HIROSHIMA MATH. J.
42 (2012), 209–231

A modified linear discriminant analysis for high-dimensional data

Masashi HYODO, Takayuki YAMADA, Tetsuto HIMENO and Takashi SEO

(Received June 14, 2010)
(Revised August 2, 2011)

ABSTRACT. We deal with the problem of classifying a new observation vector into one of two known multivariate normal populations. Linear discriminant analysis (LDA) is now widely available. However, for high-dimensional data classification problem, due to the small number of samples and the large number of variables, classical LDA has poor performance corresponding to the singularity and instability of the sample covariance matrix. Recently, Xu et al. [10] suggested modified linear discriminant analysis (MLDA). This method is based on the shrink type estimator of the covariance matrix derived by Ledoit and Wolf [6]. This estimator was proposed under the asymptotic framework $\mathrm{A}_0 : n = O(p)$ and $p = O(n)$ when $p \to \infty$. In this paper, we propose a shrink type estimator under more flexible high-dimensional framework. Using this estimator, we define the new MLDA. Through the numerical simulation, the expected correct classification rate of our MLDA is larger than the ones of other discrimination methods when $p > n$. In addition, we consider the limiting value of the expected probability of misclassification (EPMC) under some assumptions.

1. Introduction

We deal with the problem of classifying a $p \times 1$ observation vector $\boldsymbol{x}$ into one of two groups Π_1 and Π_2. Let the group Π_i have p-dimensional normal distribution $N_p(\boldsymbol{\mu}_i, \Sigma)$ with the mean vector $\boldsymbol{\mu}_i$ and the common positive definite covariance matrix Σ, where $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$. Assume that, for $i = 1, 2$, the random vectors x_{ij}, $j = 1, \dots, N_i$, are taken from Π_i. Linear discriminant analysis (LDA) is one of the standard classical methods for classifying $\boldsymbol{x}$ into either Π_1 or Π_2, which is as follows:

$$W = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{-1} \left\{ \boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \right\} > 0 \ \ (\text{resp.}, \ < 0)$$
$$\Rightarrow \boldsymbol{x} \in \Pi_1 \ \ (\text{resp.}, \ \Pi_2). \tag{1}$$

2010 *Mathematics Subject Classification*. Primary 62H30, 62H12; Secondary 62E20.
Key words and phrases. asymptotic approximations, expected probability of misclassification, linear discriminant function.

Here, S is the pooled sample covariance matrix, which is given by

$$S = n^{-1}\sum_{i=1}^{2}\sum_{j=1}^{N_i}(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{ij} - \bar{\boldsymbol{x}}_i)', \qquad \bar{\boldsymbol{x}}_i = N_i^{-1}\sum_{j=1}^{N_i}\boldsymbol{x}_{ij}, \qquad i = 1,2,$$

$$n = N_1 + N_2 - 2.$$

Recently, needs of discriminating high-dimensional data have increased. LDA is one of the most popular methods of discrimination. However, LDA cannot be used when $p > n$. For, in this case, S becomes singular, and so W cannot be defined. A simple way to adapt LDA to the high dimensional case is to use the Moore-Penrose inverse of S in (1) (MPLDA). Another simple way is to use the diagonal matrix with the same diagonal elements as S in (1) (DLDA). These analyses are applied to microarray data of leukemia in Dudoit et al. [1]. On the other hand, Xu et al. [10] proposed modified linear discriminant analysis (MLDA) as a more efficient method than DLDA and MPLDA. This method is based on the estimator of covariance matrix given by Ledoit and Wolf [6]. The basic idea is to find the linear combination $\Sigma^* = \rho_1 I + \rho_2 S$ such that the expected quadratic loss $E[\|\Sigma^* - \Sigma\|^2]$ becomes minimum. Here, $\|A\| = \sqrt{\mathrm{tr}\, AA'/p}$ is the normalized version of the Frobenious norm. Ledoit and Wolf [6] proved that the solution for $\min_{\rho_1,\rho_2} E[\|\Sigma^* - \Sigma\|^2]$ satisfies

$$\Sigma^* = \frac{\beta}{\delta}a_1 I + \frac{\alpha}{\delta}S,$$

where $\alpha = a_2 - a_1^2$, $\beta = (p/n)\{(a_2/p) + a_1^2\}$, $\delta = \{1 + (1/n)\}a_2 + \{(p/n) - 1\}a_1^2$, $a_i = \mathrm{tr}\,\Sigma^i/p$, $i = 1,\dots,4$, and

$$E[\|\Sigma^* - \Sigma\|^2] = \frac{\alpha\beta}{\delta}.$$

Here, the notation Σ^i denotes i-th power of matrix Σ. In practice, it is necessary to replace the unknown parameters with their consistent estimators. For this reason, Ledoit and Wolf [6] derived the consistent estimators under the asymptotic framework A_0: $n = O(p)$ and $p = O(n)$ when $p \to \infty$. However, it should be noticed that, in many microarray data, n is much smaller than p. Consistent estimators should be proposed under more flexible high-dimensional framework. In this paper, we propose such consistent estimators under the following high-dimensional framework:

$$\mathrm{C1} : n, p \to \infty \ \text{ with } \ p^{\zeta}/n \to c_2 \ \text{ for some } \ c_2 \in (0,\infty) \ \text{ and } \ \zeta \in (1/3, 1].$$

In addition to C1, we assume the following condition:

$$\mathrm{C2} : 0 < \lim_{p \to \infty} a_i < \infty, \qquad i = 1, \dots, 4.$$

Under C1 and C2 the consistent estimators $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\delta}$ of α, β and δ are, respecitvely,

$$\hat{\alpha} = \hat{a}_2 - \hat{a}_1^2, \qquad \hat{\beta} = (p/n)\{(\hat{a}_2/p) + \hat{a}_1^2\},$$
$$\hat{\delta} = \{1 + (1/n)\}\hat{a}_2 + \{(p/n) - 1\}\hat{a}_1^2, \tag{2}$$

where $\hat{a}_1$ and $\hat{a}_2$ are the consistent estimators of a_1 and a_2, respectively, given by

$$\hat{a}_1 = \frac{\mathrm{tr}\, S}{p}, \qquad \hat{a}_2 = \frac{n^2}{(n-1)(n+2)} \left(\frac{\mathrm{tr}\, S^2}{p} - \frac{(\mathrm{tr}\, S)^2}{np} \right).$$

The derivation of these consistent estimators are given in Section 2. Let

$$S^* = \frac{\hat{\beta}}{\hat{\delta}} \hat{a}_1 I + \frac{\hat{\alpha}}{\hat{\delta}} S. \tag{3}$$

We define the adapted version of MLDA as follows:

$$W^* = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' S^{*-1} \left\{ \boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \right\} > 0 \ \ (\text{resp.}, \ < 0)$$
$$\Rightarrow \boldsymbol{x} \in \Pi_1 \ \ (\text{resp.}, \ \Pi_2).$$

Our simulation results, which are omitted in this paper, show that the above adapted version of MLDA has better performance than that used in Xu et al. [10]. Hereafter, we refer the former simply as MLDA. In Section 3, we present the comparison between the performance of MLDA method and those of other LDA methods, based on expected correct classification rate. In Section 4, we consider the expected probability of misclassification (EPMC) of MLDA. However, it is generally difficult to obtain an explicit expression for the EPMC. So we approximate it by its limiting value. The approximation under a framework such that N_1 and N_2 are large and p is fixed has been studied by many authors. For a review of the results, see, e.g., Siotani [9]. See also, e.g., Fujikoshi and Seo [3] for the asymptotic approximation when all of N_1, N_2 and p are large. These results were given when $p < n$, and were based on asymptotic expansions. When $p > n$, we cannot use the classical theory of asymptotic expansions, so some other theory is needed. It should be noted that the EPMC can be expressed as $E[\Phi(Q_n)]$ with a statistic Q_n, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. In Section 4, we find the constant q such that Q_n converges in probability to q, and give the limiting value of the EPMC as $\Phi(q)$. In Section 5, we evaluate our result numerically by Monte Carlo simulations. Section 6 presents the conclusions of this paper. The proofs of key lemmas are given in Appendix.

2. Asymptotic results related to the estimator S^*

In this section, we show some asymptotic results related to S^* given in (3). The following lemma shows the convergence in quadratic mean of $\alpha - \hat{\alpha}$, $\beta - \hat{\beta}$ and $\delta - \hat{\delta}$ under the conditions C1 and C2.

