Heliyon 10 (2024) e26157

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Numerically stable locality-preserving partial least squares discriminant analysis for efficient dimensionality reduction and classification of high-dimensional data

Noor Atinah Ahmad

School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia

ARTICLE INFO

Dataset link: https://csr.quadram.ac.uk/example-datasets-for-download/

Keywords:
Partial least squares discriminant analysis
Fisher discriminant analysis
Dimensionality reduction
Classification
Locality preserving

ABSTRACT

Dimensionality reduction plays a pivotal role in preparing high-dimensional data for classification and discrimination tasks by eliminating redundant features and enhancing the efficiency of classifiers. The effectiveness of a dimensionality reduction algorithm hinges on its numerical stability. When data projections are numerically stable, they lead to enhanced class separability in the lower-dimensional embedding, consequently yielding higher classification accuracy. This paper investigates the numerical attributes of dimensionality reduction and discriminant subspace learning, with a specific focus on Locality-Preserving Partial Least Squares Discriminant Analysis (LPPLS-DA). High-dimensional data frequently introduce singularity in the scatter matrices, posing a significant challenge. To tackle this issue, the paper explores two robust implementations of LPPLS-DA. These approaches not only optimize data projections but also capture more discriminative features, resulting in a marked improvement in classification accuracy. Empirical evidence supports these findings through numerical experiments conducted on synthetic and spectral datasets. The results demonstrate the superior performance of the proposed methods when compared to several state-of-the-art dimensionality reduction techniques in terms of both classification accuracy and dimension reduction.

1. Introduction

The availability of large-scale, high-dimensional data is on the rise. However, constructing an effective and efficient prediction model for tasks such as clustering and classification using high-dimensional data presents a formidable challenge, primarily due to the abundance of irrelevant and redundant attributes. For instance, consider satellite hyperspectral images, which boast millions of pixels and hundreds of spectral bands, resulting in a significant redundancy of information and imposing colossal computational overheads [1]. To address this issue, dimensionality reduction serves as a crucial technique. It simplifies the model structure by eliminating superfluous features, ultimately yielding a lower-dimensional representation of the data that exclusively retains the most pertinent features from the original dataset [2,3].

Among the conventional techniques for dimensionality reduction, supervised methods like Fisher's linear discriminant analysis (LDA) and partial least squares discriminant analysis (PLS-DA) are considered more suitable when classification is the primary goal. This is because they integrate dimensionality reduction and discriminant analysis into a single unified algorithm. In the case of LDA, dimensionality reduction and the enhancement of class separability are achieved through a delicate interplay between the

E-mail address: nooratinah@usm.my.

https://doi.org/10.1016/j.heliyon.2024.e26157
Received 16 January 2023; Received in revised form 29 January 2024; Accepted 8 February 2024

Available online 12 February 2024

overall separation of data belonging to different classes and the separation of data within each class. Despite its many successes, LDA has been known to suffer from the small sample size problem [4], a situation in which the number of samples is much smaller than the dimension of the sample space, resulting in the singularity of the within-class scatter matrix in LDA. Additionally, LDA assumes a global Euclidean structure of the data, which restricts its ability to uncover local nonlinear structures [5]. It has also been observed in [6] that LDA, which measures data distance using the L_2 norm, can be sensitive to outliers, potentially resulting in inaccurate estimations of class means and the calculation of within-class and between-class scatter. In response to these challenges, numerous enhancements and extensions of LDA have been introduced in recent years. Notable developments include Robust Linear Discriminant Analysis (RLDA) [7], the Sparse Discriminant Feature Selection (SDFS) [8], the $l_{2,1}$-LDA method which learns robust discriminative projections by minimizing and maximizing $l_{2,1}$-norm simultaneously [6], the 3E-LDA method [9] and the Robust Sparse LDA (RSLDA) [10].

PLS-DA aims to find a projection of high-dimensional data into a lower-dimensional subspace that maximizes the total covariance between the data vectors and their class membership. PLS-DA has demonstrated significant success in discriminating and classifying high-dimensional spectral data, as documented in several studies [11–16]. However, a major drawback of PLS-DA is its neglect of the within-class data structure. This can lead to suboptimal performance when within-class variability predominates over between-class variability. As observed in [17], PLS-DA fails to account for scenarios where each class exhibits a distinct structure. In cases where one class has greater dispersion than the other, PLS-DA tends to bias the separation boundary towards the larger and more dispersed class. A recent study by [18,19] pointed out that only the between-class scatter matrix emerges in the eigenstructure of PLS-DA. The within-class scatter matrix which encapsulates the class-specific variation is ignored. Consequently, PLS-DA may yield unsatisfactory results when samples within a class form multiple separate clusters.

A more recent class of dimensionality reduction methods incorporates manifold learning and locality-based methods to preserve the specific structure of the original data as much as possible [5,20,21]. LPPLS-DA belongs to this class of methods where the dimension reduction capability of PLS-DA is enhanced by preserving the local class structure of the original data [19]. LPPLS-DA method seeks to find a low-dimensional embedding that not only preserves the class structure but also maximizes class separation. This is achieved by constructing an affinity graph using class labels to capture within-class structures while inheriting the global between-class structure from the underlying PLS-DA framework. In this regard, LPPLS-DA shares similarities with LDA [22] as it involves solving a multiobjective optimization that maximizes between-class scatter and minimizes within-class scatter at the same time. The typical approach to formulate this multi-objective optimization is through a generalized eigenvalue problem that involves both the between-class and within-class scatter matrices.

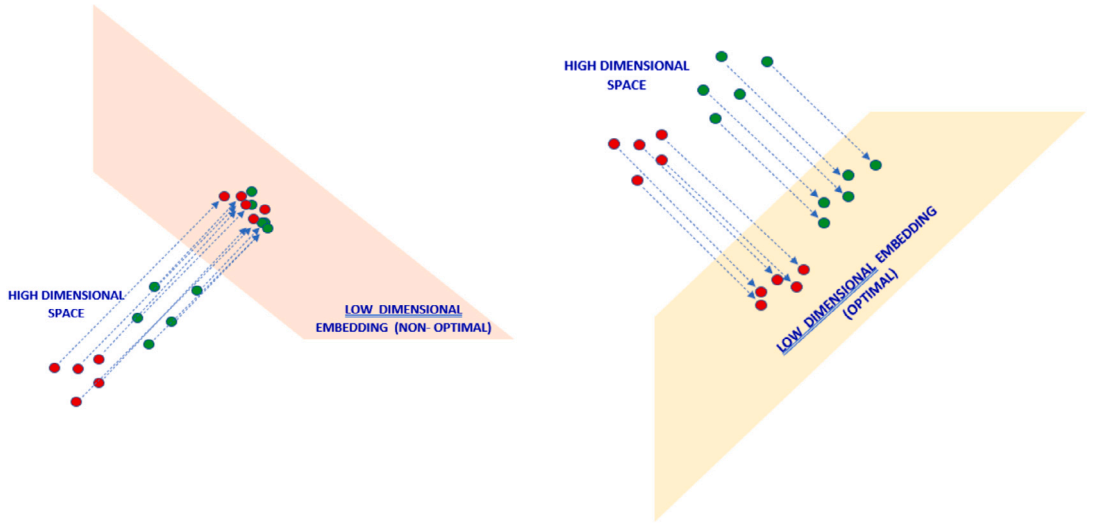
Indeed, as highlighted in [23], LPPLS-DA shares the same eigenstructure as LDA which is of the form

$$Aw = \lambda Bw, \tag{1}$$

where A represents the between-class scatter matrix, and B refers to the within-class scatter matrix. It is worth noting that this generalized eigenvalue problem becomes ill-posed when the within-class scatter matrix is singular. In the case of LPPLS-DA, the problem is ill-posed when the locality-preserving within-class scatter matrix is singular. Consequently, the small sample size problem is not only an issue for LDA but also affects LPPLS-DA. A typical way to handle the singularity of the within-class scatter matrix is by using the Tikhonov regularization [24] where a constant value is added to the diagonal entries of the matrix [4]. Various numerical implementations of regularized LDA have been proposed using this technique [25–27]. However, it is essential to recognize that the ill-posed nature of the generalized eigenvalue problem described in (1) is not solely a result of the singularity of matrix B. According to [28,29], this issue may also stem from the singularity of both matrices B and A. This situation is relevant in practice because, for high-dimensional data, on top of the small sample size problem, high correlation levels among the variables (i.e. multicollinearity) can also lead to the singularity of the between-class scatter matrix [30].

The singularity issues encountered in LPPLS-DA can introduce instability in the numerical algorithm that computes the low-dimensional subspace where classification models are derived. When the algorithm is unstable, the computed projection matrices may fail to adequately capture discriminative features of a dataset. This, in turn, can result in suboptimal classification outcomes. This paper aims to propose two numerically stable implementations of LPPLS-DA. The first, termed Enhanced LPPLS-DA (En-LPPLSDA), applies diagonal shifts to both the PLS-DA between-class scatter matrix and the locality-preserving within-class scatter matrix. These shifts serve to stabilize the diagonalization procedures of these matrices. Stable diagonalization ensures improved orthogonalization of the eigenvectors needed for data projection, ultimately increasing the efficiency of dimensionality reduction. The second approach involves reformulating the LPPLS-DA objective using the maximum scatter difference criterion [31]. This gives rise to LPPLS-DA with maximum scatter difference (LPPLSDAMSD), which reduces the problem to the conventional eigenvalue problem. In this approach, the need to stabilize two scatter matrices is replaced by the task of stabilizing the scatter difference matrix. It is shown that, both En-LPPLSDA and LPPLSDAMSD methods yield the same eigenspace as the traditional LPPLS-DA, but they are better conditioned and offer improved numerical stability.

The rest of the paper is organized as follows: Section 2, provides an overview of the dimensionality reduction problem and presents the mathematical formulations of several supervised linear dimensionality reductions are given. Section 3 discusses the numerical properties of LPPLS-DA and introduces the newly proposed En-LPPLSDA and LPPLSDAMSD approaches. This section also furnishes detailed algorithms for implementing these methods. The experimental results in Section 4 showcase experiments conducted on synthetic datasets, establishing a connection between numerical stability and the performance of LPPLS-DA-based techniques. In Section 5, the comparative performance of LPPLS-DA, En-LPPLSDA, LPPLSDAMSD and other competing methods in the dimensionality reduction and classification of four spectra datasets are presented. The conclusion of the paper is given in Section 6.



(a) Non-optimal projections (b) Optimal projections

Fig. 1. A simplified geometrical perspective of linear dimensionality reduction: Projection of high dimensional data belonging to two classes, Class 1 (red) and Class 2 (green). (a) Non-optimal projections fail to separate the two classes. (b) Optimal projection produces a good separation of data from different classes.

2. Related work

2.1. Dimensionality reduction for data classification: problem description

Given a set of data $\{x_1, x_2, \ldots, x_n\}$ where each $x_k \in R^m$ (m is the dimension of the data space) is an observation vector from m variables. The linear dimensionality reduction problem is a problem of finding a $m \times d$ transformation matrix W that maps $x_1, x_2, \ldots, x_n$ to a set of points $\{y_1, y_2, \ldots, y_n\}$, where $y_k \in R^d$. Here, we expect $d << m$ with d being the dimension of the low-dimensional embedding. The projected data y_k is an image of x_k in the low dimensional embedding and we write $y_k = W^T x_k$. Fig. 1 depicts a geometrical representation of linear dimensionality reduction of a dataset that contains two classes. Non-optimal data projection may not be able to discriminate data from different classes, thus it can produce poor class separation in the low-dimensional embedding (Fig. 1(a)), A good dimensionality reduction algorithm should be able to project data onto the low-dimensional embedding such that class separation is optimum Fig. 1(b). A classifier algorithm such as the k nearest neighbor (kNN) can be applied to the projected data to classify the data into respective classes. Because the classifier algorithm is applied in the low-dimensional space, it is expected to be more efficient than if it were applied to the original data in the m-dimensional space.

2.2. Supervised linear dimensionality reduction and the eigen structures

Supervised dimensionality reduction assumes that class labels of a dataset are known. To include the class labels in the mathematical formulations, throughout this paper, a data set of n data vectors in an m-dimensional space is denoted as

$$X = [x_1, x_2, \ldots, x_n] = [X_1, X_2, \ldots, X_G] \in R^{m \times n},$$

where the data is clustered into G classes and each block matrix $X_g \in R^{m \times n_g}$ has n_g data vectors, where N_g $(1 \le g \le G)$ be the set of column indices that belong to the class g. The between-class scatter matrix S_b and the within-class scatter matrix S_w are defined as

$$S_b = \sum_{g=1}^{G} n_g (\mu^{(g)} - \mu)(\mu^{(g)} - \mu)^T \qquad \text{and} \qquad S_w = \sum_{g=1}^{G} n_g \sum_{k \in N_g} (x_k \mu^{(g)})(x_k - \mu^{(g)})^T,$$

where $\mu = \frac{1}{n} \sum_{k=1}^{n} x_k$ denotes the total sample mean vector and $\mu^{(g)} = \frac{1}{n_g} \sum_{k \in N_g} x_k$ denotes the gth class mean vector. The separability of classes in a data set is often measured by using the traces of S_b and S_w.

