Selection of the linear and the quadratic discriminant functions when the difference between two covariance matrices is small

Tomoyuki Nakagawa and Hirofumi Wakaki
Department of Mathematics, Graduate School of Science, Hiroshima University

(Last modified: April 19, 2016)

Abstract

We consider selecting of the linear and the quadratic discriminant functions in two normal populations. We do not know which of two discriminant functions lowers the expected probability of misclassification. When difference of the covariance matrices is large, it is known that the expected probability of misclassification of the quadratic discriminant functions is smaller than that of linear discriminant function. Therefore, we should consider only the selection when the difference between covariance matrices is small. In this paper we suggest a selection method using asymptotic expansion for the linear and the quadratic discriminant functions when the difference between the covariance matrices is small.

1 Introduction

We consider classifying an individual coming from one of two populations Π_1 and Π_2. We assume that Π_j is p-variate normal population with mean vector $\boldsymbol{\mu}_j$ and the covariance matrix $\boldsymbol{\Sigma}_j (j = 1, 2)$, that is,

$$\Pi_1 : \mathrm{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \quad \Pi_2 : \mathrm{N}_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2).$$

When population parameters are known, the optimal Bayes discriminant function is given as follows (see Anderson, 2003).
In the case that the covariance matrices are equal, that is, $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, the optimal Bayes discriminant function is

$$L(\boldsymbol{X}; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1} \left\{\boldsymbol{X} - (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)/2\right\},$$

where $\boldsymbol{a}'$ is the transpose of $\boldsymbol{a}$. In the case that the covariance matrices are unequal, that is, $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, the optimal Bayes discriminant function is

$$Q(\boldsymbol{X}; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) = (\boldsymbol{X} - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_1) - (\boldsymbol{X} - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_2) + \log |\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1|,$$

where $|\boldsymbol{A}|$ is the determinant of $\boldsymbol{A}$. $L(\boldsymbol{X})$ and $Q(\boldsymbol{X})$ are called the linear and the quadratic discriminant functions, respectively.

However, the population parameters are unknown in practice. Therefore, it is necessary for us to estimate the population parameters. A sample of size N_j coming from Π_j, $\boldsymbol{X}_{j1}, \ldots, \boldsymbol{X}_{jN_j}$ is available to estimate these parameters. Let $\bar{\boldsymbol{X}}_j$ and $\boldsymbol{S}_j$ be the sample mean and the sample covariance matrix of Π_j, respectively, and let $\boldsymbol{S}$ be the pooled sample covariance matrix, that is,

$$\begin{aligned} \bar{\boldsymbol{X}}_j &= \frac{1}{N_j}\sum_{k=1}^{N_j} \boldsymbol{X}_{jk}, \quad \boldsymbol{S}_j = \frac{1}{n_j}\sum_{k=1}^{N_i}(\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j)(\boldsymbol{X}_{jk} - \bar{\boldsymbol{X}}_j)', \\ \boldsymbol{S} &= k_1\boldsymbol{S}_1 + k_2\boldsymbol{S}_2, \end{aligned} \tag{1.1}$$

where $n = n_1 + n_2$ and $k_j = n_j/n$ with $n_j = N_j - 1$ $(j = 1, 2)$. Replacing the unknown parameters with these estimators, we obtain $\hat{L}(\boldsymbol{X}) = L(\boldsymbol{X}; \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S})$ and $\hat{Q}(\boldsymbol{X}) = Q(\boldsymbol{X}; \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S}_1, \boldsymbol{S}_2)$, respectively. $\hat{L}(\boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X})$ are called the sample linear and the sample quadratic discriminant functions, respectively. An observation $\boldsymbol{X}$ is classified, for example into Π_1 for negative value of these functions.

If the covariance matrices are unequal, $\hat{Q}(\boldsymbol{X})$ is better than $\hat{L}(\boldsymbol{X})$ for large samples, since $\hat{Q}(\boldsymbol{X})$ is a consistent estimator of the optimal Bayes discriminant function while $\hat{L}(\boldsymbol{X})$ is not consistent. But even if the covariance matrices are unequal, $\hat{Q}(\boldsymbol{X})$ is not always better than $\hat{L}(\boldsymbol{X})$ for small samples. When the difference between the covariance matrices is large, $\hat{Q}(\boldsymbol{X})$ is better than $\hat{L}(\boldsymbol{X})$ (see Marks and Dunn, 1974; Wahl and Kronmal, 1977). These previous works are simulation study. Wakaki (1990) investigated performance of the two discriminant functions for moderately large samples using asymptotic expansions of the distributions of two discriminant functions under covariance matrices are proportional, that is, $\boldsymbol{\Sigma}_2 = c\boldsymbol{\Sigma}_1$ (c: constant).

We investigate performance of the discriminant functions using asymptotic expansions of the distributions of two discriminant functions under the difference between the covariance matrices is small, that is

$$\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2 = \frac{1}{\sqrt{n}}\boldsymbol{A} \quad (\boldsymbol{A}\text{: constant matrix}). \tag{1.2}$$

We suggest a selection method for the discriminant functions using these asymptotic expansions.

The remainder of the present paper is organized as follows. In Section 2, we give asymptotic expansions for two discriminant functions when difference between two covariance matrices is small. In Section 3, we will suggest a selection method for the linear and the quadratic discriminant functions by using these asymptotic expansions. In Section 4, we will perform numerical study for investigating the performance of our selection method. In Section 5, we present a discussion and our conclusions.

2 Asymptotic expansion of the linear and the quadratic discriminant functions

We consider the following asymptotic framework:

$$N_1 \to \infty, \quad N_2 \to \infty, \quad \frac{N_1}{N_2} = O(1), \quad \frac{N_2}{N_1} = O(1).$$

We can assume the following condition without loss of generality,

$$\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 = \mathbf{0}, \quad k_1\boldsymbol{\Sigma}_1 + k_2\boldsymbol{\Sigma}_2 = \boldsymbol{I}_p. \tag{2.1}$$

2.1 Asymptotic expansion for the linear discriminant function

Wakaki (1990) proposed Theorem 2.1 which derived an asymptotic expansion of the sample linear discriminant function in a general case where covariance matrices are unequal.

Theorem 2.1. *Let $\boldsymbol{X}$ be an observation vector coming from $\Pi_j (j = 1, 2)$ and let*

$$L_j^* = (\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{d})^{-1/2}\{\hat{L}(\boldsymbol{X}) + \boldsymbol{\mu}_j'\boldsymbol{d}\},$$

where $\boldsymbol{d} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. Then for large N_1 and N_2,

$$P(L_j^* \le x) = \Phi(x) + \sum_{i=1}^{2} N_i^{-1}\sum_{s=1}^{4}(\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{d})^{-s/2}a_{jis}^* H_{s-1}(x)\phi(x) + O_2, \tag{2.2}$$

where $\Phi(.)$ and $\phi(.)$ are the distribution function and the density function of $\mathrm{N}(0, 1)$, $H_s(.)$'s are the Hermite polynomials, and O_m stands for the terms of the m-th order with respect to N_1^{-1} and N_2^{-1}. The coefficient a_{jis}^*'s are given by

$$\begin{aligned}
a_{ji1}^* &= -\frac{1}{2}(-1)^{i+1}\mathrm{tr}(\boldsymbol{\Sigma}_i) + k_i^2\{\boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i^2\boldsymbol{d} + \mathrm{tr}(\boldsymbol{\Sigma}_i)\boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{d}\},\\
a_{ji2}^* &= -\frac{1}{2}\{\mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i) + (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)\}\\
&\quad -\frac{1}{2}k_i^2\{\mathrm{tr}(\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i)\boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{d} + 2\mathrm{tr}(\boldsymbol{\Sigma}_i)\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{d} + 2\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i^2\boldsymbol{d} + \frac{1}{2}(\boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{d})^2\},\\
a_{ji3}^* &= (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{d} + 2k_i^2\boldsymbol{\mu}_j'\boldsymbol{\Sigma}_i\boldsymbol{d}\boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{d},\\
a_{ji4}^* &= -\frac{1}{2}\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{d} - \frac{1}{2}k_i^2\{\boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{d}\boldsymbol{d}'\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{d} + (\boldsymbol{d}'\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{d})^2\}.
\end{aligned}$$

We can easily obtain the following corollary under the assumption (1.2).

