Linear Dimension Reduction for Multiple Heteroscedastic Multivariate Normal Populations

Songthip T. Ounpraseuth¹, Phil D. Young², Johanna S. van Zyl², Tyler W. Nelson², Dean M. Young²

¹Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AK, USA
²Department of Statistical Science, Baylor University, Waco, TX, USA
Email: stounpraseuth@uams.edu, philip_young@baylor.edu, tyler_nelson@baylor.edu, johanna_vanzyl@baylor.edu, dean_young@baylor.edu

Received 16 April 2015; accepted 19 June 2015; published 24 June 2015

Abstract

For the case where all multivariate normal parameters are known, we derive a new linear dimension reduction (LDR) method to determine a low-dimensional subspace that preserves or nearly preserves the original feature-space separation of the individual populations and the Bayes probability of misclassification. We also give necessary and sufficient conditions which provide the smallest reduced dimension that essentially retains the Bayes probability of misclassification from the original full-dimensional space in the reduced space. Moreover, our new LDR procedure requires no computationally expensive optimization procedure. Finally, for the case where parameters are unknown, we devise a LDR method based on our new theorem and compare our LDR method with three competing LDR methods using Monte Carlo simulations and a parametric bootstrap based on real data.

Keywords

Linear Transformation, Bayes Classification, Feature Extraction, Probability of Misclassification

1. Introduction

The fact that the Bayes probability of misclassification (BPMC) of a statistical classification rule does not increase as the dimension or feature space increases, provided the class-conditional probability densities are known, is well-known. However, in practice when parameters are estimated and the feature-space dimension is

large relative to the training-sample sizes, the performance or efficacy of a sample discriminant rule may be considerably degraded. This phenomenon gives rise to a paradoxical behavior that [1] has called the *curse of dimensionality*.

An exact relationship between the expected probability of misclassification (EPMC), training-sample sizes, feature-space dimension, and actual parameters of the class-conditional densities is challenging to obtain. In general, as the classifier becomes more complex, the ratio of sample size to dimensionality must increase at an exponential rate to avoid the curse of dimensionality. The authors [2] have suggested a ratio of at least ten times as many training samples per class as the feature dimension increases. Hence, as the number of feature variables $p$ becomes large relative to the training-sample sizes $n_i$, $i = 1, \cdots, m$, where $m$ is the number of classes, one might wish to use a smaller number of the feature variables to improve the classifier performance or computational efficiency. This approach is called feature subset selection.

Another effective approach to obtain a reduced dimension to avoid the curse of dimensionality is linear dimension reduction (LDR). Perhaps the most well-known LDR procedure for the $m$-class problem is linear discriminant analysis (LDA) from [3], which is a generalization of the linear discriminant function (LDF) derived in [4] for the case $m = 2$. The LDA LDR method determines a vector $a$ that maximizes the ratio of between-class scatter to average within-class scatter in the lower-dimensional space. The sample within-class scatter matrix is

$$ S_w = \sum_{i=1}^{m} a_i S_i, \quad (1) $$

where $S_i$ is the estimated sample covariance matrix for the $i$th class, and the sample between-class scatter matrix is

$$ S_b = \sum_{i=1}^{m} a_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})', \quad (2) $$

where $\bar{x}_i$ is the sample mean vector for class $\Pi_i$, $a_i$ is its *a priori* probability of class membership, $\bar{x} = \sum_{i=1}^{m} a_i \bar{x}_i$ is the estimated overall mean, and $i = 1, \cdots, m$. For $m = 2$ with $a_i = a_2$, [4] determined the vector $a$ that maximizes the criterion function

$$ J_f(a) = \frac{a'S_b a}{a'S_w a}, $$

which is achieved by an eigenvalue decomposition of $S_w^{-1}S_b$. An attractive feature of LDA as a LDR method is that it is computationally appealing; however, it attempts to maximally separate class means and does not incorporate the discriminatory information contained in the differences of the class covariance matrices. Many alternative approaches to the LDA LDR method have been proposed. For example, canonical variables can be viewed as an extension to the LDF when $m > 2$ and $q > 1$ (see [5]). Other straightforward extensions of the LDF include those in [6]-[7].

Extensions of LDA that incorporate information on the differences in covariance matrices are known as heteroscedastic linear dimension reduction (HLDR) methods. The authors [8] have proposed an eigenvalue-based HLDR approach for $m > 2$, utilizing the so-called Chernoff criterion, and have extended the well-known LDA method using directed distance matrices that can be considered a generalization of (2). Additional HLDR methods have been proposed by authors such as [9]-[10].

Using results by [13] that characterize linear sufficient statistics for multivariate normal distributions, we develop an explicit LDR matrix $B \in \mathbb{R}_{q \times p}$ such that

$$ x \rightarrow y = Bx, $$

where $x \in \mathbb{R}_{p \times 1}$, $y \in \mathbb{R}_{q \times 1}$, and $\mathbb{R}_{m \times m}$ denotes the space of all $m \times m$ real matrices. Using the Bayes classification procedure in which we assume equal costs of misclassification and that all class parameters are known, we determine the reduced dimension $q < p$ that is the smallest reduced dimension for which there exists a LDR matrix $B \in \mathbb{R}_{q \times p}$ that preserves all of the classification information originally contained in the $p$-dimensional feature space. We then derive a linear transformation that assigns $x$ to $\Pi_k$ if and only if the corres-
ponding Bayes classification rule assigns $Bx$ to $\Pi_k$, where $k \in \{1, \ldots, m\}$. We refer to this method as the SY LDR procedure.

Moreover, we use Monte Carlo simulations to compare the classification efficacy of the BE method, sliced inverse regression (SIR), and sliced average variance estimation (SAVE) found in [14]-[16], respectively, with the SY method.

The remainder of this paper is organized as follows. We begin with a brief introduction to the Bayes quadratic classifier in Section 2 and introduce some preliminary results that we use to prove our new LDR method in Section 3. In Section 4, we provide conditions under which the Bayes quadratic classification rule is preserved in the low-dimensional space and derive a new LDR matrix. We establish a SVD-based approximation to our LDR procedure along with an example of low-dimensional graphical representations in Section 5. We describe the four LDR methods that we compare using Monte Carlo simulations in Section 6. We present five Monte Carlo simulations in which we compare the competing LDR procedures for various population parameter configurations in Section 7. In Section 8, we compare the four methods using bootstrap simulations for a real data example, and, finally, we offer a few concluding remarks in Section 9.

2. The Bayes Quadratic Classifier

The Bayesian statistical classifier discriminates based on the probability density functions $p(\Pi_i|x)$, $i = 1, \ldots, m$, of each class. The Bayes classifier is optimal in the sense that it maximizes the class $a$ posteriori probability provided all class distributions and corresponding parameters are known. That is, suppose we have $m$ classes, $\Pi_1, \ldots, \Pi_m$, with assumed known $a$ priori probabilities $\alpha_i$, respectively. Also, let $p(\Pi_i)$ denote the $p$-dimensional multivariate normal density corresponding to population $\Pi_i$, $i = 1, \ldots, m$. The goal of statistical decision theory is to obtain a decision rule that assigns an unlabeled observation $x$ to $\Pi_i$ if $p(\Pi_i|x)$ is the maximum overall $a$ posteriori $p(\Pi_i|x)$, $i = 1, \ldots, m$. Then,

$$
\lambda(\Pi_i|\Pi_j) = \begin{cases} 0, & i = j \\ 1, & i \neq j \\ \end{cases},
$$

the Bayes classifier assigns $x$ to class $\Pi_i$ if

$$
p(\Pi_i|x) > p(\Pi_j|x), \quad j = 1, \ldots, m; j \neq k.
$$

This decision rule partitions the measurement or feature space into $m$ disjoint regions $R_{\Pi_i}$, where $i = 1, \ldots, m$, such that $x$ is assigned to class $\Pi_i$ if $x \in R_{\Pi_i}$. Using Bayes’ rule, the $a$ posteriori probabilities of class membership $p(\Pi_i|x)$ can be defined as

$$
p(\Pi_i|x) = \frac{\alpha_i p(x|\Pi_i)}{p(x)}.
$$

One can re-express the Bayes classification as the following:

Assign $x$ to $\Pi_i$ if

$$
\alpha_i p(x|\Pi_i) > \alpha_j p(x|\Pi_j), \quad j = 1, \ldots, m; j \neq k.
$$

This decision rule is known as the Bayes’ classification rule. Let $\Pi_i$ be modeled as a $p$-dimensional multivariate normal distribution, and let

$$
d_i(x) = \ln |\Sigma_i| - 2[\ln(\alpha_i)] + (x - \mu)_i^{\top}\Sigma_i^{-1}(x - \mu)_i, \quad i = 1, \ldots, m.
$$

The Bayes decision rule (4) is to classify the unlabeled observation $x$ into the class $\Pi_i$ such that $d_i(x) = \min \{d_i(x); i = 1, \ldots, m\}$. The classification rule defined by (4) is known as the quadratic discriminant function (QDF), or the quadratic classifier.

3. Preliminary Results

The following notation will be used throughout the remainder of the paper. We let $\mathbb{R}_p^+$ denote the set of $p \times p$
positive definite matrices and \( \mathbb{R}^{\times}_{+} \) denote the set of \( p \times p \) symmetric matrices. Also, we let \( A^\dagger \in \mathbb{R}^{m \times n} \) represent the Moore-Penrose pseudo-inverse of \( A \in \mathbb{R}^{m \times n} \).

