Open Journal of Statistics, 2015, 5, 311-333
Published Online June 2015 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2015.54033

Linear Dimension Reduction for Multiple Heteroscedastic Multivariate Normal Populations

Songthip T. Ounpraseuth[1], Phil D. Young[2], Johanna S. van Zyl[2], Tyler W. Nelson[2], Dean M. Young[2]

[1]Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AK, USA
[2]Department 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

How to cite this paper: Ounpraseuth, S.T., Young, P.D., van Zyl, J.S., Nelson, T.W. and Young, D.M. (2015) Linear Dimension Reduction for Multiple Heteroscedastic Multivariate Normal Populations. *Open Journal of Statistics*, **5**, 311-333.
http://dx.doi.org/10.4236/ojs.2015.54033

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 $\boldsymbol{a}$ 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

$$\boldsymbol{S}_W \equiv \sum_{i=1}^{m} \alpha_i \boldsymbol{S}_i, \tag{1}$$

where $\boldsymbol{S}_i$ is the estimated sample covariance matrix for the ith class, and the sample between-class scatter matrix is

$$\boldsymbol{S}_B \equiv \sum_{i=1}^{m} \alpha_i \left(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}\right)\left(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}}\right)', \tag{2}$$

where $\bar{\boldsymbol{x}}_i$ is the sample mean vector for class Π_i, α_i is its *a priori* probability of class membership, $\bar{\boldsymbol{x}} = \sum_{i=1}^{m} \alpha_i \bar{\boldsymbol{x}}_i$ is the estimated overall mean, and $i = 1, \cdots, m$. For $m = 2$ with $\alpha_1 = \alpha_2$, [4] determined the vector $\boldsymbol{a}$ that maximizes the criterion function

$$\boldsymbol{J}_F\left(\boldsymbol{a}\right) = \frac{\boldsymbol{a}'\boldsymbol{S}_B\boldsymbol{a}}{\left(\boldsymbol{a}'\boldsymbol{S}_W\boldsymbol{a}\right)},$$

which is achieved by an eigenvalue decomposition of $\boldsymbol{S}_W^{-1}\boldsymbol{S}_B$. 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 $\boldsymbol{B} \in \mathbb{R}_{q\times p}$ such that

$$\boldsymbol{x} \to \boldsymbol{y} = \boldsymbol{B}\boldsymbol{x},$$

where $\boldsymbol{x} \in \mathbb{R}_{p\times 1}$, $\boldsymbol{y} \in \mathbb{R}_{q\times 1}$, and $\mathbb{R}_{m\times n}$ denotes the space of all $m \times n$ 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 $\boldsymbol{B} \in \mathbb{R}_{q\times p}$ that preserves all of the classification information originally contained in the p-dimensional feature space. We then derive a linear transformation that assigns $\boldsymbol{x}$ to Π_k if and only if the corres-

ponding Bayes classification rule assigns $\boldsymbol{Bx}$ to Π_k, where $k \in \{1,\cdots,m\}$. 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 $p\left(\Pi_i \middle| \boldsymbol{x}\right)$, $i=1,\cdots,m$, 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, $\Pi_1,\cdots,\Pi_m$, with assumed known *a priori* probabilities $\alpha_1,\cdots,\alpha_m$, respectively. Also, let $p\left(\cdot\middle|\Pi_i\right)$ denote the p-dimensional multivariate normal density corresponding to population Π_i, $i=1,\cdots,m$. The goal of statistical decision theory is to obtain a decision rule that assigns an unlabeled observation $\boldsymbol{x}$ to Π_k if $p\left(\Pi_k \middle| \boldsymbol{x}\right)$ is the maximum overall *a posteriori* $p\left(\Pi_i \middle| \boldsymbol{x}\right)$, $i=1,\cdots,m$. Then,

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

the Bayes classifier assigns $\boldsymbol{x}$ to class Π_k if

$$p\left(\Pi_k \middle| \boldsymbol{x}\right) > p\left(\Pi_j \middle| \boldsymbol{x}\right), \quad j=1,\cdots,m; j \neq k.$$

This decision rule partitions the measurement or feature space into m disjoint regions R_{Π_i}, where $i=1,\cdots,m$, such that $\boldsymbol{x}$ is assigned to class Π_k if $\boldsymbol{x} \in R_{\Pi_k}$. Using Bayes' rule, the *a posteriori* probabilities of class membership $p\left(\Pi_k \middle| \boldsymbol{x}\right)$ can be defined as

$$p\left(\Pi_k \middle| \boldsymbol{x}\right) = \frac{\alpha_k p\left(\boldsymbol{x} \middle| \Pi_k\right)}{p\left(\boldsymbol{x}\right)}.$$

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

Assign $\boldsymbol{x}$ to Π_k if

$$\alpha_k p\left(\boldsymbol{x} \middle| \Pi_k\right) > \alpha_j p\left(\boldsymbol{x} \middle| \Pi_j\right), \quad j=1,\cdots,m; j \neq k. \tag{3}$$

This decision rule is known as the *Bayes' classification rule*. Let Π_i be modeled as a p-dimensional multivariate normal distribution, and let

$$d_i\left(\boldsymbol{x}\right) \equiv \ln\left|\boldsymbol{\Sigma}_i\right| - 2\left[\ln\left(\alpha_i\right)\right] + \left(\boldsymbol{x}-\boldsymbol{\mu}_i\right)' \boldsymbol{\Sigma}_i^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_i\right), \quad i=1,\cdots,m. \tag{4}$$

The Bayes decision rule (4) is to classify the unlabeled observation $\boldsymbol{x}$ into the class Π_k such that $d_k\left(\boldsymbol{x}\right) = \min\left\{d_i\left(\boldsymbol{x}\right); i=1,\cdots,m\right\}$. 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 $\mathbb{R}^{S}_{p\times p}$ denote the set of $p\times p$ symmetric matrices. Also, we let $\boldsymbol{A}^{+}\in\mathbb{R}_{m\times n}$ represent the Moore-Penrose pseudo-inverse of $\boldsymbol{A}\in\mathbb{R}_{n\times m}$.

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

$$\boldsymbol{T}\equiv\left[\boldsymbol{d}_2-\boldsymbol{d}_1\,|\cdots|\,\boldsymbol{d}_m-\boldsymbol{d}_1\,|\,\boldsymbol{E}_2-\boldsymbol{E}_1\,|\cdots|\,\boldsymbol{E}_m-\boldsymbol{E}_1\right], \tag{5}$$

where $\boldsymbol{d}_i\in\mathbb{R}_{p\times 1}$, $\boldsymbol{E}_i\in\mathbb{R}^{S}_{p\times p}$ such that $rank(\boldsymbol{E}_i)=p$, and $\boldsymbol{d}_1\neq\boldsymbol{d}_k$ and $\boldsymbol{E}_1\neq\boldsymbol{E}_k$ for at least one value of *k*, where $2\le k\le m$ and $i=1,\cdots,m$. Also, let $rank(\boldsymbol{T})=1\le q<p$, and let $\boldsymbol{F}\in\mathbb{R}_{p\times q}$ and $\boldsymbol{G}\in\mathbb{R}_{q\times(m-1)(p+1)}$ be matrix components of a full-rank decomposition of $\boldsymbol{T}$ so that $\boldsymbol{T}=\boldsymbol{FG}$ with $rank(\boldsymbol{F})=rank(\boldsymbol{G})=\boldsymbol{q}$. Then, the Moore-Penrose pseudoinverse of $\boldsymbol{T}$ is $\boldsymbol{T}^{+}=\boldsymbol{G}^{+}\boldsymbol{F}^{+}$, and, also, $\boldsymbol{TT}^{+}=\boldsymbol{FF}^{+}$ and $\boldsymbol{TT}^{+}\boldsymbol{T}=\boldsymbol{FF}^{+}\boldsymbol{T}=\boldsymbol{T}$. This result implies that for $i=1,\cdots,m$,

(i) $\left(\boldsymbol{I}-\boldsymbol{FF}^{+}\right)\left(\boldsymbol{d}_i-\boldsymbol{d}_1\right)=0$ and

(ii) $\left(\boldsymbol{I}-\boldsymbol{FF}^{+}\right)\left(\boldsymbol{E}_i-\boldsymbol{E}_1\right)=0$.

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

Lemma 1 *For* $\boldsymbol{T}=\boldsymbol{FG}$ *in* (5), *where* $\boldsymbol{E}_i\in\mathbb{R}^{S}_{p\times p}$, $\boldsymbol{d}_i\in\mathbb{R}_{p\times 1}$, $\boldsymbol{F}\in\mathbb{R}_{p\times q}$, *and* $\boldsymbol{G}\in\mathbb{R}_{q\times(m-1)(p+1)}$ *such that* $rank(\boldsymbol{F})=rank(\boldsymbol{G})=q$, *and* $i=1,\cdots,m$, *we have that*

(a) $\boldsymbol{FF}^{+}\left(\boldsymbol{E}_i-\boldsymbol{E}_1\right)=\left(\boldsymbol{E}_i-\boldsymbol{E}_1\right)\boldsymbol{FF}^{+}$,

(b) $\boldsymbol{FF}^{+}\boldsymbol{E}_i=\boldsymbol{E}_i\boldsymbol{FF}^{+}$, and

(c) $\left(\boldsymbol{I}-\boldsymbol{FF}^{+}\right)\boldsymbol{E}_i=\boldsymbol{E}_1\left(\boldsymbol{I}-\boldsymbol{FF}^{+}\right)$.

Proof. Part (a) follows from the fact that $\boldsymbol{E}_i=\boldsymbol{E}_i'$, $i=1,\cdots,m$, and from (ii) above. Parts (b) and (c) follow directly from (a).

Lemma 2 *For* $\boldsymbol{T}=\boldsymbol{FG}$ *in* (5), *where* $\boldsymbol{E}_i\in\mathbb{R}^{S}_{p\times p}$, $\boldsymbol{d}_i\in\mathbb{R}_{p\times 1}$, $\boldsymbol{F}\in\mathbb{R}_{p\times q}$, *and* $\boldsymbol{G}\in\mathbb{R}_{q\times(m-1)(p+1)}$ *such that* $rank(\boldsymbol{F})=rank(\boldsymbol{G})=q$, *and* $i=1,\cdots,m$, *we have that* $\left[\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right]^{-1}=\boldsymbol{F}'\boldsymbol{E}_i^{-1}\boldsymbol{F}$.