One of the most popular supervised linear dimensionality reduction methods is LDA. LDA seeks to find a linear transformation $y_k = W^T x_k$, such that in the projected subspace, the within-class scatter, S_w is minimized and the between-class scatter, S_b is maximized. To do so, the transformation W is determined by maximizing Fisher's criterion, i.e.,

$$\max_W (W^T S_w W)^{-1} (W^T S_b W),$$

that leads to the eigenvalue problem of the form [32]

$$(S_w)^{-1} S_b w = \lambda w, \tag{2}$$

Equivalently, (2) can be expressed as a generalized eigenvalue of the form

$$S_b w = \lambda S_w w,$$

In many applications such as hyperspectral classification [33], cancer classification with gene expression profiling [34] and face recognition [35], there is insufficient data to determine the within-class scatter matrix reliably, thus leading to an unstable inverse of the matrix and hence an ill-conditioned eigenvalue problem. This problem is more commonly known as *the small sample size problem* in which the number of available samples is less than the dimensionality of data.

PLS-DA is considered an alternative to LDA in that it combines dimensionality reduction and discriminant analysis into one algorithm. PLS-DA is derived from the well-known PLS algorithm for modeling the linear relationship between two sets of observed variables. The main idea behind PLS is to find projection matrices $W = [w_1, w_2, \ldots, w_n]^T \in R^{n\times m}$ and $V = [v_1, v_2, \ldots, v_n]^T \in R^{n\times N}$, where each column projection vector pair (w, v) maximizes the co-variance of the projected data. Mathematically, this is represented by a subspace optimization problem of the form

$$\max_{W\in R^{n\times d}, W^T W = I} tr(W^T \bar{X}^T \bar{Y} \bar{Y}^T \bar{X} W), \tag{3}$$

where $\bar{X}$ is the mean-centered data matrix, and the columns of W are the d eigenvectors of $\bar{X}^T \bar{Y} \bar{Y}^T \bar{X}$ (with $d < m$) associated with the d largest eigenvalues. Here, the matrix W defines the projection matrix that transforms high-dimensional data into its low-dimensional representation. For discrimination and classification purposes, the input data matrix $\bar{Y}$ is replaced by a dummy (class membership) matrix defined as

$$\bar{Y} = \begin{pmatrix} 1_{n_1} & 0_{n_1} & \cdots & 0_{n_1} \\ 0_{n_2} & 1_{n_2} & \cdots & 0_{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_G} & 0_{n_G} & \cdots & 1_{n_G} \end{pmatrix} \tag{4}$$

where 0_{n_g} and 1_{n_g} are $n_g \times 1$ vectors of zeros and ones respectively.

It was shown in [19] that, given $\bar{Y}$ as in (4),

$$\hat{S}_b = \bar{X}^T \bar{Y} \bar{Y}^T \bar{X} = \frac{1}{(n-1)^2} \sum_{g=1}^{G} n_g^2 (\mu^{(g)} - \mu)(\mu^{(g)} - \mu)^T, \tag{5}$$

and the optimization problem (3) is equivalent to the eigenvalue problem

$$\hat{S}_b w = \lambda w. \tag{6}$$

Further scrutiny reveals that the expression in (5) can be written as

$$\hat{S}_b = \check{S}_w^{-1} S_b,$$

where $\acute{S}_w = (n-1)^2 diag(\check{S}_w^{(1)}, \check{S}_w^{(2)}, \ldots, \check{S}_w^{(G)})$, such that $\check{S}_w^{(g)}$ is an $n_g \times n_g$ diagonal matrix whose diagonal entries are all equal to $\frac{1}{n_g}$. By comparing (6) with (2), the total within-class distribution is simplified as the diagonal matrix $\check{S}_w$. Thus in comparison with LDA, although PLS-DA may not suffer from instability due to the small sample size problem, the representation of the local within-class structure is rather limited, and significant information can be overlooked.

2.3. Dimensionality reduction based on LPPLS-DA

In an effort to improve PLS-DA, [19] combined PLS-DA with a manifold learning technique to capture the local nonlinear structure of within-class distribution. Locality-preserving manifold learning involves constructing an affinity graph optimized to preserve local neighborhood structures in the embedding space. LPPLS-DA method employs class information while constructing the neighborhood graph to enhance the discriminating ability of the method further. Although both LDA and LPPLS-DA aimed at preserving within-class distances, the within-class structure preserving feature in LPPLS-DA is somewhat different from LDA. LDA assumes data to be distributed normally, whereas, in LPPLS-DA, local data distribution is captured via the affinity graph and this graph structure is preserved in the low-dimensional embedding. In this way, intrinsic information on local data distribution can be preserved during dimensionality reduction.

The overall objective of LPPLS-DA is to find a low-dimensional subspace in which the intrinsic geometry and discriminant structure of data are preserved. It finds a projection matrix W that transforms the high-dimensional data set X into a low-dimensional subspace $Z = XW$ such that the overall local distance of data samples from the same class is preserved. The local/class structure for samples in class g is coded via an adjacency graph whose edges are weighted by the entries of a matrix $S^{(g)}$ and the combined structure for the entire dataset is in the form of a sparse $n \times n$ matrix S with the following structure:

$$S = \begin{pmatrix} S^{(1)} & 0_{n_1 \times n_2} & \cdots & 0_{n_1 \times n_G} \\ 0_{n_2 \times n_1} & S^{(2)} & \cdots & 0_{n_2 \times n_G} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_G \times n_1} & 0_{n_G \times n_2} & \cdots & S^{(G)} \end{pmatrix}. \tag{7}$$

The entries of S are given by:

$$S_{ij} = \begin{cases} \exp\left(-\frac{\|x_i - x_j\|^2}{t}\right); & \text{if } x_i \text{ and } x_j \text{ both belong to the same class,} \\ 0; & \text{Otherwise,} \end{cases} \tag{8}$$

where t in (8) is a user-specified parameter. The choice of the weights in (7) guarantees that samples from the same class are connected by the adjacency graph. Furthermore, if the distance between two samples in the same class is small, a large weight is assigned to the edge connecting them to highlight the importance of the two samples in the class structure. This attribute enhances the discriminative feature for each class in the dataset so that when the conventional PLS-DA objective is maximized, class separation is enhanced at the same time.

A reasonable criterion for preserving the local structure is to minimize the following objective function [36]:

$$\sum_{ij} (z_i - z_j)^2 S_{ij}, \tag{9}$$

where z_i is the projection of sample x_i onto the lower-dimensional subspace Z. The structure of S turns the sum in (9) into a weighted within-class scatter distance. Further, it is shown in [19] that minimizing (9) is equivalent to

$$\min_{W^T W = I} tr(W^T X L X^T W),$$

where $L = D - S$ is the graph Laplacian, and D is a diagonal matrix whose entries are equal to the column (or row) sum of S. To combine the PLS-DA approach of maximizing the between-class separation and minimizing within-class scatter with locality preserving, two distinct objectives are solved:

$$\max_{W^T W = I} tr\left(W^T \hat{S}_b W\right), \tag{10}$$

$$\min_{W^T W = I} tr\left(W^T \hat{S}_w W\right), \tag{11}$$

where $\hat{S}_w = \bar{X}^T L \bar{X}$ is called the LPPLS-DA within-class scatter matrix that codes the local/class structures. In LPPLS-DA, the objectives above are combined by maximizing a discriminative ratio of the form:

$$J(W) = tr\left(\frac{W^T \hat{S}_b W}{W^T \hat{S}_w W}\right). \tag{12}$$

The objective function in (12) tries to maximize the between-class scatter and minimize the within-class scatter simultaneously, similar to the objective of LDA (2). Likewise, the optimum d-dimensional projection matrix W is the matrix whose columns are the eigenvectors corresponding to the first d principal eigenvalues of a generalized eigenvalue problem of the form:

$$\hat{S}_b w = \lambda \hat{S}_w w. \tag{13}$$

Because both LPPLS-DA and LDA have the same eigenstructure, the numerical properties of LPPLS-DA are also affected by the conditioning of $\hat{S}_w$. The small sample size problem can also lead to the ill-conditioning of $\hat{S}_w$ because in such a case, the graph Laplacian is singular.

3. LPPLS-DA with improved numerical properties

Computational methods for solving the generalized eigenvalue problem in (13) are designed based on the simultaneous diagonalization of the form [37,32]

$$W^T \hat{S}_w W = I, \tag{14}$$

$$W^T \hat{S}_b W = \Lambda, \tag{15}$$

where W is an orthogonal matrix containing the generalized eigenvectors. To achieve (14) and (15), a series of orthogonalization procedures are implemented, and if the eigenvector basis has poor conditioning, even a relatively small perturbation can lead to instability. Thus, accurate computation of the eigenvectors of (13) can be especially challenging when the matrices $\hat{S}_b$ and $\hat{S}_w$ are singular or nearing singularity. In such cases, it is reasonable to expect that the computed eigenvectors are not quite orthogonal and may not lead to an optimal projection.

Algorithm 1: Enhanced LPPLS-DA.

Input:
1. Training set with class labels $x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)}, \ldots, x_1^{(G)}, x_2^{(G)}, \ldots, x_{n_G}^{(G)}$
2. Dimensionality of embedding space d

Output:
1. The optimal transformation matrix $W \in R^{m \times d}$
2. The d-dimensional embedding coordinates Z for the original input data X

Step 1: Compute $\widehat{S}_b$ as follows:
(i) Form mean-centered data matrix $\bar{X}$.
(ii) Form the class membership matrix $\bar{Y}$ according to (4).
(iii) Compute $\widehat{S}_b$ according to (5).

Step 2: Construct $\widehat{S}_w$ as follows:

(i) Construct the weight matrix S according to (7).
(ii) Compute the graph Laplacian $L = D - S$ where the diagonal entries of D are the column/row sum of S, i.e.,

$$D_{jj} = \sum_{i=1}^{n} S_{ij}, j = 1, 2, \ldots, n;$$

(iii) Compute $\widehat{S}_w = X^T L X$.

Step 3: Form the diagonally shifted matrices:

$$\widetilde{S}_b = \widehat{S}_b + \sigma_b I,$$

$$\widetilde{S}_w = \widehat{S}_w + \sigma_w I.$$

Step 4: Compute the d largest eigenvalues and the corresponding eigenvectors of the generalized eigenvalue problem $\widetilde{S}_b w = \lambda \widetilde{S}_w w$.
Step 5: Construct the matrix W whose columns are the generalized eigenvectors calculated from Step 4 above.
Step 6: Compute the d-dimensional embedding $Z = \bar{X} W$.

3.1. Diagonal shifts

In the conventional LPPLS-DA and LDA method, the singularity of $\widehat{S}_w$ (or the within-class scatter matrix S_w in LDA) is often the major concern and because it is linked to the small sample size problem [38]. To address this issue, in the LPPLS-DA algorithm, the locality preserving within-class scatter matrix $\widehat{S}_w$ is stabilized by a diagonal shift which is equivalent to adding a Tikhonov regularization [24] to the objective function. The generalized eigenvalue problem becomes

$$\widehat{S}_b w = \lambda(\widehat{S}_w + \sigma_w I) w, \tag{16}$$

where $\sigma_w > 0$ is a scalar determined empirically to achieve optimum performance. However, when working with a large dataset of samples with high dimensionality, $\widehat{S}_b$ (or the between-class scatter matrix S_b in LDA) can also become singular, or close to becoming singular, particularly when discrimination of different classes in the dataset is poor. According to [28,29], ill-conditioning of the generalized eigenvalue problem (13) may result from the ill-conditioning of both, the locality-preserving within-class scatter matrix $\widehat{S}_w$ and the between-class scatter matrix $\widehat{S}_b$. Thus, in addition to stabilizing $\widehat{S}_w$ as in (16), further enhancement to LPPLS-DA can be achieved by also adding a diagonal shift of σ_b to $\widehat{S}_b$.

Diagonal-shifted $\widehat{S}_w$ and $\widehat{S}_b$ transform the eigenvalue problem in (13) to

$$\widetilde{S}_b w = \lambda \widetilde{S}_w w, \tag{17}$$

where $\widetilde{S}_b = \widehat{S}_b + \sigma_b I$ and $\widetilde{S}_w = \widehat{S}_w + \sigma_w I$. Simultaneous diagonalization problem associated with (17) is

$$W^T(\widehat{S}_w + \sigma_w I) W = W^T \widetilde{S}_w W = (1 + \sigma_w) I, \tag{18}$$

$$W^T(\widehat{S}_b + \sigma_b I) W = W^T \widetilde{S}_b W = \Lambda + \sigma_b I. \tag{19}$$

Consequently, if there are eigenvalues of $\widehat{S}_b$ ($\widehat{S}_w$) that are zero (i.e. singular) or close to zero (i.e. near singularity), the diagonal shift σ_b (σ_w) shifts the eigenvalues toward the positive direction, thereby improving the conditioning of the matrices. Notice that from (18) and (19), the eigenspace of ($\widehat{S}_b, \widehat{S}_w$) is the same as the eigenspace of ($\widetilde{S}_b, \widetilde{S}_w$). Only the eigenvalues (i.e. diagonal entries of Λ) are shifted by σ_b (σ_w). The shift in eigenvalues can always be adjusted within the algorithm if needed. But since the eigenvalues are not needed to determine the projection matrix W, the discussions on the eigenvalues are not pursued further in this paper. The procedures for computing W based on (18) and (19) are given in Algorithm 1.