The proof of the main result for the derivation of our new LDR method requires the following notation and lemmas. Let \( M \in \mathbb{R}^{p(m-1)(p+1)} \) be

\[ T \equiv [d_2 - d_1 | \cdots | d_m - d_1 | E_2 - E_1 | \cdots | E_m - E_1], \tag{5} \]

where \( d_i \in \mathbb{R}^{p+1} \), \( E_i \in \mathbb{R}^{p \times p} \) such that \( \text{rank}(E_i) = p \), and \( d_i \neq d_k \) and \( E_i \neq E_k \) for at least one value of \( k \), where \( 2 \leq k \leq m \) and \( i = 1, \cdots, m \). Also, let \( \text{rank}(T) = 1 \leq q < p \), and let \( F \in \mathbb{R}^{p \times q} \) and \( G \in \mathbb{R}^{(m-1)(p+1)} \) be matrix components of a full-rank decomposition of \( T \) so that \( T = FG \) with \( \text{rank}(F) = \text{rank}(G) = q \). Then, the Moore-Penrose pseudoinverse of \( T \) is \( T^+ = G^+F^+ \), and, also, \( TT^+ = FF^+ \) and \( TT^T = FF^T = T \). This result implies that for \( i = 1, \cdots, m \),

(i) \( (I - FF^+)(d_i - d_i) = 0 \) and

(ii) \( (I - FF^+)(E_i - E_i) = 0 \).

We now state and prove three lemmas that we use in the proof of our main result.

**Lemma 1** For \( T = FG \) in (5), where \( E_i \in \mathbb{R}^{p \times p} \), \( d_i \in \mathbb{R}^{p+1} \), \( F \in \mathbb{R}^{p \times q} \), and \( G \in \mathbb{R}^{(m-1)(p+1)} \) such that \( \text{rank}(F) = \text{rank}(G) = q \), and \( i = 1, \cdots, m \), we have that

(a) \( FF^+(E_i - E_i) = (E_i - E_i)FF^+ \),

(b) \( FF^+E_i = E_i FF^+ \), and

(c) \( (I - FF^+)(E_i - E_i) = E_i(I - FF^+) \).

**Proof.** Part (a) follows from the fact that \( E_i = E_i' \), \( i = 1, \cdots, m \), and from (ii) above. Parts (b) and (c) follow directly from (a).

**Lemma 2** For \( T = FG \) in (5), where \( E_i \in \mathbb{R}^{p \times p} \), \( d_i \in \mathbb{R}^{p+1} \), \( F \in \mathbb{R}^{p \times q} \), and \( G \in \mathbb{R}^{(m-1)(p+1)} \) such that \( \text{rank}(F) = \text{rank}(G) = q \), and \( i = 1, \cdots, m \), we have that

\[ \left[ F^+E_iF^+ \right]^{-1} = F^+E_i^{-1}F. \]

**Proof.** Because \( F^+E_iF^+ \in \mathbb{R}^{pq} \) and \( \text{rank}(F^+E_iF^+) = q \), the result follows because

\[ \left( F^+E_iF^+ \right) \left( F^+E_i^{-1}F \right) = F^+E_i \left[ F^+E_i^{-1}F \right] E_i^{-1}F = F^+ FF^+E_iE_i^{-1}F = I_q. \]

**Lemma 3** Let \( T = FG \) in (5), where \( E_i \in \mathbb{R}^{p \times p} \), \( d_i \in \mathbb{R}^{p+1} \), \( F \in \mathbb{R}^{p \times q} \), and \( G \in \mathbb{R}^{(m-1)(p+1)} \) such that \( \text{rank}(F) = \text{rank}(G) = q \), and \( i = 1, \cdots, m \). Also, let \( C = R[I - FF^+] \in \mathbb{R}^{(p-q) \times p} \), where \( R \in \mathbb{R}^{(p-q) \times p} \), such that \( \text{rank}(C) = p - q \). Then,

(a) \( Cd_i = Cd_i \),

(b) \( Cd_i + CE_iF^+ \left( F^+E_iF^+ \right)^{-1} \left( y-F^+d_i \right) = Cd_i \), where \( y \in \mathbb{R}^{p+1} \), and

(c) \( CE_iC' - CE_iF^+ \left( F^+E_iF^+ \right)^{-1} F^+E_iC' = CE_iC' \).

**Proof.** The proof of part (a) of Lemma 3 follows from (i), and we have that

\[ CE_iF^+ \left( F^+E_iF^+ \right)^{-1} = CE_iF^+F'E_i^{-1}F = CF = 0, \tag{6} \]

for each \( i = 1, \cdots, m \). Hence, (b) and (c) follow from (6).

### 4. Linear Dimension Reduction for Multiple Heteroscedastic Nonsingular Normal Populations

We now derive a new LDR method that is motivated by results on linear sufficient statistics derived by [13] and by a linear feature selection theorem given in [17]. The theorem provides necessary and sufficient conditions for which a low-dimensional linear transformation of the original data will preserve the BPMc in the original feature space. Also, the theorem provides a representation of the LDR matrix.

**Theorem 1** Let \( \Sigma_i \) be a \( p \)-dimensional multivariate normal population with a priori probability \( \alpha > 0 \), mean \( \mu \), and covariance matrix \( \Sigma_j \), \( i = 1, \cdots, m \), such that \( \text{rank}(\Sigma_j) = p \) and \( \Sigma_i \neq \Sigma_j \) for some \( j \in \{2, \cdots, m\} \).
Next, let
\[ M = \left[ \Sigma_1^{-1} \mu_1 \mid \cdots \mid \Sigma_n^{-1} \mu_n \right] \cdot \Sigma_1^{-1} \mu_1 \mid \cdots \mid \Sigma_n^{-1} \mu_n \cdot \Sigma_1^{-1} \mu_1 \mid \cdots \mid \Sigma_n^{-1} \mu_n . \] (7)

Finally, let \( M = FG \) be a full-rank decomposition of \( M \), where \( \text{rank}(M) = q < p \). Then, the \( p \)-dimensional Bayes procedure assigns \( x \) to \( \Pi_k \) if and only if the \( q \)-dimensional Bayes procedure assigns \( F'x \) to \( \Pi_k \) for \( k \in \{1, \ldots, m\} \).

Proof. Let
\[
 w_i = \begin{bmatrix} y_i \\ u_i \end{bmatrix} = \begin{bmatrix} F_{q \times p}^+ \\ C_{(p-q) \times p} \end{bmatrix} x_i,
\]
where \( x_i \sim N(\mu_i, \Sigma_i) \), \( i = 1, \ldots, m \). Let \( H = \begin{bmatrix} F' \\ C \end{bmatrix} \in \mathbb{R}_{p \times p} \) be full-rank with \( C = R \left( I - FF' \right) \), where \( R \in \mathbb{R}_{(p-q) \times p} \) such that \( \text{rank}(R) = p-q \). Then \( w_i \sim N(H \mu_i, H \Sigma_i H') \), where
\[
 H \mu_i = \begin{bmatrix} F' \\ C \end{bmatrix} \mu_i \quad \text{and} \quad H \Sigma_i H' = \begin{bmatrix} F' \Sigma_i F' & F' \Sigma_i C' \\ C \Sigma_i F' & C \Sigma_i C' \end{bmatrix}.
\]

By Lemma 3, we conclude that \( E(u_i|y_i) = C \mu_i \) and \( \text{Var}(u_i|y_i) = C \Sigma_i C' \) for \( i = 1, \ldots, m \). Thus, \( w_i \sim N(F' \mu_i, F' \Sigma_i F') \cdot N(C \mu_i, C \Sigma_i C') \). That is, \( p(u_i|y_i, \Pi_k) \) does not depend on class \( k \in \{2, \ldots, m\} \).

Recall that for \( i, j = 1, \ldots, m \), \( i \neq j \), the \( p \)-variate Bayes procedure defined by (3) assigns \( x \) to \( \Pi_j \) if and only if
\[
 \alpha_j p(y|\Pi_j) > \alpha_j p(y|\Pi_i) \Leftrightarrow \alpha_j p(w|\Pi_j) > \alpha_j p(w|\Pi_i) \\
 \Rightarrow \alpha_j p(u|y, \Pi_j) p(y|\Pi_j) > \alpha_j p(u|y, \Pi_i) p(y|\Pi_i) \\
 \Rightarrow \alpha_j p(y|\Pi_j) > \alpha_j p(y|\Pi_i).
\]

Therefore, the original \( p \)-variate Bayes classification assignment is preserved by the linear transformation \( y = F'x \).

Theorem 1 is important in that if its conditions hold, we obtain a LDR matrix for the reduced \( q \)-dimensional subspace such that the BPMC in the \( q \)-dimensional space is equal to the BPMC for the original \( p \)-dimensional feature space. In other words, provided the conditions in Theorem 1 hold, we have that the LDR matrix \( F' \in \mathbb{R}_{q \times p} \) exists and that \( \text{BPMC}_q = \text{BPMC}_p \), where \( \text{rank}(M) = q < p \).