Proof. Because $\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\in\mathbb{R}_{q\times q}$ and $rank\left(\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right)=q$, the result follows because

$$\left(\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right)\left(\boldsymbol{F}'\boldsymbol{E}_i^{-1}\boldsymbol{F}\right)=\boldsymbol{F}^{+}\boldsymbol{E}_i\left[\boldsymbol{FF}^{+}\right]\boldsymbol{E}_i^{-1}\boldsymbol{F}=\boldsymbol{F}^{+}\boldsymbol{FF}^{+}\boldsymbol{E}_i\boldsymbol{E}_i^{-1}\boldsymbol{F}=\boldsymbol{I}_q.$$

Lemma 3 *Let* $\boldsymbol{T}=\boldsymbol{FG}$ *in* (5), *where* $\boldsymbol{E}_i\in\mathbb{R}^{S}_{p\times p}$, $\boldsymbol{d}_i\in\mathbb{R}_{p\times 1}$, $\boldsymbol{F}\in\mathbb{R}_{p\times q}$ *and* $\boldsymbol{G}\in\mathbb{R}_{q\times(m-1)(p+1)}$ *such that* $rank(\boldsymbol{F})=rank(\boldsymbol{G})=q$, *and* $i=1,\cdots,m$. *Also, let* $\boldsymbol{C}=\boldsymbol{R}\left[\boldsymbol{I}-\boldsymbol{FF}^{+}\right]\in\mathbb{R}_{(p-q)\times p}$, *where* $\boldsymbol{R}\in\mathbb{R}_{(p-q)\times p}$ *such that* $rank(\boldsymbol{C})=p-q$. *Then,*

(a) $\boldsymbol{Cd}_i=\boldsymbol{Cd}_1$,

(b) $\boldsymbol{Cd}_i+\boldsymbol{CE}_i\boldsymbol{F}^{+\prime}\left(\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right)^{-1}\left(\boldsymbol{y}-\boldsymbol{F}^{+}\boldsymbol{d}_i\right)=\boldsymbol{Cd}_1$, where $\boldsymbol{y}\in\mathbb{R}_{p\times 1}$, and

(c) $\boldsymbol{CE}_i\boldsymbol{C}'-\boldsymbol{CE}_i\boldsymbol{F}^{+\prime}\left(\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right)^{-1}\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{C}'=\boldsymbol{CE}_1\boldsymbol{C}'$.

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

$$\boldsymbol{CE}_i\boldsymbol{F}^{+\prime}\left(\boldsymbol{F}^{+}\boldsymbol{E}_i\boldsymbol{F}^{+\prime}\right)^{-1}=\boldsymbol{CE}_i\boldsymbol{F}^{+\prime}\boldsymbol{F}'\boldsymbol{E}_i^{-1}\boldsymbol{F}=\boldsymbol{CF}=0, \tag{6}$$

for each $i\in\{1,\cdots,m\}$. 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* Π_i *be a p-dimensional multivariate normal population with a priori probability* $\alpha_i>0$, *mean* $\boldsymbol{\mu}_i$, *and covariance matrix* $\boldsymbol{\Sigma}_i$, $i=1,\cdots,m$, *such that* $rank(\boldsymbol{\Sigma}_i)=p$ *and* $\boldsymbol{\Sigma}_1\neq\boldsymbol{\Sigma}_j$ *for some* $j\in\{2,\cdots,m\}$.

Next, let

$$\boldsymbol{M} \equiv \left[\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \,|\cdots|\, \boldsymbol{\Sigma}_m^{-1}\boldsymbol{\mu}_m - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 \,|\, \boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1 \,|\cdots|\, \boldsymbol{\Sigma}_m - \boldsymbol{\Sigma}_1 \right]. \tag{7}$$

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

Proof. Let

$$\boldsymbol{w}_i = \begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{u}_i \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}^+_{q\times p} \\ \boldsymbol{C}_{(p-q)\times p} \end{bmatrix} \boldsymbol{x}_i,$$

where $\boldsymbol{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1,\cdots,m$. Let $\boldsymbol{H} \equiv \begin{bmatrix} \boldsymbol{F}^{+\prime} & \boldsymbol{C}' \end{bmatrix}' \in \mathbb{R}_{p\times p}$ be full-rank with $\boldsymbol{C} \equiv \boldsymbol{R}\left(\boldsymbol{I} - \boldsymbol{FF}^+\right)$, where $\boldsymbol{R} \in \mathbb{R}_{(p-q)\times p}$ such that $rank(\boldsymbol{R}) = p - q$. Then $\boldsymbol{w}_i \sim N\left(\boldsymbol{H}\boldsymbol{\mu}_i, \boldsymbol{H}\boldsymbol{\Sigma}_i\boldsymbol{H}'\right)$, where

$$\boldsymbol{H}\boldsymbol{\mu}_i = \begin{bmatrix} \boldsymbol{F}^+\boldsymbol{\mu}_i \\ \boldsymbol{C}\boldsymbol{\mu}_i \end{bmatrix} \text{ and } \boldsymbol{H}\boldsymbol{\Sigma}_i\boldsymbol{H}' = \begin{bmatrix} \boldsymbol{F}^+\boldsymbol{\Sigma}_i\boldsymbol{F}^{+\prime} & \boldsymbol{F}^+\boldsymbol{\Sigma}_i\boldsymbol{C}' \\ \boldsymbol{C}\boldsymbol{\Sigma}_i\boldsymbol{F}^{+\prime} & \boldsymbol{C}\boldsymbol{\Sigma}_i\boldsymbol{C}' \end{bmatrix},$$

$i = 1,\cdots,m$. By Lemma 3, we conclude that $E(\boldsymbol{u}_i|\boldsymbol{y}_i) = \boldsymbol{C}\boldsymbol{\mu}_1$ and $Var(\boldsymbol{u}_i|\boldsymbol{y}_i) = \boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}'$ for $i = 1,\cdots,m$. Thus, $\boldsymbol{w}_i \sim N\left(\boldsymbol{F}^+\boldsymbol{\mu}_i, \boldsymbol{F}^+\boldsymbol{\Sigma}_i\boldsymbol{F}^{+\prime}\right) \cdot N\left(\boldsymbol{C}\boldsymbol{\mu}_1, \boldsymbol{C}\boldsymbol{\Sigma}_1\boldsymbol{C}'\right)$. That is, $p(\boldsymbol{u}|\boldsymbol{y}, \Pi_k)$ does not depend on class $k \in \{2,\cdots,m\}$.

Recall that for $i, j = 1,\cdots,m$, $i \neq j$, the p-variate Bayes procedure defined by (3) assigns $\boldsymbol{x}$ to Π_j if and only if

$$\begin{aligned} \alpha_j p(\boldsymbol{x}|\Pi_j) > \alpha_i p(\boldsymbol{x}|\Pi_i) &\Leftrightarrow \alpha_j p(\boldsymbol{w}|\Pi_j) > \alpha_i p(\boldsymbol{w}|\Pi_i) \\ &\Leftrightarrow \alpha_j p(\boldsymbol{u}|\boldsymbol{y}, \Pi_j)\, p(\boldsymbol{y}|\Pi_j) > \alpha_i p(\boldsymbol{u}|\boldsymbol{y}, \Pi_i)\, p(\boldsymbol{y}|\Pi_i) \\ &\Leftrightarrow \alpha_j p(\boldsymbol{y}|\Pi_j) > \alpha_i p(\boldsymbol{y}|\Pi_i). \end{aligned}$$

Therefore, the original p-variate Bayes classification assignment is preserved by the linear transformation $\boldsymbol{y} = \boldsymbol{F}^+\boldsymbol{x}$.

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 $\boldsymbol{F}^+ \in \mathbb{R}_{q\times p}$ exists and that $BPMC_p = BPMC_q$, where $rank(\boldsymbol{M}) = q < p$.

With the following corollary, we demonstrate that for two multivariate normal populations such that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ and $\boldsymbol{\mu}_2 \neq \boldsymbol{\mu}_1$, our *LDR* matrix derived in Theorem 1 reduces to the *LDF* of [4].

Corollary 1 *Assuming we have two multivariate normal populations* $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ *and* $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, *the proposed LDR matrix in Theorem* 1 *reduces to* $\boldsymbol{M} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$, *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 $rank(\boldsymbol{M}) = p$, 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 $1 \le r < q < p$. 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 $1 \le r < q < p$. 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, $1 \le r \le q < p$, is the *SVD* approximation to $\boldsymbol{M}$ in (7). We use the following theorem from [17] to determine such an r-dimensional *LDR* matrix.

Theorem 2 *Let* $\boldsymbol{C}_{s,t}^{(p)}$ *denote the class of all* $s \times t$ *real matrices of rank p, and let* $\boldsymbol{C}_{s,t}^{(r)}$ *denote the class of all* $s \times t$ *real matrices of rank r, where* $1 \le r < q < p$. *If* $\boldsymbol{A}_p \in \boldsymbol{C}_{s,t}^{(p)}$ *and* $\boldsymbol{A}_r \in \boldsymbol{C}_{s,t}^{(r)}$, *given by* $\boldsymbol{A}_r = \boldsymbol{U}\boldsymbol{D}_r\boldsymbol{V}'$, *then*

$$\left\|\boldsymbol{A}_p - \boldsymbol{A}_r\right\| < \left\|\boldsymbol{A}_p - \boldsymbol{X}\right\| \quad \text{for all } \boldsymbol{X} \in \boldsymbol{C}_{s,t}^{(r)},$$

where $\boldsymbol{A}_p = \boldsymbol{U}\boldsymbol{D}_p\boldsymbol{V}'$, $\boldsymbol{D}_p = \operatorname{diag}\left(d_1, \cdots, d_p\right)$, $\boldsymbol{D}_r = \operatorname{diag}\left(d_1, \cdots, d_r, 0_{r+1}, \cdots, 0_p\right)$, and $\left\|\boldsymbol{A}_p\right\|$ is the usual Euclidean or Frobenius norm of a matrix $\boldsymbol{A}_p$, given by

$$\left\|\boldsymbol{A}_p\right\| = \left(\sum_{i=1}^{s}\sum_{j=1}^{t}\left|a_{ij}\right|^2\right)^{1/2} = \left(\sum_{i=1}^{p} d_i^2\right)^{1/2}.$$

Furthermore, $\left\|\boldsymbol{A}_p - \boldsymbol{A}_r\right\| = \left(\sum_{i=r+1}^{p} d_i^2\right)^{1/2}$.