Lemma 1. *Under the conditions* C1 *and* C2, *it holds that*

$$\hat{a}_1^2 - a_1^2 \xrightarrow{q.m.} 0, \qquad \hat{\alpha} - \alpha \xrightarrow{q.m.} 0, \qquad \hat{\beta} - \beta \xrightarrow{q.m.} 0, \qquad \hat{\delta} - \delta \xrightarrow{q.m.} 0.$$

Here, the notation "$\xrightarrow{q.m.} 0$" *means the convergence to zero in quadratic mean.*

Proof. Let a and b be (arbitrary) constants. Using Lemma A.2,

$$\begin{aligned} E[|(a\hat{a}_1^2 + b\hat{a}_2) - (aa_1^2 + ba_2)|^2] &= \left\{\frac{48a^2}{n^3p^3} + \frac{48ab}{n^2p^2} + \frac{(2n^2+3n-6)b^2}{pn(n-1)(n+2)}\right\}a_4 + \left\{\frac{12a^2}{n^2p^2} + \frac{4b^2}{(n-1)(n+2)}\right\}a_2^2 \\ &\quad + \left\{\frac{32a^2}{n^2p^2} + \frac{16ab}{np}\right\}a_1a_3 + \frac{8a^2}{np}a_1^2a_2. \end{aligned} \tag{4}$$

When $a = 1$ and $b = 0$, (4) becomes

$$E[|\hat{a}_1^2 - a_1^2|^2] = \frac{48}{n^3p^3}a_4 + \frac{12}{n^2p^2}a_2^2 + \frac{32}{n^2p^2}a_1a_3 + \frac{8}{np}a_1^2a_2,$$

which is $O(p^{-1-\zeta})$ under the conditions C1 and C2. In the similar manner,

$$\begin{aligned} &E[|\hat{\alpha} - \alpha|^2] = O(p^{-2\zeta}), \qquad E[|\hat{\beta} - \beta|^2] = O(p^{-3\zeta+1}), \\ &E[|\hat{\delta} - \delta|^2] = O(p^{-3\zeta+1}). \end{aligned} \tag{5}$$

Thus, under the conditions C1 and C2,

$$E[|\hat{\alpha} - \alpha|^2] \to 0, \qquad E[|\hat{\beta} - \beta|^2] \to 0, \qquad E[|\hat{\delta} - \delta|^2] \to 0,$$

which prove Lemma 1. □

Using Lemma 1, we can prove $\|S^* - \Sigma^*\| \xrightarrow{p} 0$ under the conditions C1 and C2. This means that the loss of S^* converges in probability to that of Σ^*.

3. Comparison of linear discriminant analyses

In this section, we do simulation study to compare the performance of our proposed discrimination method with the MPLDA and DLDA methods. Let

$$W(A) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' A \left\{ \boldsymbol{x} - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2) \right\},$$

where $A = A(S)$ is a certain symmetric matrix which depends only on S. Then, MPLDA, DLDA and MLDA are defined as follows:

$$\text{MPLDA}: W(S^+) > 0 \ \ (\text{resp.}, \ < 0) \Rightarrow \boldsymbol{x} \in \Pi_1 \ \ (\text{resp.}, \ \Pi_2),$$

$$\text{DLDA}: W(D^{-1}) > 0 \ \ (\text{resp.}, \ < 0) \Rightarrow \boldsymbol{x} \in \Pi_1 \ \ (\text{resp.}, \ \Pi_2),$$

$$\text{MLDA}: W(S^{*-1}) = W^* > 0 \ \ (\text{resp.}, \ < 0) \Rightarrow \boldsymbol{x} \in \Pi_1 \ \ (\text{resp.}, \ \Pi_2),$$

where S^+ denotes the Moore-Penrose inverse matrix of S and D denotes the diagonal matrix with the same diagonal elements as those of S. Error rates are given as follows:

$$r_A(2|1) = \Pr(W(A) \le 0 \,|\, \boldsymbol{x} \in \Pi_1, \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, A),$$

$$r_A(1|2) = \Pr(W(A) > 0 \,|\, \boldsymbol{x} \in \Pi_2, \bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, A).$$

Since $r_A(1|2)$ can be obtained from $r_A(2|1)$ by interchanging N_1 and N_2, we deal with only $r_A(2|1)$. Let

$$V(A) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' A \Sigma A (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2),$$

$$Z(A) = V^{-1/2}(A)(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' A (\boldsymbol{x} - \boldsymbol{\mu}_1),$$

$$U(A) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' A (\bar{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) - \frac{1}{2} D^2(A),$$

where $D^2(A) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' A (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$. Then, $W(A)$ can be expressed as

$$W(A) = V^{1/2}(A) Z(A) - U(A). \tag{6}$$

Since $Z(A)$ and $(U(A), V(A))$ are independent, and $Z(A)$ is distributed according to the standard normal distribution under the condition $\boldsymbol{x} \in \Pi_1$ (hereafter, denoted by $Z(A) \sim N(0,1)$),

$$r_A(2|1) = \Phi(V^{-1/2}(A) U(A)), \tag{7}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of $N(0,1)$. Using (7), we can calculate the following three expected correct classification rates:

$$\text{MLDA (solid line)}: 1 - \frac{1}{2}\{E[r_{S^{*-1}}(2|1)] + E[r_{S^{*-1}}(1|2)]\},$$

$$\text{MPLDA (dotted line)}: 1 - \frac{1}{2}\{E[r_{S^{+}}(2|1)] + E[r_{S^{+}}(1|2)]\},$$

$$\text{DLDA (dashed line)}: 1 - \frac{1}{2}\{E[r_{D^{-1}}(2|1)] + E[r_{D^{-1}}(1|2)]\}.$$

Now, we compare these rates for some values of p. The data sets are generated as follows:

$$\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \dots, \boldsymbol{x}_{1N_1} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_1, \Sigma),$$

$$\boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \dots, \boldsymbol{x}_{2N_2} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_2, \Sigma),$$

where

$$\boldsymbol{\mu}_1 = p^{-1/2}(\mu_1, \mu_2, \dots, \mu_p)', \qquad \boldsymbol{\mu}_2 = (0, 0, \dots, 0)',$$

$$\Sigma = (\rho^{|i-j|}).$$

We set $N_1 = N_2 = 40$ and $\rho = 0.2, 0.5$. In each simulation, we set values drawn from a uniform distribution on the interval $[-5, 5]$ for μ_i, $i = 1, \dots, p$. Figure 1 (resp., Figure 2) illustrates the graphs of expected correct classification rate of three discrimination methods for $p = 1, \dots, 400$ and $\rho = 0.2$ (resp., $\rho = 0.5$).

In Figure 1, we can see that DLDA is as good as MLDA, and MPLDA is much inferior to the other two methods for most p. In Figure 2, MLDA is

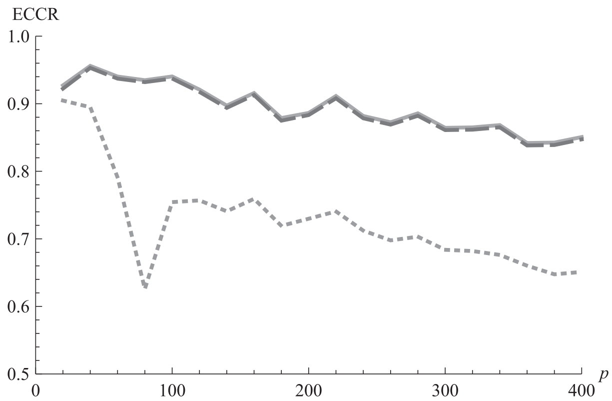

Fig. 1. The expected correct classification rate when $\rho = 0.2$.

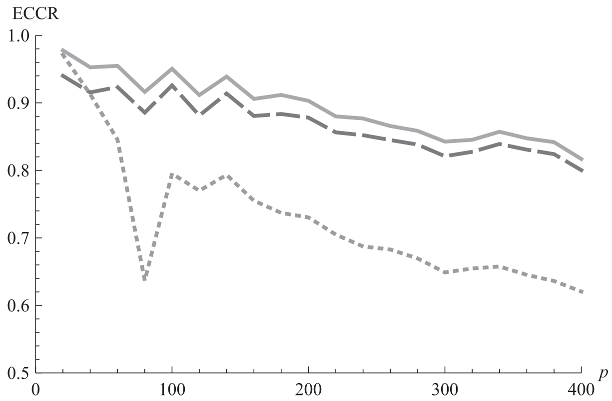

Fig. 2. The expected correct classification rate when $\rho = 0.5$.

much superior to the other methods. From these results we can see that MLDA gives a good performance even if the variables are strongly correlated.