Corollary 2.1. *Suppose that the condition of Theorem2.1 and (1.2) hold. Let*

$$L_j = (\boldsymbol{d}'\boldsymbol{d})^{-1/2}\{\hat{L}(\boldsymbol{X}) + \boldsymbol{\mu}_j'\boldsymbol{d}\}.$$

Then for large N_1 and N_2,

$$\begin{aligned}
P(L_j \le x) = \Phi(x) + \phi(x)\Bigg[&-\frac{k_{\sigma(j)}}{2\sqrt{n}}D_0^{-1}D_1H_1(x) - \frac{k_{\sigma(j)}}{8n}D_0^{-2}D_1^2H_3(x)\\
&+\sum_{i=1}^{2} n_i^{-1}\sum_{s=1}^{4}(D_0)^{-s/2}a_{jis}H_{s-1}(x)\Bigg] + O_{3/2},
\end{aligned} \tag{2.3}$$

where $D_k = \boldsymbol{d}'\boldsymbol{A}^k\boldsymbol{d}$*,* $\sigma(1) = \sigma(2) + 1 = 2$*. The coefficient* a_{jis}*'s are given by*

$$\begin{aligned}
a_{ji1} &= -(-1)^{i+1}\frac{p}{2} + k_i^2\{\boldsymbol{\mu}_j'\boldsymbol{d} + p\boldsymbol{\mu}_j'\boldsymbol{d}\},\\
a_{ji2} &= \frac{1}{2}\{p + (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)'(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)\} - \frac{1}{2}k_i^2\{3(p+1)D_0 + D_0^2/2\},\\
a_{ji3} &= (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)'\boldsymbol{d} + 2k_i^2\boldsymbol{\mu}_j'\boldsymbol{d}D_0,\\
a_{ji4} &= -\frac{1}{2}D_0 - k_i^2 D_0^2.
\end{aligned}$$

2.2 Asymptotic expansion for the quadratic discriminant function

We derive an asymptotic expansion of the sample quadratic discriminant function under the assumption (1.2). The following lemmas are very important.

Lemma 2.1. *If* $\boldsymbol{W}$ *is a random matrix distributed as Wishart distribution* $W_p(n, \boldsymbol{\Sigma})$*, where* n *is a positive integer and* $\boldsymbol{Y}$ *is any* p*-variate random vector which is independent of* $\boldsymbol{W}$ *with* $P(\boldsymbol{Y} = \boldsymbol{0}) = 0$ *then*

$$\boldsymbol{Y}'\boldsymbol{W}^{-1}\boldsymbol{Y} + \log|\boldsymbol{W}| \sim V_{n-p+1}^{-1}\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} + \log|\boldsymbol{\Sigma}| + \sum_{i=1}^{p}\log V_{n-i+1}, \quad V_{n-i+1} \sim \chi^2_{n-i+1},$$

where χ^2_m *is the chi-square distribution with* m *degrees of freedom. Moreover,* $\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}$ *and* $V_{n-i+1}(i = 1, \ldots, p)$ *are mutually independent.*

Lemma 2.2 (characteristic function of noncentral chi-square distribution)**.** *Let* Z *be a random variable distributed as normal distribution with mean* μ *and variance* 1*. Then*

$$\mathrm{E}[\exp(itZ^2)] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}(z-\mu)^2 + itz^2\right\}dz = (1-2it)^{-1/2}\exp\left\{\frac{\mu^2 it}{1-2it}\right\}.$$

The proofs of the above two lemmas can be seen in Muirhead (1982) and Fujikoshi et al. (2010).

Lemma 2.3. *Let* $\boldsymbol{X}$ *be* p*-variate normal random vector with mean vector* $\boldsymbol{0}$ *and covariance matrix* $\boldsymbol{I}_p$*, and let*

$$\begin{aligned}
g(t_1,t_2) &= g(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = \mathrm{E}[\exp\{t_1(\boldsymbol{X}-\boldsymbol{\eta}_1)\boldsymbol{\Gamma}_1(\boldsymbol{X}-\boldsymbol{\eta}_1) + t_2(\boldsymbol{X}-\boldsymbol{\eta}_2)\boldsymbol{\Gamma}_2(\boldsymbol{X}-\boldsymbol{\eta}_2)\}],\\
L(t_1,t_2) &= L(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = \mathrm{tr}[(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1]\\
&\quad + \{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}.
\end{aligned}$$

Then

$$\begin{aligned}
g(t_1,t_2) &= |\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\eta}_1'\boldsymbol{\eta}_1 + t_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)'\boldsymbol{\Gamma}_2\;(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\right.\\
&\qquad\left. + \frac{1}{2}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}\right\},\\
\frac{\partial g(t_1,t_2)}{\partial t_1} &= g(t_1,t_2)L(t_1,t_2),\\
\frac{\partial L(t_1,t_2)}{\partial t_1} &= L(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = 2\mathrm{tr}[\{(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1\}^2]\\
&\qquad + 4\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\Gamma}_1(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\}^2\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\},\\
g(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) &= g(t_2,t_1;\boldsymbol{\eta}_2,\boldsymbol{\eta}_1,\boldsymbol{\Gamma}_2,\boldsymbol{\Gamma}_1),\\
\frac{\partial g(t_1,t_2)}{\partial t_2} &= g(t_1,t_2)L(t_2,t_1;\boldsymbol{\eta}_2,\boldsymbol{\eta}_1,\boldsymbol{\Gamma}_2,\boldsymbol{\Gamma}_1).
\end{aligned}$$

The proof is given in Appendix. Using these lemmas, We obtain an asymptotic expansion of $\hat{Q}(\boldsymbol{X})$ under the assumption (1.2) as in the following theorem.

Theorem 2.2. *Let* $\boldsymbol{X}$ *be an observation vector coming from* $\Pi_j(j = 1, 2)$*, and let* $Q_j = (\boldsymbol{d}'\boldsymbol{d})^{-1/2}\{\hat{Q}(\boldsymbol{X})/2 + \boldsymbol{\mu}_j'\boldsymbol{d}\}$*. Then for large* N_1 *and* N_2*,*

$$P(Q_j \le x) = \Phi(x) + \phi(x)n^{-1/2}\sum_{s=1}^{3}b_{js}H_{s-1}(x) + \sum_{i=0}^{2}n_i^{-1}\sum_{s=1}^{6}(\boldsymbol{d}'\boldsymbol{d})^{-s/2}b_{jis}H_{s-1}(x) + O_{3/2},$$

where $n_0 = n$. *The coefficient* b_{js}*'s and* b_{jis}*'s are given by*

$$\begin{aligned}
&b_{j1} = (-1)^{j-1}k_jD_1/2, \quad b_{j2} = -(1+k_j)D_1/2, \quad b_{j3} = (-1)^{j-1}D_1/2,\\
&b_{j01} = (-1)^{j-1}\{T_2/4 + k_j^2D_2/2\}, \quad b_{j02} = -T_2/4 - (k_j^2+2k_j)D_2/2 - k_{\sigma(j)}^2D_1^2/8,\\
&b_{j03} = (-1)^{j-1}\left\{(1+2k_j)D_2/2 + k_j(1+k_j)D_1^2/4\right\}, \quad b_{j04} = D_2/2 - (1+4k_j+k_j^2)D_1^2/8,\\
&b_{j05} = (-1)^{j-1}(1+k_j)D_1^2/4, \quad b_{j06} = -D_1^2/8,\\
&b_{ji1} = (-1)^j(p^2+3p)/4 + (p+1)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)'\boldsymbol{d}/2,\\
&b_{ji2} = -(p^2+3p)/4 - (4p+5)((\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i))^2/4 - (p+2)(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)/2,\\
&b_{ji3} = -(p+1)\boldsymbol{\mu}_j'\boldsymbol{d}D_0 + (p+4)(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'\boldsymbol{d}D_0/2 + (p+2)(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'\boldsymbol{d}/2,\\
&b_{ji4} = -(p+1)D_0/2 - 3((\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i))^2/2,\\
&b_{ji5} = (\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'\boldsymbol{d}D_0,\\
&b_{ji6} = ((\boldsymbol{\mu}_j-\boldsymbol{\mu}_i)'(\boldsymbol{\mu}_j-\boldsymbol{\mu}_i))^2/4,
\end{aligned}$$

where $T_k = \mathrm{tr}(A^k)$.

