsymbol{W}^{-2}] =& \frac{p^2n^4}{m^6} + \frac{p^2n^3(4n-m)}{m^7} + O(n^{-2}),
\end{aligned}
$$

which are utilized to evaluate J_2 as

$$
\begin{aligned}
J_2 =& n^2E[\{\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\}^2] \\
=& n^2E[2\mathrm{tr}\,\boldsymbol{W}^{-4} + (\mathrm{tr}\,\boldsymbol{W}^{-2})^2 - 2\frac{pn}{m^3}\mathrm{tr}\,\boldsymbol{W}^{-2} + \frac{p^2n^2}{m^6}] \\
=& 2\frac{pn^3\{np + (n+p)^2\}}{m^7} + O(n^{-2}).
\end{aligned}
$$

Similarly, for J_3, it is observed that

$$
\begin{aligned}
J_3 =& nE[\{n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\}\{\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\}] \\
=& nE[n^2(\mathrm{tr}\,\boldsymbol{W}^{-2})(\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi}) - \frac{n^3}{m^3}\Delta^2\mathrm{tr}\,\boldsymbol{W}^{-2} - \frac{pn}{m^3}n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} + \frac{pn^4}{m^6}\Delta^2] \\
=& O(n^{-2}),
\end{aligned}
$$

by noting (A.22) and

$$n^3 E[(\mathrm{tr}\, \boldsymbol{W}^{-2})(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{\xi}\boldsymbol{\xi}')] = \frac{pn^5}{m^6}\Delta^2 + 2\frac{pn^4(4n-m)}{m^7}\Delta^2 + O(n^{-2}).$$

It can be seen from Proposition A.3 that J_4 is approximated as

$$J_4 = n^3 E[\boldsymbol{\xi}'\boldsymbol{W}^{-4}\boldsymbol{\xi}] + O(n^{-2}) = \frac{n^4\{np + (n+p)^2\}}{m^7}\Delta^2 + O(n^{-2}).$$

Combining the above evaluations gives that

$$\begin{aligned} E[V_1^2] =& 2\frac{n^5(4n+p)}{m^7}\Delta^4 + 4\frac{Nn^5}{N_1N_2m^7}\{pn + (n+p)^2\}\Delta^2 \\ &+ 2\frac{N^2n^5p}{(N_1N_2)^2m^7}\{pn + (n+p)^2\} + O(n^{-2}). \end{aligned} \tag{A.32}$$

[6] Evaluation of $E[U_1V_1]$. Finally, we shall evaluate the moment $E[U_1V_1]$, which is expressed as

$$\begin{aligned} E[U_1V_1] =& -\frac{1}{2}E\Big[\big\{n\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n}{m}\Delta^2\big\} \\ &\times \big\{\big[n^2\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n^3}{m^3}\Delta^2\big] + \frac{Nn^2}{N_1N_2}\big[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\big]\big\}\Big] \\ &- \frac{1}{2}\frac{N_1-N_2}{N_1N_2}E\Big[\big\{n\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m}\big\} \\ &\times \big\{\big[n^2\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{\delta} - \frac{n^3}{m^3}\Delta^2\big] + \frac{Nn^2}{N_1N_2}\big[\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m^3}\big]\big\}\Big] \\ &- \frac{2}{N_2}E\big[\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta}\big] \\ =& -\frac{1}{2}K_1 - \frac{1}{2}\frac{(N_1-N_2)n}{N_1N_2}K_2 - 2\frac{n}{N_2}K_3 + O(n^{-3/2}), \quad \text{(say)}. \end{aligned}$$

For K_1, it is noted that

$$\begin{aligned} n^3 E[(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}')(\mathrm{tr}\, \boldsymbol{W}^{-2}\boldsymbol{\xi}\boldsymbol{\xi}')] =& \frac{n^4}{m^4}\Delta^4 + \frac{n^3(9n-m)}{m^5}\Delta^4 + O(n^{-2}), \\ n^2 E[(\mathrm{tr}\, \boldsymbol{W}^{-2})(\mathrm{tr}\, \boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}')] =& \frac{n^3p}{m^4}\Delta^2 + \frac{n^2p(5n-m)}{m^5}\Delta^2 + O(n^{-2}), \end{aligned}$$

which are used to evaluate K_1 as

$$\begin{aligned} K_1 =& E[\big\{n\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} - \frac{n}{m}\Delta^2\big\}\big\{(n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2) + \frac{Nn^2}{N_1N_2}(\mathrm{tr}\, \boldsymbol{W}^{-2} - \frac{pn}{m^3})\big\}] \\ =& E[n^3\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi}\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^4}{m^3}\Delta^2\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} - \frac{n^3}{m}\Delta^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} + \frac{n^4}{m^4}\Delta^4] \\ &+ \frac{Nn^2}{N_1N_2}E[n(\mathrm{tr}\, \boldsymbol{W}^{-2})\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} - \frac{n}{m}\Delta^2\mathrm{tr}\, \boldsymbol{W}^{-2} - \frac{pn^2}{m^3}\boldsymbol{\xi}'\boldsymbol{W}^{-1}\boldsymbol{\xi} + \frac{pn^2}{m^4}\Delta^2] \\ =& 4\frac{n^4}{m^5}\Delta^4 + O(n^{-2}). \end{aligned}$$