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 $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{D}_p\boldsymbol{V}$ be the *SVD* of the matrix $\boldsymbol{M}$, where $\boldsymbol{D}_p \equiv \operatorname{diag}\left(\lambda_1, \cdots, \lambda_p\right)$ for λ_i, $i = 1, \cdots, p$, which are the singular values of $\boldsymbol{M}$ with $\lambda_i \ge \lambda_j$ for $1 \le i < j \le p$, $\lambda_q > 0$, $\lambda_{q+1} \ge 0$ for $1 \le q \le p$. Let $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{D}_p$ and define $\boldsymbol{D}_r \equiv \operatorname{diag}\left(\lambda_1, \cdots, \lambda_r, 0_{r+1}, \cdots, 0_p\right)$ with $\lambda_i \ge \lambda_j$ for $1 \le i < j \le q$. From Theorem 2, we have that $\boldsymbol{M}_r = \boldsymbol{U}\boldsymbol{D}_r\boldsymbol{V}'$ is a rank-*r* approximation of $\boldsymbol{M}$, and, therefore, a rank-*r* approximation to $\boldsymbol{F}$ is $\boldsymbol{F}_r = \boldsymbol{U}\boldsymbol{D}_r$. Thus, $\boldsymbol{F}_r'$ is an $r \times p$ *LDR* matrix that yields an *r*-dimensional representation of the original *p*-dimensional class models. One can also use $\boldsymbol{F}_r^+$ 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 $N\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right)$, $N\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\right)$, and $N\left(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3\right)$, where

$$\boldsymbol{\mu}_1 = [0,0,0,0,0,0],$$

$$\boldsymbol{\mu}_2 = [1,1,1,1,1,1],$$

$$\boldsymbol{\mu}_3 = [2,2,2,2,2,2],$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 2 \end{bmatrix}, \text{ and } \boldsymbol{\Sigma}_3 = \begin{bmatrix} 2 & 1 & 0 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 & 1 \\ 0 & 0 & 7 & 0 & 0 & 0 \\ 1 & 1 & 0 & 2 & 1 & 1 \\ 1 & 1 & 0 & 1 & 2 & 1 \\ 1 & 1 & 0 & 1 & 1 & 2 \end{bmatrix}.$$

We have $rank\left(\boldsymbol{M}\right) = 2$, 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 $Col\left(\boldsymbol{F}\right)$, where

$$\boldsymbol{F} = \begin{bmatrix} 0.3800 & 0.3800 & 0.5272 & 0.3800 & 0.3800 & 0.3800 \\ 0.2358 & 0.2358 & -0.8497 & 0.2358 & 0.2358 & 0.2358 \end{bmatrix}'. \tag{8}$$

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 $\boldsymbol{M}$ given in (7). A one-dimensional representation space is column one of the matrix $\boldsymbol{F}$ 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.

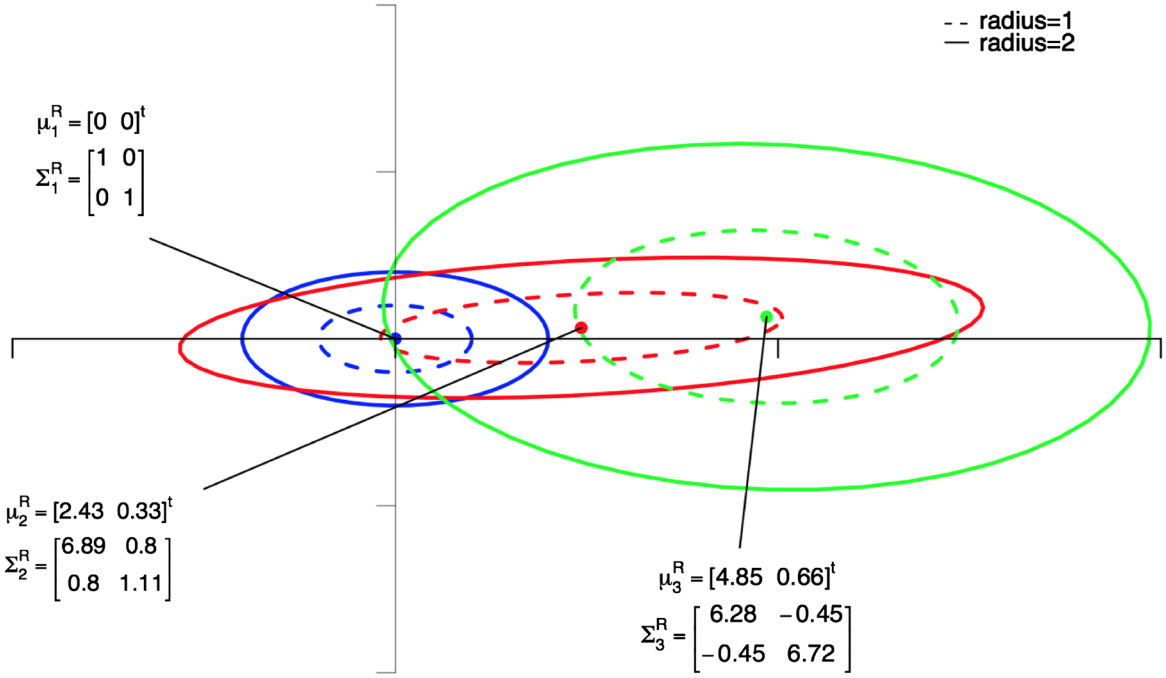

Figure 1. The optimal two-dimensional representation for normal densities from the example in Section 5.2.

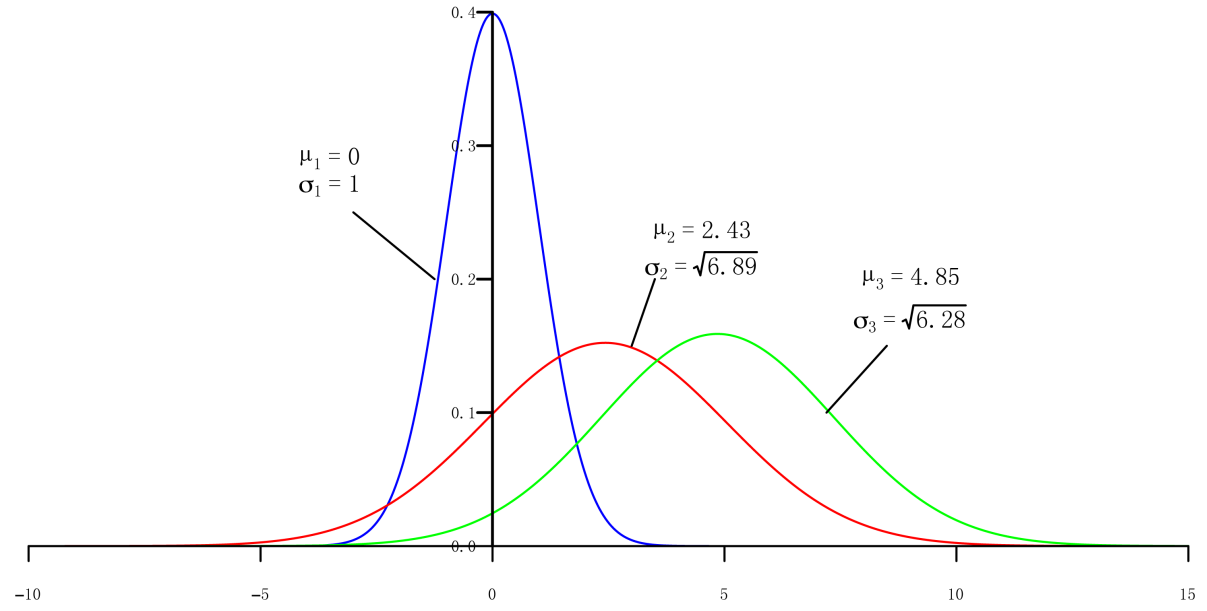

Figure 2. The optimal one-dimensional representation for normal densities from the example in Section 5.2.

6.1. The *SY* Method

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

$$\hat{\boldsymbol{M}} \equiv \left[\boldsymbol{S}_2^{-1}\overline{\boldsymbol{x}}_2 - \boldsymbol{S}_1^{-1}\overline{\boldsymbol{x}}_1 \,|\, \boldsymbol{S}_3^{-1}\overline{\boldsymbol{x}}_3 - \boldsymbol{S}_1^{-1}\overline{\boldsymbol{x}}_1 \,|\cdots|\, \boldsymbol{S}_m^{-1}\overline{\boldsymbol{x}}_m - \boldsymbol{S}_1^{-1}\overline{\boldsymbol{x}}_1 \,|\, \boldsymbol{S}_2 - \boldsymbol{S}_1 \,|\cdots|\, \boldsymbol{S}_m - \boldsymbol{S}_1 \right],$$

provided $n_i > p$, $i=1,\cdots,m$. Our estimator $\hat{\boldsymbol{M}}$, 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 $\hat{\boldsymbol{M}}$.