With the following corollary, we demonstrate that for two multivariate normal populations such that \( \Sigma_i = \Sigma_2 = \Sigma \) and \( \mu_i \neq \mu_2 \), our LDR matrix derived in Theorem 1 reduces to the LDF of [4].

Corollary 1. Assuming we have two multivariate normal populations \( N(\mu_i, \Sigma) \) and \( N(\mu_2, \Sigma) \), the proposed LDR matrix in Theorem 1 reduces to \( M = \Sigma^{-1}(\mu_2 - \mu_1) \), which is the well-known Fisher’s LDF.

Proof. The proof is immediate from (7).

5. Low-Dimensional Graphical Representations for Heteroscedastic Multivariate Normal Populations

5.1. Low-Dimensional LDR Using the SVD

If \( \text{rank}(M) = p \), one cannot use Theorem 1 to directly obtain a \( q \times p \) LDR matrix that preserves the full-feature BPMC. Also, in some situations when Theorem 1 holds, we may desire to determine a low-dimensional representation with dimension less than \( q \), say \( r \), where \( 1 \leq r < q < p \). Even if the required conditions of Theorem 1 hold, we may desire to determine a low-dimensional representation with dimension less than \( q \), say \( r \), where \( 1 \leq r < q < p \). Thus, we seek to construct an \( r \)-dimensional representation which preserves as much of the original \( p \)-dimensional BPMC as possible.

One method of obtaining an \( r \)-dimensional LDR matrix, \( 1 \leq r \leq q < p \), is the SVD approximation to \( M \) in (7). We use the following theorem from [17] to determine such an \( r \)-dimensional LDR matrix.
Theorem 2 Let $C_{ps}^{(p)}$ denote the class of all $s \times t$ real matrices of rank $p$, and let $C_{st}^{(r)}$ denote the class of all $s \times t$ real matrices of rank $r$, where $1 \leq r < q < p$. If $A_p \in C_{ps}^{(p)}$ and $A_r \in C_{st}^{(r)}$, given by $A_p = UD V'$, then

$$\|A_p - A_r\| < \|A_p - X\|$$

for all $X \in C_{st}^{(r)}$, where $A_p = UD V'$, $D_p = \text{diag}(d_1, \ldots, d_p)$, $D_r = \text{diag}(d_1, \ldots, d_r, 0_{s \times r-p})$, and $\|A_p\|$ is the usual Euclidean or Frobenius norm of a matrix $A_p$, given by

$$\|A_p\| = \left(\sum_{i=1}^{p} \sum_{j=1}^{t} |a_{ij}|^2\right)^{1/2}.$$

Furthermore, $\|A_p - A_r\| = \left(\sum_{i=1}^{r} d_i^2\right)^{1/2}$.

Using Theorems 1 and 2, we now construct a linear transformation for projecting high-dimensional data onto a low-dimensional subspace when all class distribution parameters are known. Let $M = UD V'$ be the SVD of the matrix $M$, where $D_p = \text{diag}(\lambda_1, \ldots, \lambda_p)$ for $\lambda_i \geq \lambda_i$ for $1 \leq i < j \leq p$, $\lambda_q > 0$, $\lambda_p > 0$. Let $F = UD_r$ and define $D_r = \text{diag}(\lambda_1, \ldots, \lambda_r, 0_{s \times r-p})$ with $\lambda_i \geq \lambda_i$ for $1 \leq i < j \leq q$. From Theorem 2, we have that $M_r = UD V'$ is a rank-$r$ approximation of $M$, and, therefore, a rank-$r$ approximation to $F$ is $F_r = UD_r$. Thus, $F_r$ is an $r \times p$ LDR matrix that yields an $r$-dimensional representation of the original $p$-dimensional class models. One can also use $F_r$ to construct low-dimensional representations of high-dimensional class densities.

We next provide an example to demonstrate the efficacy of Theorems 1 and 2 to determine low-dimensional representations for multiple multivariate normal populations with known mean vectors and covariance matrices. In the example, we display the simplicity of Theorem 1 to formulate a low-dimensional representation for three populations ($m = 3$) with unequal covariance matrices and original dimension $p = 6$. Note that unlike the low-dimensional representation of [18], our Theorem 1 does not restrict the reduced dimension to be $r = 1$.

5.2. Example

Consider the configuration $N\left(\mu_1, \Sigma_1\right)$, $N\left(\mu_2, \Sigma_2\right)$, and $N\left(\mu_3, \Sigma_3\right)$, where

$$\mu_1 = \left[0,0,0,0,0,0\right],$$

$$\mu_2 = \left[1,1,1,1,1,1\right],$$

$$\mu_3 = \left[2,2,2,2,2,2\right].$$

$$\Sigma_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad \text{and} \quad \Sigma_3 = \begin{bmatrix} 2 & 1 & 0 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 & 1 & 2 \end{bmatrix}. $$

We have $\text{rank}(M) = 2$, and, thus, by Theorem 1, the six-dimensional multivariate normal densities can be compressed to the dimension $q = 2$ without increasing the BPMC.

Using Theorem 1, we have that an optimal two-dimensional representation space is $\text{Col}(F)$, where

$$F = \begin{bmatrix} 0.3800 & 0.3800 & 0.5272 & 0.3800 & 0.3800 & 0.3800 \\ 0.2358 & 0.2358 & -0.8497 & 0.2358 & 0.2358 & 0.2358 \end{bmatrix}. $$

The optimal two-dimensional ellipsoidal representation is shown in Figure 1.
We can also determine a one-dimensional representation of the three multivariate normal populations through application of the SVD described in Theorem 2 applied to the matrix $M$ given in (7). A one-dimensional representation space is column one of the matrix $F$ in (8), and the graphical representation of this configuration of univariate normal densities is depicted in Figure 2.

6. Four LDR Methods for Statistical Discrimination

In this section, we present and describe the four LDR methods that we wish to compare and contrast in Sections 7 and 8.

Figure 1. The optimal two-dimensional representation for normal densities from the example in Section 5.2.

Figure 2. The optimal one-dimensional representation for normal densities from the example in Section 5.2.
6.1. The SY Method

In Theorem 1, we assume the parameters \( \mu_i \) and \( \Sigma_j \), \( i = 1, \cdots, m \), are known, but in reality this assumption is rarely the case. In a sampling situation, we can replace the columns of the \( \mathbf{M} \) matrix in (7) by their sample estimators yielding

\[
\hat{\mathbf{M}} = \left[ \mathbf{S}_1 \mathbf{x}_1 - \mathbf{S}_1 \mathbf{x}_1 | \mathbf{S}_1 \mathbf{x}_1 - \mathbf{S}_1 \mathbf{x}_1 | \cdots | \mathbf{S}_n \mathbf{x}_1 - \mathbf{S}_n \mathbf{x}_1 \right],
\]

provided \( n_i > p \), \( i = 1, \cdots, m \). Our estimator \( \hat{\mathbf{M}} \), along with Theorems 1 and 2, yields a LDR technique based on the selection of an \( r \)-dimensional hyperplane determined from a rank-\( r \) approximation to the full-rank matrix \( \hat{\mathbf{M}} \).

Thus, using the SVD, we let \( \hat{\mathbf{M}} = \mathbf{UDV}^\prime \), where \( \mathbf{D}_p = \text{diag}(\lambda_1, \cdots, \lambda_p) \), \( \lambda_j > 0 \) are the singular values of \( \mathbf{M} \) for \( j = 1, \cdots, p \), and let \( \hat{\mathbf{F}} = \mathbf{UD}_p \). Also, let

\[
\mathbf{D}_r = \text{diag}(\lambda_1, \cdots, \lambda_r, 0_{r+1}, \cdots, 0_p)
\]

with \( \lambda_j \geq \lambda_1 \) for \( 1 \leq j < l \leq p \) and \( 1 \leq r < p \). From Theorem 2, we have that \( \hat{\mathbf{M}} = \mathbf{UDV}^\prime \) is a rank-\( r \) approximation of \( \mathbf{M} \), and a rank-\( r \) approximation to \( \mathbf{F} \) is \( \hat{\mathbf{F}} = \mathbf{UD}_p \). Thus, \( \hat{\mathbf{F}}_r \) is our new \( r \times p \) LDR matrix. Provided \( \frac{(\sum_{j=1}^r \lambda_j)}{(\sum_{j=1}^p \lambda_j)} \) is relatively small, \( \hat{\mathbf{F}}_r \) will yield an EPMC(\( r \)) such that EPMC(\( p \)) \( \approx \) EPMC(\( r \)), and in certain population parameter configurations, we may have that EPMC(\( r \)) \( < \) EPMC(\( p \)). We refer to our new LDR matrix as the SY LDR method.