4. Asymptotic approximations of EPMC for MLDA

In this section, we consider the following EPMC for MLDA:

$$e(2|1) = \Pr(W^* \le 0 \,|\, \boldsymbol{x} \in \Pi_1), \qquad e(1|2) = \Pr(W^* > 0 \,|\, \boldsymbol{x} \in \Pi_2).$$

It is generally difficult to obtain an explicit expression for the EPMC. So, we derive asymptotic approximations of EPMC for MLDA under the following high dimensional asymptotic framework:

$$\text{A1} : N_1, N_2, \;\; p \to \infty \;\text{ with }\; p^{\gamma}/n \to c_3 \;\text{ and }\; N_1/N_2 \to c_4$$
$$\text{for some } c_3, c_4 \in (0, \infty) \text{ and } \gamma \in (1/2, 1).$$

Here recall from Section 1 that $n = N_1 + N_2 - 2$. In addition, we assume the following:

$$\text{A2} : 0 < \lim_{p\to\infty} \lambda_1 < \infty,$$

$$\text{A3} : 0 < \lim_{p\to\infty} \Delta_0 = \lim_{p\to\infty} \boldsymbol{\delta}'\boldsymbol{\delta}/p^{1-\gamma} < \infty.$$

Here, λ_1 is the maximum eigenvalue of Σ and $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Since $e(1|2)$ can be obtained from $e(2|1)$ simply by interchanging N_1 and N_2, we only deal with $e(2|1)$. Using (7),

$$e(2|1) = E[\Phi(V^{-1/2}(S^{*-1})U(S^{*-1}))].$$

To evaluate $U(S^{*-1})$ and $V(S^{*-1})$, set

$$z_1 = N^{-1/2}(N_1\bar{\boldsymbol{x}}_1 + N_2\bar{\boldsymbol{x}}_2 - N_1\boldsymbol{\mu}_1 - N_2\boldsymbol{\mu}_2),$$

$$z_2 = \left(\frac{N}{N_1N_2}\right)^{-1/2}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2 - \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2),$$

where $N = N_1 + N_2$. Note that $z_i \sim N_p(\mathbf{0}, \Sigma)$, $i = 1, 2$. In addition, z_1 and z_2 are independent. We can express $U(S^{*-1})$ and $V(S^{*-1})$ in terms of z_1 and z_2 as

$$U(S^{*-1}) = -\frac{1}{2}\boldsymbol{\delta}'S^{*-1}\boldsymbol{\delta} + \frac{1}{N^{1/2}}\boldsymbol{\delta}'S^{*-1}z_1 - \left(\frac{N_1}{NN_2}\right)^{1/2}\boldsymbol{\delta}'S^{*-1}z_2$$
$$+ \frac{1}{(N_1N_2)^{1/2}}z_1'S^{*-1}z_2 - \frac{N_1 - N_2}{2N_1N_2}z_2'S^{*-1}z_2,$$

$$V(S^{*-1}) = \boldsymbol{\delta}'S^{*-1}\Sigma S^{*-1}\boldsymbol{\delta} + 2\left(\frac{N}{N_1N_2}\right)^{1/2}\boldsymbol{\delta}'S^{*-1}\Sigma S^{*-1}z_2$$
$$+ \frac{N}{N_1N_2}z_2'S^{*-1}\Sigma S^{*-1}z_2.$$

For notational simplicity, we write $U = U(S^{*-1})$, $V = V(S^{*-1})$. Define the following constant ξ_0:

$$\xi_0 = p^{(1-\gamma)/2}\left\{-\frac{1}{2}\Delta_0 - \frac{(N_1 - N_2)p^{\gamma}}{2N_1N_2}a_1\right\}\left\{\Delta_1 + \frac{Np^{\gamma}}{N_1N_2}a_2\right\}^{-1/2},$$

where $\Delta_1 = \boldsymbol{\delta}'\Sigma\boldsymbol{\delta}/p^{1-\gamma}$. Since $\Phi(\xi_0) = 1/2$ when $\xi_0 = 0$, we deal only with the case $\xi_0 \neq 0$. Using Lemma A.6 in Appendix, we derive the following lemma.

Lemma 2. *Under the assumptions* A1–3, *it holds that*

$$\left|\frac{U}{\sqrt{V}} - \xi_0\right| = o_p(1).$$

The proof of Lemma 2 is given in Appendix. Using Lemma 2, we get the following lemma.

Lemma 3. *Under the assumptions* A1–3,

$$|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| = o_p(e^{-p^{\nu}})$$

holds for any $\nu \in (0, 1-\gamma)$.

This result means that the conditional probability of misclassification for MLDA converges to $\Phi(\xi_0)$ in probability. The proof of Lemma 3 is given in Appendix. Now, we turn to the evaluation. It holds that

$$\begin{aligned}|e(2|1) - \Phi(\xi_0)| &= |\mathrm{E}[\Phi(U/\sqrt{V})] - \Phi(\xi_0)| \\ &= |\mathrm{E}[\Phi(U/\sqrt{V}) - \Phi(\xi_0)]| \\ &\leq \mathrm{E}[|\Phi(U/\sqrt{V}) - \Phi(\xi_0)|].\end{aligned}$$

Lemma 3 and Lebesgue's dominated convergence theorem yield that the sequence $\{|\Phi(U/\sqrt{V}) - \Phi(\xi_0)|\}$ is uniformly integrable. Hence, under the assumptions A1–3,

$$\mathrm{E}[|\Phi(U/\sqrt{V}) - \Phi(\xi_0)|] \to 0.$$

Thus, we can see that $\Phi(\xi_0)$ provides an asymptotic approximation to $e(2|1)$ which is given in the following theorem.

THEOREM 1. *Under the assumptions* A1–3, *the following statement holds:*

$$|e(2|1) - \Phi(\xi_0)| \to 0.$$

5. Accuracy of approximation

To investigate the accuracy of asymptotic approximation proposed in Theorem 1, a numerical experiment was performed. We calculated $e(2|1)$ by a 10000 iterations of simulation. In each step, the data sets are generated as follows:

$$\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \ldots, \boldsymbol{x}_{1N_1} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_1, \Sigma),$$

$$\boldsymbol{x}_{21}, \boldsymbol{x}_{22}, \ldots, \boldsymbol{x}_{2N_2} \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}_2, \Sigma),$$

where the structure of Σ was assumed to have serial correlation $\Sigma = (\rho^{|i-j|})$, and $\boldsymbol{\mu}_1 = (n)^{-1/2}(\mu, \mu, \ldots, \mu)'$ and $\boldsymbol{\mu}_2 = (0, 0, \ldots, 0)'$. The sample sizes N_1 and N_2 were assumed to be equal to $N = 2 \cdot [p^\gamma]$, where $[a]$ denotes the integer part of a. The values of $e(2|1)$ in tables are the average values of $\Phi(U/\sqrt{V})$ calculated by 10000 repetitions. We consider the asymptotic approximation for EPMC: $e^*(2|1) = \Phi(\xi_0)$. Here, ξ_0 is given in Section 4. Tables 1–4 give the values of $e(2|1)$ and $e^*(2|1)$ when $\gamma = 0.4, 0.5, 0.6$ and 0.7, respectively. In each table, values were shown for the cases in which $p = 100, 200$, $\rho = 0.2, 0.4$,

$\mu = 1, 1.5$. Through these simulation results, we can see that $e^*(2|1)$ has better approximation for large γ than for small γ.