Thus, using the *SVD*, we let $\hat{\boldsymbol{M}} = \boldsymbol{U}\boldsymbol{D}_p\boldsymbol{V}$, where $\boldsymbol{D}_p \equiv \operatorname{diag}\left(\lambda_1,\cdots,\lambda_p\right)$, $\lambda_j > 0$ are the singular values of $\hat{\boldsymbol{M}}$ for $j=1,\cdots,p$, and let $\hat{\boldsymbol{F}} = \boldsymbol{U}\boldsymbol{D}_p$. Also, let

$$\boldsymbol{D}_r \equiv \operatorname{diag}\left(\lambda_1,\cdots,\lambda_r,0_{r+1},\cdots,0_p\right) \tag{9}$$

with $\lambda_j \geq \lambda_l$ for $1 \leq j < l \leq p$ and $1 \leq r < p$. From Theorem 2, we have that $\tilde{\boldsymbol{M}}_r = \boldsymbol{U}\boldsymbol{D}_r\boldsymbol{V}'$ is a rank-*r* approximation of $\hat{\boldsymbol{M}}$, and a rank-*r* approximation to $\boldsymbol{F}$ is $\tilde{\boldsymbol{F}}_r = \boldsymbol{U}\boldsymbol{D}_r$. Thus, $\tilde{\boldsymbol{F}}_r'$ is our new $r \times p$ *LDR* matrix. Provided $\left[\left(\sum_{j=r+1}^{p}\lambda_j\right)\Big/\left(\sum_{i=1}^{p}\lambda_j\right)\right]$ is relatively small, $\tilde{\boldsymbol{F}}_r'$ 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

$$\boldsymbol{U} \equiv \left[\left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right)^{-1}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) \,|\cdots|\, \left(\boldsymbol{\Sigma}_i + \boldsymbol{\Sigma}_j\right)^{-1}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\right) \,|\cdots|\, \left(\boldsymbol{\Sigma}_{m-1} + \boldsymbol{\Sigma}_m\right)^{-1}\left(\boldsymbol{\mu}_{m-1} - \boldsymbol{\mu}_m\right) \right]$$

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

$$\hat{\boldsymbol{U}} \equiv \left[\left(\boldsymbol{S}_1 + \boldsymbol{S}_2\right)^{-1}\left(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2\right) \,|\cdots|\, \left(\boldsymbol{S}_i + \boldsymbol{S}_j\right)^{-1}\left(\overline{\boldsymbol{x}}_i - \overline{\boldsymbol{x}}_j\right) \,|\cdots|\, \left(\boldsymbol{S}_{m-1} + \boldsymbol{S}_m\right)^{-1}\left(\overline{\boldsymbol{x}}_{m-1} - \overline{\boldsymbol{x}}_m\right) \right]. \tag{10}$$

The *LDR* matrix derived in [14], which we refer to as the *BE* matrix, was also obtained when we determined low-rank approximation to $\hat{\boldsymbol{U}}$ by using the *SVD*. That is, let $\hat{\boldsymbol{U}} = \boldsymbol{R}\boldsymbol{D}_p\boldsymbol{S}'$ be the *SVD* of $\hat{\boldsymbol{U}}$, where

$$\boldsymbol{D}_p \equiv \operatorname{diag}\left(\lambda_1,\cdots,\lambda_p\right) \tag{11}$$

with $\lambda_j \geq \lambda_l$ for $1 \leq j < l \leq p$, and let $\hat{\boldsymbol{H}} = \boldsymbol{R}\boldsymbol{D}_p$. Define $\boldsymbol{D}_r$ as in (9) with $\lambda_j \geq \lambda_l$ for $1 \leq j \leq r$. Then, $\tilde{\boldsymbol{U}} = \boldsymbol{R}\boldsymbol{D}_r\boldsymbol{S}'$ is a rank-*r* approximation of $\hat{\boldsymbol{U}}$, and $\tilde{\boldsymbol{H}} = \boldsymbol{R}\boldsymbol{D}_r = \boldsymbol{R}_r \in \mathbb{R}_{p\times r}$ is a rank-*r* approximation of $\hat{\boldsymbol{H}}$. Thus, the *BE LDR* matrix is $\tilde{\boldsymbol{H}}_r'$, or, equivalently, $\boldsymbol{R}'_r$, 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 $\boldsymbol{\Sigma}_W \equiv \sum_{i=1}^{m}\boldsymbol{\Sigma}_i \big/ m$ as the within-group covariance matrix and $\boldsymbol{\Sigma}_B \equiv \sum_{i=1}^{m}\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_i - \boldsymbol{\mu}\right)' \big/ m$ as the between-group covariance matrix, where $\boldsymbol{\mu} \equiv \sum_{i=1}^{m}\boldsymbol{\mu}_i \big/ m$ is the overall

population mean. As its criterion matrix, the *SIR LDR* method uses

$$\boldsymbol{M}_{SIR} \equiv \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Sigma}_B\boldsymbol{\Gamma}^{-1/2}, \tag{12}$$

where $\boldsymbol{\Gamma} \equiv \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W$ is the marginal covariance matrix of $\boldsymbol{x}$. When population parameters must be estimated, an estimator of (12) is

$$\hat{\boldsymbol{M}}_{SIR} \equiv \hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{S}_B\hat{\boldsymbol{\Gamma}}^{-1/2}, \tag{13}$$

where $\hat{\boldsymbol{\Gamma}} \equiv \boldsymbol{S}_B + \boldsymbol{S}_W$ with $\boldsymbol{S}_W$ and $\boldsymbol{S}_B$ given in (1) and (2), respectively. Let $\hat{\boldsymbol{M}}_{SIR} = \boldsymbol{G}\boldsymbol{D}_p\boldsymbol{P}'$ be the *SVD* of (13), and let $\hat{\boldsymbol{K}} = \hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{G}_p$, where $\boldsymbol{G}_p$ 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, $\hat{\boldsymbol{K}}_r = \hat{\boldsymbol{\Gamma}}^{-1/2}G_r$ is a rank-*r* approximation of $\hat{\boldsymbol{K}}$, and the $r \times p$ *SIR LDR* matrix is $\hat{\boldsymbol{K}}_r'$, which is composed of the eigenvectors corresponding to the *r* largest singular values of $\boldsymbol{K}$.

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 $\boldsymbol{M}^*_{SAVE} \equiv \sum_{i=1}^{m}\left(\boldsymbol{I}_p - \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Sigma}_i\boldsymbol{\Gamma}^{-1/2}\right)^2 \big/ m$, where $\boldsymbol{\Gamma} \equiv \boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_W$. We use a form of *SAVE* given in [21], which is

$$\boldsymbol{M}_{SAVE} \equiv \left(\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Sigma}_B\boldsymbol{\Gamma}^{-1/2}\right)^2 + \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{-1/2}, \tag{14}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}} \equiv \sum_{i=1}^{m}\left(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_W\right)\boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_W\right)\big/ m$. An estimator of M_{SAVE} is

$$\hat{\boldsymbol{M}}_{SAVE} \equiv \left(\hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{S}_B\hat{\boldsymbol{\Gamma}}^{-1/2}\right)^2 + \hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{S}_{\hat{\boldsymbol{\Gamma}}}\hat{\boldsymbol{\Gamma}}^{-1/2}, \tag{15}$$

where, $\boldsymbol{S}_{\hat{\boldsymbol{\Gamma}}} \equiv \sum_{i=1}^{m}\left(\boldsymbol{S}_i - \boldsymbol{S}_W\right)\hat{\boldsymbol{\Gamma}}^{-1}\left(\boldsymbol{S}_i - \boldsymbol{S}_W\right)\big/ m$ with $\boldsymbol{S}_W$ and $\boldsymbol{S}_B$ given in (1) and (2), respectively. Next, let $\hat{\boldsymbol{M}}_{SAVE} = \boldsymbol{B}\boldsymbol{D}_p\boldsymbol{Q}'$ be the *SVD* of (15), and let $\hat{\boldsymbol{L}} = \hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{B}_p$, where $\boldsymbol{B}_p$ is composed of the eigenvectors of (15) such that the j^{th} column of $\boldsymbol{B}_p$ corresponds to the j^{th} of (15) with the j^{th} largest singular value, $j = 1, \cdots, p$. Then, $\hat{\boldsymbol{L}}_r = \hat{\boldsymbol{\Gamma}}^{-1/2}\boldsymbol{B}_r$ is a rank-r approximation of $\hat{\boldsymbol{L}}$, and, thus, the $r \times p$ *SAVE LDR* matrix is $\hat{\boldsymbol{L}}_r'$. An alternative representation of the *r*-dimensional *SAVE LDR* matrix is $\boldsymbol{B}_r$, which is composed of the eigenvectors corresponding to the *r* largest singular values of $\boldsymbol{B}$.

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 Π_1 or Π_2 when $m = 2$ or Π_1, Π_2, or Π_3 when $m = 3$ using the sample *QDF* corresponding to (4). We have applied the four competing *LDR* matrices $\tilde{\boldsymbol{F}}_r'$, $\tilde{\boldsymbol{H}}_r'$, $\tilde{\boldsymbol{K}}_r'$, and $\tilde{\boldsymbol{L}}_r'$ and calculated the *EPMC*s for the full-dimensional *QDF* and for the four reduced-dimensional *QDF*s 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 $n_i = (2.5)p$ and $n_i = 5p$, $i = 1, 2$ 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 $\hat{\boldsymbol{U}} \in \mathbb{R}_{p \times r}$. 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 $1 \le r < p$, 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 *EPMC*s of the four competing *LDR* methods for the various population configurations and values of n_i and *r*, where $i = 1, \cdots, m$. The estimated standard error of all *EPMC*s in the following tables was less than 0.001. For each configuration, we also calculated $\hat{\boldsymbol{M}}$, $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{K}}$, and $\hat{\boldsymbol{L}}$ 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: $EPMC_r(SY)$, $EPMC_r(BE)$, $EPMC_r(SIR)$, and $EPMC_r(SAVE)$ denote the estimated *EPMC*s 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 $N_p(\mathbf{0}, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where $p = 10$,

$$\boldsymbol{\mu}_1 = [0,0,0,0,0,0,0,0,0,0]',$$

$$\boldsymbol{\mu}_2 = [5,5,5,5,5,5,5,5,5,5]',$$

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

			Rank			
Configuration	Means	*m*	***M***	***U***	***K***	***L***
1	Moderately separated unequal means	2	3	1	1	2
2	Relatively close unequal means	3	2	1	1	2
3	Relatively close unequal means	2	2	1	1	2
4	Relatively close unequal means	3	2	1	1	2
5	Close unequal means	2	4	3	2	4

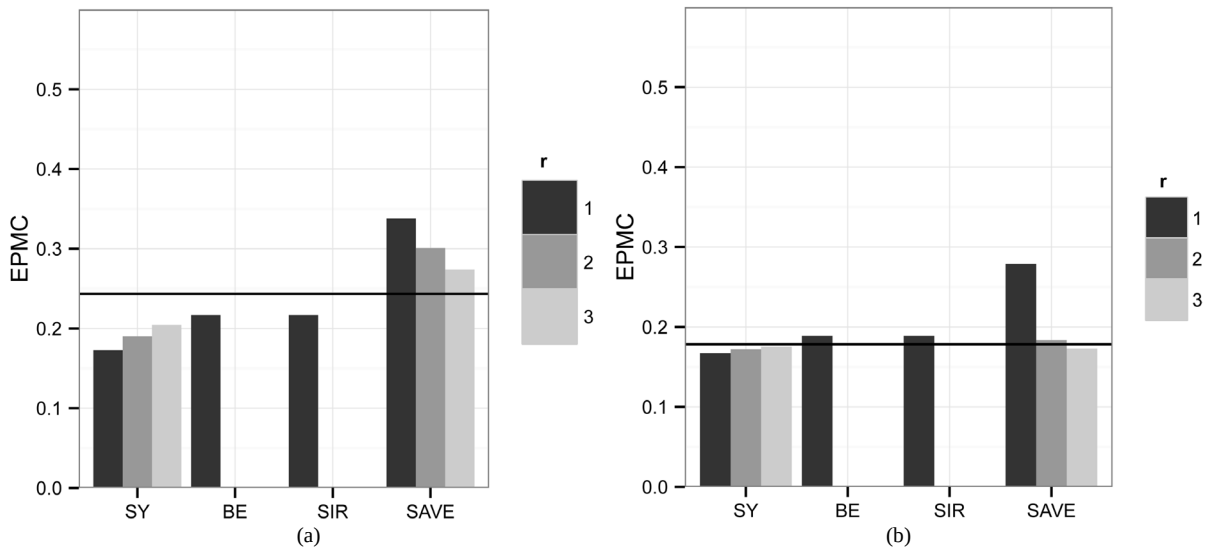

Figure 3. Simulation results from Section 7.1. The horizontal bar in each graph represents the $EPMC_{10}$. (a) $n_i = 25$; (b) $n_i = 50$.