Algorithm 2: LPPLSDAMSD.

Input:
1. Training set with class labels $x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)}, \ldots, x_1^{(G)}, x_2^{(G)}, \ldots, x_{n_G}^{(G)}$
2. Dimensionality of embedding space d

Output:
1. The optimal transformation matrix $W \in R^{m\times d}$
2. The d-dimensional embedding coordinates Z for the original input data X

Do Steps 1-2 of Algorithm 1

Step 3: Form the scatter-difference matrix $\bar{S} = \hat{S}_b - \theta\hat{S}_w$

Step 4: Add a diagonal shift for improved stability: $\widetilde{S} = \bar{S} + \sigma_\theta I$

Step 5: Compute the d largest eigenvalues and the corresponding eigenvectors of the generalized eigenvalue problem $\widetilde{S}w = \lambda w$.

Step 6: Construct the matrix W whose columns are the generalized eigenvectors calculated from Step 4 above.

Step 7: Compute the d-dimensional embedding $Z = \bar{X}W$.

3.2. Maximum scatter difference criterion

The LPPLSDAMSD method is designed based on the maximum scatter difference (MSD) criterion. Based on this criterion, the simultaneous optimization of objective functions (10) and (11) is achieved by maximizing the following function:

$$J(W) = trW^T(\hat{S}_b - \theta\hat{S}_w)W,$$

subject to the constraint $W^TW = I$. The parameter θ is the scatter-difference parameter that will be empirically determined. It can be shown that the optimum d-dimensional projection matrix W_{opt} is the matrix whose columns are the eigenvectors corresponding to the first d principal eigenvalues of an eigenvalue problem of the form:

$$(\hat{S}_b - \theta\hat{S}_w)w = \pi w. \tag{20}$$

Notice that, if the orthogonal matrix W satisfies the simultaneous diagonalization in (14) and (15), then

$$W^T(\hat{S}_b - \theta\hat{S}_w)W = \Lambda - \theta I. \tag{21}$$

The matrix form for (20) is

$$(\hat{S}_b - \theta\hat{S}_w)W = W\Pi, \tag{22}$$

where Π is a diagonal matrix containing the eigenvalues of (20). By comparing (21) with (22), we can interpret the generalized eigenvectors of $(\hat{S}_b, \hat{S}_w)$ as the eigenvectors of (22) associated with eigenvalues $\Lambda - \theta I$. Based on this observation, it is deduced that the projection matrix of the conventional LPPLSDA is equivalent to the projection matrix of the LPPLSDAMSD.

The stability of the eigenvalue problem in (20) depends on the conditioning of a single matrix which is $\hat{S}_b - \theta\hat{S}_w$. Singularity (or near-singularity) of either ($\hat{S}_b$ or $\hat{S}_w$) can be compensated in the matrix pencil $\bar{S} = \hat{S}_b - \theta\hat{S}_w$. Further stability can be achieved by adding a diagonal shift σ_θ to $\bar{S}$ to give $\widetilde{S} = \bar{S} + \sigma_\theta I$. The procedures for the LPPLSDAMSD method are given in Algorithm 2.

4. Experiments A: synthetic datasets

While a dimension reduction method may be based on sound modeling principles, its numerical implementation can introduce instability, preventing the realization of its full potential. To emphasize the impact of this issue on the performance of LPPLS-DA, this section presents a systematic approach to study the numerical properties of LPPLS-DA. The aim is to provide an intuitive and visual understanding of the dimensionality reduction problem and the effects of numerical instability.

Two sets of experiments are conducted on specifically designed synthetic data. in Section 4.1, experiments are performed using four different supervised methods namely PLS-DA, LDA, supervised LPP (SLPP), and LPPLS-DA to highlight the dimension reduction properties of these methods and underscore the true capabilities of LPPLS-DA. In Section 4.2, the effects of singularities of the between-class and within-class scatter matrices on the performance of LPPLS-DA are demonstrated. Furthermore, the effectiveness of En-LPPLSDA and LPPLSDAMSD in overcoming the numerical instability resulting from ill-conditioned datasets is discussed.

4.1. Supervised dimensionality reduction via PLS-DA, LPPLS-DA, LDA and SLPP

To demonstrate visually the comparative dimension reduction properties of the PLS-DA, LPPLS-DA, LDA, and SLPP, a synthetic dataset containing two classes with Gaussian distribution is used. Each class is composed of 60 samples. The two classes of samples form two separate (main) clusters with samples in Class 1 forming three small clusters with 20 samples in each cluster. Samples in the dataset have three features and the specific distribution is shown in Fig. 2. The data set is designed so that the optimal projection is vertical onto the horizontal plane. For the LPPLS-DA method, the condition number of $\hat{S}_w$ for this dataset is 2.3817 while the condition number of $\hat{S}_b$ is 7.6942×10^{16}. In the case of LDA, the condition number of S_w is 9.4902 and the condition number of S_b is 5.8904×10^16. Therefore, it is reasonable to expect that the conditioning of LPPLS-DA and LDA for the chosen dataset is comparable.

The methods PLS-DA, LPPLS-DA, LDA, and SLPP are applied to the dataset and two-dimensional embeddings obtained for each method are depicted in Fig. 3. Among the four methods, PLS-DA, LDA, and LPPLS-DA emphasize class separation and this property

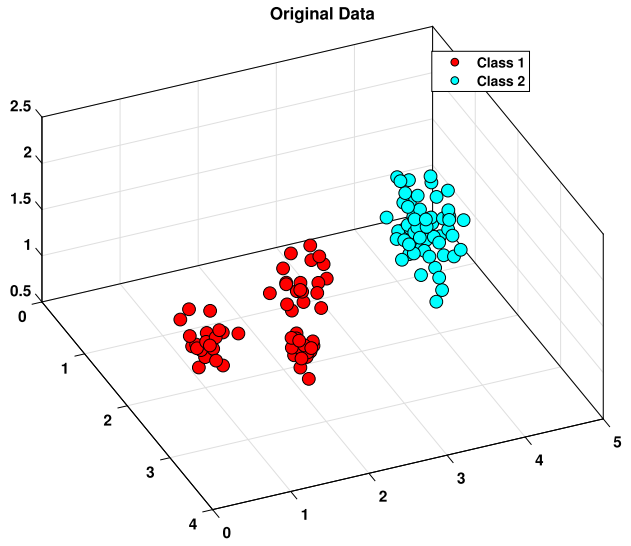


Fig. 2. Synthetic dataset composed of two classes. Dataset from Class 1 is composed of three small clusters.

is clearly captured in Figs. 3(a), (b), and (d), where projected data from Class 1 and Class 2 are shown to be well-separated. For SLPP, between-class separation is not emphasized, and it can be seen in Fig. 3(c) that class separation is almost completely ignored. A closer observation of the within-class structures reveals that, only LPPLS-DA manages to capture and preserves the three-clustered distribution of Class 1 data in its two-dimensional embedding while maximizing class separation. SLPP manages to capture and preserve the class structures but with limited ability to discriminate the different classes. PLS-DA and LDA captured some of the multi-clustered structure of Class 1 data but with limited effect.

The experiment in this section allows us to assess visually the dimensionality reduction properties of PLS-DA, LDA, SLPP, and LPPLS-DA. Intuitively, the performance of each method is restricted by how it is designed. Compared to generic methods like PCA and SLPP, the discriminant methods PLS-DA, LDA and LPPLS-DA perform better in separating the different classes in a dataset. But it is evident that, if there are unique features within a specific class, for example in this case, a class with multi-clustered distribution, the structure-preserving property in LPPLS-DA is proven to be more effective in capturing the intrinsic information. Capturing more discriminative features in the low-dimensional embedding is desirable because it leads to a more accurate representation of the dataset's class structure in the classification model derived from the low-dimensional space.

4.2. Effects of numerical instability on LPPLS-DA

The generalized eigenvalue problem (13) is ill-posed and potentially numerically unstable when dealing with small-sample size problems. To portray this situation, a data matrix X with n rows and m columns is generated, where n corresponds to the total number of samples in the dataset and m refers to the number of variables. A small sample size problem is a situation where $n < m$ and in the experiments in this section, m is set to equal $2n$. Three different datasets are created, each of which has $n = 120$ samples from three different classes.

1. Dataset Synthetic I contains 120 samples from three classes generated from the same Gaussian distribution. Each class contains 60 samples with 240 variables, and variables 120 to 240 of each class are affected by noise (different noise levels for each class).
2. Dataset Synthetic II contains 120 samples from three classes generated from the different Gaussian distributions. Class 1 has a standard deviation of 0.1, Class 2 has a standard deviation of 0.05, and Class 3 has a standard deviation of 0.01.
3. Dataset Synthetic III contains 120 samples from three classes generated from different Gaussian distributions. Class 1 is composed of two small clusters with 30 samples each, and the clusters have different Gaussian distributions.

The conditioning of the dimensionality reduction problem with respect to these datasets is determined by matrices $\widehat{S}_b$ and $\widehat{S}_w$ (LPPLS-DA), $\widetilde{S}_b$ and $\widetilde{S}_w$ (En-LPPLSDA), and $\widetilde{S}$ (LPPLSDAMSD). Table 1 summarizes the condition numbers of these matrices for each of the datasets.

Condition numbers in Table 1 highlight the ill-conditioned property of the LPPLS-DA method. Both matrices $\widehat{S}_b$ and $\widehat{S}_w$ exhibited very high condition numbers in all three synthetic datasets, indicating that ill-posedness and instability result not only from the ill-conditioning of $\widehat{S}_w$ but also from the ill-conditioning of $\widehat{S}_b$. Table 1 reveals that adding the diagonal shifts to both $\widehat{S}_b$ and $\widehat{S}_w$ significantly reduces the condition numbers of the shifted matrices $\widetilde{S}_b$ and $\widetilde{S}_w$ to approximately 1. The condition number of $\widetilde{S}$ for LPPLSDAMSD also hovers around 1. Thus it is expected that the eigenvalue problems associated with En-LPPLSDA and LPPLSDAMSD are more stable compared to the eigenvalue problem associated with the conventional LPPLS-DA.

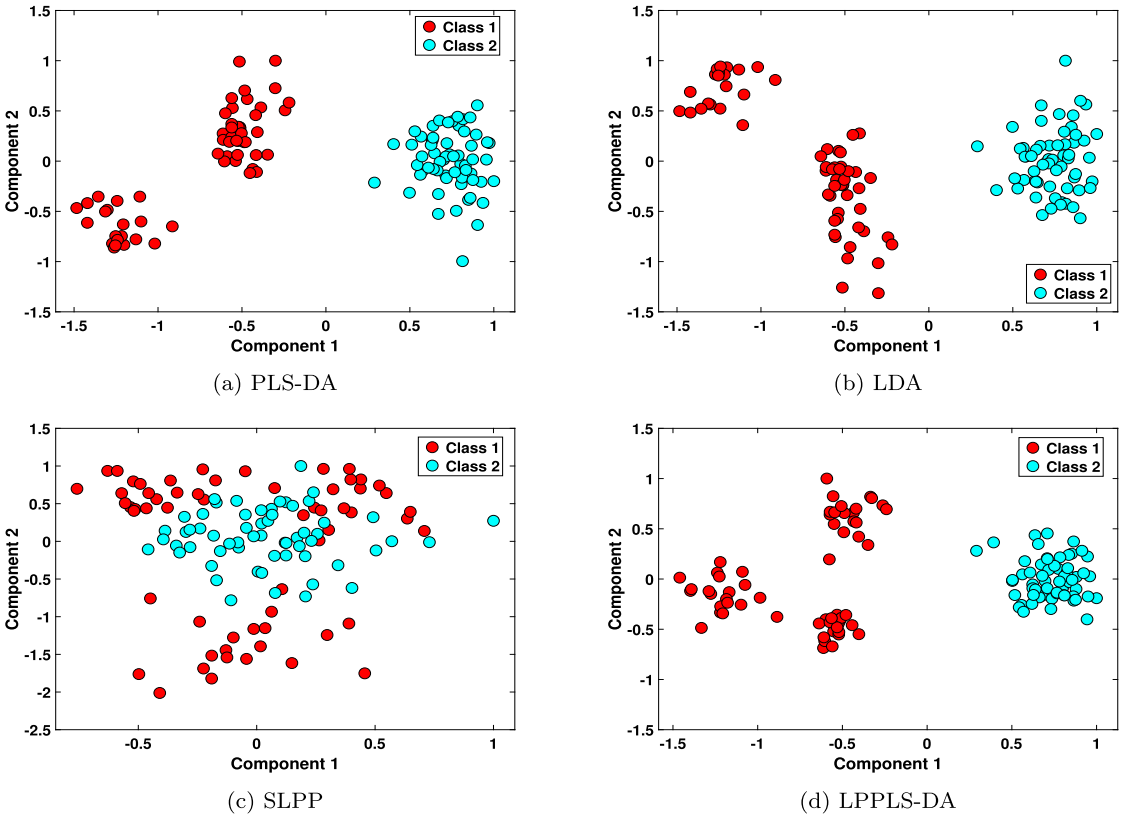


Fig. 3. Two-dimensional embeddings resulted from (a) PLS-DA, (b) LDA, (c) SLPP, and (d) LPPLS-DA.