Table 1. Accuracy of approximation when $\gamma = 0.4$

(p, N)	ρ	$\mu = 1.0$ $e(2\|1)$	$e^*(2\|1)$	$\mu = 1.5$ $e(2\|1)$	$e^*(2\|1)$
(100, 12)	0.2	0.290	0.275	0.140	0.120
	0.4	0.324	0.304	0.186	0.160
(200, 16)	0.2	0.216	0.199	0.059	0.049
	0.4	0.250	0.233	0.099	0.080

Table 2. Accuracy of approximation when $\gamma = 0.5$

(p, N)	ρ	$\mu = 1.0$ $e(2\|1)$	$e^*(2\|1)$	$\mu = 1.5$ $e(2\|1)$	$e^*(2\|1)$
(100, 20)	0.2	0.334	0.322	0.193	0.182
	0.4	0.355	0.344	0.240	0.221
(200, 28)	0.2	0.265	0.256	0.112	0.100
	0.4	0.302	0.286	0.155	0.139

Table 3. Accuracy of approximation when $\gamma = 0.6$

(p, N)	ρ	$\mu = 1.0$ $e(2\|1)$	$e^*(2\|1)$	$\mu = 1.5$ $e(2\|1)$	$e^*(2\|1)$
(100, 30)	0.2	0.359	0.352	0.234	0.229
	0.4	0.382	0.372	0.280	0.269
(200, 48)	0.2	0.306	0.296	0.158	0.148
	0.4	0.334	0.322	0.202	0.188

Table 4. Accuracy of approximation when $\gamma = 0.7$

(p, N)	ρ	$\mu = 1.0$ $e(2\|1)$	$e^*(2\|1)$	$\mu = 1.5$ $e(2\|1)$	$e^*(2\|1)$
(100, 50)	0.2	0.388	0.384	0.289	0.283
	0.4	0.408	0.400	0.325	0.313
(200, 80)	0.2	0.342	0.339	0.213	0.209
	0.4	0.370	0.360	0.257	0.246

6. Conclusions

We introduced a new method for linear discriminant analysis based on shrink type estimators of covariance matrices for high-dimensional data. From Section 3, for $p \geq N$, MLDA is much superior to other discrimination methods in the sense of expected correct classification rate. In addition, we considered the EPMC of MLDA. Since it is generally difficult to obtain an explicit expression for the EPMC, we derived a limiting value of EPMC under the assumption: $n, p \to \infty$ with $p^{\gamma}/n \to c_3$ for some $c_3 \in (0, \infty)$ and $\gamma \in (1/2, 1)$ in Section 4. By simulation results, $\Phi(\xi_0)$ is close to the true value for high-dimensional data. It should be noted that, in MLDA, the EPMC becomes smaller as $|\xi_0|$ becomes larger, while, in LDA, the EPMC becomes smaller as Mahalanobis distance becomes larger. The explicit formula of error bounds for limiting value of EPMC is not obtained. This problem should be examined in a high-dimensional case, which is left as a future problem.

Appendix

In this appendix, we prove Lemmas 2 and 3 stated in Section 4. We begin with some preliminary results.

Lemma A.1. *Let $\boldsymbol{P}$ and $\boldsymbol{Q}$ be non-singular matrices of proper order. Then, if $\boldsymbol{Q} = \boldsymbol{P} + \boldsymbol{U}\boldsymbol{V}$,*

$$\boldsymbol{Q}^{-1} = \boldsymbol{P}^{-1} - \boldsymbol{P}^{-1}\boldsymbol{U}(\boldsymbol{I} + \boldsymbol{V}\boldsymbol{P}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{P}^{-1}.$$

For the proof, see Schott [8].

Lemma A.2. *Let $V = XX'$, where $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ and the components $\boldsymbol{x}_i$ are iid $N_p(\boldsymbol{0}, \Sigma)$. Then the following assertions hold:*

(i) $E[(\mathrm{tr}\ V)^2] = n^2(\mathrm{tr}\ \Sigma)^2 + 2n\ \mathrm{tr}\ \Sigma^2$,

(ii) $E[\mathrm{tr}\ V^2] = (n^2 + n)\ \mathrm{tr}\ \Sigma^2 + n(\mathrm{tr}\ \Sigma)^2$,

(iii) $$E[\mathrm{tr}\ V^2\ \mathrm{tr}\ V\Sigma] = 4n(n+1)\ \mathrm{tr}\ \Sigma^4 + n^2(n+1)(\mathrm{tr}\ \Sigma^2)^2 + 4n\ \mathrm{tr}\ \Sigma^3\ \mathrm{tr}\ \Sigma + n^2(\mathrm{tr}\ \Sigma)^2\ \mathrm{tr}\ \Sigma^2,$$

(iv) $E[(\mathrm{tr}\ V\Sigma)^2] = n^2(\mathrm{tr}\ \Sigma^2)^2 + 2n\ \mathrm{tr}\ \Sigma^4$,

(v) $$E[(\mathrm{tr}\ V)^4] = n^4(\mathrm{tr}\ \Sigma)^4 + 12n^3(\mathrm{tr}\ \Sigma)^2\ \mathrm{tr}\ \Sigma^2 + 12n^2(\mathrm{tr}\ \Sigma^2)^2 + 32n^2\ \mathrm{tr}\ \Sigma\ \mathrm{tr}\ \Sigma^3 + 48n\ \mathrm{tr}\ \Sigma^4,$$

(vi) $E[(\mathrm{tr}\ V)^2\ \mathrm{tr}\ V^2] = n^3(\mathrm{tr}\ \Sigma)^4 + (n^4 + n^3 + 10n^2)(\mathrm{tr}\ \Sigma)^2\ \mathrm{tr}\ \Sigma^2$

$$+ 2(n^3 + n^2 + 4n)(\mathrm{tr}\ \Sigma^2)^2$$

$$+ 8(n^3 + n^2 + 2n)\ \mathrm{tr}\ \Sigma\ \mathrm{tr}\ \Sigma^3$$

$$+ 24(n^2 + n)\ \mathrm{tr}\ \Sigma^4,$$

(vii) $E[(\mathrm{tr}\ V^2)^2] = n^2(\mathrm{tr}\ \Sigma)^4 + 2(n^3 + n^2 + 4n)(\mathrm{tr}\ \Sigma)^2\ \mathrm{tr}\ \Sigma^2$

$$+ (n^4 + 2n^3 + 5n^2 + 4n)(\mathrm{tr}\ \Sigma^2)^2$$

$$+ 16(n^2 + n)\ \mathrm{tr}\ \Sigma\ \mathrm{tr}\ \Sigma^3$$

$$+ 4(2n^3 + 5n^2 + 5n)\ \mathrm{tr}\ \Sigma^4,$$

(viii) $E[\mathrm{tr}(AV^2)^2] = (n^4 + 4n^3 + 7n^2 + 4n)\ \mathrm{tr}(A\Sigma^2)^2$

$$+ (2n^3 + 6n^2 + 8n)\ \mathrm{tr}\ A\Sigma^2 A\Sigma\ \mathrm{tr}\ \Sigma$$

$$+ (2n^3 + 14n^2 + 16n)\ \mathrm{tr}\ A\Sigma^3 A\Sigma$$

$$+ (2n^3 + 6n^2 + 8n)\ \mathrm{tr}\ A\Sigma^3\ \mathrm{tr}\ A\Sigma$$

$$+ (4n^2 + 4n)\ \mathrm{tr}\ A\Sigma^2\ \mathrm{tr}\ A\Sigma\ \mathrm{tr}\ \Sigma$$

$$+ (2n^3 + 4n^2 + 2n)(\mathrm{tr}\ A\Sigma^2)^2$$

$$+ (n^2 + n)\ \mathrm{tr}(A\Sigma)^2(\mathrm{tr}\ \Sigma)^2$$

$$+ (n^2 + 3n)\ \mathrm{tr}(A\Sigma)^2\ \mathrm{tr}\ \Sigma^2 + n(\mathrm{tr}\ A\Sigma)^2(\mathrm{tr}\ \Sigma)^2$$

$$+ (n^2 + n)(\mathrm{tr}\ A\Sigma)^2\ \mathrm{tr}\ \Sigma^2,$$

(xi) $E[(\mathrm{tr}\ AV^2)^2] = (n^4 + 2n^3 + 3n^2 + 2n)(\mathrm{tr}\ A\Sigma^2)^2$

$$+ (2n^3 + 2n^2 + 4n)\ \mathrm{tr}\ \Sigma\ \mathrm{tr}\ A\Sigma\ \mathrm{tr}\ A\Sigma^2$$

$$+ (4n^3 + 8n^2 + 4n)\ \mathrm{tr}(A\Sigma^2)^2$$

$$+ (4n^3 + 12n^2 + 16n)\ \mathrm{tr}\ A\Sigma A\Sigma^3$$

$$+ (8n^2 + 8n)\ \mathrm{tr}\ \Sigma\ \mathrm{tr}\ A\Sigma A\Sigma^2$$

$$+ (8n^2 + 8n)\ \mathrm{tr}\ A\Sigma\ \mathrm{tr}\ A\Sigma^3$$

$$+ (2n^2 + 2n)\ \mathrm{tr}\ \Sigma^2\ \mathrm{tr}(A\Sigma)^2 + 2n(\mathrm{tr}\ \Sigma)^2\ \mathrm{tr}(A\Sigma)^2$$

$$+ n^2(\mathrm{tr}\ \Sigma)^2(\mathrm{tr}\ A\Sigma)^2 + 2n(\mathrm{tr}\ A\Sigma)^2\ \mathrm{tr}\ \Sigma^2.$$