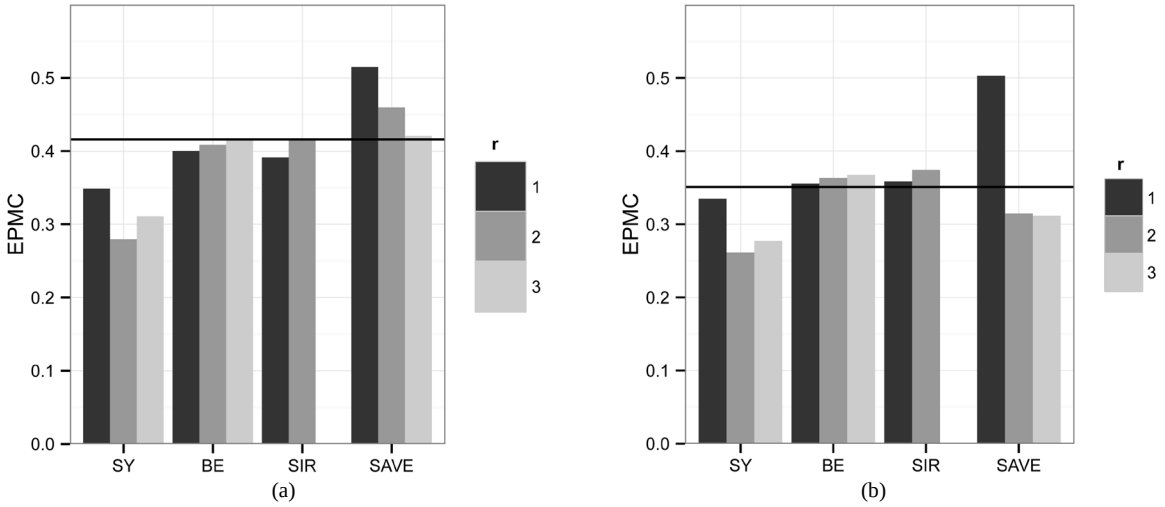

Figure 4. Simulation results from Configuration 2. The horizontal bar in each graph represents the $EPMC_{10}$ with no dimension reduction. (a) $n_i = 25$; (b) $n_i = 50$.

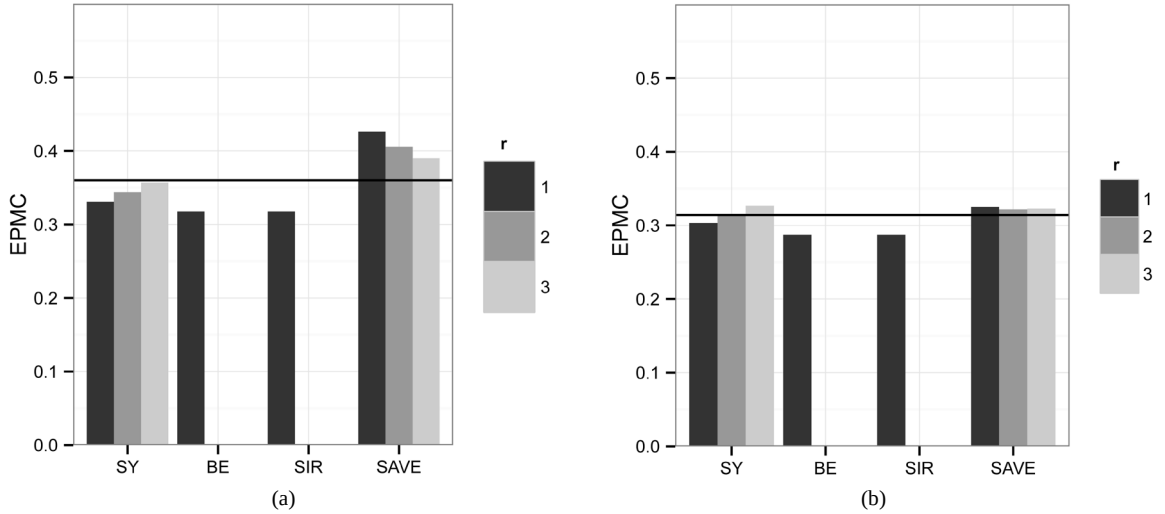

Figure 5. Simulation results from Section 7.3. The horizontal bar in each graph represents the $EPMC_{10}$. (a) $n_i = 25$; (b) $n_i = 50$.

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 10.00 & 4.00 & 5.00 & 4.00 & 3.00 & 4.00 & 4.00 & 5.00 & 4.00 & 3.00 \\ 4.00 & 10.00 & 5.00 & 2.00 & 4.00 & 3.00 & 3.00 & 5.00 & 4.00 & 3.00 \\ 5.00 & 5.00 & 10.00 & 5.00 & 5.00 & 3.00 & 4.00 & 4.00 & 4.00 & 4.00 \\ 4.00 & 2.00 & 5.00 & 10.00 & 3.00 & 4.00 & 2.00 & 3.00 & 4.00 & 3.00 \\ 3.00 & 4.00 & 5.00 & 3.00 & 10.00 & 3.00 & 4.00 & 5.00 & 3.00 & 3.00 \\ 4.00 & 3.00 & 3.00 & 4.00 & 3.00 & 12.00 & 3.00 & 4.00 & 4.00 & 4.00 \\ 4.00 & 3.00 & 4.00 & 2.00 & 4.00 & 3.00 & 14.00 & 2.00 & 2.00 & 2.00 \\ 5.00 & 5.00 & 4.00 & 3.00 & 5.00 & 4.00 & 2.00 & 12.00 & -0.50 & -0.50 \\ 4.00 & 4.00 & 4.00 & 4.00 & 3.00 & 4.00 & 2.00 & -0.50 & 14.00 & -1.00 \\ 3.00 & 3.00 & 4.00 & 3.00 & 3.00 & 4.00 & 2.00 & -0.50 & -1.00 & 11.00 \end{bmatrix},$$

and

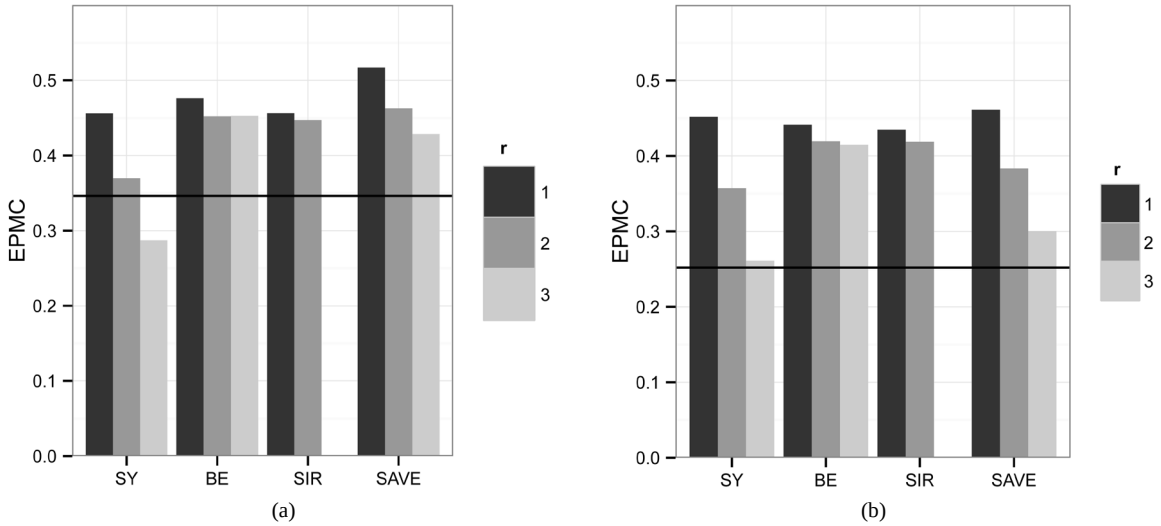

Figure 6. Simulation results from Configuration 4. The horizontal bar in each graph represents the $EPMC_{10}$ with no dimension reduction. (a) $n_i = 25$; (b) $n_i = 50$.