Proof.
Since $\boldsymbol{X}_{j1}, \ldots, \boldsymbol{X}_{jN_j}$ are normal random vectors with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$ and are mutually independent, $\bar{\boldsymbol{X}}_j$ is distributed as p-variate normal distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $N_j^{-1}\boldsymbol{\Sigma}_j$, and $n_j\boldsymbol{S}_j$ is distributed as Wishart distribution $W_p(n_j, \boldsymbol{\Sigma}_j)$. Suppose that V_{jk} is the chi-square random variable with degrees of freedom $n_j - k + 1$. From Lemma 2.1 and 2.2,

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\frac{it}{2}\left((\boldsymbol{X}-\bar{\boldsymbol{X}}_j)'\boldsymbol{S}_j^{-1}(\boldsymbol{X}-\bar{\boldsymbol{X}}_j) + \log|\boldsymbol{S}_j|\right)\right\}\middle| \boldsymbol{X}, \boldsymbol{S}_j\right]\\
&= \mathrm{E}\left[\exp\left\{\frac{it}{2}\left(n_jV_{jp}^{-1}(\boldsymbol{X}-\bar{\boldsymbol{X}}_j)'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{X}-\bar{\boldsymbol{X}}_j) + \log|n_j^{-1}\boldsymbol{\Sigma}_j| + \sum_{k=1}^p \log V_{jk}\right)\right\}\middle| \boldsymbol{X}, V_{jk}, k=1,\ldots,p\right]\\
&= \exp\left[\frac{itn_j}{2V_{jp}}\Omega_j\left(1 - \frac{itn_j}{N_jV_{jp}}\right)\right]\left(1 - \frac{itn_j}{N_jV_{jp}}\right)^{-p/2}\left(\exp\left\{\frac{it}{2}\left(\log|n_j^{-1}\boldsymbol{\Sigma}_j| + \sum_{k=1}^p\log V_{jk}\right)\right\}\right).
\end{aligned}$$

Let

$$v_{jk} = \sqrt{\frac{n_j-k+1}{2}}\left(\frac{V_{jk}}{n_j-k+1} - 1\right),$$

then $v_{jk} = O_p(1)$ follows from the central limit theorem. The following result is given by using above formulae.

$$\begin{aligned}
&\exp\left[\frac{itn_j}{2V_{jp}}\Omega_j\left(1-\frac{itn_j}{N_jV_{jp}}\right)\right]\left(1-\frac{itn_j}{N_jV_{jp}}\right)^{-p/2}\\
&= \mathrm{e}^{it\Omega_j/2}\left[1 + \frac{it}{2}\left\{\frac{p}{n_j} + \Omega_j\left(\frac{p-1}{n_j} + \frac{it}{n_j} - \sqrt{\frac{2}{n_j}}v_{jp} + \frac{2}{n_j}v_{jp}^2\right)\right\} + \Omega_j^2\frac{(it)^2}{4n_j}v_{jp}^2\right] + O_p(n_j^{-3/2}),
\end{aligned} \tag{2.4}$$

where $\Omega_j = (\boldsymbol{\mu}_j - \boldsymbol{X})'\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\mu}_j - \boldsymbol{X})$.

$$\exp\left\{\frac{it}{2}\left(\sum_{k=1}^p \log V_{jk}\right)\right\} = 1 + \frac{it}{2}\sum_{k=1}^p\left(\frac{v_{jk}}{\sqrt{n_j}} - \frac{v_{jk}^2}{n_j}\right) + \frac{(it)^2}{8}\left(\sum_{k=1}^p\frac{v_{jk}}{\sqrt{n_j}}\right)^2 + O_p(n_j^{-3/2}). \tag{2.5}$$

From (2.4), (2.5), $\mathrm{E}[v_{jk}] = 0$, $\mathrm{E}[v_{jk}^2] = 1$ and Lemma 2.1,

$$\begin{aligned}
&\mathrm{E}\left[\exp\left\{\frac{it}{2}\left((\boldsymbol{X}-\bar{\boldsymbol{X}}_j)'\boldsymbol{S}_j^{-1}(\boldsymbol{X}-\bar{\boldsymbol{X}}_j) + \log|\boldsymbol{S}_j|\right)\right\}\middle| \boldsymbol{X}\right]\\
&= \exp(it\Omega_j)\left[1 + \frac{1}{n_j}\left\{\frac{it}{2}\left(-\frac{p(p-1)}{2} + (p+1)\Omega_j\right) + \frac{(it)^2}{4}(p+\Omega_j^2)\right\}\right] + O_p(n_j^{-3/2}).
\end{aligned} \tag{2.6}$$

Suppose that $\boldsymbol{X}$ belongs to Π_j, and let

$$\psi_j(t) = \mathrm{E}[\mathrm{e}^{itQ(\boldsymbol{X})/2}].$$

From Lemma 2.3 and (2.6),

$$\psi_j(t) = g(it/2, -it/2)|\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}|^{it/2}\left[1 + \frac{1}{n_1}\left\{\frac{it}{2}\left(-\frac{p(p-1)}{2} + (p+1)L_1\right) + \frac{(it)^2}{4}(p + (L_1^2 + L_{11}))\right\}\right.$$
$$\left. + \frac{1}{n_2}\left\{-\frac{it}{2}\left(-\frac{p(p-1)}{2} + (p+1)L_2\right) + \frac{(it)^2}{4}(p + (L_2^2 + L_{22}))\right\}\right] + O_{3/2}, \quad (k = 1, 2),$$

where

$$L_k = \frac{1}{g(t_1,t_2)}\frac{\partial g(t_1,t_2)}{\partial t_k}\bigg|_{(t_1,t_2)=(it/2,-it/2)}, \quad L_{kk} = \frac{\partial}{\partial t_k}\frac{1}{g(t_1,t_2)}\frac{\partial g(t_1,t_2)}{\partial t_k}\bigg|_{(t_1,t_2)=(it/2,-it/2)}, \quad (k = 1, 2).$$

From (1.2) and (2.1), we have

$$\boldsymbol{\Sigma}_1 = \boldsymbol{I}_p + \frac{k_2}{\sqrt{n}}\boldsymbol{A}, \quad \boldsymbol{\Sigma}_2 = \boldsymbol{I}_p - \frac{k_1}{\sqrt{n}}\boldsymbol{A}. \tag{2.7}$$

For $j = 1$, the parameters of g and L are given as follows.

$$\boldsymbol{\eta}_1 = \boldsymbol{0}, \quad \boldsymbol{\eta}_2 = \boldsymbol{\Sigma}_1^{-1/2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1), \quad \boldsymbol{\Gamma}_1 = \boldsymbol{I}_p, \quad \boldsymbol{\Gamma}_2 = \boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_1^{1/2}. \tag{2.8}$$

From (2.7) and (2.8), we obtain the following expansion.

$$g(it/2, -it/2)|\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{-1}|^{it/2} = \exp\left[-\frac{it}{2}(1-it)D_0 + \frac{1}{\sqrt{n}}\left\{\frac{k_2(1-it)}{2}D_1 - \frac{(k_2+it)(1-it)}{2}D_1\right\}\right.$$
$$\left. + \frac{1}{n}\left\{\frac{(it)^2 - it}{4}T_2 - \frac{k_2^2(1-it)}{2}D_2 + \frac{(1-it)^2\{(it)^2 + (k_2-k_1)it + k_2^2\}}{2}D_2\right\}\right] + O_{3/2},$$

$$L_1 = p + (it)^2 D_0 + O_{1/2}, \quad L_2 = p + (1-it)^2 D_0 + O_{1/2},$$
$$L_{11} = 2p + 4(it)^2 D_0 + O_{1/2}, \quad L_{22} = 2p + 4(1-it)^2 D_0 + O_{1/2}.$$

Hence, we obtain the following expansion of ψ_1.

$$\psi_1(t) = \exp\left\{-\frac{it(1-it)}{2}D_0\right\}\left[1 + \frac{1}{\sqrt{n}}\left\{\frac{1}{2}(-k_1 it + (1+k_1)(it)^2 - (it)^3\right\}D_1\right.$$
$$+ \frac{1}{n}\left\{\frac{(it)^2 - it}{4}T_2 + \frac{1}{2}\left\{-k_1^2 it + (k_1^2 + 2k_1)(it)^2 - (1+2k_1)(it)^3 + (it)^4\right\}D_2\right.$$
$$\left. + \frac{1}{8}\left\{k_2 it + (1+k_1)(it)^2 - \frac{(it)^3}{2}\right\}^2 D_1^2\right\}$$
$$+ \frac{1}{n_1}\left\{\frac{it}{2}\left(\frac{p^2+3p}{2} + (p+1)(it)^2 D_0\right) + \frac{(it)^2}{4}\left\{p + (p + (it)^2 D_0)^2 + 2p + 4(it)^2 D_0\right\}\right\}$$
$$\left. + \frac{1}{n_2}\left\{-\frac{it}{2}\left(\frac{p^2+3p}{2} + (p+1)(1-it)^2 D_0\right) + \frac{(it)^2}{4}\left\{p + (p + (1-it)^2 D_0)^2 + 2p + 4(1-it)^2 D_0\right\}\right\}\right] + O_{3/2}.$$

The expansion of ψ_2 is given by replacing the parameters (it, k_1, k_2, n_1, n_2) of ψ_1 with $(-it, k_2, k_1, n_2, n_1)$. Inverting $\psi_j (j = 1, 2)$ formally, we obtain the desired result. □

3 The criterion for selecting between the linear and quadratic discriminant functions

3.1 Derivation of the criterion

In this section, we consider the expected misclassification probabilities P_L and P_Q of $\hat{L}(\boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X})$, respectively. We assume that a priori probabilities are $1/2$. Then P_L and P_Q are given by

$$P_L = \frac{1}{2}\left\{1 - P\left(L_1 \le \frac{1}{2}(\boldsymbol{d}'\boldsymbol{d})^{1/2}\right) + P\left(L_2 \le -\frac{1}{2}(\boldsymbol{d}'\boldsymbol{d})^{1/2}\right)\right\},$$
$$P_Q = \frac{1}{2}\left\{1 - P\left(Q_1 \le \frac{1}{2}(\boldsymbol{d}'\boldsymbol{d})^{1/2}\right) + P\left(Q_2 \le -\frac{1}{2}(\boldsymbol{d}'\boldsymbol{d})^{1/2}\right)\right\},$$

respectively.