6.2. The BE Method

A second LDR method presented by [14] is

\[
\mathbf{U} = \left[ (\Sigma_i + \Sigma_j)^{\frac{1}{2}} (\mu_i - \mu_j) | \cdots | (\Sigma_i + \Sigma_j)^{\frac{1}{2}} (\mu_i - \mu_j) \right]
\]

for \( 1 \leq i < j \leq m \), provided all multivariate normal population parameters are known. For the unknown parameter case, an estimator of \( \mathbf{U} \) is then

\[
\hat{\mathbf{U}} = \left[ (\hat{\Sigma}_i + \hat{\Sigma}_j)^{\frac{1}{2}} (\hat{\mu}_i - \hat{\mu}_j) | \cdots | (\hat{\Sigma}_i + \hat{\Sigma}_j)^{\frac{1}{2}} (\hat{\mu}_i - \hat{\mu}_j) \right].
\]

The LDR matrix derived in [14], which we refer to as the BE matrix, was also obtained when we determined low-rank approximation to \( \hat{\mathbf{U}} \) by using the SVD. That is, let \( \hat{\mathbf{U}} = \mathbf{RD}_p \mathbf{S}_p^\prime \) be the SVD of \( \hat{\mathbf{U}} \), where

\[
\mathbf{D}_p = \text{diag}(\lambda_1, \cdots, \lambda_p)
\]

with \( \lambda_j \geq \lambda_1 \) for \( 1 \leq j < l \leq p \), and let \( \hat{\mathbf{H}} = \mathbf{RD}_p \). Define \( \mathbf{D}_r \) as in (9) with \( \lambda_j \geq \lambda_1 \) for \( 1 \leq j \leq r \). Then, \( \hat{\mathbf{U}} = \mathbf{RD}_p \mathbf{S}_p^\prime \) is a rank-\( r \) approximation of \( \hat{\mathbf{U}} \), and \( \hat{\mathbf{H}} = \mathbf{RD}_p = \mathbf{R}_p \in \mathbb{R}_{p \times r} \) is a rank-\( r \) approximation of \( \hat{\mathbf{H}} \). Thus, the BE LDR matrix is \( \hat{\mathbf{H}}_r \), or, equivalently, \( \mathbf{R}_p \), where the \( j^{th} \) eigenvector corresponds to the \( j^{th} \) largest singular value and \( j = 1, \cdots, r \).

The BE LDR approach is based on the rotated differences in the means. The LDR matrix (10) uses a type of pooled covariance matrix estimator for the precision matrices. However, the BE method does not incorporate all of the information contained in the different individual covariance matrices. Another disadvantage of the BE LDR approach is that it is limited to a reduced dimension that depends on the number of classes, \( m \). For \( m = 2 \), BE allows one to reduce the data to only one dimension, regardless of the full-feature vector dimension. Therefore, one may lose some discriminatory information with the application of the BE LDR method when the covariance matrices are considerably different.

6.3. Sliced Inverse Regression (SIR)

The next LDR method we consider is sliced inverse regression (SIR), which was proposed in [16]. Assuming all population parameters are known, we first define \( \mathbf{\Sigma}_w = \sum_{i=1}^m \mathbf{\Sigma}_i / m \) as the within-group covariance matrix and \( \mathbf{\Sigma}_g = \sum_{i=1}^m (\mu_i - \mu)(\mu_i - \mu) / m \) as the between-group covariance matrix, where \( \mu = \sum_{i=1}^m \mu_i / m \) is the overall
population mean. As its criterion matrix, the SIR LDR method uses
\[ M_{\text{SIR}} = \Gamma^{1/2} \Sigma_m \Gamma^{-1/2}, \]  
(12)
where \( \Gamma = \Sigma + \Sigma_m \) is the marginal covariance matrix of \( x \). When population parameters must be estimated, an estimator of (12) is
\[ \hat{M}_{\text{SIR}} = \hat{\Gamma}^{1/2} S_m \hat{\Gamma}^{-1/2}, \]
(13)
where \( \hat{\Gamma} = S_m + S_w \) with \( S_m \) and \( S_w \) given in (1) and (2), respectively. Let \( \hat{M}_{\text{SIR}} = GDp' \) be the SVD of (13), and let \( \hat{K} = \hat{\Gamma}^{1/2} G_r \), where \( G_r \) is composed of eigenvectors of (13) such that the \( j^{th} \) eigenvector of (13) corresponds to the \( j^{th} \) largest singular value of (13), \( j = 1, \cdots, p \). Then, \( \hat{K}_j = \hat{\Gamma}^{1/2} G_r \) is a rank-\( r \) approximation of \( \hat{K} \), and the \( r \times p \) SIR LDR matrix is \( \hat{K}' \), which is composed of the eigenvectors corresponding to the \( r \) largest singular values of \( \hat{K} \).

6.4. Sliced Average Variance Estimation (SAVE)

The last LDR method we consider is sliced average variance estimation (SAVE), which has been proposed in [19]-[20]. The SAVE method uses the \( p \times p \) criterion matrix \( M_{\text{SAVE}} = \sum_{i=1}^{p} (I_p - \Gamma^{1/2} \Sigma_i \Gamma^{-1/2})^2 / m, \) where \( \Gamma = \Sigma + \Sigma_m \). We use a form of SAVE given in [21], which is
\[ M_{\text{SAVE}} = \left( \Gamma^{1/2} \Sigma_m \Gamma^{-1/2} \right)^2 + \Gamma^{1/2} \Sigma_r \Gamma^{-1/2}, \]
(14)
where \( \Sigma_r = \sum_{i=1}^{m} (\Sigma_i - \Sigma_m) \Gamma^{-1} (\Sigma_i - \Sigma_m) / m \). An estimator of \( M_{\text{SAVE}} \) is
\[ \hat{M}_{\text{SAVE}} = \left( \hat{\Gamma}^{1/2} S_m \hat{\Gamma}^{-1/2} \right)^2 + \hat{\Gamma}^{1/2} S_r \hat{\Gamma}^{-1/2}, \]
(15)
where, \( S_r = \sum_{i=1}^{m} (S_i - S_m) \hat{\Gamma}^{-1} (S_i - S_m) / m \) with \( S_m \) and \( S_w \) given in (1) and (2), respectively. Next, let \( \hat{M}_{\text{SAVE}} = BDp' \) be the SVD of (15), and let \( \hat{L} = \hat{\Gamma}^{1/2} B_r \), where \( B_r \) is composed of the eigenvectors of (15) such that the \( j^{th} \) column of \( B_r \) corresponds to the \( j^{th} \) of (15) with the \( j^{th} \) largest singular value, \( j = 1, \cdots, p \). Then, \( \hat{L}_j = \hat{\Gamma}^{1/2} B_r \) is a rank-\( r \) approximation of \( \hat{L} \), and, thus, the \( r \times p \) SAVE LDR matrix is \( \hat{L}' \). An alternative representation of the \( r \)-dimensional SAVE LDR matrix is \( B' \), which is composed of the eigenvectors corresponding to the \( r \) largest singular values of \( B \).

7. A Monte Carlo Comparison of Four LDR Methods for Statistical Classification

Here, we compare our new SY LDR method derived above to the BE, SIR, and SAVE LDR methods. Specifically, we evaluate the classification efficacy in terms of the EPMC for the SY, BE, SIR, and SAVE LDR methods using Monte Carlo simulations for five different configurations of multivariate normal populations with \( p = 10 \). We have generated 10,000 training-sample and test datasets from the appropriate multivariate normal distributions for each parameter configuration. The test data were assigned to either population class \( \Pi_1 \) or \( \Pi_2 \) when \( m = 2 \) or \( \Pi_1 \), \( \Pi_2 \), or \( \Pi_3 \) when \( m = 3 \) using the sample QDF corresponding to (4). We have applied the four competing LDR matrices \( \hat{F}' \), \( \hat{H}' \), \( \hat{K}' \), and \( \hat{L}' \) and calculated the EPMCs for the full-dimensional QDF and for the four reduced-dimensional QDFs by averaging the estimated conditional error rate over all training samples. We examined the effect of the training-sample sizes on the four LDR methods using sample sizes \( n_1 = (2.5) p \) and \( n_2 = 5 p, \) \( i = 1, 2 \) or \( i = 1, 2, 3 \).

For the SY and SAVE LDR approaches, we reduce the dimension to \( r = 1, 2, 3 \). In general, for \( m \) populations, we remark that the BE LDR method can reduce the feature vector to at most the dimension \( r = m(m - 1)/2 \) because that is the column dimension of the matrix \( \hat{U} \in \mathbb{R}_{m \times m} \). This limitation is potentially a major drawback when the ratio \( m/p \) is small and especially when \( m = 2 \). In particular, for the BE and SIR LDR approaches, we can reduce only to the dimension \( r = 1 \) when \( m = 2 \). When \( m = 3 \), the BE LDR method can be applied to reduce the dimensions to at most \( r = 2, 3 \), and for the SIR LDR method, we can reduce the number of original features to at most \( r = m - 1 \). However, the SY LDR method avoids this shortcoming and allows one to reduce the original feature vector to our choice of reduced dimension \( r \), where \( 1 \leq r < p \), for any finite number of
populations $m$.
The five Monte Carlo simulations were generated using the programming language R. Table 1 gives a description of the number of populations and the theoretically optimal rank of the four LDR indices for each configuration. In Figures 3-7, we display the corresponding EPMCs of the four competing LDR methods for the various population configurations and values of $n_i$ and $r$, where $i = 1, \ldots, m$. The estimated standard error of all EPMCs in the following tables was less than 0.001. For each configuration, we also calculated $M$, $U$, $K$, and $L$ along with their respective singular values using the SVD. These singular values contain information concerning the amount of discriminatory information available in each reduced dimension. In the subsequent subsections, we use the following notation: $EPMC_{(SY)}$, $EPMC_{(BE)}$, $EPMC_{(SIR)}$, and $EPMC_{(SAVE)}$ denote the estimated EPMCs for the $SY$, $BE$, $SIR$, and $SAVE$ LDR methods, respectively, for each of the appropriate reduced dimensions.