PROOF. The results in (i) and (ii) are well known (see, e.g., Gupta and Nagar [4]). For the proofs of (v)–(vii), see Hyodo and Yamada [5]. We give the proof of (iii). Let $U = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$, where $\boldsymbol{u}_i$ are iid $N_p(\mathbf{0}, \Sigma)$. Suppose P is an orthogonal matrix such that $\Sigma = P\Lambda P'$, where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$ and, for $i = 1, \dots, p$, λ_i is the i-th eigenvalue of Σ. Then $X = \Sigma^{1/2}U$ and

$$
\begin{aligned}
&E[\mathrm{tr}(\Sigma^{1/2}UU'\Sigma^{1/2})^2\ \mathrm{tr}(\Sigma^{1/2}UU'\Sigma^{1/2})\Sigma] \\
&\quad = E[\mathrm{tr}(Z'\Lambda Z)^2\ \mathrm{tr}(Z'\Lambda^2 Z)] \\
&\quad = E\left[\left(\sum_{i=1}^p \lambda_i^2 (\boldsymbol{z}_i'\boldsymbol{z}_i)^2 + \sum_{i \neq j}^p \lambda_i\lambda_j(\boldsymbol{z}_i'\boldsymbol{z}_j)^2\right)\left(\sum_{i=1}^p \lambda_i^2(\boldsymbol{z}_i'\boldsymbol{z}_i)\right)\right] \\
&\quad = E\left[\sum_{i=1}^p \lambda_i^4(\boldsymbol{z}_i'\boldsymbol{z}_i)^3 + \sum_{i \neq j} \lambda_i^2\lambda_j^2(\boldsymbol{z}_i'\boldsymbol{z}_i)(\boldsymbol{z}_j'\boldsymbol{z}_j)^2\right. \\
&\qquad\qquad \left. + 2\sum_{i \neq j}^p \lambda_i^3\lambda_j(\boldsymbol{z}_i'\boldsymbol{z}_i)(\boldsymbol{z}_i'\boldsymbol{z}_j)^2 + \sum_{i \neq j, j \neq k, k \neq i}^p \lambda_i^2\lambda_j\lambda_k(\boldsymbol{z}_i'\boldsymbol{z}_j)^2(\boldsymbol{z}_k'\boldsymbol{z}_k)\right] \\
&\quad = 4n(n+1)\ \mathrm{tr}\,\Sigma^4 + n^2(n+1)(\mathrm{tr}\,\Sigma^2)^2 + 4n\ \mathrm{tr}\,\Sigma^3\ \mathrm{tr}\,\Sigma + n^2(\mathrm{tr}\,\Sigma)^2\ \mathrm{tr}\,\Sigma^2,
\end{aligned}
$$

where $U'P' = Z' = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_p)$, and $\boldsymbol{z}_i$ are iid $N_n(\mathbf{0}, I)$. The proof of (iv) can be similarly derived. Next, we prove (viii). Let $\Lambda^{1/2}P'AP\Lambda^{1/2} = B = (b_{ij})_{i,j=1,\dots,p}$. Then

$$
\begin{aligned}
E[\mathrm{tr}(AV^2)^2] &= E\left[\mathrm{tr}\left(BZ\left(\sum_{i=1}^p \lambda_i\boldsymbol{z}_i\boldsymbol{z}_i'\right)Z'BZ\left(\sum_{i=1}^p \lambda_i\boldsymbol{z}_i\boldsymbol{z}_i'\right)Z'\right)\right] \\
&= \mathrm{tr}\left(B\left(\sum_{k,l,q,r} \lambda_k\lambda_l b_{q,r} E[\boldsymbol{z}_i'\boldsymbol{z}_k\boldsymbol{z}_k'\boldsymbol{z}_q\boldsymbol{z}_r'\boldsymbol{z}_l\boldsymbol{z}_l'\boldsymbol{z}_j]\right)_{i,j=1,\dots,p}\right) \\
&= (n^4 + 4n^3 + 7n^2 + 4n)\ \mathrm{tr}\,B\Lambda B\Lambda \\
&\quad + (n^3 + 7n^2 + 8n)(\mathrm{tr}\,B\Lambda^2 B + \mathrm{tr}\,B^2\Lambda^2) \\
&\quad + (n^2 + 3n)\ \mathrm{tr}\,B^2\ \mathrm{tr}\,\Lambda^2 + (n^3 + 3n^2 + 4n)\ \mathrm{tr}\,B\Lambda^2\ \mathrm{tr}\,B \\
&\quad + (n^3 + 3n^2 + 4n)\ \mathrm{tr}\,B\ \mathrm{tr}\,B\Lambda^2 + (n^2 + n)(\mathrm{tr}\,B)^2\ \mathrm{tr}\,\Lambda^2 \\
&\quad + (n^3 + 3n^2 + 4n)(\mathrm{tr}\,B\Lambda B\ \mathrm{tr}\,\Lambda + \mathrm{tr}\,B^2\Lambda\ \mathrm{tr}\,\Lambda) \\
&\quad + (n^2 + n)\ \mathrm{tr}\,B^2(\mathrm{tr}\,\Lambda)^2 + (2n^3 + 4n^2 + 2n)(\mathrm{tr}\,B\Lambda)^2 \\
&\quad + (2n^2 + 2n)\ \mathrm{tr}\,B\Lambda\ \mathrm{tr}\,B\ \mathrm{tr}\,\Lambda + (2n^2 + 2n)\ \mathrm{tr}\,B\ \mathrm{tr}\,B\Lambda\ \mathrm{tr}\,\Lambda \\
&\quad + n(\mathrm{tr}\,B)^2(\mathrm{tr}\,\Lambda)^2
\end{aligned}
$$

$$
\begin{aligned}
&= (n^4 + 4n^3 + 7n^2 + 4n) \operatorname{tr}(A\Sigma^2)^2 \\
&\quad + (2n^3 + 6n^2 + 8n) \operatorname{tr} A\Sigma^2 A\Sigma \operatorname{tr} \Sigma \\
&\quad + (2n^3 + 14n^2 + 16n) \operatorname{tr} A\Sigma^3 A\Sigma \\
&\quad + (2n^3 + 6n^2 + 8n) \operatorname{tr} A\Sigma^3 \operatorname{tr} A\Sigma \\
&\quad + (4n^2 + 4n) \operatorname{tr} A\Sigma^2 \operatorname{tr} A\Sigma \operatorname{tr} \Sigma \\
&\quad + (2n^3 + 4n^2 + 2n)(\operatorname{tr} A\Sigma^2)^2 \\
&\quad + (n^2 + n) \operatorname{tr}(A\Sigma)^2 (\operatorname{tr} \Sigma)^2 \\
&\quad + (n^2 + 3n) \operatorname{tr}(A\Sigma)^2 \operatorname{tr} \Sigma^2 + n(\operatorname{tr} A\Sigma)^2 (\operatorname{tr} \Sigma)^2 \\
&\quad + (n^2 + n)(\operatorname{tr} A\Sigma)^2 \operatorname{tr} \Sigma^2.
\end{aligned}
$$

The proof of (ix) can be similarly derived. □

Lemma A.3. *Let* $\boldsymbol{y} \sim N_p(\mathbf{0}, \Sigma)$ *and* $Q = \boldsymbol{y}'B\boldsymbol{y}$. *Then, the following assertions hold:*

$$\text{(i)} \quad E[Q] = \operatorname{tr} B\Sigma,$$

$$\text{(ii)} \quad E[Q^2] = 2 \operatorname{tr}(B\Sigma)^2 + (\operatorname{tr} B\Sigma)^2.$$

For the proofs, see Mathai et al. [7].