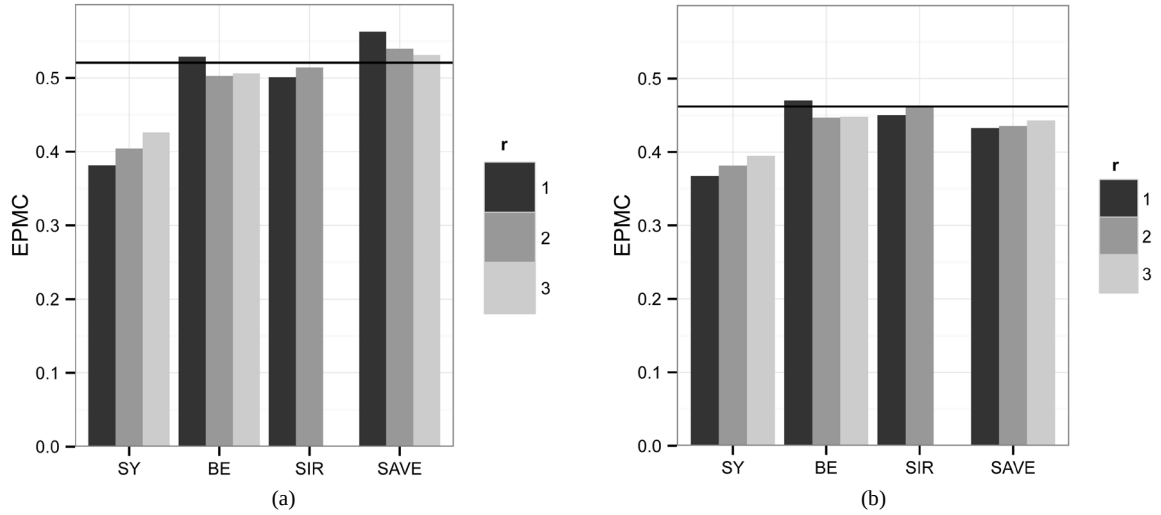

Figure 7. Simulation results from Section 7.5. The horizontal bar in each graph represents the $EPMC_{10}$. (a) $n_i = 25$; (b) $n_i = 50$.

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 16.00 & 10.00 & 11.00 & 10.00 & 9.00 & 10.00 & 10.00 & 11.00 & 10.00 & 9.00 \\ 10.00 & 16.00 & 11.00 & 8.00 & 10.00 & 9.00 & 9.00 & 11.00 & 10.00 & 9.00 \\ 11.00 & 11.00 & 16.00 & 11.00 & 11.00 & 9.00 & 10.00 & 10.00 & 10.00 & 10.00 \\ 10.00 & 8.00 & 11.00 & 16.00 & 9.00 & 10.00 & 8.00 & 9.00 & 10.00 & 9.00 \\ 9.00 & 10.00 & 11.00 & 9.00 & 16.00 & 9.00 & 10.00 & 11.00 & 9.00 & 9.00 \\ 10.00 & 9.00 & 9.00 & 10.00 & 9.00 & 18.00 & 9.00 & 10.00 & 10.00 & 10.00 \\ 10.00 & 9.00 & 10.00 & 8.00 & 10.00 & 9.00 & 20.00 & 8.00 & 8.00 & 8.00 \\ 11.00 & 11.00 & 10.00 & 9.00 & 11.00 & 10.00 & 8.00 & 22.00 & 9.50 & 9.50 \\ 10.00 & 10.00 & 10.00 & 10.00 & 9.00 & 10.00 & 8.00 & 9.50 & 24.00 & 9.00 \\ 9.00 & 9.00 & 10.00 & 9.00 & 9.00 & 10.00 & 8.00 & 9.50 & 9.00 & 21.00 \end{bmatrix}.$$

Here, $rank(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1) = 1$, which implies $rank(\boldsymbol{M}) = 2$ because $(\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1) \notin span(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)$. The singular values of $\hat{\boldsymbol{M}}$ 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 $r \ll 10$, 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 $\left[EPMC_{10} - EPMC_1(SY)\right] \approx 0.06$ and for $n_i = 50$, $\left[EPMC_{10} - EPMC_1(SY)\right] \approx 0.01$, 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 $EPMC_r(\cdot)$ approached $EPMC_p(\cdot)$ 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 $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, and $N_p(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$ where $p = 10$,

$$\boldsymbol{\mu}_1 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]',$$

$$\boldsymbol{\mu}_2 = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]',$$

$$\boldsymbol{\mu}_3 = [2, 2, 2, 2, 2, 2, 2, 2, 2, 2]',$$

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

Singular value	$n_i = 25$				$n_i = 50$			
	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$
e_1	72.16	0.38	0.57	0.74	67.98	0.19	0.52	0.66
e_2	16.27			0.54	11.67			0.36
e_3	12.21			0.43	8.93			0.25
e_4	9.31			0.34	6.88			0.18
e_5	7.22			0.25	5.34			0.13
e_6	5.46			0.18	4.06			0.09
e_7	3.99			0.11	2.98			0.05
e_8	2.73			0.07	2.05			0.03
e_9	1.67			0.03	1.23			0.01
e_{10}	0.75			0.01	0.53			<0.01

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 2.00 & 1.00 & 0.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 2.00 & 0.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 0.00 & 0.00 & 10.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & 1.00 & 0.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 \\ 1.00 & 1.00 & 0.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 \end{bmatrix}.$$

In this configuration, the population means are unequal but relatively close. Moreover, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$ are unequal but notably more similar to one another than to $\boldsymbol{\Sigma}_1$. However, the variance of the third feature in Π_3 is significantly greater than the population variances of the third feature for either Π_1 or Π_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 *SY 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* methords 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.

Singular value	$n_i = 25$				$n_i = 50$			
	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$
e_1	15.18	1.31	0.55	1.25	14.27	1.15	0.51	1.18
e_2	8.92	0.90	0.11	0.61	8.89	0.47	0.06	0.43
e_3	4.07	0.43		0.49	2.09	0.29		0.28
e_4	2.51			0.40	1.50			0.21
e_5	1.84			0.32	1.21			0.16
e_6	1.49			0.25	1.01			0.13
e_7	1.22			0.19	0.84			0.10
e_8	1.00			0.14	0.70			0.07
e_9	0.79			0.10	0.56			0.05
e_{10}	0.58			0.06	0.42			0.03

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 *SY LDR* in particular.

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

In this configuration, we have $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ with $p = 10$,

$$\boldsymbol{\mu}_1 = [0,0,0,0,0,0,0,0,0,0]',$$

$$\boldsymbol{\mu}_2 = [2,2,2,2,2,2,2,2,2,2]',$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 10.00 & 4.00 & 5.00 & 4.00 & 3.00 & 4.00 & 4.00 & 5.00 & 4.00 & 3.00 \\ 4.00 & 10.00 & 5.00 & 2.00 & 4.00 & 3.00 & 3.00 & 5.00 & 4.00 & 3.00 \\ 5.00 & 5.00 & 10.00 & 5.00 & 5.00 & 3.00 & 4.00 & 4.00 & 4.00 & 4.00 \\ 4.00 & 2.00 & 5.00 & 10.00 & 3.00 & 4.00 & 2.00 & 3.00 & 4.00 & 3.00 \\ 3.00 & 4.00 & 5.00 & 3.00 & 12.00 & 3.00 & 4.00 & 5.00 & 3.00 & 3.00 \\ 4.00 & 3.00 & 3.00 & 4.00 & 3.00 & 9.00 & 3.00 & 4.00 & 4.00 & 4.00 \\ 4.00 & 3.00 & 4.00 & 2.00 & 4.00 & 3.00 & 14.00 & 2.00 & 2.00 & 2.00 \\ 5.00 & 5.00 & 4.00 & 3.00 & 5.00 & 4.00 & 2.00 & 12.00 & 1.00 & -0.50 \\ 4.00 & 4.00 & 4.00 & 4.00 & 3.00 & 4.00 & 2.00 & 1.00 & 14.00 & -1.00 \\ 3.00 & 3.00 & 4.00 & 3.00 & 3.00 & 4.00 & 2.00 & -0.50 & -1.00 & 11.00 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 8.00 & 2.00 & 3.00 & 4.00 & 1.00 & 2.00 & 2.00 & 3.00 & 2.00 & 1.00 \\ 2.00 & 8.00 & 3.00 & 2.00 & 2.00 & 1.00 & 1.00 & 3.00 & 2.00 & 1.00 \\ 3.00 & 3.00 & 8.00 & 5.00 & 3.00 & 1.00 & 2.00 & 2.00 & 2.00 & 2.00 \\ 4.00 & 2.00 & 5.00 & 10.00 & 3.00 & 4.00 & 2.00 & 3.00 & 4.00 & 3.00 \\ 1.00 & 2.00 & 3.00 & 3.00 & 10.00 & 1.00 & 2.00 & 3.00 & 1.00 & 1.00 \\ 2.00 & 1.00 & 1.00 & 4.00 & 1.00 & 7.00 & 1.00 & 2.00 & 2.00 & 2.00 \\ 2.00 & 1.00 & 2.00 & 2.00 & 2.00 & 1.00 & 12.00 & 0.00 & 0.00 & 0.00 \\ 3.00 & 3.00 & 2.00 & 3.00 & 3.00 & 2.00 & 0.00 & 10.00 & -1.00 & -2.50 \\ 2.00 & 2.00 & 2.00 & 4.00 & 1.00 & 2.00 & 0.00 & -1.00 & 12.00 & -3.00 \\ 1.00 & 1.00 & 2.00 & 3.00 & 1.00 & 2.00 & 0.00 & -2.50 & -3.00 & 9.00 \end{bmatrix}.$$

As in Configuration 1, we have that $rank\left(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1\right) = 1$, and, hence, $rank\left(\boldsymbol{M}\right) = 2$. 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 matrices, which resulted in less-variable covariance matrix estimators and, therefore, smaller values of $EPMC_1\left(BE\right)$ and $EPMC_1\left(SIR\right)$. From **Table 4**, we see that the first singular value for $\hat{\boldsymbol{M}}$ was considerably less predominant than the first singular value for $\hat{\boldsymbol{M}}$ 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 $\left[EPMC_{10} - EPMC_1\left(SIR\right)\right] \approx \left[EPMC_{10} - EPMC_1\left(BE\right)\right] \approx 0.04$ for $n_i = 25$ and $\left[EPMC_{10} - EPMC1\left(SIR\right)\right] \approx \left[EPMC_{10} - EPMC_1\left(BE\right)\right] \approx 0.06$ 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 $\left[EPMC_{10} - EPMC_1\left(BE\right)\right]$ and $\left[EPMC_{10} - EPMC_1\left(SIR\right)\right]$ 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, $EPMC_1\left(SY\right)$ was not much greater than either $EPMC_1\left(BE\right)$ or $EPMC_1\left(SIR\right)$. 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: $N_p\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right)$, $N_p\left(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\right)$, and $N_p\left(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3\right)$,

Table 4. Summary of singular values for the four competing *LDR* methods for Configuration 3.