Table 1
Condition numbers of the matrices associated with LPPLS-DA, En-LPPLSDA and LP-PLSDAMSD for synthetic datasets.

Datasets	Condition number				
	$\widehat{S}_b$	$\widehat{S}_w$	$\widetilde{S}_w$	$\widetilde{S}_b$	$\widetilde{S}$
Synthetic I	4.7655×10^{20}	1.7626×10^{18}	1.0023	1.0022	1.0322
Synthetic II	4.7655×10^{20}	1.7626×10^{18}	1.000	1.000	1.000
Synthetic III	4.7655×10^{20}	1.7626×10^{18}	1.000	1.000	1.000

The effects of conditioning on the quality of dimensionality reduction can be seen in Figs. 4–6. Figs. 4(a)-(d) illustrates the dimensionality reduction of dataset Synthetic I, revealing no significant difference between the two-dimensional embeddings of PLS-DA and LPPLS-DA. Conversely, the two-dimensional embedding of En-LPPLSDA (Fig. 4(c)) demonstrates improved between-class separation and more compact within-class structures. This observation emphasizes the fact that discrimination of the different classes in the dataset is suboptimal when the problem is ill-conditioned. Improving the conditioning of the problem as in the case of En-LPPLSDA enables more robust discrimination. Optimal discrimination is also achieved with LPPLSDAMSD as shown in Fig. 4(d). Here, the robustness of the method depends on the condition number of the scatter difference matrix $\widetilde{S}$. Figs. 5 and 6 bring to light the effects of conditioning on the structure-preserving feature of each method. For dataset Synthetic II, the two-dimensional embeddings in Fig. 5(a) and Fig. 5(b) show that between-class separation is almost non-apparent in the case of PLS-DA and LPPLS-DA. But for En-LPPLSDA and LPPLSDAMSD (Fig. 5(c) and Fig. 5(d) respectively), we observe good between-class separation as well as distinct structure-preserving. The distribution of samples from Class 3 which has the smallest variance, appears in the two-dimensional embedding as having the most compact structure. The enhanced structure-preserving feature is also visible in Figs. 6(a)-(d) for dataset Synthetic III. Here, Class 1 samples are composed of two separate clusters and this property is preserved well by the numerically robust En-LPPLSDA and LPPLSDAMSD.

In conclusion, the conditioning of both the between-class scatter matrix and the locality-preserving within-class scatter matrix significantly impacts the numerical stability of the LPPLS-DA method. Improving the conditioning of the method by either adding diagonal shifts to the between-class scatter matrix and the locality-preserving within-class scatter matrix or reformulating the problem as a maximum scatter-difference problem, gives rise to two numerically robust methods namely the En-LPPLSDA and LPPLSDAMSD respectively which provide more optimal discrimination of the classes and at the same time preserve intrinsic class information.

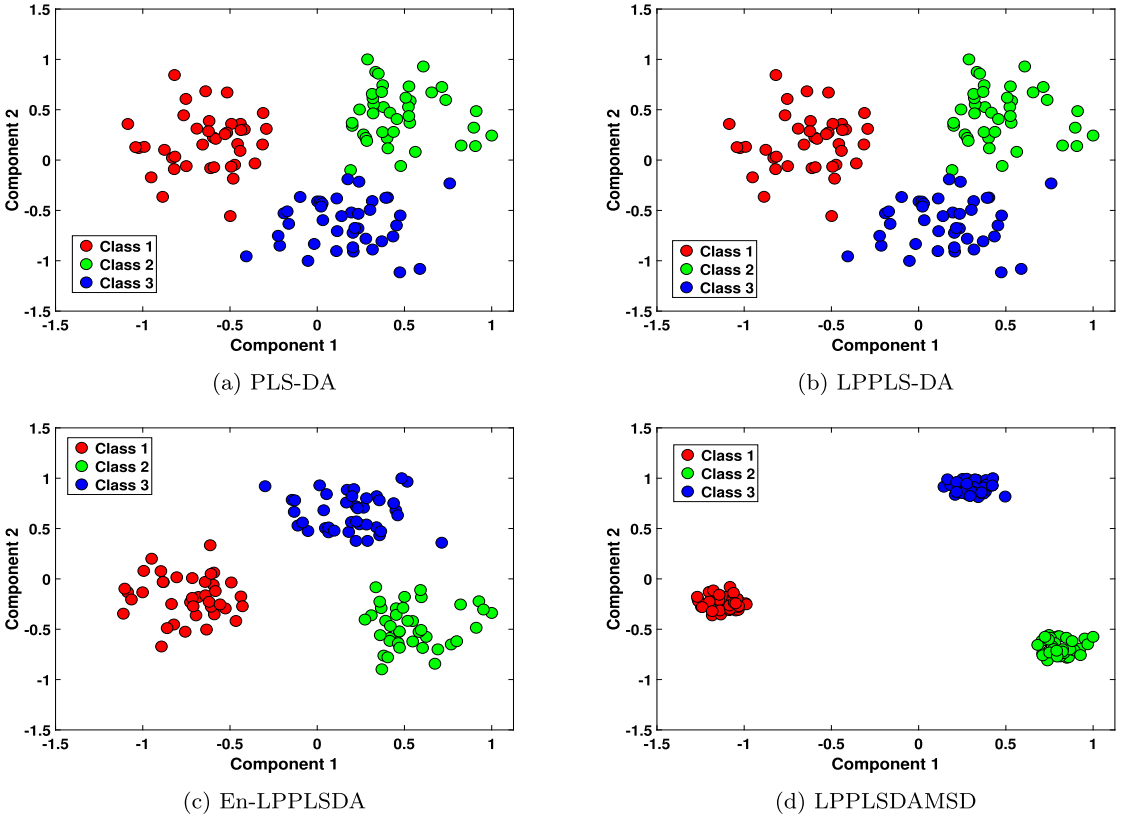


Fig. 4. Ill-conditioned dataset Synthetic I. Two-dimensional projections resulted from (a) PLS-DA, (b) LPPLSDA, (c) En-LPPLSDA, and (d) LPPLSDAMSD.

5. Experiments B: real-life datasets

The aim of this section is to provide an overall understanding of the effect of numerical instability and ill-conditioning on dimension reduction algorithms and their impact on subsequent classification accuracy. Experiments are conducted using four publicly available spectra datasets. All experiments have been performed in MATLAB R2019a on an Intel core i7 3.20 GHz Windows 10 machine with 8 GB memory.

5.1. Datasets

The datasets used in the experimental studies in this section are Coffee [39,40], Meat [41,40], Oil [42,40] and Fruit [43,40] datasets. Table 2 summarizes the statistics of these datasets.

1. The Coffee dataset contains 56 samples belonging to two different species; arabica and robusta species with 29 and 27 samples, respectively. Spectra were acquired from the sample by Fourier transform infrared spectroscopy with diffuse reflectance sampling where each spectrum contains 286 variables in the range of $810 - 1910$ cm^{-1}.
2. The Meat dataset contains infrared spectra of 60 independent samples of fresh minced meats: turkey, chicken, and pork, with 20 samples belonging to each class. The spectra were obtained by Fourier transform infrared spectroscopy with attenuated total reflectance (ATR) sampling and each spectrum contain 448 variables in the range of $1005 - 1867$ cm^{-1}.
3. The Oil dataset contains infrared spectra of 60 samples of authenticated extra virgin olive oils from 4 different countries of origin: Spain, Portugal, Italy, and Greece with $25, 8, 17$, and 10 samples, respectively. The spectra were acquired by Fourier transform infrared spectroscopy with attenuated total reflectance (ATR) sampling. Each spectrum contains 570 variables in the range of $798 - 1896$ cm^{-1}.
4. The Fruit dataset contains 983 samples of authenticated fruit purees belonging to two different classes: “Strawberry” and “Non-Strawberry” with 351 and 632 samples, respectively. The class labeled “Strawberry” contains purees prepared from whole fruits, while the class labeled “Non-Strawberry” contains a diverse collection of other purees including strawberry adulterated with other fruits and sugar solutions, apricot, cherry, plum, blackberry, blackcurrant, apple, raspberry, grape juice and mixtures of these. Mid-infrared spectra were obtained from each puree using attenuated total reflectance (ATR) sampling. Each spectrum contains 235 variables in the range of $899 - 1802$ cm^{-1}.

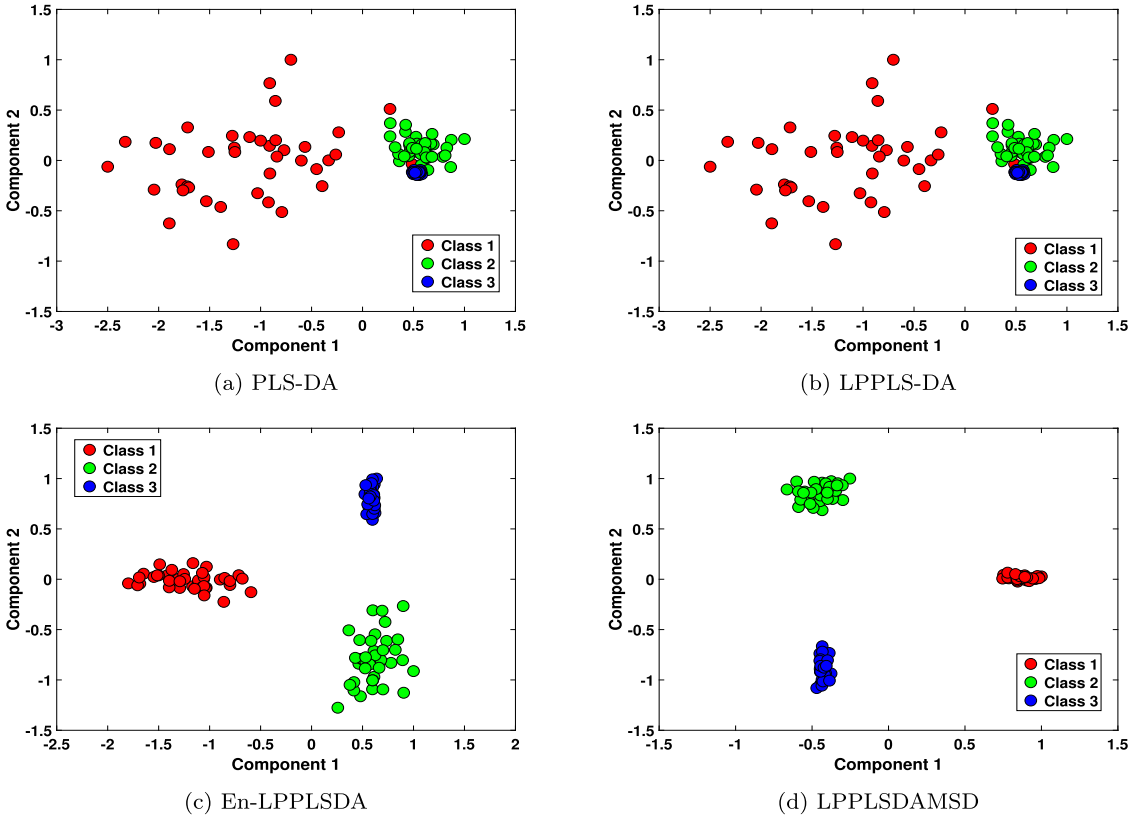


Fig. 5. Ill-conditioned dataset Synthetic II. Two-dimensional projections resulted from (a) PLS-DA, (b) LPPLSDA, (c) En-LPPLSDA, and (d) LPPLSDAMSD.

Table 2
Summary of datasets used in the experiments.

Datasets	Sample size	No. of features	No. of classes	Spectral region (cm^{-1})
Coffee	56	286	2	810 – 1910
Meat	60	448	3	1005 – 1867
Oil	60	570	4	789 – 1896
Fruit	983	235	2	899 – 1802

5.2. Numerically unstable dimension reduction and the effects on classification

To be able to perform effective classification tasks in the low-dimensional subspace, it is important to ascertain that projection matrices computed using the dimensionality reduction algorithms project samples optimally onto a lower-dimensional embedding. The optimal projection here means class separation is optimized and innate features of each class such as local class structures are preserved in the low-dimensional embedding. Classification models developed using optimally projected training samples have a better chance of capturing significant class information of the original dataset. The accuracy of classification using the trained classification model is also expected to be higher if the distribution of projected test samples (i.e., unknown data) is comparatively similar to the distribution of projected training samples. Numerically unstable dimensionality reduction algorithms may lead to test samples being projected outside the distribution of training samples. To gain some understanding of how conditioning affects classification, we take a closer look at the distribution of projected samples from the datasets introduced in the previous section.