Theorem 3.1. *We obtain the limiting misclassification probabilities of P_L and P_Q which are equal. Moreover, the term of 1/2-th order with respect to N_1^{-1} and N_2^{-1} of $P_Q - P_L$ is vanished. Let*

$$D(\boldsymbol{A}, \boldsymbol{d}) = \lim_{n\to\infty} n(P_Q - P_L).$$

Then $D(\boldsymbol{A}, \boldsymbol{d})$ is given by following formula.

$$D(\boldsymbol{A}, \boldsymbol{d}) = \phi(D_0^{1/2}/2) \sum_{s=1}^{6} D_0^{-s/2} H_{s-1}(D_0^{1/2}/2) c_s,$$

where the coefficient c_s's are given by

$$\begin{aligned}
c_1 &= -T_2/2 - (k_1^2 + k_2^2)D_2 + \{k_1^2(p+1)D_0 - (p+1)D_0^2\}/k_1 + \{k_2^2(p+1)D_0 - (p+1)D_0^2\}/k_2, \\
c_2 &= T_2/2 + (k_1^2 + k_2^2 + 2)D_2/2 + (k_1^2 + k_2^2)D_1^2/8 \\
&\quad + \{(p^2+p)/2 + (4p+5)D_0^2/4 + (p+1)D_0^2 - k_1^2(3(p+1)D_0 + D_0^2/2)\}/k_1 \\
&\quad + \{(p^2+p)/2 + (4p+5)D_0^2/4 + (p+1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2/2)\}/k_2, \\
c_3 &= -2D_2 - (1 + k_1^2 + k_2^2)D_1^2/4 + \{-D_0^2 - (p+1)D_0 + 2k_1 D_0^2\}/k_1 + \{-D_0^2 - (p+1)D_0 + 2k_2 D_0^2\}/k_2, \\
c_4 &= D_2 + 3D_1^2/4 + \{(p+1)D_0 + 3D_0^2/2 - 2k_1^2 D_0^2\}/k_1 + \{(p+1)D_0 + 3D_0^2/2 - 2k_2^2 D_0^2\}/k_2, \\
c_5 &= -3D_1^2/4 - D_0^2/k_1 - D_0^2/k_2, \\
c_6 &= D_1^2/4 + D_0^2/2k_1 + D_0^2/2k_2.
\end{aligned}$$

Proof. The result is easily obtained from Theorem 2.1 and 2.2. □

We obtain a criterion for selecting between the linear and the quadratic discriminant functions as $D(\boldsymbol{A}, \boldsymbol{d})$. Then if $D(\boldsymbol{A}, \boldsymbol{d})$ is negative, we can consider that $\hat{Q}(\boldsymbol{X})$ is better than $\hat{L}(\boldsymbol{X})$. Otherwise, we can consider that $\hat{L}(\boldsymbol{X})$ is better than $\hat{Q}(\boldsymbol{X})$. However, $\boldsymbol{A}$ and $\boldsymbol{d}$ are unknown parameters which should be estimated. We may consider to use simple estimators,

$$\hat{\boldsymbol{d}} = \boldsymbol{S}^{-1/2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2), \quad \hat{\boldsymbol{A}} = \sqrt{n}\boldsymbol{S}^{-1/2}(\boldsymbol{S}_1 - \boldsymbol{S}_2)\boldsymbol{S}^{-1/2}. \tag{3.1}$$

But it is insufficient for a criterion for selecting between the linear and the quadratic discriminant functions only by replacing the unknown parameters with these estimators. Because, $\hat{\boldsymbol{A}}$ is not consistent, $D(\hat{\boldsymbol{A}}, \hat{\boldsymbol{d}})$ does not converge in probability to $D(\boldsymbol{A}, \boldsymbol{d})$ for large samples. Moreover, $\mathrm{E}[D(\hat{\boldsymbol{A}}, \hat{\boldsymbol{d}})]$ do not converge to $D(\boldsymbol{A}, \boldsymbol{d})$. Therefore, we correct the bias in the next section.

3.2 Correcting the bias

We have the criterion for selecting between the linear and the quadratic discriminant functions in Theorem 3.1. But this include the unknown parameters $\boldsymbol{A}$, $\boldsymbol{d}$. When replacing these parameters with the estimators (3.1), $D(\hat{\boldsymbol{A}}, \hat{\boldsymbol{d}})$ have the asymptotic bias for $D(\boldsymbol{A}, \boldsymbol{d})$. So we will correct the bias of the criterion. The criterion can be given as a linear combination of the following terms.

$$\mathrm{tr}(\boldsymbol{A}^2)(\boldsymbol{d}'\boldsymbol{d})^{l/2}, \quad (\boldsymbol{d}'\boldsymbol{d})^{l/2}\boldsymbol{d}'\boldsymbol{A}^2\boldsymbol{d}, \quad (\boldsymbol{d}'\boldsymbol{d})^{l/2}(\boldsymbol{d}'\boldsymbol{A}\boldsymbol{d})^2, \quad (\boldsymbol{d}'\boldsymbol{d})^{l/2}. \quad (l \in \mathbb{Z}). \tag{3.2}$$

Theorem 3.2. *We obtain the criterion D^* which correct the bias as the following formula. Define*

$$D^*(\boldsymbol{A}, \boldsymbol{d}) = \phi(D_0^{1/2}/2) \sum_{s=1}^{6} D_0^{-s/2} H_{s-1}(D_0^{1/2}/2) c_s^*,$$

where the coefficient c_s^*'s are given by

$$\begin{aligned}
c_1^* &= -\{T_2/2 - (p^2+p)/k_1 - (p^2+p)/k_2\} - (k_1^2+k_2^2)\{D_2 - (p+1)D_0/k_1 - (p+1)D_0/k_2\} \\
&\quad + \{k_1^2(p+1)D_0 - (p+1)D_0^2\}/k_1 + \{k_2^2(p+1)D_0 - (p+1)D_0^2\}/k_2, \\
c_2^* &= \{T_2/2 - (p^2+p)/k_1 - (p^2+p)/k_2\}/2 + (k_1^2+k_2^2+2)\{D_2 - (p+1)D_0/k_1 - (p+1)D_0/k_2\}/2 \\
&\quad + (k_1^2+k_2^2)\{D_1^2 - 2D_0^2/k_1 - 2D_0^2/k_2\}/8 \\
&\quad + \{(p^2+p)/2 + (4p+5)D_0^2/4 + (p+1)D_0^2 - k_1^2(3(p+1)D_0 + D_0^2/2)\}/k_1 \\
&\quad + \{(p^2+p)/2 + (4p+5)D_0^2/4 + (p+1)D_0^2 - k_2^2(3(p+1)D_0 + D_0^2/2)\}/k_2, \\
c_3^* &= -2\{D_2 - (p+1)D_0/k_1 - (p+1)D_0/k_2\} - (1+k_1^2+k_2^2)\{D_1^2 - 2D_0^2/k_1 - 2D_0^2/k_2\}/4 \\
&\quad + \{-D_0^2 - (p+1)D_0 + 2k_1D_0^2\}/k_1 + \{-D_0^2 - (p+1)D_0 + 2k_2D_0^2\}/k_2, \\
c_4^* &= \{D_2 - (p+1)D_0/k_1 - (p+1)D_0/k_2\} + 3\{D_1^2 - 2D_0^2/k_1 - 2D_0^2/k_2\}/4 \\
&\quad + \{(p+1)D_0 + 3D_0^2/2 - 2k_1^2D_0^2\}/k_1 + \{(p+1)D_0 + 3D_0^2/2 - 2k_2^2D_0^2\}/k_2, \\
c_5^* &= -3\{D_1^2 - 2D_0^2/k_1 - 2D_0^2/k_2\}/4 - D_0^2/k_1 - D_0^2/k_2, \\
c_6^* &= \{D_1^2 - 2D_0^2/k_1 - 2D_0^2/k_2\}/4 + D_0^2/2k_1 + D_0^2/2k_2.
\end{aligned}$$