Lemma A.4. *Let* $\boldsymbol{z} \sim N_p(\mathbf{0}, \Sigma)$. *Under the assumptions* A1 *and* A2, *it holds that*

(i) $\left| \dfrac{\boldsymbol{z}'\boldsymbol{z}}{p} - a_1 \right| = O_p(p^{-1/2})$,

(ii) $\left| \dfrac{\boldsymbol{z}'\Sigma\boldsymbol{z}}{p} - a_2 \right| = O_p(p^{-1/2})$.

Proof. At first, we show (i). It is sufficient to show that

$$pE\left[\left| \frac{\boldsymbol{z}'\boldsymbol{z}}{p} - a_1 \right|^2 \right] = O(1).$$

Using Lemma A.3, we have

$$E\left[\left| \frac{\boldsymbol{z}'\boldsymbol{z}}{p} - a_1 \right|^2 \right] = \frac{2a_2}{p} (= O(p^{-1})).$$

This proves (i).

Secondly, we show (ii). It is sufficient to show that

$$pE\left[\left|\frac{\boldsymbol{z}'\Sigma\boldsymbol{z}}{p} - a_2\right|^2\right] = O(1).$$

Using Lemma A.3, we obtain

$$E\left[\left|\frac{\boldsymbol{z}'\Sigma\boldsymbol{z}}{p} - a_2\right|^2\right] = \frac{2a_4}{p}(= O(p^{-1})).$$

This proves (ii). □

Lemma A.5. *Let $\boldsymbol{z}$ be a random vector distributed as $N_p(\boldsymbol{0}, \Sigma)$. Under the assumptions* A1–3*, it holds that*

(i) $\dfrac{\boldsymbol{\delta}'S\boldsymbol{\delta}}{p^{1-\gamma}} = O_p(1)$,

(ii) $\dfrac{\boldsymbol{z}'S\boldsymbol{z}}{p} = O_p(1)$,

(iii) $\dfrac{\boldsymbol{\delta}'S\Sigma\boldsymbol{\delta}}{p^{1-\gamma}} = O_p(1)$,

(iv) $\dfrac{\boldsymbol{\delta}'S^2\boldsymbol{\delta}}{p^{2-2\gamma}} = O_p(1)$,

(v) $\dfrac{\boldsymbol{\delta}'\Sigma S^2\Sigma\boldsymbol{\delta}}{p^{2-2\gamma}} = O_p(1)$,

(vi) $\dfrac{\boldsymbol{z}'S\Sigma\boldsymbol{z}}{p} = O_p(1)$,

(vii) $\dfrac{\boldsymbol{z}'S^2\boldsymbol{z}}{p^{2-\gamma}} = O_p(1)$,

(viii) $\dfrac{\boldsymbol{z}'\Sigma S^2\Sigma\boldsymbol{z}}{p^{2-\gamma}} = O_p(1)$.

Proof. Using Lemma A.2, we obtain

$$\frac{1}{p^{1-\gamma}}E[\boldsymbol{\delta}'S\boldsymbol{\delta}] = \Delta_1,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{\delta}'S\boldsymbol{\delta}}{p^{1-\gamma}}\right] = \frac{2\Delta_1^2}{n} = O(p^{-\gamma}),$$

which proves (i). Using Lemmas A.2 and A.3, we obtain

$$\frac{1}{p}E[\boldsymbol{z}'S\boldsymbol{z}] = a_2,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{z}'S\boldsymbol{z}}{p}\right] = \frac{2(n^2+2n)}{pn^2}a_4 + \frac{2}{n}a_2^2 = O(p^{-\gamma}).$$

Thus, (ii) follows. By Lemma A.2, we get

$$E\left[\frac{\boldsymbol{\delta}' S\Sigma\boldsymbol{\delta}}{p^{1-\gamma}}\right] = \Delta_2,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{\delta}' S\Sigma\boldsymbol{\delta}}{p^{1-\gamma}}\right] = \frac{\Delta_2^2 + \Delta_1\Delta_3}{n} = O(p^{-\gamma}),$$

$$E\left[\frac{\boldsymbol{\delta}' S^2\boldsymbol{\delta}}{p^{2-2\gamma}}\right] = \left(\frac{p^{\gamma}}{n}\right)a_1\Delta_1 + \left(\frac{1}{p^{1-\gamma}} + \frac{1}{np^{1-\gamma}}\right)\Delta_2,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{\delta}' S^2\boldsymbol{\delta}}{p^{2-2\gamma}}\right] = \frac{2p^{2\gamma}a_1^2\Delta_1^2}{n^3} + o(p^{-\gamma})$$

$$= O(p^{-\gamma}),$$

$$E\left[\frac{\boldsymbol{\delta}'\Sigma S^2\Sigma\boldsymbol{\delta}}{p^{2-2\gamma}}\right] = \left(\frac{p^{\gamma}}{n}\right)a_1\Delta_3 + \left(\frac{1}{p^{1-\gamma}} + \frac{1}{np^{1-\gamma}}\right)\Delta_4,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{\delta}'\Sigma S^2\Sigma\boldsymbol{\delta}}{p^{2-2\gamma}}\right] = \frac{2p^{2\gamma}a_1^2\Delta_3^2}{n^3} + o(p^{-\gamma})$$

$$= O(p^{-\gamma}),$$

where $\Delta_i = \boldsymbol{\delta}'\Sigma^i\boldsymbol{\delta}/p^{1-\gamma}$, $i = 2, 3, 4$. Consequently, (iii)–(v). Using Lemma A.2, we obtain

$$E\left[\frac{\boldsymbol{z}' S\Sigma\boldsymbol{z}}{p}\right] = a_3,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{z}' S\Sigma\boldsymbol{z}}{p}\right] = \frac{2}{n}a_3^2 + \frac{2a_6}{p}\left(1 + \frac{2}{n}\right)$$

$$= O(p^{-\gamma}),$$

$$E\left[\frac{\boldsymbol{z}' S^2\boldsymbol{z}}{p^{2-\gamma}}\right] = \left(\frac{p^{\gamma}}{n}\right)a_1a_2 + \left(\frac{1}{p^{1-\gamma}} + \frac{1}{np^{1-\gamma}}\right)a_3,$$

$$\mathrm{Var}\left[\frac{\boldsymbol{z}' S^2\boldsymbol{z}}{p^{2-\gamma}}\right] = \frac{2p^{2\gamma}a_1^2a_2^2}{n^3} + o(p^{-\gamma})$$

$$= O(p^{-\gamma}),$$

$$E\left[\frac{\boldsymbol{z}'\Sigma S^2\Sigma\boldsymbol{z}}{p^{2-\gamma}}\right] = \left(\frac{p^{\gamma}}{n}\right)a_1a_4 + \left(\frac{1}{p^{1-\gamma}} + \frac{1}{np^{1-\gamma}}\right)a_5,$$

$$\mathrm{Var}\left[\frac{z'\Sigma S^2\Sigma z}{p^{2-\gamma}}\right] = \frac{2p^{2\gamma}a_1^2a_4^2}{n^3} + o(p^{-\gamma})$$
$$= O(p^{-\gamma}).$$

Thus, (vi)–(viii) follow. □

Lemma A.6. *Let z be a random vector distributed as $N_p(\mathbf{0},\Sigma)$. Under the assumptions* A1–3*, it holds that*

(i) $\dfrac{1}{p^{1-\gamma}}\left(\boldsymbol{\delta}'S^{*-1}\boldsymbol{\delta} - \dfrac{\hat{\delta}}{\hat{\beta}\hat{a}_1}\boldsymbol{\delta}'\boldsymbol{\delta}\right) = O_p(p^{-1+\gamma})$,

(ii) $\dfrac{1}{p}\left(z'S^{*-1}z - \dfrac{\hat{\delta}}{\hat{\beta}\hat{a}_1}z'z\right) = O_p(p^{-1+\gamma})$,

(iii) $\dfrac{1}{p^{1-\gamma}}\left(\boldsymbol{\delta}'S^{*-1}\Sigma S^{*-1}\boldsymbol{\delta} - \dfrac{\hat{\delta}^2}{\hat{\beta}^2\hat{a}_1^2}\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}\right) = O_p(p^{-1+\gamma})$,

(iv) $\dfrac{1}{p}\left(z'S^{*-1}\Sigma S^{*-1}z - \dfrac{\hat{\delta}^2}{\hat{\beta}^2\hat{a}_1^2}z'\Sigma z\right) = O_p(p^{-1+\gamma})$.