Table 3 summarizes the condition numbers of matrices $\widehat{S}_b$, $\widehat{S}_w$, $\widetilde{S}_b$, $\widetilde{S}_w$, and $\widetilde{S}$ for each of the datasets used in this section. It is evident that matrices $\widehat{S}_b$ and $\widehat{S}_w$ exhibit very large condition numbers for all real-life datasets, whereas condition numbers of matrices associated with the En-LPPLSDA and LPPLSDAMSD methods are very close to 1. Based on these numbers it is expected that the dimension reduction problems of En-LPPLSDA and LPPLSDAMSD are better conditioned and more stable compared to the dimension reduction problem of the conventional LPPLS-DA.

In Figs. 7–10, we illustrate the distribution of projected samples from the four datasets under consideration, each divided into a training set and a test set. The projection matrices are computed using LPPLS-DA, En-LPPLSDA, and LPPLSDAMSD for the training sets. These projection matrices are then applied to project both the training and test samples onto the lower-dimensional embeddings.

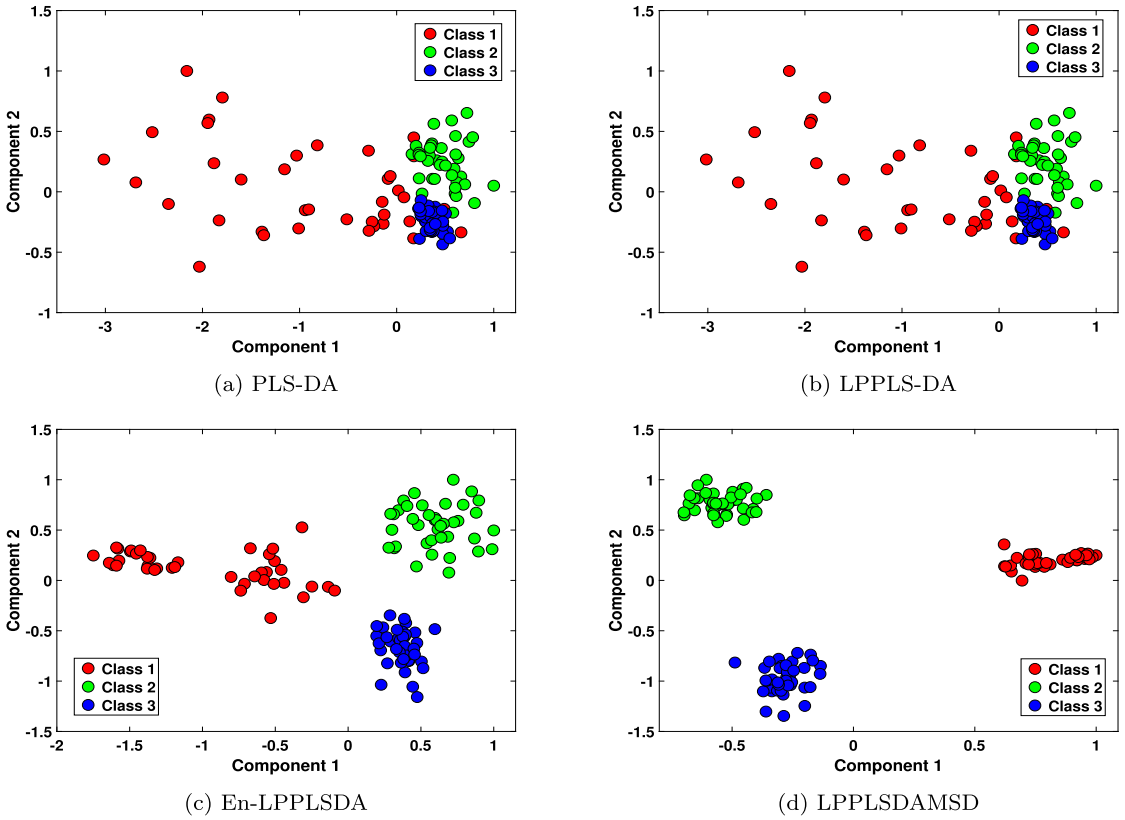


Fig. 6. Ill-conditioned dataset Synthetic III. Two-dimensional projections resulted from (a) PLS-DA, (b) LPPLSDA, (c) En-LPPLSDA, and (d) LPPLSDAMSD.

Table 3
Condition numbers of the matrices associated with LPPLS-DA, En-LPPLSDA and LPPLSDAMSD for real-life datasets.

Datasets	Condition number				
	$\widehat{S}_b$	$\widehat{S}_w$	$\widetilde{S}_w$	$\widetilde{S}_b$	$\widetilde{S}$
Coffee	8.3755×10^{21}	6.0466×10^{19}	1.0000	1.0000	1.0000
Meat	3.9673×10^{23}	1.9382×10^{20}	1.1794	1.0413	1.1931
Oil	3.4168×10^{21}	6.2982×10^{20}	1.0292	1.0206	1.0897
Fruit	1.0386×10^{22}	2.5376×10^{10}	1.001	1.002	1.0117

The objective of this experiment is to compare the distribution of projected training and test samples in terms of class separation and to analyze the effect of conditioning on these distributions. The observations for each dataset are outlined below:

- **Coffee dataset (Figs. 7(a)-(c)):** Training and test samples from the Coffee dataset are projected onto two-dimensional embeddings i.e., dimension reduction from 286 to 2. Results from each method suggest that optimum separation between Class 1 and Class 2 is achievable with only one component. Class separability of projected training samples is significantly better for En-LPPLSDA and LPPLSDAMSD than LPPLS-DA which shows that conditioning of the methods affects projection quality. LPPLS-DA exhibits a slightly different distribution of projected test samples compared to training samples, with less well-defined class separation. On the contrary, En-LPPLSDA and LPPLSDAMSD maintain a distribution of projected test samples similar to that of the training samples, with more distinct class separation.
- **Meat dataset (Figs. 8(a)-(c)):** Training and test samples from the Meat dataset are projected onto two-dimensional embeddings i.e., dimension reduction from 448 to 2. LPPLS-DA performs rather poorly in terms of class separability. A significant overlap is observed between Class 1 and Class 3 samples in the distribution of projected test samples. The numerical instability in LPPLS-DA is quite pronounced, and more severe than the Coffee dataset. En-LPPLSDA and LPPLSDAMSD significantly improve class separability compared to LPPLS-DA. The distribution of projected test samples by these algorithms is also similar to the projected training samples.
- **Oil dataset (Figs. 9(a)-(c)):** Training and test samples are projected onto three-dimensional embeddings i.e., dimension reduction from 570 to 3. This dataset has four classes thus it is expected that more dimension is needed to achieve good class

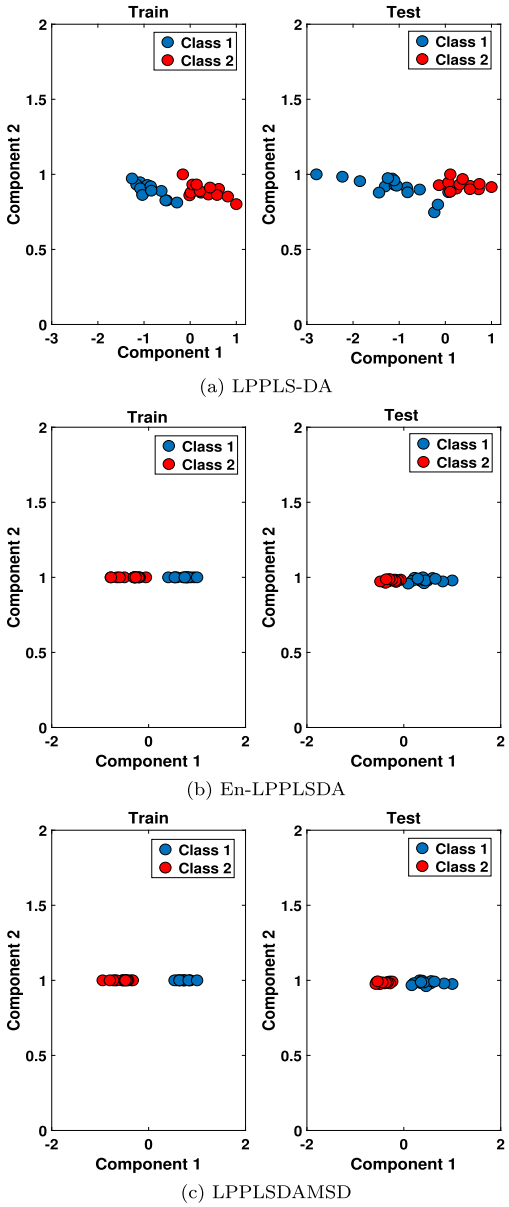
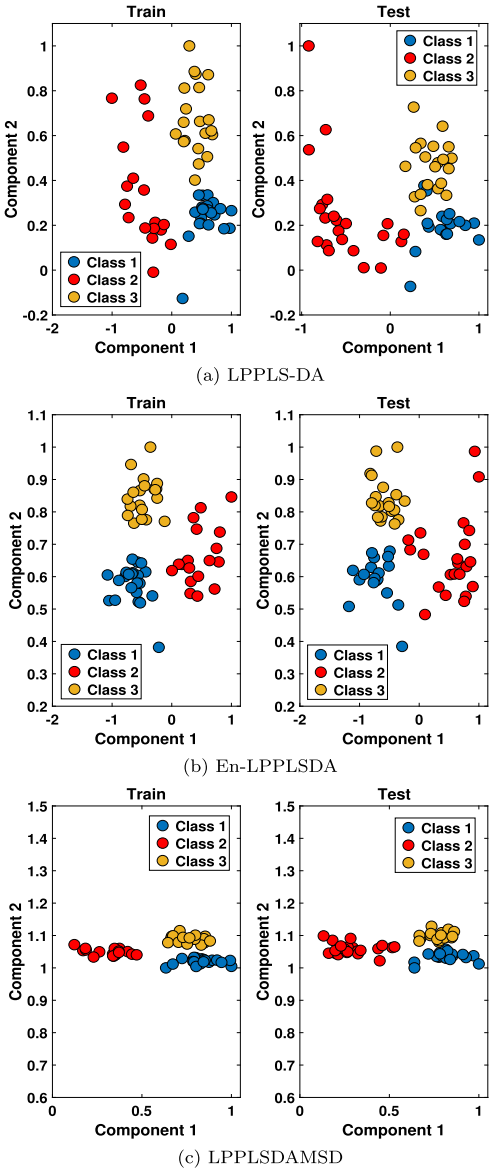


Fig. 7. Visualization of the two-dimensional embedding of the Coffee dataset. Distribution of projected training samples (left) and test samples (right) from (a) LPPLS-DA, (b) En-LPPLSDA, (c) LPPLSDAMSD.

separability. En-LPPLSDA and LPPLSDAMSD, which are numerically stable, achieve better class separability in the training samples compared to the ill-conditioned LPPLS-DA. Class separability of projected test samples, while somewhat inferior to previous datasets, still shows samples from each class projected close to the locality of the training samples.

- **Fruit dataset (Figs. 10(a)-(c)):** Training and test samples are projected onto two-dimensional embeddings i.e., dimension reduction from 235 to 3. Class separation for this dataset is not as good as for the other datasets. However, the plane of separation of Class 1 and Class 2 is quite visible in the En-LPPLSDA and LPPLSDAMSD embeddings for both training and test samples. The En-LPPLSDA and LPPLSDAMSD embeddings reveal that optimal class separation can be achieved with Component 1 alone.

Based on the results in this section, there is substantial evidence that ill-conditioning of the dimensionality reduction method induces numerical instability and can affect the accuracy of classification in the lower-dimensional space. The numerically stable En-LPPLSDA and LPPLSDAMSD produce better class separability of training samples compared to the ill-conditioned LPPLS-DA. Numerical stability results in the distribution structures of projected test samples being similar to those of training samples. Similar distribution structures enable higher classification accuracy in the lower-dimensional space.



(a) LPPLS-DA

(b) En-LPPLSDA

(c) LPPLSDAMSD

Fig. 8. Visualization of the two-dimensional embedding of the Meat dataset. Distribution of projected training samples (left) and test samples (right) from (a) LPPLS-DA, (b) En-LPPLSDA, (c) LPPLSDAMSD.

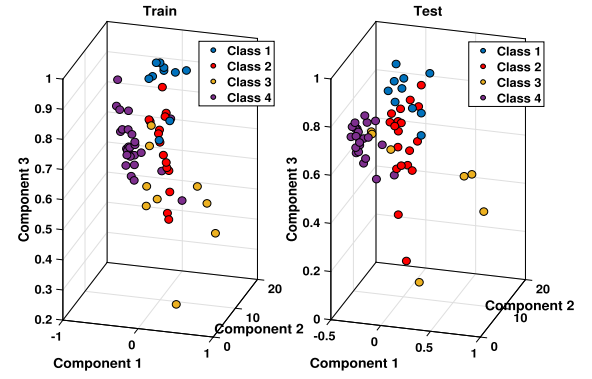
5.3. Dimensionality reduction for classification of real-life datasets

In this section, several numerical experiments are conducted to validate the effect of numerical instability and ill-conditioning of dimensionality reduction algorithms on classification accuracies. A comparison of the proposed LPPLSDAMSD and En-LPPLSDA with state-of-the-art dimension reduction methods is also conducted.