7.1. Configuration 1: $m = 2$ with Moderately Different Covariance Matrices
The first population configuration we examined was composed of two multivariate normal populations $N_p(0, \Sigma_1)$ and $N_p(\mu_2, \Sigma_2)$, where $p = 10$,

$$\mu_1 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T,$$

$$\mu_2 = [5, 5, 5, 5, 5, 5, 5, 5, 5, 5]^T.$$

Table 1. A description of Monte Carlo simulation parametric configurations and singular values in Section 2 with unequal covariance matrices and $p = 10$.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Means</th>
<th>Rank</th>
<th>$m$</th>
<th>$M$</th>
<th>$U$</th>
<th>$K$</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Moderately separated unequal means</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Relatively close unequal means</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Relatively close unequal means</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Relatively close unequal means</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Close unequal means</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Simulation results from Section 7.1. The horizontal bar in each graph represents the $EPMC_{10}$. (a) $n_i = 25$; (b) $n_i = 50$. 
Figure 4. Simulation results from Configuration 2. The horizontal bar in each graph represents the \( EPMC_{10} \) with no dimension reduction. (a) \( n_i = 25 \); (b) \( n_i = 50 \).

Figure 5. Simulation results from Section 7.3. The horizontal bar in each graph represents the \( EPMC_{10} \). (a) \( n_i = 25 \); (b) \( n_i = 50 \).

\[
\Sigma_i = \begin{bmatrix}
10.00 & 4.00 & 5.00 & 4.00 & 3.00 & 4.00 & 4.00 & 5.00 & 4.00 & 3.00 \\
4.00 & 10.00 & 5.00 & 2.00 & 4.00 & 3.00 & 3.00 & 5.00 & 4.00 & 3.00 \\
5.00 & 5.00 & 10.00 & 5.00 & 5.00 & 3.00 & 4.00 & 4.00 & 4.00 & 4.00 \\
4.00 & 2.00 & 5.00 & 10.00 & 3.00 & 4.00 & 2.00 & 3.00 & 4.00 & 3.00 \\
3.00 & 4.00 & 5.00 & 3.00 & 10.00 & 3.00 & 4.00 & 5.00 & 3.00 & 3.00 \\
4.00 & 3.00 & 3.00 & 4.00 & 3.00 & 12.00 & 3.00 & 4.00 & 4.00 & 4.00 \\
4.00 & 3.00 & 4.00 & 2.00 & 4.00 & 3.00 & 14.00 & 2.00 & 2.00 & 2.00 \\
5.00 & 5.00 & 4.00 & 3.00 & 5.00 & 4.00 & 2.00 & 12.00 & -0.50 & -0.50 \\
4.00 & 4.00 & 4.00 & 4.00 & 3.00 & 4.00 & 2.00 & -0.50 & 14.00 & -1.00 \\
3.00 & 3.00 & 4.00 & 3.00 & 3.00 & 4.00 & 2.00 & -0.50 & -1.00 & 11.00
\end{bmatrix}
\]
Figure 6. Simulation results from Configuration 4. The horizontal bar in each graph represents the \( EPMC \) with no dimension reduction. (a) \( n_i = 25 \); (b) \( n_i = 50 \).

Figure 7. Simulation results from Section 7.5. The horizontal bar in each graph represents the \( EPMC_{18} \). (a) \( n_i = 25 \); (b) \( n_i = 50 \).

\[
\Sigma_2 = \begin{bmatrix}
16.00 & 10.00 & 11.00 & 10.00 & 9.00 & 10.00 & 10.00 & 11.00 & 10.00 & 9.00 \\
10.00 & 16.00 & 11.00 & 8.00 & 9.00 & 10.00 & 9.00 & 11.00 & 10.00 & 9.00 \\
11.00 & 11.00 & 16.00 & 11.00 & 11.00 & 9.00 & 10.00 & 10.00 & 10.00 & 10.00 \\
10.00 & 8.00 & 11.00 & 16.00 & 9.00 & 10.00 & 8.00 & 9.00 & 10.00 & 9.00 \\
9.00 & 10.00 & 11.00 & 9.00 & 16.00 & 9.00 & 10.00 & 11.00 & 9.00 & 9.00 \\
10.00 & 9.00 & 9.00 & 10.00 & 9.00 & 18.00 & 9.00 & 10.00 & 10.00 & 10.00 \\
10.00 & 9.00 & 10.00 & 8.00 & 10.00 & 9.00 & 20.00 & 8.00 & 8.00 & 8.00 \\
11.00 & 11.00 & 10.00 & 9.00 & 11.00 & 10.00 & 8.00 & 22.00 & 9.50 & 9.50 \\
10.00 & 10.00 & 10.00 & 10.00 & 9.00 & 10.00 & 8.00 & 9.50 & 24.00 & 9.00 \\
9.00 & 9.00 & 10.00 & 9.00 & 9.00 & 10.00 & 8.00 & 9.50 & 9.00 & 21.00
\end{bmatrix}
\]
Here, \( \text{rank}(\Sigma_2 - \Sigma_i) = 1 \), which implies \( \text{rank}(\mathbf{M}) = 2 \) because \( (\Sigma_1^{-1} \mu_i - \Sigma_i^{-1} \mu) \notin \text{span}(\Sigma_2 - \Sigma_i) \). The singular values of \( \hat{\mathbf{M}} \) in Table 2 indicate that most of the classificatory information can be captured when \( r = 1 \), because the subsequent singular values are small relative to the first. Also, the BE LDR technique loses classificatory information from pooling the acutely dissimilar pair of covariance matrices.

When \( n_i = 25 \), the EPMC was reduced by the SY, BE, and SIR LDR methods but not for the SAVE LDR method. This effect occurred because the training-sample size \( n_i = 25 \), \( i = 1, 2 \), are small relative to the full-feature dimensionality \( p = 10 \), and, therefore, insufficient data were available to accurately estimate the \( p(p + 3) \) total population parameters. Thus, by reducing the full-feature dimension \( p = 10 \) to dimension \( r \approx 1 \), we considerably increased the ratio of the training-sample size relative to the original new dimension so that \( n_i/r \gg n_i/p \) to \( n_i/r \), where \( r \ll p \). Thus, we achieved improved parameter estimates in the \( r \)-dimensional subspaces. Not surprisingly, for \( n_i = 25 \), \( i = 1, 2 \), we found that \( [\text{EPMC}_{10} - \text{EPMC}_i \left( \text{SY} \right)] \approx 0.06 \) and for \( n_i = 50 \), \( [\text{EPMC}_{10} - \text{EPMC}_i \left( \text{SY} \right)] \approx 0.01 \), which demonstrated the advantage of employing LDR in the classification process, and, more specifically, demonstrated the value of the SY LDR method. Additionally, as \( n_i \) increased, the \( \text{EPMC}_i \) approached \( \text{EPMC}_p \) for all four LDR methods as \( r \) increased.

In addition, the SIR LDR method did not utilize discriminatory information contained in the differences of the covariance matrices when \( n_i = 25 \), \( i = 1, 2 \). However, for \( n_i = 50 \), \( i = 1, 2 \), all four LDR methods yielded essentially the same EPMC, except for SAVE when \( r = 1 \).

7.2. Configuration 2: \( m = 3 \) with Two Similar Covariance Matrices and One Spherical Covariance Matrix

The second Monte Carlo simulation used a configuration with the three multivariate normal populations \( N_p(\mu_i, \Sigma_i) \), \( N_p(\mu_2, \Sigma_2) \), and \( N_p(\mu_3, \Sigma_3) \) where \( p = 10 \),

\[
\mu_1 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]',
\mu_2 = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]',
\mu_3 = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2]',
\]

Table 2. Summary of singular values for the four competing LDR methods for Configuration 1.

<table>
<thead>
<tr>
<th>Singular value</th>
<th>( n_i = 25 )</th>
<th>( n_i = 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\mathbf{M}} )</td>
<td>( \hat{\mathbf{U}} )</td>
</tr>
<tr>
<td>( e_1 )</td>
<td>72.16</td>
<td>0.38</td>
</tr>
<tr>
<td>( e_2 )</td>
<td>16.27</td>
<td>0.54</td>
</tr>
<tr>
<td>( e_3 )</td>
<td>12.21</td>
<td>0.43</td>
</tr>
<tr>
<td>( e_4 )</td>
<td>9.31</td>
<td>0.34</td>
</tr>
<tr>
<td>( e_5 )</td>
<td>7.22</td>
<td>0.25</td>
</tr>
<tr>
<td>( e_6 )</td>
<td>5.46</td>
<td>0.18</td>
</tr>
<tr>
<td>( e_7 )</td>
<td>3.99</td>
<td>0.11</td>
</tr>
<tr>
<td>( e_8 )</td>
<td>2.73</td>
<td>0.07</td>
</tr>
<tr>
<td>( e_9 )</td>
<td>1.67</td>
<td>0.03</td>
</tr>
<tr>
<td>( e_{10} )</td>
<td>0.75</td>
<td>0.01</td>
</tr>
</tbody>
</table>
In this configuration, the population means are unequal but relatively close. Moreover, \( \Sigma_2 \) and \( \Sigma_3 \) are unequal but notably more similar to one another than to \( \Sigma_1 \). However, the variance of the third feature in \( \Pi_1 \) is significantly greater than the population variances of the third feature for either \( \Pi_1 \) or \( \Pi_2 \).