Then

$$\mathrm{E}[D^*(\hat{\boldsymbol{A}}, \hat{\boldsymbol{d}})] = D(\boldsymbol{A}, \boldsymbol{d}) + O(n^{-1/2}).$$

Proof.
From the central limit theorem, we can obtain the following,

$$\boldsymbol{Z}_j = \sqrt{N_j}(\bar{\boldsymbol{X}}_j - \boldsymbol{\mu}_j) = O_p(1), \quad \boldsymbol{W}_j = \sqrt{n_j}(\boldsymbol{S}_j - \boldsymbol{\Sigma}_j) = O_p(1), \quad (j = 1, 2).$$

Then the following statistics are expanded as

$$\begin{aligned}
\mathrm{tr}(\hat{\boldsymbol{A}})(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2} &= \mathrm{tr}\left\{\sqrt{n}\boldsymbol{S}^{-1}(\boldsymbol{S}_1 - \boldsymbol{S}_2)\right\}\left\{(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\right\}^{l/2} \\
&= \mathrm{tr}\left\{(\boldsymbol{A} + k_1^{-1/2}W_1 - k_2^{-1/2}W_2)^2\right\}(\boldsymbol{d}'\boldsymbol{d})^{l/2} + O_p(n^{-1/2}), \\
(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}\hat{\boldsymbol{d}}'\hat{\boldsymbol{A}}^k\hat{\boldsymbol{d}} &= \left\{(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\right\}^{l/2}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}\{\sqrt{n}(\boldsymbol{S}_1 - \boldsymbol{S}_2)\}\boldsymbol{S}^{-1}\{\sqrt{n}(\boldsymbol{S}_1 - \boldsymbol{S}_2)\}\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2) \\
&= (\boldsymbol{d}'\boldsymbol{d})^{l/2}\boldsymbol{d}\left\{\boldsymbol{A} + k_1^{-1/2}W_1 - k_2^{-1/2}W_2\right\}^2\boldsymbol{d} + O_p(n^{-1/2}), \\
(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}(\hat{\boldsymbol{d}}'\hat{\boldsymbol{A}}\hat{\boldsymbol{d}})^2 &= \left\{(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\right\}^{l/2}\left\{(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}\{\sqrt{n}(\boldsymbol{S}_1 - \boldsymbol{S}_2)\}\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\right\}^2 \\
&= (\boldsymbol{d}'\boldsymbol{d})^{l/2}\left\{\boldsymbol{d}(\boldsymbol{A} + k_1^{-1/2}W_1 - k_2^{-1/2}W_2)\boldsymbol{d}\right\}^2 + O_p(n^{-1/2}), \\
(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2} &= \left\{(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)'\boldsymbol{S}^{-1}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)\right\}^{l/2} = \boldsymbol{d}'\boldsymbol{d} + O_p(n^{-1/2}).
\end{aligned}$$

Moreover, $\boldsymbol{W}_1$ is independent of $\boldsymbol{W}_2$ and the following formulae can be seen in Fujikoshi et al. (2010).

$$\mathrm{E}[\boldsymbol{W}_j] = \boldsymbol{O}, \quad \mathrm{E}[\boldsymbol{W}_j\boldsymbol{B}\boldsymbol{W}_j] = \mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}_j)\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_j\boldsymbol{B}\boldsymbol{\Sigma}_j,$$

where $\boldsymbol{B}$ is an arbitrary constant matrix. Thus we obtain the following,

$$\begin{aligned}
\mathrm{E}\left[\mathrm{tr}(\hat{\boldsymbol{A}})(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}\right] &= (\boldsymbol{d}'\boldsymbol{d})^{l/2}\{\mathrm{tr}(\boldsymbol{A}^2) + (p^2+p)/k_1 + (p^2+p)/k_2\} + O(n^{-1/2}), \\
\mathrm{E}\left[(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}\hat{\boldsymbol{d}}'\hat{\boldsymbol{A}}^k\hat{\boldsymbol{d}}\right] &= (\boldsymbol{d}'\boldsymbol{d})^{l/2}\{\boldsymbol{d}'\boldsymbol{A}^2\boldsymbol{d} + (p+1)\boldsymbol{d}'\boldsymbol{d}/k_1 + (p+1)\boldsymbol{d}'\boldsymbol{d}/k_2\} + O(n^{-1/2}), \\
\mathrm{E}\left[(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}(\hat{\boldsymbol{d}}'\hat{\boldsymbol{A}}\hat{\boldsymbol{d}})^2\right] &= (\boldsymbol{d}'\boldsymbol{d})^{l/2}\{(\boldsymbol{d}'\boldsymbol{A}\boldsymbol{d})^2 + 2(\boldsymbol{d}'\boldsymbol{d})^2/k_1 + 2(\boldsymbol{d}'\boldsymbol{d})^2/k_2\} + O(n^{-1/2}), \\
\mathrm{E}\left[(\hat{\boldsymbol{d}}'\hat{\boldsymbol{d}})^{l/2}\right] &= (\boldsymbol{d}'\boldsymbol{d})^{l/2} + O(n^{-1/2}).
\end{aligned}$$

Using these formulae, replacing $\mathrm{tr}(\boldsymbol{A}^2)(\boldsymbol{d}'\boldsymbol{d})^{l/2}$, $(\boldsymbol{d}'\boldsymbol{d})^{l/2}\boldsymbol{d}'\boldsymbol{A}^2\boldsymbol{d}$, $(\boldsymbol{d}'\boldsymbol{d})^{l/2}(\boldsymbol{d}'\boldsymbol{A}\boldsymbol{d})^2$ with

$$\begin{aligned}
&(\boldsymbol{d}'\boldsymbol{d})^{l/2}\{\mathrm{tr}(\boldsymbol{A}^2) - (p^2+p)/k_1 - (p^2+p)/k_2\}, \\
&(\boldsymbol{d}'\boldsymbol{d})^{l/2}\{\boldsymbol{d}'\boldsymbol{A}^2\boldsymbol{d} - (p+1)\boldsymbol{d}'\boldsymbol{d}/k_1 - (p+1)\boldsymbol{d}'\boldsymbol{d}/k_2\}, \\
&(\boldsymbol{d}'\boldsymbol{d})^{l/2}\{(\boldsymbol{d}'\boldsymbol{A}\boldsymbol{d})^2 - 2(\boldsymbol{d}'\boldsymbol{d})^2/k_1 - 2(\boldsymbol{d}'\boldsymbol{d})^2/k_2\},
\end{aligned}$$

respectively, we obtain the desired result. □

4 Numerical study

In this section, We perform numerical study for investigating the performance of D^* and other selection method. We assume that

$$\boldsymbol{d} = (\sqrt{p}, \dots, \sqrt{p})'/p, \quad \boldsymbol{\Sigma}_1 = \boldsymbol{I}_p, \quad \boldsymbol{\Sigma}_2 = \boldsymbol{I}_p + C \cdot \mathrm{diag}(1, \dots, p)/p,$$

in numerical studies. The expected misclassification probabilities are calculated by the Monte Carlo simulation with 100,000 iterations.