	$n_i = 25$				$n_i = 50$			
Singular value	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$
e_1	27.91	0.45	0.42	0.61	22.98	0.370	0.35	0.44
e_2	14.84			0.49	10.48			0.28
e_3	10.73			0.39	7.69			0.21
e_4	8.12			0.30	5.86			0.15
e_5	6.18			0.23	4.50			0.11
e_6	4.65			0.16	3.39			0.08
e_7	3.41			0.10	2.50			0.50
e_8	2.41			0.06	1.78			0.03
e_9	1.55			0.03	1.16			0.01
e_{10}	0.76			0.01	0.60			<0.01

where $p = 10$,

$$\boldsymbol{\mu}_1 = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]',$$

$$\boldsymbol{\mu}_2 = [1, 0, 1, 0, 1, 0, 1, 0, 1, 0]',$$

$$\boldsymbol{\mu}_3 = [-1, -1, -1, -1, -1, -1, -1, -1, -1, -1]',$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 2.00 & 0.80 & 1.00 & 0.80 & 0.60 & 0.80 & 0.80 & 1.00 & 0.80 & 0.60 \\ 0.80 & 2.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\ 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\ 0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 1.00 & 0.60 & 0.60 \\ 0.80 & 0.60 & 0.60 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\ 1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\ 0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\ 0.60 & 0.60 & 0.80 & 0.60 & 0.60 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 \end{bmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 20.00 & 0.80 & 1.00 & 0.80 & 0.60 & 0.80 & 0.80 & 1.00 & 0.80 & 0.60 \\ 0.80 & 40.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\ 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\ 0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 1.00 & 0.60 & 0.60 \\ 0.80 & 0.60 & 0.60 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\ 1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\ 0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\ 0.60 & 0.60 & 0.80 & 0.60 & 0.60 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 35.00 & 0.80 & 1.00 & 0.80 & 0.60 & 0.80 & 0.80 & 1.00 & 0.80 & 0.60 \\ 0.80 & 40.00 & 1.00 & 0.40 & 0.80 & 0.60 & 0.60 & 1.00 & 0.80 & 0.60 \\ 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.60 & 0.80 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.40 & 1.00 & 2.00 & 0.60 & 0.80 & 0.40 & 0.60 & 0.80 & 0.60 \\ 0.60 & 0.80 & 1.00 & 0.60 & 2.00 & 0.60 & 0.80 & 1.00 & 0.60 & 0.60 \\ 0.80 & 0.60 & 0.60 & 0.80 & 0.60 & 2.40 & 0.60 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.60 & 0.80 & 0.40 & 0.80 & 0.60 & 2.80 & 0.40 & 0.40 & 0.40 \\ 1.00 & 1.00 & 0.80 & 0.60 & 1.00 & 0.80 & 0.40 & 2.40 & -0.10 & -0.10 \\ 0.80 & 0.80 & 0.80 & 0.80 & 0.60 & 0.80 & 0.40 & -0.10 & 2.80 & -0.20 \\ 0.60 & 0.60 & 0.80 & 0.60 & 0.60 & 0.80 & 0.40 & -0.10 & -0.20 & 2.20 \end{bmatrix}.$$

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 $\boldsymbol{S}_j - \boldsymbol{S}_1$ and in $\boldsymbol{S}_j^{-1}\overline{\boldsymbol{x}}_j - \boldsymbol{S}_1^{-1}\overline{\boldsymbol{x}}_1$, $j = 2, 3$. The largest $EPMC_r$ for all four *LDR* methods occurred when $r = 1$. Generally, the $EPMC_r$ de-

creased as r increased. For the *SY* method, we have $rank(\boldsymbol{M}) = 4$, 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 $\hat{\boldsymbol{M}}$ 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 $EPMC_p$. However, we can often reduce the $EPMC_p$ 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: $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, and $N_p(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3)$, where $p = 10$,

$$\boldsymbol{\mu}_1 = [0,0,0,0,0,0,0,0,0,0]',$$

$$\boldsymbol{\mu}_2 = [4.43,4.43,4.43,4.43,4.43,4.43,4.43,4.43,4.43,4.43]',$$

$$\boldsymbol{\mu}_3 = [8,8,8,8,8,8,8,8,8,8]',$$

with

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

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

Singular value	$n_i = 25$				$n_i = 50$			
	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$	$\hat{\boldsymbol{M}}$	$\hat{\boldsymbol{U}}$	$\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{L}}$
e_1	59.76	1.93	0.48	0.90	56.56	0.86	0.44	0.72
e_2	36.37	0.66	0.16	0.69	37.02	0.60	0.11	0.56
e_3	7.72	0.31		0.52	5.36	0.16		0.31
e_4	5.21			0.42	3.53			0.24
e_5	3.97			0.34	2.65			0.19
e_6	3.12			0.27	2.09			0.14
e_7	2.46			0.21	1.66			0.11
e_8	1.91			0.16	1.31			0.08
e_9	1.44			0.11	1.00			0.06
e_{10}	0.98			0.07	0.70			0.03

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 45.01 & 30.81 & 31.25 & 31.13 & 27.90 & 27.70 & 32.43 & 26.70 & 29.13 & 28.89 \\ 30.81 & 56.10 & 31.51 & 29.26 & 30.89 & 34.35 & 31.25 & 31.85 & 29.50 & 29.61 \\ 31.25 & 31.51 & 54.55 & 24.43 & 33.96 & 31.62 & 29.73 & 30.49 & 24.03 & 30.87 \\ 31.13 & 29.26 & 24.43 & 59.36 & 26.42 & 29.11 & 32.21 & 26.29 & 29.48 & 27.81 \\ 27.90 & 30.89 & 33.96 & 26.42 & 50.17 & 36.05 & 24.80 & 32.22 & 29.20 & 27.31 \\ 27.70 & 34.35 & 31.62 & 29.11 & 36.05 & 70.18 & 24.82 & 33.83 & 31.94 & 30.51 \\ 32.43 & 31.25 & 29.73 & 32.21 & 24.80 & 24.82 & 47.93 & 29.83 & 26.91 & 30.49 \\ 26.70 & 31.85 & 30.49 & 26.29 & 32.22 & 33.83 & 29.83 & 56.05 & 29.46 & 33.34 \\ 29.13 & 29.50 & 24.03 & 29.48 & 29.20 & 31.94 & 26.91 & 29.46 & 46.30 & 28.96 \\ 28.89 & 29.61 & 30.87 & 27.81 & 27.31 & 30.51 & 30.49 & 33.34 & 28.96 & 56.84 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 75.01 & 60.81 & 61.25 & 61.13 & 57.90 & 57.70 & 62.43 & 56.70 & 59.13 & 58.89 \\ 60.81 & 86.10 & 61.51 & 59.26 & 60.89 & 64.35 & 61.25 & 61.85 & 59.50 & 59.61 \\ 61.25 & 61.51 & 84.55 & 54.43 & 63.96 & 61.62 & 59.73 & 60.49 & 54.03 & 60.87 \\ 61.13 & 59.26 & 54.43 & 89.36 & 56.42 & 59.11 & 62.21 & 56.29 & 59.48 & 57.81 \\ 57.90 & 60.89 & 63.96 & 56.42 & 80.17 & 66.05 & 54.80 & 62.22 & 59.20 & 57.31 \\ 57.70 & 64.35 & 61.62 & 59.11 & 66.05 & 100.18 & 54.82 & 63.83 & 61.94 & 60.51 \\ 62.43 & 61.25 & 59.73 & 62.21 & 54.80 & 54.82 & 77.93 & 59.83 & 56.91 & 60.49 \\ 56.70 & 61.85 & 60.49 & 56.29 & 62.22 & 63.83 & 59.83 & 86.05 & 59.46 & 63.34 \\ 59.13 & 59.50 & 54.03 & 59.48 & 59.20 & 61.94 & 56.91 & 59.46 & 76.30 & 58.96 \\ 58.89 & 59.61 & 60.87 & 57.81 & 57.31 & 60.51 & 60.49 & 63.34 & 58.96 & 86.84 \end{bmatrix}.$$

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 $\hat{\boldsymbol{M}}$, we see that most of the discriminatory information was contained in the first transformed dimension. This fact was illustrated with how $EPMC_r(SY)$ and r were directly related, regardless of the training-sample size n_i, $i = 1, 2, 3$.

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 *EMPC* was for the *SY LDR* method where $\left[EPMC_{10} - EPMC_1(SY)\right] \approx 0.09$ for $n_i = 50$ and $\left[EPMC_{10} - EPMC_1(SY)\right] \approx 0.14$ 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 Califorina 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: $N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where

$$\boldsymbol{\mu}_1 = \left[61.68, 65.85, 70.37, 70.48, 66.01, 68.21, 60.18, 73.78, 71.68, 63.66\right]',$$

Table 6. Summary of singular values for the four competing *LDR* methods for Configuration 6.