Proof. At first, we show (i). Using Lemma A.1,

$$\frac{1}{p^{1-\gamma}}\left(\boldsymbol{\delta}'S^{*-1}\boldsymbol{\delta} - \frac{\hat{\delta}}{\hat{\beta}\hat{a}_1}\boldsymbol{\delta}'\boldsymbol{\delta}\right) = -\left(\frac{\hat{\alpha}\hat{\delta}}{\hat{\beta}^2\hat{a}_1^2}\right)\frac{\boldsymbol{\delta}'S^{1/2}\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S^{1/2}\boldsymbol{\delta}}{p^{1-\gamma}}. \tag{8}$$

From (5), under the assumptions A1 and A2, it holds that

$$\hat{a}_1^2 = a_1^2 + O_p(p^{-(\gamma+1)/2}), \qquad \left(\frac{n}{p}\right)\hat{\alpha} = \left(\frac{n}{p}\right)(a_2 - a_1^2) + O_p(p^{-1}),$$
$$\left(\frac{n}{p}\right)\hat{\beta} = a_1^2 + O_p(p^{-(\gamma+1)/2}),$$
$$\left(\frac{n}{p}\right)\hat{\delta} = \left(1 - \frac{n}{p}\right)a_1^2 + \left(\frac{n}{p}\right)a_2 + O_p(p^{-(\gamma+1)/2}). \tag{9}$$

We remark that $\alpha = a_2 - a_1^2 = 0$ when $\Sigma = \sigma I$, where σ is some positive constant. Hence,

$$\frac{\hat{\alpha}\hat{\delta}}{\hat{\beta}^2\hat{a}_1^2} = \begin{cases} O_p(p^{-1+\gamma}), & \Sigma \neq \sigma I, \\ O_p(p^{-1}), & \Sigma = \sigma I. \end{cases} \tag{10}$$

On the other hand,

$$\frac{\boldsymbol{\delta}'S^{1/2}\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S^{1/2}\boldsymbol{\delta}}{p^{1-\gamma}} \le \Psi_1\left(\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\right)\frac{\boldsymbol{\delta}'S\boldsymbol{\delta}}{p^{1-\gamma}} \quad a.a.s. \tag{11}$$

Here, the notation "*a.a.s*" means asymptotically almost surely. Using Theorem 2 in El Karoui [2], we obtain

$$\Psi_1\left(\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\right)=\frac{1}{1+O_p(p^{-\gamma})}<\infty \quad a.a.s. \tag{12}$$

when $\Sigma=\sigma I$. And we obtain

$$\Psi_1\left(\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\right)=\frac{1}{1+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}\Psi_p(S)}\leq 1<\infty \quad a.a.s. \tag{13}$$

when $\Sigma\neq\sigma I$. Here, $\Psi_i(A)$ denotes the i-th largest eigenvalue of a matrix A. Combining (11)–(13) and (i) of Lemma A.5, we obtain

$$\frac{\boldsymbol{\delta}'S^{1/2}\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S^{1/2}\boldsymbol{\delta}}{p^{1-\gamma}}=\begin{cases}O_p(1), & \Sigma\neq\sigma I,\\ O_p(1), & \Sigma=\sigma I.\end{cases} \tag{14}$$

So, (i) follows from (8), (10) and (14).

Secondly, we show (ii). By the similar evaluation method of (8),

$$\frac{1}{p}\left(\boldsymbol{z}'S^{*-1}\boldsymbol{z}-\frac{\hat{\delta}}{\hat{\beta}\hat{a}_1}\boldsymbol{z}'\boldsymbol{z}\right)=-\left(\frac{\hat{\alpha}\hat{\delta}}{\hat{\beta}^2\hat{a}_1^2}\right)\frac{\boldsymbol{z}'S^{1/2}\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S^{1/2}\boldsymbol{z}}{p^{1-\gamma}}. \tag{15}$$

Using (11), (12) and (ii) of Lemmas A.5, we obtain

$$\frac{\boldsymbol{z}'S^{1/2}\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S^{1/2}\boldsymbol{z}}{p^{1-\gamma}}=\begin{cases}O_p(1), & \Sigma\neq\sigma I,\\ O_p(1), & \Sigma=\sigma I.\end{cases} \tag{16}$$

So, (ii) follows from (10), (15) and (16).

Thirdly, we show (iii). Using Lemma A.1,

$$\begin{aligned}
&\frac{1}{p^{1-\gamma}}\left(\boldsymbol{\delta}'S^{*-1}\Sigma S^{*-1}\boldsymbol{\delta}-\frac{\hat{\delta}^2}{\hat{\beta}^2\hat{a}_1^2}\boldsymbol{\delta}'\Sigma\boldsymbol{\delta}\right)\\
&=\frac{p^{1-\gamma}}{p^{2-2\gamma}}\left\{-\frac{2\hat{\alpha}\hat{\delta}^2}{\hat{\beta}^3\hat{a}_1^3}\boldsymbol{\delta}'S\Sigma\boldsymbol{\delta}+\left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{\delta}'S\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S\Sigma\boldsymbol{\delta}\right.\\
&\qquad+\left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{\delta}'\Sigma S\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S\boldsymbol{\delta}\\
&\qquad\left.+\left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{\delta}'S\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\Sigma\left(I_p+\frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}S\boldsymbol{\delta}\right\}.
\end{aligned} \tag{17}$$

In addition, using (9), we get

$$\frac{\hat{\alpha}\hat{\delta}^2}{\hat{\beta}^3\hat{a}_1^3} = \begin{cases} O_p(p^{-1+\gamma}), & \Sigma \neq \sigma I, \\ O_p(p^{-1}), & \Sigma = \sigma I, \end{cases} \tag{18}$$

and

$$\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4} = \begin{cases} O_p(p^{-2+2\gamma}), & \Sigma \neq \sigma I, \\ O_p(p^{-2}), & \Sigma = \sigma I. \end{cases} \tag{19}$$

Moreover, using (iii)–(v) in Lemma A.5, (12) and (13), we obtain

$$\frac{1}{p^{1-\gamma}}\boldsymbol{\delta}' S\Sigma\boldsymbol{\delta} = O_p(1), \tag{20}$$

$$\begin{aligned} &\frac{2}{p^{2-2\gamma}}\left|\boldsymbol{\delta}' S\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1} S\Sigma\boldsymbol{\delta}\right| \\ &\quad \leq 2\sqrt{\frac{\boldsymbol{\delta}' S^2\boldsymbol{\delta}}{p^{2-2\gamma}}}\sqrt{\frac{\boldsymbol{\delta}'\Sigma S(I_p + \hat{\alpha}/(\hat{\beta}\hat{a}_1)S)^{-2} S\Sigma\boldsymbol{\delta}}{p^{2-2\gamma}}} \\ &\quad = \begin{cases} O_p(1), & \Sigma \neq \sigma I, \\ O_p(1), & \Sigma = \sigma I, \end{cases} \end{aligned} \tag{21}$$

$$\begin{aligned} &\frac{1}{p^{2-2\gamma}}\left|\boldsymbol{\delta}' S\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\Sigma\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1} S\boldsymbol{\delta}\right| \\ &\quad = \begin{cases} O_p(1), & \Sigma \neq \sigma I, \\ O_p(1), & \Sigma = \sigma I. \end{cases} \end{aligned} \tag{22}$$

So, (iii) follows from (17)–(22).