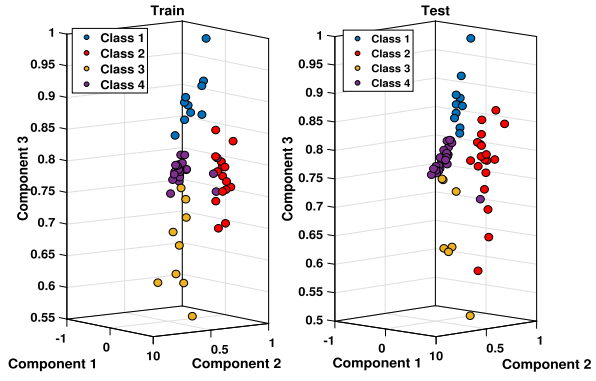
5.3.1. Experimental setup

Classification accuracies are obtained for different values of the number of extracted components/feature d where d also determines the dimension of the low-dimensional embedding in which the classification task is performed. A 10-fold cross-validation approach is applied in each experiment to obtain an average classification accuracy for each method. The cross-validation procedure involves

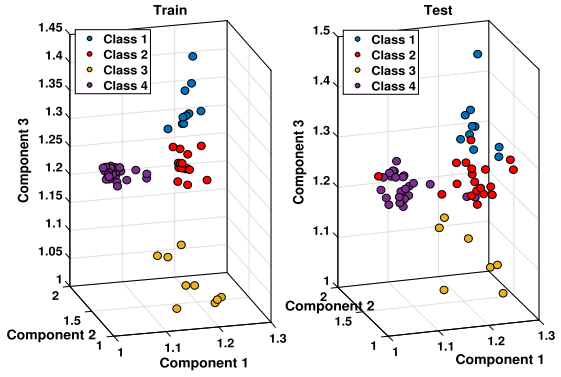
- Random partitioning of datasets into training and test sets.
- The training set is used to compute the projection matrix W.
- Projected training samples are used to build the k-NN classification model.
- The test set is projected onto the low-dimensional embedding using W.



(a) LPPLS-DA



(b) En-LPPLSDA



(c) LPPLSDAMSD

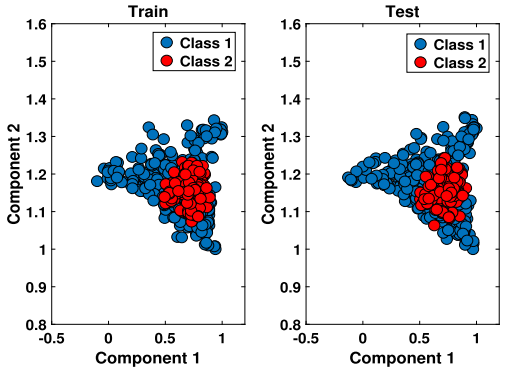
Fig. 9. Visualization of the two-dimensional embedding of the Oil dataset. Distribution of projected training samples (left) and test samples (right) from (a) LPPLS-DA, (b) En-LPPLSDA, (c) LPPLSDAMSD.

- k-NN classification model is applied to classify projected test samples.

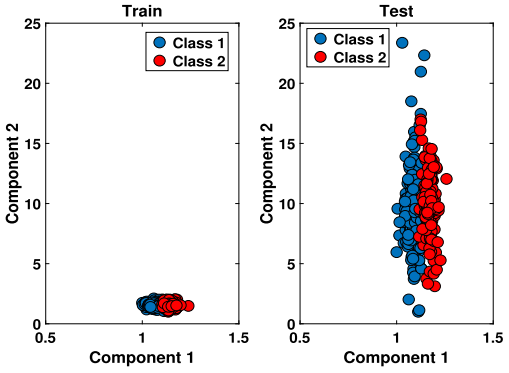
To reduce variability in the results, the random partitioning of datasets into training sets and test sets is repeated 10 times, and the average classification accuracy rates are reported. The basic procedure for dimensionality reduction for classification is shown in the form of a flow chart in Fig. 11.

The regularization parameters in LPPLS-DA, En-LPPLSDA, and LPPLSDAMSD and the scatter-difference parameter in LPPLS-DAMSD are fixed in all experiments in this section. These values are $\sigma_w = \sigma_b = \sigma_\theta = 0.9$ and $\theta = 5$.

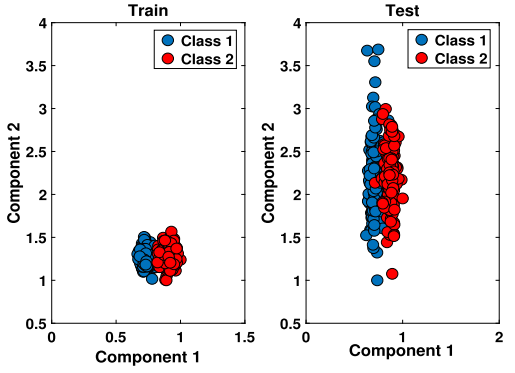
The optimal number of neighbors (i.e. k) to be used in the k-NN classifier could be specific to different data sets and dimensionality reduction methods. Therefore, cross-validation procedures were carried out to investigate the effect of different values k on the maximum classification accuracy achievable by each method. The outcome of these procedures is explained in the following sections.



(a) LPPLS-DA



(b) En-LPPLSDA



(c) LPPLSDAMSD

Fig. 10. Visualization of the two-dimensional embedding of the Fruit dataset. Distribution of projected training samples (left) and test samples (right) from (a) LPPLS-DA, (b) En-LPPLSDA, (c) LPPLSDAMSD.

5.3.2. Classification accuracies: effect of numerical stability and conditioning

In this section the effect of numerical stability and conditioning of dimensionality reduction algorithms on the classification of the four datasets described in Section 5.1 is evaluated. Condition numbers in Table 3 indicate that, LPPLSDAMSD and En-LPPLSDA are expected to be better conditioned and more numerically stable compared to PLS-DA and LPPLS-DA. The classification accuracies of LPPLSDAMSD, En-LPPLSDA, LPPLS-DA, and PLS-DA are analyzed with respect to the dimension of the low-dimensional embedding space (i.e. the number of extracted components, d). Ideally, maximum classification accuracy is achieved for the value of d that corresponds to the true dimensionality of the dataset. The purpose of the analysis is to determine whether En-LPPLSDA and LPPLSDAMSD provide a significant improvement to LPPLS-DA and PLS-DA.

To further highlight the effect of numerical stability and conditioning, a similar comparison is made between LDA and its numerically stable versions, namely the enhanced LDA (En-LDA) and LDA with maximum scatter difference (LDA-MSD) [31]. The En-LDA algorithm is designed similarly to En-LPPLSDA where diagonal shifts σI (σ is the regularization parameter) are added to the matrices S_b and S_w in (2) to improve their conditioning. Table 4 shows the condition numbers of the matrices associated with LDA, En-LDA, and LDA-MSD. Based on these numbers it is expected that En-LDA and LDA-MSD are more numerically stable and better conditioned compared to the conventional LDA.

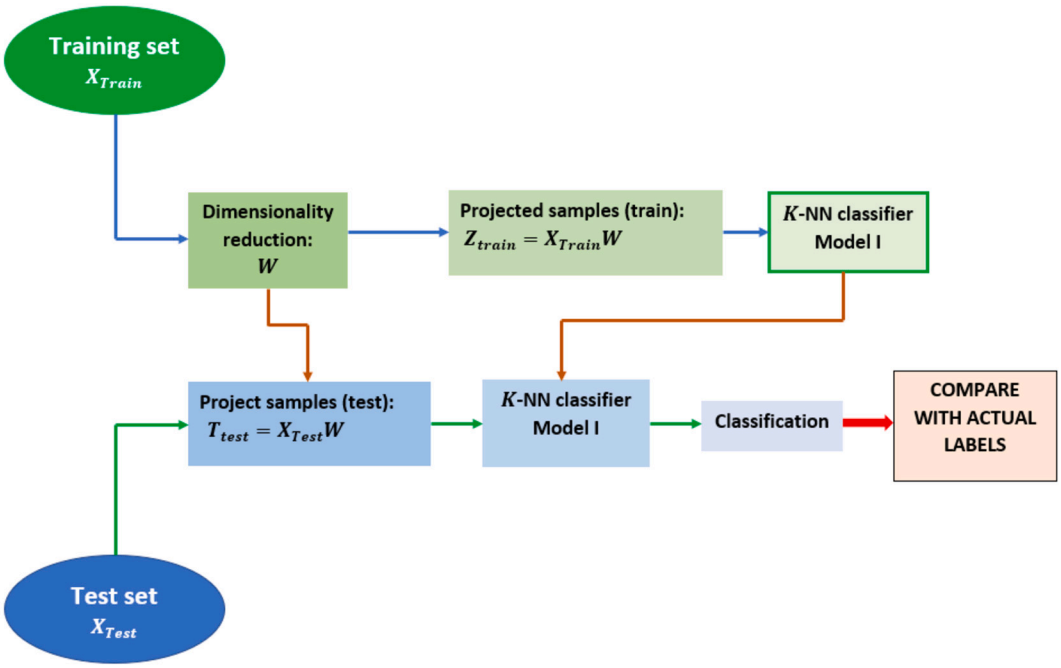


Fig. 11. Dimensionality reduction for classification.

Table 4
Condition number of the matrices associated with LDA, En-LDA and LDA-MSD for real-life datasets.

Datasets	Condition number				
	Between-class (LDA)	Within-class (LDA)	Between-class (En-LDA)	Within-class (En-LDA)	$\widetilde{S}$ (LDA-MSD)
Coffee	1.5978×10^{21}	3.3345×10^{19}	1.3103	1.004	1.3103
Meat	1.1881×10^{21}	1.5864×10^{20}	1.6665	1.000	1.6667
Oil	7.3791×10^{21}	2.2181×10^{20}	1.6666	1.000	1.6667
Fruit	4.8131×10^{23}	2.5343×10^{10}	6.3423	1.0800	10.5210

The outcome of the cross-validation procedure to determine the number of neighbors (i.e., k) to be used in the k-NN classifier in the case of dimension reduction by the PLS-DA family of algorithms is summarized in Figs. 12(a)-(d) for data sets Coffee, Meat, Oil and Fruit respectively. It is revealed that for all the data sets concerned, the best average classification accuracy associated with dimension reduction by the PLS-DA family of algorithms remains steady for small values of k and shows a downward trend as k increases. Based on this observation, in the subsequent experiments, the k-NN classifier with $k = 3$ is used in all the classification tasks involving PLS-DA, LPPLS-DA, En-LPPLSDA, and LPPLSDAMSD. In Figs. 13(a)-(d), the outcome of the cross-validation procedure involving the LDA-based algorithms for the four data sets involved is displayed. For all data sets used, a slight fluctuation in the accuracy is observed in all three LDA-based methods for the smaller k values. All algorithms appear to settle on a maximum at $k = 7$ so this is the value chosen in all the classification tasks involving the LDA, En-LDA and LDA-MSD.

The classification accuracies resulting from the reduction of the dimensionality by all the algorithms mentioned above are shown in Figs. 14(a)-(d). For Coffee and Fruit datasets, binary classification is involved, whereas multiclass classification is involved for Meat and Oil datasets (with 3 and 4 classes, respectively). Observations for each dataset are outlined below:

- **Coffee dataset (Fig. 14(a))** All the PLS-DA-based algorithms achieve accuracies of above 95% with $d \geq 1$. The more stable algorithms (that is, En-LPPLSDA and LPPLSDAMSD) provide slightly better accuracies, where, with $d = 1$, LPPLSDAMSD achieves an accuracy of close to 99% and En-LPPLSDA achieves almost 98%. On the other hand, with $d = 1$, LPPLS-DA achieves an accuracy of slightly less than 97%, and PLS-DA achieves an accuracy of about 96%. For the LDA-based algorithm, the more stable versions show marked improvement to the conventional LDA. It is noted that LDA-MSD achieves comparable performance to both En-LPPLSDA and LPPLSDAMSD for $d \geq 2$, and En-LDA achieves comparable performance to LPPLS-DA also for $d \geq 2$.
- **Meat dataset (Fig. 14(b))** In this three-class classification, the modified algorithms (En-LPPLSDA and LPPLSDAMSD) show significantly better performance compared to their conventional counterparts. All algorithms achieve maximum accuracy with only two extracted components; LPPLSDAMSD achieves an accuracy of over 98%, En-LPPLSDA achieves over 97%, and both LPPLS-DA and PLS-DA achieve an accuracy of only 94%. The LDA-based algorithms achieve maximum accuracy with $d \geq 3$, thus requiring one more component than the PLS-DA counterparts.