As a result of markedly different covariance matrices, the \( BE \) and \( SIR \) \( LDR \) methods are not ideal because both methods aggregated the sample covariance matrices. The \( S Y \) \( LDR \) method, however, attempts to estimate each individual covariance matrix for all three populations, and uses this information that yielded \( SY \) as the superior \( LDR \) procedure. Table 3 gives the singular values of each of the \( LDR \) methods considered here.

For the larger sample-size scenario, \( n_i = 50 \), \( i = 1, 2 \), the \( SAVE \) \( LDR \) was more competitive with \( SY \) than either \( BE \) or \( SIR \) when \( r > 1 \). However, the \( SY \) \( LDR \) remained the preferred \( LDR \) method. In Figure 4, we added noise to the \( SY \) method when we used \( r > 2 \), and, we eliminated essential discriminatory information for the \( SY \) procedure when we chose \( r = 1 \). Thus, the optimal choice for the reduced dimension for the \( SY \) \( LDR \) method was \( r = 2 \), regardless of the training-sample size of \( n_i \), \( i = 1, 2, 3 \). Furthermore, in our simulation we determined...
Table 3. Summary of singular values for the four competing LDR methods applied to Configuration 2.

<table>
<thead>
<tr>
<th>Singular value</th>
<th>$n_i = 25$</th>
<th>$n_i = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>15.18</td>
<td>14.27</td>
</tr>
<tr>
<td>$e_2$</td>
<td>8.92</td>
<td>8.89</td>
</tr>
<tr>
<td>$e_3$</td>
<td>4.07</td>
<td>2.90</td>
</tr>
<tr>
<td>$e_4$</td>
<td>2.51</td>
<td>1.50</td>
</tr>
<tr>
<td>$e_5$</td>
<td>1.84</td>
<td>1.21</td>
</tr>
<tr>
<td>$e_6$</td>
<td>1.49</td>
<td>1.01</td>
</tr>
<tr>
<td>$e_7$</td>
<td>1.22</td>
<td>0.84</td>
</tr>
<tr>
<td>$e_8$</td>
<td>1.00</td>
<td>0.70</td>
</tr>
<tr>
<td>$e_9$</td>
<td>0.79</td>
<td>0.56</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.58</td>
<td>0.42</td>
</tr>
</tbody>
</table>

that $\left[ EPMC_{10} (SY) - EPMC_{2} (SY) \right] \approx 0.10$ for $n_i = 50$ and $\left[ EPMC_{10} (SY) - EPMC_{2} (SY) \right] \approx 0.15$ for $n_i = 25$, $i = 1, 2, 3$. This significant reduction from $EPMC_{10}$ demonstrated a significant benefit of dimension reduction in general and the $SYLDR$ in particular.

7.3. Configuration 3: $m = 2$ with Relatively Close Means and Different But Similar Covariance Matrices

In this configuration, we have $N_p (\mathbf{\mu}, \Sigma)$ and $N_p (\mathbf{\mu}_z, \Sigma_z)$ with $p = 10$,

$$\mathbf{\mu} = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]',$$

$$\mathbf{\mu}_z = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2]'$$

and $\Sigma_i$
As in Configuration 1, we have that $rank(\Sigma_2 - \Sigma_1) = 1$, and, hence, $rank(M) = 2$. In Configuration 3, the BE and SIR LDR methods outperformed the SY and SAVE LDR methods because of the similarity in the covariance matrices. This phenomenon occurred because both LDR methods aggregated the sample covariance matrix estimators and, therefore, smaller values of $EPMC_{BE}^1$ and $EPMC_{SIR}^1$. From Table 4, we see that the first singular value for $\hat{M}$ was considerably less predominant than the first singular value for $M$ in Configuration 1. This result explained the inferiority of the SY method for this configuration, regardless of the chosen $n_i$, $i = 1, 2$.

For $r = 1$, we have that $\left[ EPMC_{10}^{BE} - EPMC_{i}^{SIR} \right] \approx \left[ EPMC_{10} - EPMC_{i}^{BE} \right] \approx 0.04$ for $n_i = 25$ and $\left[ EPMC_{10} - EPMC_{i}^{SIR} \right] \approx \left[ EPMC_{10} - EPMC_{i}^{BE} \right] \approx 0.06$ for $n_i = 50$. Again, we exemplify the value of LDR as a classification tool. In this configuration, not only was $EPMC_{10}$ considerably reduced as $n_i$ increased, but the difference between $\left[ EPMC_{10} - EPMC_{i}^{BE} \right]$ and $\left[ EPMC_{10} - EPMC_{i}^{SIR} \right]$ decreased as well. This example demonstrated the fact that the SY method is not a uniformly superior LDR approach even when covariance matrices are unequal. However, $EPMC_{i}^{SY}$ was not much greater than either $EPMC_{i}^{BE}$ or $EPMC_{i}^{SIR}$. Here, SAVE is not as competitive as the three other LDR methods because it does not use all the information in the difference of the covariance matrices and means to obtain a better information-preserving subspace.

### 7.4. Configuration 4: $m = 3$ with Two Similar Covariance Matrices Except for the First Two Dimensions

In this situation, we have three multivariate normal populations: $N_p(\mu, \Sigma_1)$, $N_p(\mu, \Sigma_2)$, and $N_p(\mu, \Sigma_3)$.

<table>
<thead>
<tr>
<th>Singular value</th>
<th>$n_i = 25$</th>
<th>$n_i = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>27.91</td>
<td>22.98</td>
</tr>
<tr>
<td>$e_2$</td>
<td>14.84</td>
<td>10.48</td>
</tr>
<tr>
<td>$e_3$</td>
<td>10.73</td>
<td>7.69</td>
</tr>
<tr>
<td>$e_4$</td>
<td>8.12</td>
<td>5.86</td>
</tr>
<tr>
<td>$e_5$</td>
<td>6.18</td>
<td>4.50</td>
</tr>
<tr>
<td>$e_6$</td>
<td>4.65</td>
<td>3.39</td>
</tr>
<tr>
<td>$e_7$</td>
<td>3.41</td>
<td>2.50</td>
</tr>
<tr>
<td>$e_8$</td>
<td>2.41</td>
<td>1.78</td>
</tr>
<tr>
<td>$e_9$</td>
<td>1.55</td>
<td>1.16</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.76</td>
<td>0.60</td>
</tr>
</tbody>
</table>

As in Configuration 1, we have that $rank(\Sigma_2 - \Sigma_1) = 1$, and, hence, $rank(M) = 2$. In Configuration 3, the BE and SIR LDR methods outperformed the SY and SAVE LDR methods because of the similarity in the covariance matrices. This phenomenon occurred because both LDR methods aggregated the sample covariance matrix estimators and, therefore, smaller values of $EPMC_{BE}^1$ and $EPMC_{SIR}^1$. From Table 4, we see that the first singular value for $\hat{M}$ was considerably less predominant than the first singular value for $M$ in Configuration 1. This result explained the inferiority of the SY method for this configuration, regardless of the chosen $n_i$, $i = 1, 2$.

For $r = 1$, we have that $\left[ EPMC_{10}^{BE} - EPMC_{i}^{SIR} \right] \approx \left[ EPMC_{10} - EPMC_{i}^{BE} \right] \approx 0.04$ for $n_i = 25$ and $\left[ EPMC_{10} - EPMC_{i}^{SIR} \right] \approx \left[ EPMC_{10} - EPMC_{i}^{BE} \right] \approx 0.06$ for $n_i = 50$. Again, we exemplify the value of LDR as a classification tool. In this configuration, not only was $EPMC_{10}$ considerably reduced as $n_i$ increased, but the difference between $\left[ EPMC_{10} - EPMC_{i}^{BE} \right]$ and $\left[ EPMC_{10} - EPMC_{i}^{SIR} \right]$ decreased as well. This example demonstrated the fact that the SY method is not a uniformly superior LDR approach even when covariance matrices are unequal. However, $EPMC_{i}^{SY}$ was not much greater than either $EPMC_{i}^{BE}$ or $EPMC_{i}^{SIR}$. Here, SAVE is not as competitive as the three other LDR methods because it does not use all the information in the difference of the covariance matrices and means to obtain a better information-preserving subspace.

### 7.4. Configuration 4: $m = 3$ with Two Similar Covariance Matrices Except for the First Two Dimensions

In this situation, we have three multivariate normal populations: $N_p(\mu, \Sigma_1)$, $N_p(\mu, \Sigma_2)$, and $N_p(\mu, \Sigma_3)$.