4.1 Comparison of Cross-Validation method and D^*

We consider Cross-Validation(CV) as one of an estimation method of the expected misclassification probability. Let $\boldsymbol{X}_{i1}, \dots, \boldsymbol{X}_{iN_i}$ be a training samples from Π_i(i = 1,2), and let $d(\boldsymbol{X}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ be arbitrary discriminant function which involves unknown parameters, then the expected misclassification probabilityP_d is given as

$$P_d(\boldsymbol{X}, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S}_1, \boldsymbol{S}_2) = \frac{1}{2} \left\{ P(\, d(\boldsymbol{X}, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S}_1, \boldsymbol{S}_2) > 0 \middle| \, \boldsymbol{X} \in \Pi_1) + P(\, d(\boldsymbol{X}, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S}_1, \boldsymbol{S}_2) \le 0 \middle| \, \boldsymbol{X} \in \Pi_2) \right\},$$

where estimators of these parameters are given by (1.1). Suppose that $\bar{\boldsymbol{X}}_j^{(-k)}$, $\boldsymbol{S}_j^{(-k)}$ are the sample mean and the sample covariance matrix by using training samples which are deleted $\boldsymbol{X}_{jk}(k = 1, \dots, N_j, \;\; j = 1, 2)$, that is,

$$\bar{\boldsymbol{X}}_j^{(-k)} = \frac{1}{N_j - 1} \sum_{i=1, i \neq k}^{N_j} \boldsymbol{X}_{ji}, \quad \boldsymbol{S}_j^{(-k)} = \frac{1}{N_j - 2} \sum_{i=1, i \neq k}^{N_j} (\boldsymbol{X}_{ji} - \bar{\boldsymbol{X}}_j)(\boldsymbol{X}_{ji} - \bar{\boldsymbol{X}}_j)'.$$

Then we obtain the CV estimation of P_d as the following:

$$\begin{aligned}
\hat{P}_d^{(CV)} &= \frac{1}{2} \left\{ \hat{P}_d^{(CV)}(2|1) + \hat{P}_d^{(CV)}(1|2) \right\}, \\
\hat{P}_d^{(CV)}(2|1) &= \frac{1}{N_1} \sum_{k=1}^{N_1} \chi(d(\boldsymbol{X}_{1k}, \bar{\boldsymbol{X}}_1^{(-k)}, \bar{\boldsymbol{X}}_2, \boldsymbol{S}_1^{(-k)}, \boldsymbol{S}_2) > 0), \\
\hat{P}_d^{(CV)}(1|2) &= \frac{1}{N_2} \sum_{k=1}^{N_2} \chi(d(\boldsymbol{X}_{2k}, \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2^{(-k)}, \boldsymbol{S}_1, \boldsymbol{S}_2^{(-k)}) \le 0),
\end{aligned}$$

where $\chi(A)$ is defined as follows. If A is true, then $\chi(A) = 1$. Otherwise, $\chi(A) = 0$. By using CV estimation, we obtain a criterion D_{CV} for selecting between the linear and the quadratic discriminant functions, that is,

$$D_{CV} = \hat{P}_Q^{(CV)} - \hat{P}_L^{(CV)}.$$

Hence if D_{CV} is negative, we can consider that $\hat{Q}(\boldsymbol{X})$ is better than $\hat{L}(\boldsymbol{X})$. Otherwise, we can consider that $\hat{L}(\boldsymbol{X})$ is better than $\hat{Q}(\boldsymbol{X})$. The asymptotic bias of CV estimation is 0, that is,

$$\mathrm{E}[\hat{P}_d^{CV}] \to P_d \;\; (n \to \infty).$$

However, CV estimation takes heavy costs of calculating. Hence it is hard to use it in practice.

Table 1 gives the expected misclassification probabilities for different p, C, N_1 and N_2. Here the columns L and Q are the expected misclassification probabilities by using only $\hat{L}(\boldsymbol{X})$ and $\hat{Q}(\boldsymbol{X})$, respectively. The columns D^* and D_{CV} are the expected misclassification probability by selecting between the linear and the quadratic discriminant functions with using D^* and D_{CV}, respectively, and using the selected discriminant function. The numerical values on the double line and the line are the minimum and value without significance difference from the minimum in each row, respectively.

From Table 1, we can see that the performance of D^* is better than D_{CV} when C is small, but the performance of D^* is worse then D_{CV} when C is large.

4.2 Method of using hypothesis test and D^*

We performed the numerical study for investigating the performance of D^* and D_{CV} in the previous section. From the result of numerical study, performance of D^* is worse than D_{CV} when the difference between two covariance matrices is large. Because the setting of numerical study does not match the framework of asymptotic expansion.

p	C	N_1	N_2	L	Q	D^*	D_{CV}
2	0.0	25	25	0.320325	0.334350	0.324745	0.324765
2	0.1	25	25	0.322985	0.336480	0.32751	0.32835
2	0.5	25	25	0.333875	0.342335	0.33764	0.33803
2	1.0	25	25	0.344855	0.335850	0.34176	0.341045
2	5.0	25	25	0.374740	0.241970	0.27009	0.24881
2	9.0	25	25	0.383625	0.190830	0.207125	0.19246
2	0.0	50	50	0.313940	0.320265	0.31576	0.31678
2	0.1	50	50	0.318275	0.323405	0.32003	0.32047
2	0.5	50	50	0.326185	0.326880	0.327385	0.326675
2	1.0	50	50	0.335995	0.321925	0.32833	0.32798
2	5.0	50	50	0.364975	0.232475	0.245315	0.23346
2	9.0	50	50	0.377285	0.180920	0.185035	0.18094
5	0.0	25	25	0.342425	0.385365	0.34248	0.3532
5	0.1	25	25	0.344355	0.384790	0.34437	0.354565
5	0.5	25	25	0.353150	0.381590	0.35336	0.36343
5	1.0	25	25	0.361285	0.358925	0.361265	0.361665
5	5.0	25	25	0.384420	0.199435	0.36015	0.20249
5	9.0	25	25	0.391675	0.131620	0.34082	0.13179
5	0.0	50	50	0.325875	0.354550	0.325875	0.3321
5	0.1	50	50	0.329885	0.357905	0.329875	0.336875
5	0.5	50	50	0.335370	0.348430	0.335455	0.33994
5	1.0	50	50	0.348325	0.327540	0.348055	0.336055
5	5.0	50	50	0.374350	0.167090	0.31971	0.167125
5	9.0	50	50	0.380630	0.104740	0.284115	0.10474

Table 1: Comparison of D^* and D_{CV}

Hence it is insufficient for selecting of these discriminant functions only by using D^*. Therefore, we suggest using the hypothesis test in addition to D^*.

For testing $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ against $H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$, the modified likelihood ratio test statistics is given as follows.

$$T = -2\rho \log \Lambda,$$

where

$$\rho = 1 - \frac{2p^2 + 3p - 1}{6(p+1)n} \left(\sum_{j=1}^{2} \frac{n}{n_j} - 1 \right),$$

$$\Lambda = \frac{|\boldsymbol{S}_1|^{n_1} |\boldsymbol{S}_2|^{n_2}}{|\boldsymbol{S}|^n}.$$

Moreover, we are given following result when the null hypothesis H_0 is true.

$$P(T \leq x) = P(\chi_f^2 \leq x) + O_2$$

where $f = p(p+1)/2$. The proof of the above lemma can be seen in Muirhead (1982). We consider that the difference between two covariance matrices is large when $P(\chi_f^2 > T) < \alpha$, where α is significance level and χ_f^2 is chi-square random variable with degree of freedom f. Hence, we obtain a criterion for selecting between the linear and the quadratic discriminant functions as follows. Let

$$D_H = P(\chi_f^2 > T),$$

then if D_H is lower than α, then we consider that the difference between two covariance matrices is large. Thus we select the quadratic discriminant function. On the other hand, if D_H is not lower than α, then we consider that the difference between two covariance matrices is small. Hence we consider selecting between the linear and the quadratic discriminant functions by using D^*. So we suggest the following selection method D_H^*.

STEP 1. We decide significance level α.

STEP 2. If $D_H < \alpha$, We select the quadratic discriminant function. Otherwise, we go next step.

STEP 3. We select between the linear and quadratic discriminant functions by D^*.

We perform numerical study for comparison of performances of D^*, D_{CV}, D_H and D_H^* in the following. Here we obtain the selection method D_H as follows. If D_H is lower than α, then we select the quadratic discriminant function. Otherwise, we select the linear discriminant function. Table 2 gives the expected misclassification probabilities for different p, C, N_1 and N_2. The column D_H and D_H^* are the expected misclassification probability for selecting between the linear and quadratic discriminant functions by using D_H and D_H^*, respectively, and using the selected discriminant function. Here the number in the parentheses is the significance level.