For K_2, it is evaluated as

$$
\begin{aligned}
K_2 =& n^{-1}E[\{n\mathrm{tr}\,\boldsymbol{W}^{-1} - \frac{pn}{m}\}\{n^2\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\}] \\
&+ \frac{Nn}{N_1N_2}E[n\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn^2}{m^3}\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - \frac{pn}{m}\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{\Sigma}\boldsymbol{S}^{-1}\boldsymbol{z}_2 + \frac{p^2n^2}{m^4}] \\
=& E[n^2(\mathrm{tr}\,\boldsymbol{W}^{-1})\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} - \frac{n^3}{m^3}\Delta^2\mathrm{tr}\,\boldsymbol{W}^{-1} - \frac{pn^2}{m}\boldsymbol{\xi}'\boldsymbol{W}^{-2}\boldsymbol{\xi} + \frac{pn^3}{m^4}\Delta^2] \\
&+ \frac{Nn}{N_1N_2}E[2n\mathrm{tr}\,\boldsymbol{W}^{-3} + n(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-2}) - \frac{pn^2}{m^3}\mathrm{tr}\,\boldsymbol{W}^{-1} - \frac{pn}{m}\mathrm{tr}\,\boldsymbol{W}^{-2} + \frac{p^2n^2}{m^4}] \\
=& 2\frac{Nn^3p(n+p)}{N_1N_2m^5} + O(n^{-2}),
\end{aligned}
$$

since

$$
nE[(\mathrm{tr}\,\boldsymbol{W}^{-1})(\mathrm{tr}\,\boldsymbol{W}^{-2})] = \frac{n^2p^2}{m^4} + \frac{np^2(5n-m)}{m^5} + O(n^{-2}).
$$

Finally, K_3 is approximated as

$$
K_3 = n^2E[\boldsymbol{\xi}'\boldsymbol{W}^{-3}\boldsymbol{\xi}] + O(n^{-2}) = \frac{n^3(n+p)}{m^5}\Delta^2 + O(n^{-2}).
$$

Combining the values of K_1, K_2 and K_3 gives

$$
E[U_1V_1] = -2\frac{n^4}{m^5}\Delta^4 - 2\frac{n^4(n+p)}{N_2m^5}\Delta^2 - \frac{(N_1^2-N_2^2)n^4p(n+p)}{N_1^2N_2^2m^5} + O(n^{-2}). \tag{A.33}
$$

A.3 Proofs of Lemma 3.1 and 3.2

We here derive the expectations $E[D_1]$, $E[D_2]$ and $E[D_1^2]$. This gives a proof of Lemma 3.2. Also, from the fact that $E[D_1^2] = O(n^{-1})$, it follows that $D_1 = O_p(n^{-1/2})$ in Lemma 3.1.

Since D_1 is given by

$$
D_1 = (m\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \Delta^2) + \frac{N}{N_1N_2}(m\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - p) + 2\frac{N^{1/2}}{(N_1N_2)^{1/2}}\frac{m}{n}\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{z}_2,
$$

it is easily seen that

$$
E[D_1] = \frac{1}{m}\Delta^2 + \frac{N}{N_1N_2}\frac{p}{m} + O(n^{-2}).
$$

For $E[D_2]$, it is noted that

$$
D_2 = -m\hat{\lambda}\boldsymbol{\delta}'\boldsymbol{S}^{-2}\boldsymbol{\delta} - \frac{Nm}{N_1N_2}\hat{\lambda}\boldsymbol{z}_2'\boldsymbol{S}^{-2}\boldsymbol{z}_2.
$$

From (A.24) and (A.25), it follows that

$$
E[D_2] = -\frac{\lambda}{m^2}(m\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-2}\boldsymbol{\delta} + \mathrm{tr}\,\boldsymbol{\Sigma}^{-1}\Delta^2) + \lambda\frac{N}{N_1N_2}\frac{n}{m^2}\mathrm{tr}\,\boldsymbol{\Sigma}^{-1} + O(n^{-2}).
$$

Finally, $E[D_1^2]$ is written as

$$
\begin{aligned}
E[D_1^2] =& E[(m\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \Delta^2)^2] + \frac{N^2}{N_1^2 N_2^2} E[(m\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - p)^2] \\
&+ 4\frac{N}{N_1 N_2}\frac{m^2}{n^2} E[\boldsymbol{\delta}'\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\Sigma}\widehat{\boldsymbol{\Sigma}}_\lambda^{-1}\boldsymbol{\delta}] \\
&+ \frac{2N}{N_1 N_2} E[(m\boldsymbol{\delta}'\boldsymbol{S}^{-1}\boldsymbol{\delta} - \Delta^2)(m\boldsymbol{z}_2'\boldsymbol{S}^{-1}\boldsymbol{z}_2 - p)].
\end{aligned}
$$

Using (A.28), (A.29) and (A.30), we can see that $E[D_1^2]$ is expressed as

$$
E[D_1^2] = \frac{2}{m}\Delta^4 + 4\frac{N}{N_1 N_2}\frac{n+p}{m}\Delta^2 + 2\frac{N^2}{N_1^2 N_2^2}\frac{np}{m} + O(n^{-2}).
$$