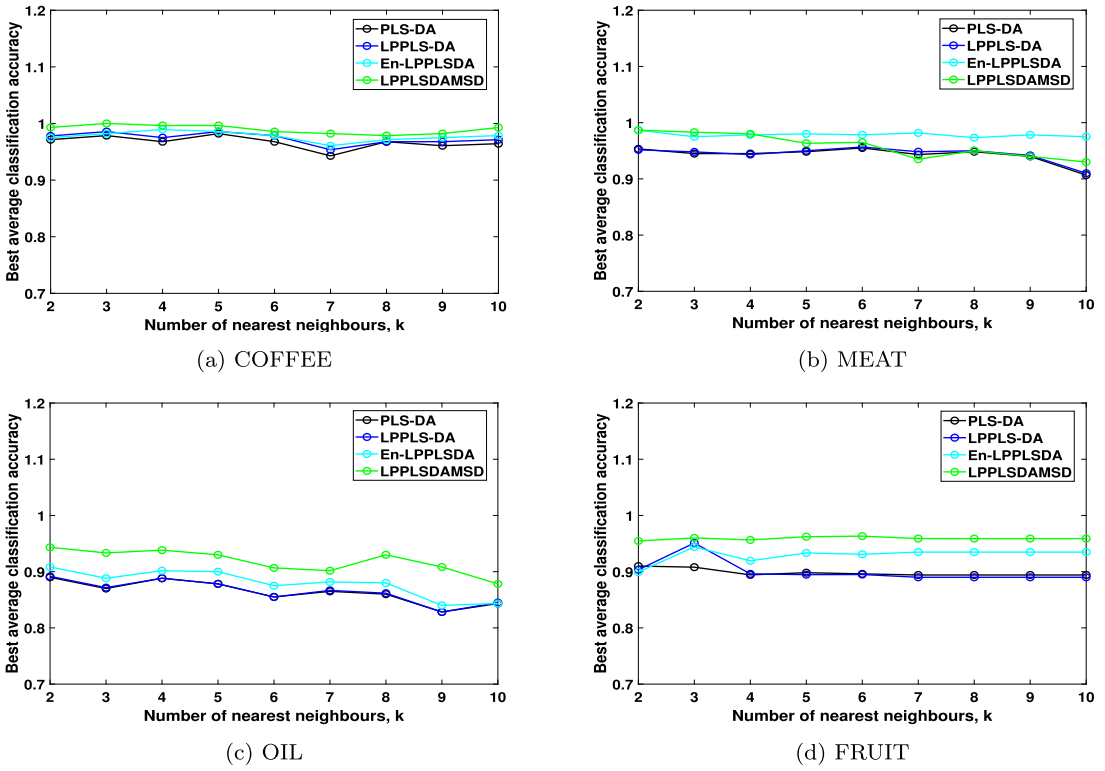


Fig. 12. Best average classification accuracies of PLS-DA, LPPLS-DA, En-LPPLSDA and LPPLSDAMSD for the four datasets using k-NN classifier with $k = 2$ to $k = 10$.

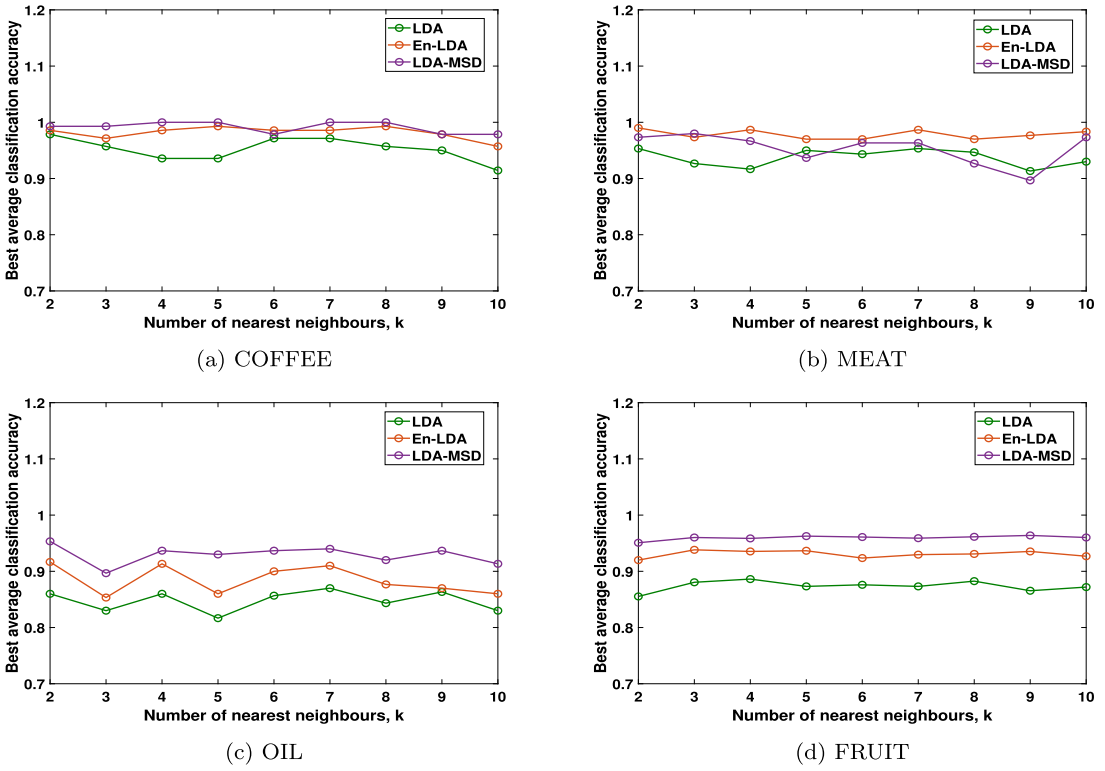


Fig. 13. Best average classification accuracies of LDA, En-LDA and LDA-MSD for the four datasets using k-NN classifier with $k = 2$ to $k = 10$.

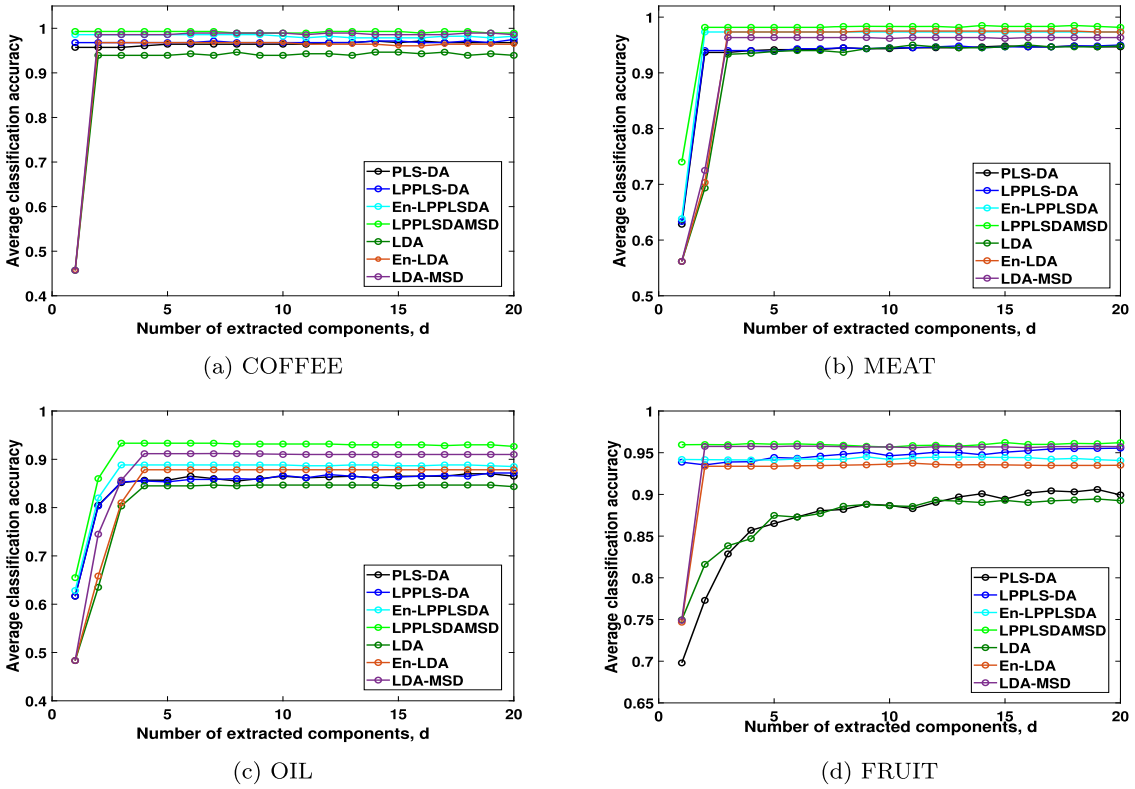


Fig. 14. Average classification accuracies associated with PLS-DA, LPPLS-DA, En-LPPLSDA, LPPLSDAMSD, LDA, En-LDA AND LDA-MSD. Here, half of the datasets are used as training sets, and the remaining half as test sets.

- **Oil dataset (Fig. 14(c))** In this four-class classification, the numerically stable algorithms (En-LPPLSDA, LPPLSDAMSD, En-LDA and LDA-MSD) show significantly better performance compared to their conventional counterparts. The PLS-DA based algorithms achieve maximum accuracy with only three components while the LDA-based methods achieve their maximum accuracy with four components. LPPLSDAMSD is shown to perform way better than En-LPPLSDA for this data set with LPPLSDAMSD achieving an accuracy of 93.33%, En-LPPLSDA achieving 88.83%, and both LPPLS-DA and PLS-DA achieving an accuracy of 85.33% (with slight improvement as d increases).
- **Fruit dataset (Fig. 14(d))** Unlike the first three datasets, the number of features of the Fruit dataset is less than the number of samples, that is, the small sample size problem is not the main issue. This is evident from Table 3 and Table 4 where the condition number of the unconditioned within-class scatter matrix is much lower compared to the other datasets. However, the condition number of the unconditioned between-class scatter matrix is still very high. For this dataset, LPPLSDAMSD, En-LPPLSDA, and LPPLS-DA show somewhat comparable performance with $\geq 94\%$ accuracy achieved with only one component. PLS-DA performs rather poorly for this dataset. The LDA-based algorithms exhibit comparable maximum accuracy to the respective PLS-DA counterparts, however, En-LDA and LDA-MSD are shown to consistently need one extra component than En-LPPLSDA and LPPLSDAMSD to achieve the maximum accuracy.

In summary, the results demonstrate that improving the numerical stability of dimension reduction methods, as seen in En-LPPLSDA and LPPLSDAMSD, leads to enhanced accuracy when compared to their less stable counterparts, LPPLS-DA and PLS-DA. These findings align with the observations in Section 5.2, indicating that numerically stable dimension reduction techniques better retain discriminant information in the low-dimensional embedding, resulting in improved k-NN classification performance. The parallel performance improvements in En-LDA and LDA-MSD compared to LDA emphasize the significance of enhancing the conditioning of scatter matrices. Thus, enhancing the stability and conditioning of both between-class and within-class scatter matrices represents a promising avenue for improving the performance of LPPLS-DA and LDA.

Next, the focus is on evaluating the performance of the algorithms with varying sizes of the training set. Three scenarios are examined, each corresponding to different fractions of the total sample size: one-third, half, and two-thirds. The results, as summarized in Table 5, highlight that En-LPPLSDA and LPPLSDAMSD consistently achieve higher average classification accuracy rates compared to PLS-DA and LPPLS-DA across all datasets. Moreover, both En-LPPLSDA and LPPLSDAMSD exhibit a gradual improvement in performance as the size of the training sets increases, progressing from one-third to half and then two-thirds of the overall sample size. These findings underline the superiority of LPPLSDAMSD in terms of both accuracy and the number of components required to attain

Table 5
Maximum average classification accuracy for PLS-DA, LPPLS-DA, En-LPPLSDA and LPPLSDAMSD for different sizes of training set.

	Accuracy (1/3 Train)	Number of extracted components	Accuracy (1/2 Train)	Number of extracted components	Accuracy (2/3 Train)	Number of extracted components
Coffee dataset:						
PLS-DA	0.937	16	0.979	18	0.990	2
LPPLS-DA	0.950	20	0.986	12	1.000	14
En-LPPLSDA	0.961	1	0.985	1	1.000	11
LPPLSDAMSD	**0.984**	**1**	**0.993**	**1**	**1.0000**	**1**
Meat dataset						
PLS-DA	0.908	16	0.945	16	0.973	10
LPPLS-DA	0.908	16	0.948	20	0.980	20
En-LPPLSDA	0.9350	2	0.973	2	0.983	2
LPPLSDAMSD	**0.943**	**2**	**0.982**	**2**	**0.995**	**2**
Oil dataset						
PLS-DA	0.849	19	0.870	19	0.883	20
LPPLS-DA	0.849	13	0.872	19	0.880	10
En-LPPLSDA	0.8513	3	0.888	3	0.900	3
LPPLSDAMSD	**0.901**	**3**	**0.933**	**3**	**0.945**	**3**
Fruit dataset						
PLS-DA	0.883	20	0.906	19	0.919	19
LPPLS-DA	0.943	20	0.955	20	0.956	10
En-LPPLSDA	0.931	3	0.942	1	0.943	1
LPPLSDAMSD	**0.955**	**1**	**0.960**	**1**	**0.966**	**1**

it. Notably, the performance of LPPLSDAMSD slightly outpaces that of En-LPPLSDA, emphasizing the overall effectiveness of the maximum scatter-difference criterion in enhancing the numerical stability of LPPLS-DA.