<table>
<thead>
<tr>
<th>Singular value</th>
<th>$n_i = 25$</th>
<th>$n_i = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>27.91</td>
<td>22.98</td>
</tr>
<tr>
<td>$e_2$</td>
<td>14.84</td>
<td>10.48</td>
</tr>
<tr>
<td>$e_3$</td>
<td>10.73</td>
<td>7.69</td>
</tr>
<tr>
<td>$e_4$</td>
<td>8.12</td>
<td>5.86</td>
</tr>
<tr>
<td>$e_5$</td>
<td>6.18</td>
<td>4.50</td>
</tr>
<tr>
<td>$e_6$</td>
<td>4.65</td>
<td>3.39</td>
</tr>
<tr>
<td>$e_7$</td>
<td>3.41</td>
<td>2.50</td>
</tr>
<tr>
<td>$e_8$</td>
<td>2.41</td>
<td>1.78</td>
</tr>
<tr>
<td>$e_9$</td>
<td>1.55</td>
<td>1.16</td>
</tr>
<tr>
<td>$e_{10}$</td>
<td>0.76</td>
<td>0.60</td>
</tr>
</tbody>
</table>
where \( p = 10 \),

\[
\mu = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^\top,
\]

\[
\mu_1 = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]^\top,
\]

\[
\mu_2 = [-1, -1, -1, -1, -1, -1, -1, -1, -1, -1]^\top,
\]

\[
\Sigma_1 = \begin{bmatrix}
2.00 & 0.80 & 1.00 & 0.80 & 0.80 & 0.80 & 1.00 & 0.80 & 0.80 & 0.60 \\
0.80 & 2.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\
1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\
0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.60 \\
0.80 & 0.60 & 0.80 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\
1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\
0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\
0.60 & 0.60 & 0.80 & 0.60 & 0.80 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 
\end{bmatrix},
\]

\[
\Sigma_2 = \begin{bmatrix}
20.00 & 0.80 & 1.00 & 0.80 & 0.80 & 0.80 & 1.00 & 0.80 & 0.80 & 0.60 \\
0.80 & 40.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\
1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\
0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 1.00 & 0.60 & 0.60 \\
0.80 & 0.60 & 0.80 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\
1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\
0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\
0.60 & 0.60 & 0.80 & 0.60 & 0.80 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 
\end{bmatrix},
\]

and

\[
\Sigma_3 = \begin{bmatrix}
35.00 & 0.80 & 1.00 & 0.80 & 0.80 & 0.80 & 1.00 & 0.80 & 0.80 & 0.60 \\
0.80 & 40.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\
1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\
0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 1.00 & 0.60 & 0.60 \\
0.80 & 0.60 & 0.80 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\
0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\
1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\
0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\
0.60 & 0.60 & 0.80 & 0.60 & 0.80 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 
\end{bmatrix},
\]

For the fourth configuration, the covariance matrices are considerably different from one another, which benefits both the SY and SAVE LDR methods. Specifically, the SY LDR procedure uses information contained in the unequal covariance matrices to determine classificatory information contained in \( S_j - S_i \) and in \( S_j^{-1} x_j - S_i^{-1} x_i \), \( j = 2,3 \). The largest EPMC, for all four LDR methods occurred when \( r = 1 \). Generally, the EPMC, de-
increased as \( r \) increased. For the SY method, we have \( \text{rank}(\mathbf{M}) = 4 \), which clarifies the reason that \( EPMC_r(SY) \) and \( r \) were inversely related. As one can see in Table 5, the first three singular values of \( \hat{\mathbf{M}} \) were relatively large. However, the pooling of the sample covariance matrices used in \( BE \) and \( SIR \) tended to obscure classificatory information in the sample covariance matrices. The only improvement from the full dimension we found for the values of \( r \) and \( n_i \) considered here was the SY method when \( r = 3 \) for \( n_i = 25, i = 1, 2, 3 \). Specifically, we have that \( \left[ EPMC_{10} - EPMC_3(SY) \right] \approx 0.05 \) when \( n_i = 25 \), although \( \left[ EPMC_{10} - EPMC_3(SY) \right] \approx -0.01 \) when \( n_i = 50, i = 1, 2, 3 \).

This population configuration illustrated the fact that we cannot always choose \( r = 1 \) and expect to see a reduction in \( EPMC_r \). However, we can often reduce the \( EPMC_r \) if we use both a judicious choice of \( r \) and an appropriate LDR method.

### 7.5. Configuration 5: \( m = 3 \) with Diverse Population Covariance Matrices

In Configuration 5, we have three multivariate normal populations: \( N_p(\mathbf{\mu}_1, \mathbf{\Sigma}_1) \), \( N_p(\mathbf{\mu}_2, \mathbf{\Sigma}_2) \), and \( N_p(\mathbf{\mu}_3, \mathbf{\Sigma}_3) \), where \( p = 10 \),

\[
\mathbf{\mu}_1 = \begin{bmatrix} 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 \end{bmatrix},
\mathbf{\mu}_2 = \begin{bmatrix} 4.43, 4.43, 4.43, 4.43, 4.43, 4.43, 4.43, 4.43, 4.43, 4.43 \end{bmatrix},
\mathbf{\mu}_3 = \begin{bmatrix} 8, 8, 8, 8, 8, 8, 8, 8, 8, 8 \end{bmatrix},
\]

with

\[
\mathbf{\Sigma}_1 = \begin{bmatrix}
15.01 & 0.81 & 1.25 & 1.13 & -2.10 & -2.30 & 2.43 & -3.30 & -0.87 & -1.11 \\
0.81 & 26.10 & 1.51 & -0.74 & 0.89 & 4.35 & 1.25 & 1.85 & -0.50 & -0.39 \\
1.25 & 1.51 & 24.55 & -5.57 & 3.96 & 1.62 & -0.27 & 0.49 & -5.97 & 0.87 \\
1.13 & -0.74 & -5.57 & 29.36 & -3.58 & -0.89 & 2.21 & -3.71 & -0.52 & -2.19 \\
-2.10 & 0.89 & 3.96 & -3.58 & 20.17 & 6.05 & -5.20 & 2.22 & -0.80 & -2.69 \\
-2.30 & 4.35 & 1.62 & -0.89 & 6.05 & 40.18 & -5.18 & 3.83 & 1.94 & 0.51 \\
2.43 & 1.25 & -0.27 & 2.21 & -5.20 & -5.18 & 17.93 & -0.17 & -3.09 & 0.49 \\
-3.30 & 1.85 & 0.49 & -3.71 & 2.22 & 3.83 & -0.17 & 26.05 & -0.54 & 3.34 \\
-0.87 & -0.50 & -5.97 & -0.52 & -0.80 & 1.94 & -3.09 & -0.54 & 16.30 & -1.04 \\
-1.11 & -0.39 & 0.87 & -2.19 & -2.69 & 0.51 & 0.49 & 3.34 & -1.04 & 26.84
\end{bmatrix}.
\]

**Table 5. Summary of singular values for the four competing LDR methods for Configuration 4.**

<table>
<thead>
<tr>
<th>Singular value</th>
<th>( n_i = 25 )</th>
<th>( n_i = 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e_1 )</td>
<td>59.76</td>
<td>0.90</td>
</tr>
<tr>
<td>( e_2 )</td>
<td>36.37</td>
<td>0.66</td>
</tr>
<tr>
<td>( e_3 )</td>
<td>7.72</td>
<td>0.52</td>
</tr>
<tr>
<td>( e_4 )</td>
<td>5.21</td>
<td>0.42</td>
</tr>
<tr>
<td>( e_5 )</td>
<td>3.97</td>
<td>0.34</td>
</tr>
<tr>
<td>( e_6 )</td>
<td>3.12</td>
<td>0.27</td>
</tr>
<tr>
<td>( e_7 )</td>
<td>2.46</td>
<td>0.21</td>
</tr>
<tr>
<td>( e_8 )</td>
<td>1.91</td>
<td>0.16</td>
</tr>
<tr>
<td>( e_9 )</td>
<td>1.44</td>
<td>0.11</td>
</tr>
<tr>
<td>( e_{10} )</td>
<td>0.98</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Here, we have three considerably different covariance matrices. For this population configuration, the SY method outperformed the three other LDR methods for the reduced dimensions $r = 1, 2, 3$. Examining the singular values of $\hat{M}$, we see that most of the discriminator information was contained in the first transformed dimension. This fact was illustrated with how $EPMC_{SY}$ and $r$ were directly related, regardless of the training-sample size $i = 1, 2, 3$. The BE and SIR LDR methods did not perform as well here because of the pooling of highly diverse estimated covariance matrices. While the SAVE LDR procedure was the least effective LDR method for $n_i = 25$, $EPMC(SAVE)$ and $n_i$ were inversely related. While each of the four LDR methods decreased from $EPMC_{10}$ for some combination of $n_i$ and $r$, the largest reduction in $EPMC$ was for the SY LDR method where $\left[ EPMC_{10} - EPMC_i(SY) \right] \approx 0.09$ for $n_i = 50$ and $\left[ EPMC_{10} - EPMC_i(SY) \right] \approx 0.14$ for $n_i = 25$, $i = 1, 2, 3$. This result implied that the SY LDR method was especially useful when $n_i/p$ was relatively small and the covariance matrices were considerably different.