p	C	N_1	N_2	L	Q	D^*	D_{CV}	$D_H(0.01)$	$D_H(0.05)$	$D_H(0.1)$	$D_H^*(0.01)$	$D_H^*(0.05)$	$D_H^*(0.1)$
2	0.0	25	25	0.3205	0.3328	0.3245	0.3251	0.3210	0.3221	0.3236	0.3246	0.3250	0.3258
2	0.1	25	25	0.3229	0.3364	0.3271	0.3276	0.3235	0.3250	0.3265	0.3273	0.3280	0.3287
2	0.5	25	25	0.3316	0.3383	0.3348	0.3348	0.3326	0.3341	0.3351	0.3351	0.3357	0.3362
2	1.0	25	25	0.3428	0.3359	0.3408	0.3404	0.3432	0.3422	0.3408	0.3407	0.3401	0.3394
2	5.0	25	25	0.3747	0.2433	0.2717	0.2501	0.2455	0.2436	0.2435	0.2443	0.2435	0.2434
2	0.0	50	50	0.3146	0.3198	0.3161	0.3165	0.3149	0.3155	0.3162	0.3163	0.3165	0.3168
2	0.1	50	50	0.3169	0.3229	0.3194	0.3196	0.3172	0.3180	0.3189	0.3195	0.3197	0.3201
2	0.5	50	50	0.3279	0.3271	0.3281	0.3273	0.3283	0.3282	0.3282	0.3281	0.3280	0.3279
2	1.0	50	50	0.3351	0.3235	0.3296	0.3281	0.3311	0.3277	0.3266	0.3275	0.3260	0.3256
2	5.0	50	50	0.3653	0.2334	0.2460	0.2342	0.2334	0.2334	0.2334	0.2334	0.2334	0.2334
2	0.0	100	100	0.3105	0.3136	0.3113	0.3118	0.3106	0.3112	0.3116	0.3114	0.3117	0.3120
2	0.1	100	100	0.3162	0.3190	0.3172	0.3173	0.3164	0.3169	0.3173	0.3172	0.3174	0.3177
2	0.5	100	100	0.3241	0.3210	0.3225	0.3225	0.3233	0.3227	0.3219	0.3222	0.3221	0.3215
2	1.0	100	100	0.3321	0.3155	0.3217	0.3204	0.3185	0.3164	0.3159	0.3170	0.3159	0.3158
2	5.0	100	100	0.3622	0.2283	0.2318	0.2283	0.2283	0.2283	0.2283	0.2283	0.2283	0.2283
5	0.0	25	25	0.3431	0.3848	0.3431	0.3533	0.3440	0.3467	0.3494	0.3440	0.3467	0.3495
5	0.1	25	25	0.3452	0.3854	0.3453	0.3556	0.3460	0.3485	0.3508	0.3461	0.3485	0.3508
5	0.5	25	25	0.3536	0.3807	0.3537	0.3627	0.3542	0.3566	0.3590	0.3542	0.3567	0.3590
5	1.0	25	25	0.3619	0.3604	0.3619	0.3626	0.3626	0.3636	0.3634	0.3627	0.3636	0.3634
5	5.0	25	25	0.3858	0.1999	0.3617	0.2028	0.2010	0.2001	0.1999	0.2009	0.2001	0.1999
5	0.0	50	50	0.3255	0.3542	0.3256	0.3326	0.3262	0.3280	0.3304	0.3262	0.3281	0.3304
5	0.1	50	50	0.3282	0.3577	0.3282	0.3353	0.3287	0.3304	0.3327	0.3287	0.3304	0.3327
5	0.5	50	50	0.3381	0.3507	0.3381	0.3436	0.3392	0.3415	0.3430	0.3392	0.3415	0.3430
5	1.0	50	50	0.3457	0.3283	0.3453	0.3354	0.3411	0.3358	0.3337	0.3409	0.3358	0.3337
5	5.0	50	50	0.3737	0.1685	0.3191	0.1686	0.1685	0.1685	0.1685	0.1685	0.1685	0.1685
5	0.0	100	100	0.3177	0.3338	0.3177	0.3215	0.3181	0.3192	0.3208	0.3181	0.3192	0.3208
5	0.1	100	100	0.3219	0.3379	0.3219	0.3257	0.3224	0.3236	0.3250	0.3224	0.3236	0.3250
5	0.5	100	100	0.3274	0.3303	0.3275	0.3291	0.3288	0.3298	0.3301	0.3289	0.3298	0.3301
5	1.0	100	100	0.3357	0.3072	0.3330	0.3134	0.3112	0.3086	0.3080	0.3112	0.3086	0.3080
5	5.0	100	100	0.3645	0.1530	0.2792	0.1530	0.1530	0.1530	0.1530	0.1530	0.1530	0.1530

Table 2: Comparison of D^*, D_{CV}, D_H and D_H^*

From Table 2, we can see that D_H^* is better then D^* and D_{CV}, and performance of D_H^* is the same as performance of D_H in the cases that C is large. In most cases that the difference between two covariance matrices is small, the performances of all selection method are about same. In the case of $C = 1$, D_H^* better than D_H. Moreover, the performances D_H and D_H^* depend on signification level α, but in the cases of $\alpha = 0.01, 0.05, 0.1$, the performances are very close.

5 Conclusion

As the first, We suggest the method D^* for selecting of the linear and the quadratic discriminant functions by using the asymptotic expansion when the difference between two covariance matrices is small. In the case that the difference between two covariance matrices is small, D^* is better then D_{CV}. However, D^* is worse then D_{CV} when the difference between two covariance matrices is large. So secondly, we suggest selection method D_H^* of using the hypothesis test in addition to D^*. We see performance of D_H^* in the numerical studies. The performance of D_H^* is better then performance other selection method, or equally. However from numerical studies, the performance of D_H^* is worse than D_{CV} when p is large. Because the asymptotic expansions usually do not give good approximation formulae for large p. It may be possible to improve our selection method with using asymptotic expansions in high-dimensional and large samples framework, that is, both N, and p become large, which is left for future work.

Appendix

A.1 Proof of lemma 2.1

Proof. Since $\boldsymbol{W} \sim W_p(n, \boldsymbol{\Sigma})$,

$$\boldsymbol{Y}'\boldsymbol{W}^{-1}\boldsymbol{Y} \sim \boldsymbol{Y}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_0^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}, \quad \log|\boldsymbol{W}| \sim \log|\boldsymbol{\Sigma}| + \log|\boldsymbol{W}_0|$$

where $\boldsymbol{W}_0 \sim W_p(n, \boldsymbol{I}_p)$. Let $\boldsymbol{W}_0$ be partition as

$$\boldsymbol{W}_0 = \begin{pmatrix} W_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22} \end{pmatrix},$$

where $W_{11} : 1 \times 1$, and let

$$M = \begin{pmatrix} 1 & -\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1} \\ \boldsymbol{O} & \boldsymbol{I}_{p-1} \end{pmatrix}.$$

Then we can obtain the following, (see e.g. Fujikoshi. et al. (2010))

$$\begin{aligned}
\boldsymbol{W}_0^{-1} = \boldsymbol{W}_0 &= \begin{pmatrix} W_{11} & \boldsymbol{W}_{12} \\ \boldsymbol{W}_{21} & \boldsymbol{W}_{22} \end{pmatrix}^{-1} \\
&= \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{22}^{-1} \end{pmatrix} + \begin{pmatrix} 1 \\ -\boldsymbol{W}_{22}^{-1}\boldsymbol{W}_{21} \end{pmatrix} W_{11.2}^{-1} \begin{pmatrix} 1 & -\boldsymbol{W}_{12}\boldsymbol{W}_{22}^{-1} \end{pmatrix}, \\
|\boldsymbol{W}_0| = |\boldsymbol{M}\boldsymbol{W}_0\boldsymbol{M}'| &= \begin{vmatrix} W_{11.2} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{W}_{22} \end{vmatrix} = W_{11.2}|\boldsymbol{W}_{22}|,
\end{aligned}$$

where $W_{11.2} = W_{11} - \boldsymbol{W}_{12}\boldsymbol{W}_{22}\boldsymbol{W}_{21}$. Moreover, since $\boldsymbol{W}_0 \sim W_p(n, \boldsymbol{I}_p)$, $W_{11.2}$ is distributed as $W_1(n-p+1, 1)$, that is the chi-square distribution with degree of freedom $n-p+1$, and $\boldsymbol{W}_{22} \sim W_{p-1}(n, \boldsymbol{I}_{p-1})$, $W_{11.2}$ is independent of $\boldsymbol{W}_{22}$. Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2)$ be the orthogonal matrix such that $\boldsymbol{H}_1 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y}$, then

$$\begin{aligned}
\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{W}_0^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} &\sim \boldsymbol{Y}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}\boldsymbol{W}_0^{-1}\boldsymbol{H}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \\
&= W_{11.2}^{-1}\boldsymbol{Y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}.
\end{aligned}$$

Therefore, we obtain the desired result. □

A.2 Proof of lemma 2.2

Proof.

$$\begin{aligned}
\mathrm{E}[\exp\{itZ^2\}] &= \int_{-\infty}^{\infty} \exp\{itz^2\} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(z-\mu)^2\right\} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{itz^2 - \frac{1}{2}(z^2 - 2z\mu + \mu^2)\right\} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}((1-2it)z^2 - 2z\mu + \mu^2)\right\} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(1-2it)\left(z - \frac{\mu}{(1-2it)}\right)^2 - \frac{it\mu^2}{(1-2it)}\right\} dz \\
&= (1-2it)^{-1/2} \exp\left\{-\frac{it\mu^2}{(1-2it)}\right\}
\end{aligned}$$

□

A.3 Proof of Lemma 2.3

Proof.