Singular value	$n_i = 25$				$n_i = 50$			
	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$
e_1	691.33	0.37	0.36	0.72	683.07	0.13	0.30	0.53
e_2	71.16	0.14	0.11	0.56	50.37	0.07	0.06	0.31
e_3	50.13	0.08		0.46	35.78	0.04		0.24
e_4	39.44			0.37	28. 39			0.19
e_5	32.07			0.30	23.19			0.15
e_6	26.27			0.24	19.10			0.12
e_7	21.43			0.18	15.67			0.09
e_8	17.22			0.14	12.13			0.07
e_9	13.41			0.09	9.89			0.05
e_{10}	9.59			0.06	7.08			0.03

$$\boldsymbol{\mu}_2 = \left[62.60, 67.90, 71.60, 71.60, 71.00, 69.40, 62.30, 75.90, 72.90, 66.00\right]',$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 131.90 & 62.83 & 6.01 & -1.82 & -6.75 & 0.79 & -16.74 & 3.69 & 30.97 & 9.27 \\ 62.83 & 127.44 & 0.44 & 21.72 & -6.27 & 10.91 & -12.34 & 4.41 & 4.48 & 26.22 \\ 6.01 & 0.44 & 42.82 & 20.46 & 27.11 & 21.93 & 26.23 & 16.45 & 16.68 & 14.73 \\ -1.82 & 21.72 & 20.46 & 43.16 & 13.24 & 22.12 & 12.11 & 15.18 & 10.83 & 17.75 \\ -6.75 & -6.27 & 27.11 & 13.24 & 71.60 & 15.62 & 20.40 & 13.33 & 26.02 & 3.17 \\ 0.79 & 10.91 & 21.93 & 22.12 & 15.62 & 48.53 & 15.54 & 8.80 & 6.24 & 13.60 \\ -16.74 & -12.34 & 26.23 & 12.11 & 20.40 & 15.54 & 49.80 & 12.69 & 3.44 & -0.59 \\ 3.69 & 4.41 & 16.45 & 15.18 & 13.33 & 8.80 & 12.69 & 47.52 & 25.45 & 5.78 \\ 30.97 & 4.48 & 16.68 & 10.83 & 26.02 & 6.24 & 3.44 & 25.45 & 62.92 & 21.23 \\ 9.27 & 26.22 & 14.73 & 17.75 & 3.17 & 13.60 & -0.59 & 5.78 & 21.23 & 87.24 \end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 33.54 & 18.26 & -1.10 & 0.69 & 3.29 & -10.69 & 10.57 & -2.99 & 9.97 & 13.07 \\ 18.26 & 27.35 & -5.60 & -1.60 & -9.43 & -7.19 & 2.95 & -0.94 & 12.64 & 24.29 \\ -1.10 & -5.60 & 9.11 & 3.83 & 0.79 & 0.46 & 7.71 & -3.77 & -3.31 & -8.71 \\ 0.69 & -1.60 & 3.83 & 23.26 & 8.36 & 10.46 & -4.50 & 4.59 & 10.97 & 11.93 \\ 3.29 & -9.43 & 0.79 & 8.36 & 18.43 & 10.29 & -0.93 & 6.64 & 10.79 & -8.50 \\ -10.69 & -7.19 & 0.46 & 10.46 & 10.29 & 29.11 & -12.57 & 4.27 & 4.03 & -5.50 \\ 10.57 & 2.95 & 7.71 & -4.50 & -0.93 & -12.57 & 28.52 & -3.88 & -6.05 & -7.57 \\ -2.99 & -0.94 & -3.77 & 4.59 & 6.64 & 4.27 & -3.88 & 19.41 & 17.56 & 9.50 \\ 9.97 & 12.64 & -3.31 & 10.97 & 10.79 & 4.03 & -6.05 & 17.56 & 46.07 & 20.93 \\ 13.07 & 24.29 & -8.71 & 11.93 & -8.50 & -5.50 & -7.57 & 9.50 & 20.93 & 46.57 \end{bmatrix}.$$

For this dataset, $rank(\boldsymbol{M}) = rank(\hat{\boldsymbol{M}}) = 10$, $rank(\boldsymbol{U}) = rank(\hat{\boldsymbol{U}}) = 1$, $rank(\boldsymbol{K}) = rank(\hat{\boldsymbol{K}}) = 2$, and $rank(\boldsymbol{L}) = rank(\hat{\boldsymbol{L}}) = 10$. **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* $EPMC_3(SY)$ was the smallest error rate and that $EPMC_{10}\approx EPMC_3(SY)$. 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 7. Summary of singular values for the four competing *LDR* methods for the parametric bootstrap example.

Singular value	$n_i=25$				$n_i=50$			
	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$	$\hat{M}$	$\hat{U}$	$\hat{K}$	$\hat{L}$
e_1	390.43	0.23	0.36	0.94	275.15	0.15	0.28	0.92
e_2	198.20			0.86	175.62			0.83
e_3	118.13			0.74	111.28			0.71
e_4	74.01			0.56	69.53			0.50
e_5	52.16			0.42	47.01			0.31
e_6	37.65			0.30	35.02			0.21
e_7	25.72			0.20	24.92			0.13
e_8	16.73			0.12	15.91			0.07
e_9	9.50			0.05	8.60			0.03
e_{10}	4.27			0.02	3.81			0.01

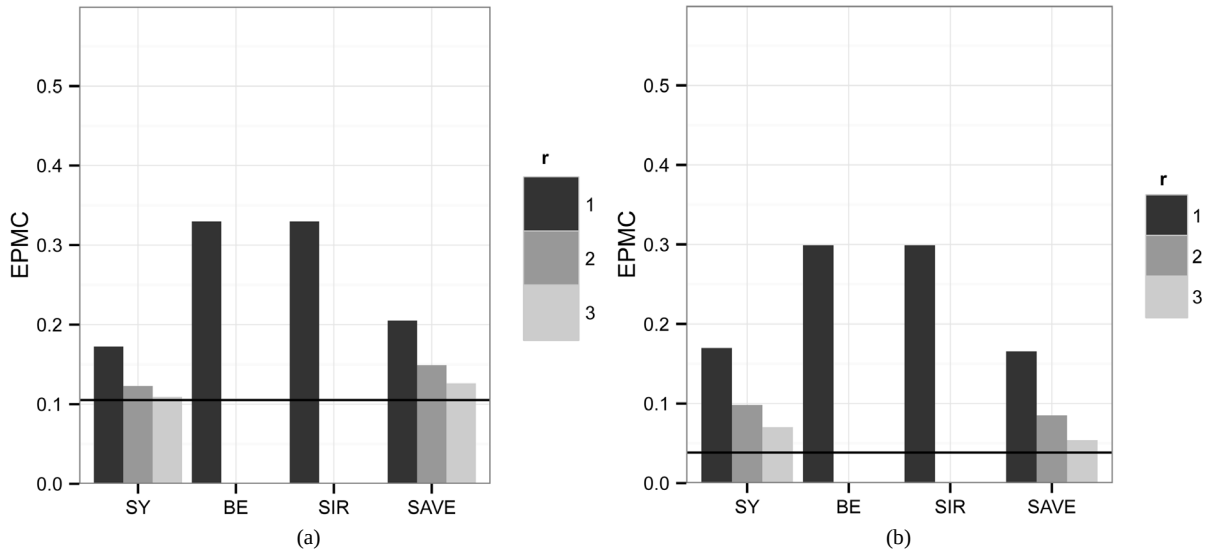

Figure 8. Simulation results from Section 8. The horizontal bar in each graph represents the $EPMC_{10}$. (a) $n_i=25$; (b) $n_i=50$.

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 $\boldsymbol{M}$ 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

[1] Bellman, R. (1961) Adaptive Control Processes: A Guide Tour. John Wiley, Princeton University, Princeton.

[2] Jain, A.K. and Chandrasekaran, B. (1982) Dimensionality and Sample Size Considerations in Pattern Recognition Practice. In: Krishnaiah, P.R. and Kanal, L.N., Eds., *Handbook of Statistics*, Volume 2, Elsevier, Amsterdam, 201-213. http://dx.doi.org/10.1016/s0169-7161(82)02042-2

[3] Rao, C.R. (1948) The Utilization of Multiple Measurements in Problems of Biological Classification. *Journal of the Royal Statistical Society*: *Series B*, **10**, 159-203.

[4] Fisher, R.A. (1936) The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**, 179-188. http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x

[5] McLachlan, G.J. (1992) Discriminant Analysis and Statistical Pattern Recognition. John Wiley, New York. http://dx.doi.org/10.1002/0471725293

[6] Decell, H.P. and Mayekar, S.M. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, *Series B*, **39**, 1-38.

[7] Okada, T. and Tomita, S. (1984) An Extended Fisher Criterion for Feature Extraction—Malina's Method and Its Problems. *Electronics and Communications in Japan*, **67**, 10-16. http://dx.doi.org/10.1002/ecja.4400670603

[8] Loog, M. and Duin, P.W. (2004) Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **26**, 732-739. http://dx.doi.org/10.1109/TPAMI.2004.13

[9] Fukunaga, K. (1990) Introduction to Statistical Pattern Recognition. 2nd Edition, Academic Press, Boston.

[10] Hennig, C. (2004) Asymmetric Linear Dimension Reduction for Classification. *Journal of Computational and Graphical Statistics*, **13**, 930-945. http://dx.doi.org/10.1198/106186004X12740

[11] Kumar, N. and Andreou, A.G. (1996) A Generalization of Linear Discriminant Analysis in a Maximum Likelihood Framework. *Proceedings of the Joint Statistical Meeting*.

[12] Young, D.M., Marco, V.R. and Odell, P.L. (1987) Quadratic Discrimination: Some Results on Optimal Low-Dimensional Representation. *The Journal of Statistical Planning and Inference*, **17**, 307-319.

http://dx.doi.org/10.1016/0378-3758(87)90122-4

[13] Peters, B.C., Redner, R. and Decell, H.P. (1978) Characterization of Linear Sufficient Statistics. *Sankhya*, **40**, 303-309.

[14] Brunzell, H. and Eriksson, J. (2000) Feature Reduction for Classification of Multidimensional Data. *Pattern Recognition*, **33**, 1741-1748. http://dx.doi.org/10.1016/S0031-3203(99)00142-9

[15] Cook, R.D. and Weisberg, S. (1991) Discussion of "Sliced Inverse Regression for Dimension Reduction" by K.-C. Li. *Journal of the American Statistical Association*, **86**, 328-332.

[16] Li, K.-C. (1991) Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, **86**, 316-327. http://dx.doi.org/10.1080/01621459.1991.10475035

[17] Tubbs, J.D., Coberly, W.A. and Young, D.M. (1982) Linear Dimension Reduction and Bayes Classification with Unknown Population Parameters. *Pattern Recognition*, **15**, 167-172. http://dx.doi.org/10.1016/0031-3203(82)90068-1

[18] Schervish, M.J. (19884) Linear Discrimination for Three Known Normal Populations. *Journal of Statistical Planning and Inference*, **10**, 167-175. http://dx.doi.org/10.1016/0378-3758(84)90068-5

[19] Cook, R.D. (2000) SAVE: A Method for Dimension Reduction and Graphics in Regression. *Communications in Statistics—Theory and Methods*, **29**, 2109-2121. http://dx.doi.org/10.1080/03610920008832598

[20] Cook, R.D. and Yin, X. (1991) Dimension Reduction and Visualization in Discriminant Analysis. *Australian and New Zealand Journal of Statistics*, **43**, 147-199. http://dx.doi.org/10.1111/1467-842X.00164

[21] Vellila, S. (2012) A Note on the Structure of the Quadratic Subspace in Discriminant Analysis. *Statistics and Probability Letters*, **82**, 739-747. http://dx.doi.org/10.1016/j.spl.2011.12.020