Finally, we show (iv). Using the similar evaluation method of (17), we obtain

$$\begin{aligned} &\frac{1}{p}\left(\boldsymbol{z}' S^{*-1}\Sigma S^{*-1}\boldsymbol{z} - \frac{\hat{\delta}^2}{\hat{\beta}^2\hat{a}_1^2}\boldsymbol{z}'\Sigma\boldsymbol{z}\right) \\ &\quad = \frac{p^{1-\gamma}}{p^{2-\gamma}}\left\{-\frac{2\hat{\alpha}\hat{\delta}^2}{\hat{\beta}^3\hat{a}_1^3}\boldsymbol{z}' S\Sigma\boldsymbol{z} + \left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{z}' S\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1} S\Sigma\boldsymbol{z}\right. \\ &\qquad + \left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{z}'\Sigma S\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1} S\boldsymbol{z} \\ &\qquad \left. + \left(\frac{\hat{\alpha}^2\hat{\delta}^2}{\hat{\beta}^4\hat{a}_1^4}\right)\boldsymbol{z}' S\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1}\Sigma\left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1}S\right)^{-1} S\boldsymbol{z}\right\}. \end{aligned} \tag{23}$$

In addition, using (vi)–(viii) in Lemma A.5, (12) and (13), we obtain

$$\frac{1}{p} \boldsymbol{z}' S \Sigma \boldsymbol{z} = O_p(1), \tag{24}$$

$$\begin{aligned} &\frac{2}{p^{2-\gamma}} \left| \boldsymbol{z}' S \left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1} S \right)^{-1} S \Sigma \boldsymbol{z} \right| \\ &\qquad \leq 2 \sqrt{\frac{\boldsymbol{z}' S^2 \boldsymbol{z}}{p^{2-\gamma}}} \sqrt{\frac{\boldsymbol{z}' \Sigma S (I_p + \hat{\alpha}/(\hat{\beta}\hat{a}_1) S)^{-2} S \Sigma \boldsymbol{z}}{p^{2-\gamma}}} \\ &\qquad = \begin{cases} O_p(1), & \Sigma \neq \sigma I, \\ O_p(1), & \Sigma = \sigma I, \end{cases} \end{aligned} \tag{25}$$

$$\begin{aligned} &\frac{1}{p^{2-\gamma}} \left| \boldsymbol{z}' S \left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1} S \right)^{-1} \Sigma \left(I_p + \frac{\hat{\alpha}}{\hat{\beta}\hat{a}_1} S \right)^{-1} S \boldsymbol{z} \right| \\ &\qquad = \begin{cases} O_p(1), & \Sigma \neq \sigma I, \\ O_p(1), & \Sigma = \sigma I. \end{cases} \end{aligned} \tag{26}$$

So, (iv) follows from (23)–(26). □

We are ready to prove Lemmas 2 and 3.

Proof of Lemma 2. Using Lemma A.6, under the assumptions A1–3,

$$\frac{1}{p^{1-\gamma}} \left| U - \left(\frac{\hat{\delta}}{\hat{\beta}\hat{a}_1} \right) \tilde{U} \right| = O_p(p^{-1+\gamma}), \tag{27}$$

$$\frac{1}{p^{1-\gamma}} \left| V - \left(\frac{\hat{\delta}}{\hat{\beta}\hat{a}_1} \right)^2 \tilde{V} \right| = O_p(p^{-1+\gamma}), \tag{28}$$

where

$$\frac{\tilde{U}}{p^{1-\gamma}} = -\frac{1}{2} \Delta_0 - \frac{(N_1 - N_2) p^\gamma}{2 N_1 N_2} \frac{\boldsymbol{z}_2' \boldsymbol{z}_2}{p}, \qquad \frac{\tilde{V}}{p^{1-\gamma}} = \Delta_1 + \frac{N p^\gamma}{N_1 N_2} \frac{\boldsymbol{z}_2' \Sigma \boldsymbol{z}_2}{p},$$

$$\Delta_0 = \frac{\boldsymbol{\delta}' \boldsymbol{\delta}}{p^{1-\gamma}}, \qquad \Delta_1 = \frac{\boldsymbol{\delta}' \Sigma \boldsymbol{\delta}}{p^{1-\gamma}}.$$

Using (27) and (28), we obtain

$$\left| \frac{U}{\sqrt{V}} - \xi \right| = O_p(p^{-(1-\gamma)/2}) (= o_p(1)), \tag{29}$$

where $\xi = p^{(1-\gamma)/2}(\tilde{U}/p^{1-\gamma})/(\tilde{V}/p^{1-\gamma})^{1/2}$. Using Lemma A.4, under the assumptions A1–3, we obtain

$$\frac{1}{p^{1-\gamma}}|\tilde{U} - \tilde{U}_0| = O_p(p^{-1/2}), \qquad \frac{1}{p^{1-\gamma}}|\tilde{V} - \tilde{V}_0| = O_p(p^{-1/2}),$$

where

$$\frac{\tilde{U}_0}{p^{1-\gamma}} = -\frac{1}{2}\Delta_0 - \frac{(N_1 - N_2)p^{\gamma}}{2N_1N_2}a_1, \qquad \frac{\tilde{V}_0}{p^{1-\gamma}} = \Delta_1 + \frac{Np^{\gamma}}{N_1N_2}a_2.$$

Hence

$$|\xi - \xi_0| = O_p(p^{-\gamma/2})(= o_p(1)). \tag{30}$$

Combining (29) and (30), we obtain Lemma 2. □

Proof of Lemma 3. For $\forall\varepsilon \in (0, \infty)$ and $\forall\nu \in (0, 1-\gamma)$,

$$P(e^{p^{\nu}}|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| > \varepsilon) = J_1 + J_2,$$

where

$$J_1 = P(\{|U/\sqrt{V} - \xi_0| > c\} \cap \{e^{p^{\nu}}|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| > \varepsilon\}),$$

$$J_2 = P(\{|U/\sqrt{V} - \xi_0| \le c\} \cap \{e^{p^{\nu}}|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| > \varepsilon\}).$$

Here, c is some positive constant which satisfies $|\xi_0| > c$. From Lemma 2, $J_1 \to 0$ under the assumptions A1–3. Now, we evaluate the order of J_2 when $\xi_0 < 0$. It can be expressed as

$$\begin{aligned} e^{p^{\nu}}|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| &= e^{p^{\nu}} \int_{\min(U/\sqrt{V}, \xi_0)}^{\max(U/\sqrt{V}, \xi_0)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\ &\le \frac{1}{\sqrt{2\pi}} e^{p^{\nu}} e^{-(\max(U/\sqrt{V}, \xi_0))^2/2} |U/\sqrt{V} - \xi_0| \\ &\le \frac{c}{\sqrt{2\pi}} e^{p^{\nu}} e^{-(\xi_0 + c)^2/2}. \end{aligned}$$

The right-hand-side of the last inequality converges to 0 under the assumptions A1–3. Thus, $J_2 \to 0$ under the assumptions A1–3. So, we proved $|\Phi(U/\sqrt{V}) - \Phi(\xi_0)| = o_p(e^{-p^{\nu}})$. Similarly, we can prove $J_2 \to 0$ when $\xi_0 > 0$. □

Acknowledgement

This work was supported by Japan Society for the Promotion of Science (JSPS). The authors would like to thank the referee for valuable comments and careful reading of the manuscript. They would also like to thank Professor Yasunori Fujikoshi for his advice and encouragement.

References

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

[2] N. El Karoui, On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity, arXiv:math.ST/0309355 (2003).

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

[4] A. K. Gupta and D. K. Nagar, Matrix Variate Distributions, Chapman and Hall, London, 1999.

[5] M. Hyodo and T. Yamada, Asymptotic distribution of studentized contribution ratio in high-dimensional principal component analysis, Comm. Statist. Simulation Comput., **38** (2009), 905–917.

[6] O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Anal., **88** (2004), 365–411.

[7] A. M. Mathai, B. P. Serge and T. Hayakawa, Bilinear Forms and Zonal Polynomials, Springer-Verlag, Lecture Notes in Statistics, No.**102**, New York, 2009.

[8] J. R. Schott, Matrix Analysis for Statistics, John Wiley and Sons, 1997.

[9] M. Siotani, Large sample approximations and asymptotic expansions of classification statistic, Handbook of Statistics, **2** (2004) (P. R. Krishnaiah and L. N. Kanal, Eds.), North-Holland Publishing Company, 47–60.

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

Masashi Hyodo
JSPS Research Fellow
Graduate School of Economics
The University of Tokyo
7-3-1 *Hongo Bunkyo-ku Tokyo* 113-0033, *Japan*
E-mail: caicmhy@gmail.com

Takayuki Yamada
Division of Biostatistics
Department of Clinical Medicine
School of Pharmacy
Kitasato University
5-9-1 *Shirokane Minato-ku Tokyo* 108-8041, *Japan*

Tetsuto Himeno
Transdisciplinary Research Integration Center
Research Organization of Information and Systems
Kamiyacho 2*F* 4-3-13 *Toranomon Minato-ku Tokyo* 105-0001*, Japan*

Takashi Seo
Department of Mathematical Information Science
Tokyo University of Science
1-3 *Kagurazaka Shinjuku-ku Tokyo* 162-8601*, Japan*