$$
\begin{aligned}
g(t_1,t_2) &= \mathrm{E}\left[\exp\left\{\boldsymbol{X}'\left(t_1\boldsymbol{\Gamma}_1+t_2\boldsymbol{\Gamma}_2\right)\boldsymbol{X}+2\left(t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\right)\boldsymbol{X}+t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right\}\right]\\
&= \int_{-\infty}^{\infty}(2\pi)^{-p/2}\exp\left\{-\boldsymbol{x}'\boldsymbol{x}/2\right\}\exp\left\{\boldsymbol{x}'\left(t_1\boldsymbol{\Gamma}_1+t_2\boldsymbol{\Gamma}_2\right)\boldsymbol{x}+2\left(t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\right)\boldsymbol{x}+t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right\}d\boldsymbol{x}\\
&= \int_{-\infty}^{\infty}(2\pi)^{-p/2}\exp\left[-\frac{1}{2}\left\{\boldsymbol{x}'(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{x}-4\left(t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\right)\boldsymbol{x}-2t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1-2t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right\}\right]d\boldsymbol{x}\\
&= \int_{-\infty}^{\infty}(2\pi)^{-p/2}\exp\left[\frac{1}{2}\left\{\boldsymbol{x}-2\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\left(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right)\right\}'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)\right.\\
&\qquad\cdot\left\{\boldsymbol{x}-2\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\left(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right)\right\}\\
&\qquad\left.+2(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2)'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\left(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right)+t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right]d\boldsymbol{x}\\
&= |\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2|^{-1/2}\exp\left[2(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2)'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\left(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right)\right.\\
&\qquad\left.+\,t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right].
\end{aligned}
$$

From the above result, we can easily obtain $g(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2)=g(t_2,t_1;\boldsymbol{\eta}_2,\boldsymbol{\eta}_1,\boldsymbol{\Gamma}_2,\boldsymbol{\Gamma}_1)$. By using $2t_1\boldsymbol{\Gamma}_1=(\boldsymbol{I}_p-2t_2\boldsymbol{\Gamma}_2)-(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)$,

$$
\begin{aligned}
&t_1\boldsymbol{\eta}_1'\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2+2(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2)'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\left(t_1\boldsymbol{\Gamma}_1\boldsymbol{\eta}_1+t_2\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\right)\\
&= \frac{1}{2}\boldsymbol{\eta}_1'(\boldsymbol{I}_p-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{\eta}_1-\frac{1}{2}\boldsymbol{\eta}_1'(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\\
&\quad+\frac{1}{2}\{\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}-(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{\eta}_1\}'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\\
&\qquad\cdot\{\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}-(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{\eta}_1\}\\
&= \frac{1}{2}\boldsymbol{\eta}_1'(\boldsymbol{I}_p-2t_2\boldsymbol{\Gamma}_2)\boldsymbol{\eta}_1+t_2\boldsymbol{\eta}_2'\boldsymbol{\Gamma}_2\boldsymbol{\eta}_2\\
&\quad+\frac{1}{2}\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'\left(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2\right)^{-1}\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}-\boldsymbol{\eta}_1'\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}\\
&= -\frac{1}{2}\boldsymbol{\eta}_1'\boldsymbol{\eta}_1+t_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)'\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\\
&\quad+\frac{1}{2}\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p-2t_1\boldsymbol{\Gamma}_1-2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1-2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}.
\end{aligned}
$$

Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{M}$ be arbitrary $p\times p$ matrix, then we obtain the following formulae.

$$
\begin{aligned}
&|\boldsymbol{I}_p+h\boldsymbol{C}|=\exp\{\log|\boldsymbol{I}_p+h\boldsymbol{C}|\}=\exp\{h\mathrm{tr}(\boldsymbol{C})+O(h^2)\}=1+h\mathrm{tr}(\boldsymbol{C})+O(h^2),\\
&|\boldsymbol{A}+(t+h)\boldsymbol{B}|=|\boldsymbol{A}+t\boldsymbol{B}+h\boldsymbol{B}|=|\boldsymbol{A}+t\boldsymbol{B}||\boldsymbol{I}_p+h(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}|\\
&\quad=|\boldsymbol{A}+t\boldsymbol{B}|\left[1+h\mathrm{tr}\{(\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{B}\}+O(h^2)\right], \qquad \text{(A.3.1)}\\
&\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{A}+(t+h)\boldsymbol{B})^{-1}\right\}=\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\{\boldsymbol{I}_p+h\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\}^{-1}\right]\\
&\quad=\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\{\boldsymbol{I}+h\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{B})^{-1}+O(h^2)\}\right]\\
&\quad=\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})\right]-h\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\right]+O(h^2). \qquad \text{(A.3.2)}
\end{aligned}
$$

From (A.3.1) and (A.3.2),

$$
\begin{aligned}
&\frac{d}{dt}|\boldsymbol{A}+t\boldsymbol{B}|=\lim_{h\to0}\frac{1}{h}\left\{|\boldsymbol{A}+(t+h)\boldsymbol{B}|-|\boldsymbol{A}+t\boldsymbol{B}|\right\}=|\boldsymbol{A}+t\boldsymbol{B}|\mathrm{tr}\left\{(\boldsymbol{A}+t\boldsymbol{B})^{-1}\boldsymbol{B}\right\},\\
&\frac{d}{dt}\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\right\}=\lim_{h\to0}\frac{1}{h}\left[\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{A}+(t+h)\boldsymbol{B})^{-1}\right\}-\mathrm{tr}\left\{\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\right\}\right]\\
&\qquad=-\mathrm{tr}\left[\boldsymbol{M}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{A}+t\boldsymbol{B})^{-1}\right].
\end{aligned}
$$

By using the these result,

$$
\begin{aligned}
\frac{\partial g(t_1,t_2)}{\partial t_1} &= \left[\frac{\partial}{\partial t_1}|\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2|^{-1/2}\right]\exp\left\{-\frac{1}{2}\boldsymbol{\eta}_1'\boldsymbol{\eta}_1 + t_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)'\boldsymbol{\Gamma}_2\ (\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\right. \\
&\qquad\qquad \left. + \frac{1}{2}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}\right\} \\
&\quad + |\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\eta}_1'\boldsymbol{\eta}_1 + t_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)'\boldsymbol{\Gamma}_2\ (\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\right. \\
&\qquad\qquad \left. + \frac{1}{2}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}\right\} \\
&\quad \cdot \frac{\partial}{\partial t_1}\left\{-\frac{1}{2}\boldsymbol{\eta}_1'\boldsymbol{\eta}_1 + t_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)'\boldsymbol{\Gamma}_2\ (\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\right. \\
&\qquad\qquad \left. + \frac{1}{2}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}\right\} \\
&= g(t_1,t_2)\mathrm{tr}[(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1] \\
&\quad + \{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}, \\
\frac{\partial L(t_1,t_2)}{\partial t_1} &= L(t_1,t_2;\boldsymbol{\eta}_1,\boldsymbol{\eta}_2,\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2) = 2\mathrm{tr}[\{(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\boldsymbol{\Gamma}_1\}^2] \\
&\quad + 4\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\}'(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\{\boldsymbol{\Gamma}_1(\boldsymbol{I}_p - 2t_1\boldsymbol{\Gamma}_1 - 2t_2\boldsymbol{\Gamma}_2)^{-1}\}^2\{\boldsymbol{\eta}_1 - 2t_2\boldsymbol{\Gamma}_2(\boldsymbol{\eta}_1-\boldsymbol{\eta}_2)\} \\
\frac{\partial g(t_1,t_2)}{\partial t_2} &= g(t_1,t_2)L(t_2,t_1;\boldsymbol{\eta}_2,\boldsymbol{\eta}_1,\boldsymbol{\Gamma}_2,\boldsymbol{\Gamma}_1).
\end{aligned}
$$

□

References

[1] Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Third Edision. Wiley, Hoboken, NJ.

[2] Fujikoshi, Y. ,Ulyanv, V. V. and Shimizu, R. (2010), *Multivariate Statistics: High-Dimensional and Large-Sample Approximations.* Wiley, Hoboken, NJ.

[3] Marks, S. and Dunn, O. J. (1974). Discriminant function when covariance matrices are unequal. *J. Am. Statist. Assoc.*, **69**, 555-559.

[4] Muirhead, R. J. (1982), *Aspects of Multivate Statistical Theory.* Wiley, New York.

[5] Wahl, P. W. and Kronmal, R. A. (1977). Discriminant function when covariance matrices are unequal and sample sizes are moderate. *Biometrics*, **33**, 479-484.

[6] Wakaki, H. (1990). Comparison of linear and quadratic discriminant functions. *Biometrika*, **77**, 227-229.