8. A Parametric Bootstrap Simulation

In the following parametric bootstrap simulation, we use a real dataset to obtain the population means and covariance matrices for three multivariate normal populations. The chosen dataset comes from the University of California at Irvine Machine Learning Repository, which describes the diagnoses of cardiac Single Proton Emission Computed Tomography (SPECT) images. Each patient in the study is classified into two categories: normal or abnormal. The dataset contains 267 SPECT image sets of patients. Each observation consists of 44 continuous features for each patient. However, for our simulation, we chose only ten of the 44 features. The ten selected features were F2R, F6R, F7S, F9S, F11R, F11S, F14S, F16R, F17R, and F19S. Hence, we performed the parametric bootstrap Monte Carlo simulation with two populations: $N(\mu, \Sigma_1)$ and $N(\mu, \Sigma_2)$, where

$$\mu = \begin{bmatrix} 61.68, 65.85, 70.37, 70.48, 66.01, 68.21, 60.18, 73.78, 71.68, 63.66 \end{bmatrix}^T,$$

and


and

$$\Sigma_2 = \begin{bmatrix} 75.01 & 60.81 & 61.25 & 61.13 & 57.90 & 57.70 & 62.43 & 56.70 & 59.13 & 58.89 \\ 60.81 & 86.10 & 61.51 & 59.26 & 60.89 & 64.35 & 61.25 & 61.85 & 59.50 & 59.61 \\ 61.25 & 61.51 & 84.55 & 54.43 & 63.96 & 61.62 & 59.73 & 60.49 & 54.03 & 60.87 \\ 61.13 & 59.26 & 54.43 & 89.36 & 56.42 & 59.11 & 62.21 & 56.29 & 59.48 & 57.81 \\ 57.90 & 60.89 & 63.96 & 65.42 & 80.17 & 66.05 & 54.80 & 62.22 & 59.20 & 57.31 \\ 57.70 & 64.35 & 61.25 & 59.73 & 62.21 & 54.80 & 54.82 & 77.93 & 59.83 & 56.91 \\ 62.43 & 61.25 & 59.73 & 62.21 & 54.80 & 54.82 & 77.93 & 59.83 & 56.91 & 60.49 \\ 56.70 & 61.85 & 60.49 & 56.29 & 62.22 & 63.83 & 59.83 & 86.05 & 59.46 & 63.34 \\ 59.13 & 59.50 & 54.03 & 59.48 & 59.20 & 61.94 & 56.91 & 59.46 & 76.30 & 58.96 \\ 58.89 & 59.61 & 60.87 & 57.81 & 57.31 & 60.51 & 60.49 & 63.34 & 58.96 & 86.84 \end{bmatrix}.$$

<table>
<thead>
<tr>
<th>Singular value</th>
<th>$n_i = 25$</th>
<th>$n_i = 50$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{M}$</td>
<td>691.33</td>
<td>683.07</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.37</td>
<td>0.13</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.36</td>
<td>0.30</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>0.72</td>
<td>0.53</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>50.13</td>
<td>0.07</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.46</td>
<td>0.31</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>35.78</td>
<td>0.24</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>39.44</td>
<td>28.39</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.37</td>
<td>0.19</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>21.43</td>
<td>0.12</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>13.41</td>
<td>0.09</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>26.27</td>
<td>19.10</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.24</td>
<td>0.12</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.30</td>
<td>0.15</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>17.22</td>
<td>0.12</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>13.41</td>
<td>9.89</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.09</td>
<td>9.89</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.13</td>
<td>0.07</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>22.67</td>
<td>21.43</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.26</td>
<td>15.67</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.14</td>
<td>16.68</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>17.22</td>
<td>14.73</td>
</tr>
<tr>
<td>$\hat{M}$</td>
<td>39.44</td>
<td>13.33</td>
</tr>
<tr>
<td>$\hat{U}$</td>
<td>0.37</td>
<td>8.80</td>
</tr>
<tr>
<td>$\hat{K}$</td>
<td>0.08</td>
<td>13.60</td>
</tr>
<tr>
<td>$\hat{L}$</td>
<td>21.43</td>
<td>17.75</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>[62.60, 67.90, 71.60, 71.00, 69.40, 62.30, 75.90, 72.90, 66.00]</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1$</td>
<td>[131.90 62.83 6.01 1.82 6.75 0.79 16.74 3.69 30.97 9.27</td>
<td></td>
</tr>
<tr>
<td>$\Sigma_2$</td>
<td>[33.54 18.26 1.10 0.69 3.29 10.69 10.57 2.99 9.97 13.07</td>
<td></td>
</tr>
</tbody>
</table>

For this dataset, $\text{rank}(M) = \text{rank}(\hat{M}) = 10$, $\text{rank}(U) = \text{rank}(\hat{U}) = 1$, $\text{rank}(K) = \text{rank}(\hat{K}) = 2$, and $\text{rank}(L) = \text{rank}(\hat{L}) = 10$. Table 6 gives the singular values for each of the LDR methods considered here.

Figure 8 gives the EPMCs for the reduced dimensions 1, 2, and 3 for each of the LDR methods. In this example, a considerable amount of discriminatory information is contained in the covariance matrices, which are ex-
tremely different. Hence, not surprisingly, neither the BE nor the SIR LDR methods performed well for \( r = 1, 2, 3 \). Furthermore, \( EPMC_r(SY) \), \( EPMC_r(SIR) \), and \( EPMC_r(SAVE) \) were inversely related to \( r \), regardless of the training-sample size \( n_i \), \( i = 1, 2 \). From Table 7, we see that one reason that \( EPMC_r(SY) \) increased as \( r \) decreased is that the singular values are relatively large up to \( r = 7 \). Thus, when we chose \( r = 1, 2, 3 \), we discarded some necessary discriminatory information.

For \( n_i = 50 \), \( i = 1, 2 \), SAVE slightly outperformed the other three LDR methods, though the SY LDR approach was very competitive. For \( n_i = 25 \), \( i = 1, 2 \), we found that SY \( EPMC_r(SY) \) was the smallest error rate and that \( EPMC_{10} \approx EPMC_r(SY) \). This particular example verified that one can implement the SY method and obtain excellent results even though the conditions of Theorem 1 do not essentially hold. This example also illustrated the prospect that, with a judicious choice of \( r \) combined with the appropriate LDR method, we can significantly reduce the feature dimension while still preserving the EMPC.

<table>
<thead>
<tr>
<th>Singular value</th>
<th>( n_i = 25 )</th>
<th>( n_i = 50 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e_1 )</td>
<td>390.43</td>
<td>275.15</td>
</tr>
<tr>
<td>( e_2 )</td>
<td>198.20</td>
<td>175.62</td>
</tr>
<tr>
<td>( e_3 )</td>
<td>118.13</td>
<td>111.28</td>
</tr>
<tr>
<td>( e_4 )</td>
<td>74.01</td>
<td>69.53</td>
</tr>
<tr>
<td>( e_5 )</td>
<td>52.16</td>
<td>47.01</td>
</tr>
<tr>
<td>( e_6 )</td>
<td>37.65</td>
<td>35.02</td>
</tr>
<tr>
<td>( e_7 )</td>
<td>25.72</td>
<td>24.92</td>
</tr>
<tr>
<td>( e_8 )</td>
<td>16.73</td>
<td>15.91</td>
</tr>
<tr>
<td>( e_9 )</td>
<td>9.50</td>
<td>8.60</td>
</tr>
<tr>
<td>( e_{10} )</td>
<td>4.27</td>
<td>3.81</td>
</tr>
</tbody>
</table>

Table 7. Summary of singular values for the four competing LDR methods for the parametric bootstrap example.

Figure 8. Simulation results from Section 8. The horizontal bar in each graph represents the \( EPMC_r \). (a) \( n_i = 25 \); (b) \( n_i = 50 \).
9. Discussion

In this paper, while all population parameters are known, we have presented a simple and flexible algorithm for a low-dimensional representation of data from multiple multivariate normal populations with different parametric configurations. Also, we have given necessary and sufficient conditions for attaining the subspace of smallest dimension $q < p$, which preserves the original Bayes classification assignments. We have provided a constructive proof for obtaining a low-dimensional representation space when certain population parameter conditions are satisfied. Under a special case for the two-class multivariate normal problem with equal nonsingular covariance structures, our proposed LDR transformation is the LDF in [4]. Moreover, we have also extended our concept proposed in Theorem 1 to cases where the conditions in Theorem 1 are not satisfied through the application of the SVD given in Theorem 2.

We have presented several advantages of our proposed low-dimensional representation method. First, our method is not restricted to a one-dimensional representation regardless of the number of populations, unlike the transformation introduced by [14]. Second, our method allows for equal and unequal covariance structures. Third, the original feature dimension $p$ does not significantly impact the computational complexity. Also, under certain conditions, one-dimensional representations of populations with unequal covariance structures can be accomplished without an appreciable increase in the EPMC.

Furthermore, we have derived a LDR method for realistic cases where the class population parameters are unknown and the linear sufficient matrix $M$ in (7) must be estimated using training data. Using Monte Carlo simulation studies, we have compared the performance of the SY LDR method with three other LDR procedures derived by [14] [16] [19]. In the Monte Carlo simulation studies, we have demonstrated that the full-dimension EMPC can sometimes be decreased by implementing the SY LDR method when the training-sample size is small relative to the total number of estimated parameters.

Finally, we have extended our concept proposed in Theorem 1 through the application of the SVD given in Theorem 2 to cases where one might wish to considerably reduce the original feature dimension. Our new LDR approach can yield excellent results provided the population covariance matrices are sufficiently different.

References