5.3.3. Comparison with state-of-the-art dimensionality reduction methods

In this section, the comparative performance of LPPLSDAMSD and En-LPPLSDA is assessed against five state-of-the-art dimensionality reduction methods, namely RLDA, SDFS, $l_{2,1}$-LDA, 3E-LDA and RSLDA. The parameters for each method are set to the suggested values provided by the respective authors. For the iterative optimization methods, a termination condition of up to 100 iterations is applied. A cross-validation procedure is utilized to determine the optimal number of nearest neighbors for the k-NN classifier. The results, as shown in Figs. 15(a)-(d), demonstrate that for all four data sets, the maximum classification accuracy is achieved by the k-NN classifier with $k = 3$. Therefore, the subsequent classification task in this section is conducted with $k = 3$.

Figs. 16(a)-(d) presents a comparison of the average classification accuracies resulting from various dimensionality reduction methods. In general, LPPLSDAMSD and En-LPPLSDA outperform other methods across all datasets. Notably, as shown in Fig. 16(a), $l_{2,1}$-LDA achieves comparable accuracy with En-LPPLSDA and LPPLSDAMSD for the Coffee dataset, but it requires at least eight components to do so, while En-LPPLSDA and LPPLSDAMSD only need one component. Similarly, for the Meat dataset in Fig. 16(b), RLDA and RSLDA reach maximum accuracy comparable to En-LPPLSDA and LPPLSDAMSD but require six and three components, respectively, while En-LPPLSDA and LPPLSDAMSD achieve it with only two components. For the Oil and Fruit datasets (Figs. 16(c) and (d) respectively), LPPLSDAMSD excels in dimension reduction performance, achieving the best classification accuracy with very few components compared to other methods.

Table 6 provides an overview of the best maximum accuracies achieved by various dimensionality reduction methods for different sizes of training sets (one-third, half, and two-thirds of the total sample size). It also includes the number of components required to attain these accuracies. The results highlight the consistency and superiority of En-LPPLSDA and LPPLSDAMSD across all datasets and training set sizes. These methods consistently achieve higher accuracies than other methods and do so with the least number of components. The effectiveness of dimension reduction by En-LPPLSDA and LPPLSDAMSD is evident, as they require only a few components to capture most of the discriminant information in each dataset. This demonstrates the robustness and efficiency of these methods in various scenarios.

Figs. 17(a)-(d) provides a comparison of the computational cost of each method, measured by their average runtime. The average runtime of En-LPPLSDA and LPPLSDAMSD is similar to that of 3E-LDA and RLDA, and significantly less than that of SDFS, RSLDA, and $l_{2,1}$-LDA. For the iterative optimization methods, increasing the number of iterations allows the methods to provide better accuracy and require less number of components. However, it comes at the cost of increased computational time.

6. Conclusions

Classification of high-dimensional datasets, often employing classifiers like the k-NN classifier, encounters challenges due to the presence of redundant features. Dimensionality reduction is a critical preprocessing step aimed at enhancing the effectiveness of

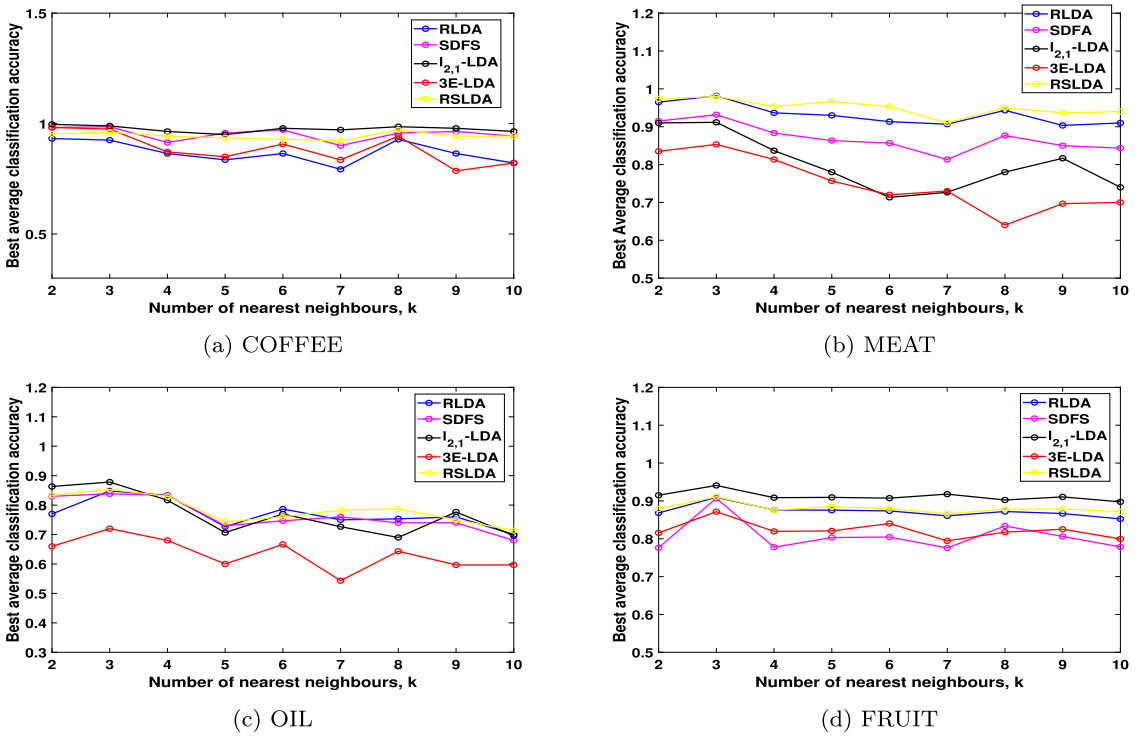


Fig. 15. Best average classification accuracies of RLDA, SDFS, $l_{2,1}$-LDA, 3E-LDA and RSLDA for the four datasets using k-NN classifier with $k = 2$ to $k = 10$.

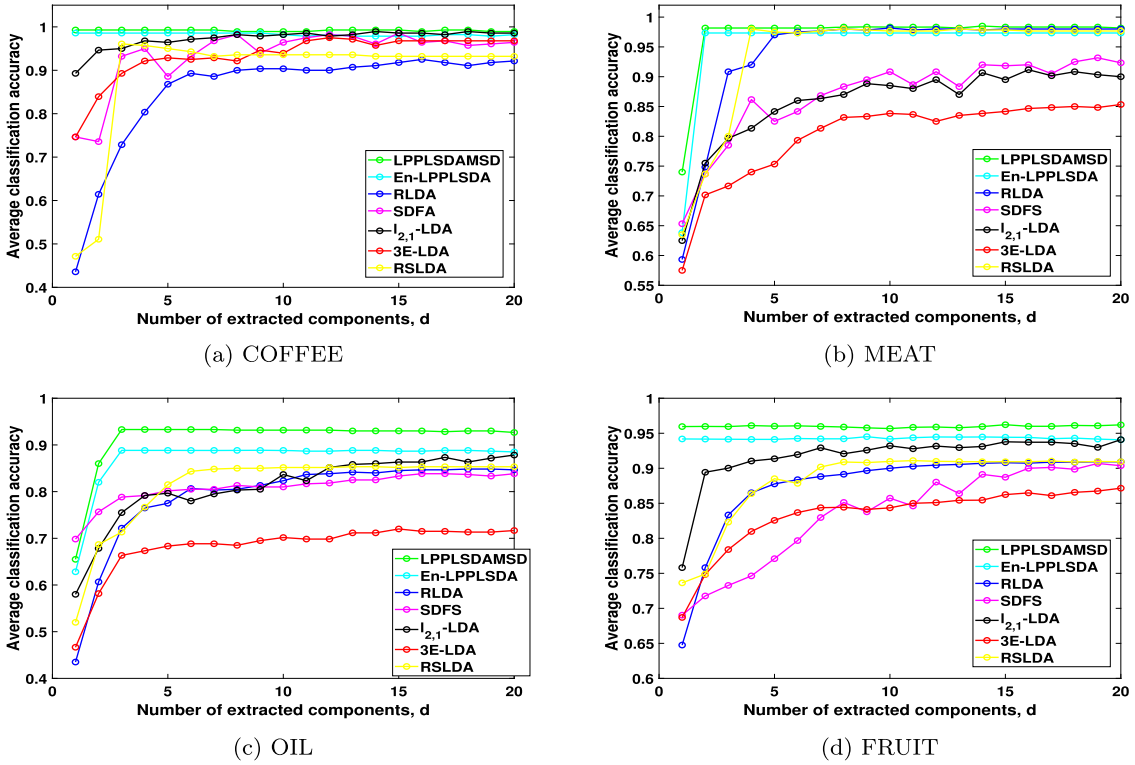


Fig. 16. Comparison of average classification accuracies of LPPLSDAMSD and En-LPPLSDA with state-of-the-art dimension reduction methods. Here, half of the datasets are used as training sets, and the remaining half as test sets.

Table 6
Maximum average classification accuracy for state-of-the-art dimension reduction methods for different sizes of training set.

	Accuracy (1/3 Train)	Number of extracted components	Accuracy (1/2 Train)	Number of extracted components	Accuracy (2/3 Train)	Number of extracted components
Coffee dataset						
RLDA	0.903	14	0.925	16	0.947	19
SDFS	0.968	13	0.986	15	1.000	18
$l_{2,1}$-LDA	0.979	18	0.989	14	1.000	6
3E-LDA	0.982	19	0.975	12	0.700	2
RSLDA	0.934	4	0.961	3	0.968	3
En-LPPLSDA	0.961	1	0.985	1	1.000	11
LPPLSDAMSD	**0.984**	**1**	**0.993**	**1**	**1.000**	**1**
Meat dataset						
RLDA	0.956	17	0.982	10	0.970	18
SDFS	0.891	15	0.932	19	0.918	16
$l_{2,1}$-LDA	0.890	15	0.912	16	0.928	16
3E-LDA	0.866	20	0.853	20	0.902	17
RSLDA	0.964	6	0.980	4	0.965	7
En-LPPLSDA	0.935	2	0.973	2	0.983	2
LPPLSDAMSD	**0.943**	**2**	**0.982**	**2**	**0.995**	**2**
Oil dataset						
RLDA	0.814	20	0.848	18	0.855	19
SDFS	0.814	20	0.838	16	0.842	18
$l_{2,1}$-LDA	0.828	18	0.878	20	0.900	16
3E-LDA	0.659	20	0.720	15	0.585	4
RSLDA	0.822	8	0.853	13	0.870	8
En-LPPLSDA	0.851	3	0.888	3	0.900	3
LPPLSDAMSD	**0.901**	**3**	**0.933**	**3**	**0.945**	**3**
Fruit dataset						
RLDA	0.904	20	0.909	18	0.930	20
SDFS	0.894	20	0.907	19	0.921	19
$l_{2,1}$-LDA	0.932	19	0.941	20	0.946	18
3E-LDA	0.873	20	0.872	20	0.902	19
RSLDA	0.905	12	0.911	11	0.925	13
En-LPPLSDA	0.931	3	0.942	1	0.943	1
LPPLSDAMSD	**0.955**	**1**	**0.960**	**1**	**0.966**	**1**

classification methods for such datasets. A proficient dimensionality reduction technique should be capable of capturing the most discriminative features of the data within a low-dimensional subspace designed for classification tasks. LPPLS-DA method stands as a promising candidate for this purpose. However, the numerical characteristics of the eigenvalue problem associated with LPPLS-DA can occasionally lead to suboptimal performance.

In this paper, we introduce two techniques designed to ameliorate the numerical properties of LPPLS-DA. The first technique, termed En-LPPLSDA, brings stability to both the PLS-DA between-class scatter matrix and the locality-preserving within-class scatter matrix. The second approach reformulates the multi-objective optimization of LPPLS-DA using the maximum scatter-difference criterion, yielding the LPPLSDAMSD method.

We analyze the numerical properties of LPPLS-DA, En-LPPLSDA, and LPPLSDAMSD using specially crafted synthetic datasets, emphasizing the impact of ill-conditioning and numerical instabilities on dimension reduction and discrimination. En-LPPLSDA and LPPLSDAMSD outperform LPPLS-DA and PLS-DA in terms of optimal discrimination among various classes in the datasets, while preserving essential class information.

Furthermore, classification experiments are carried out on four spectral datasets, demonstrating that the numerical properties of dimensionality reduction methods have a direct bearing on the accuracy of classification in the lower-dimensional space. Numerically stable dimension reduction yields two favorable outcomes: improved class separability and a more robust projection of test samples into the lower-dimensional embedding, leading to enhanced classification accuracy.

Based on our analysis, the maximum scatter difference approach emerges as the more effective means to enhance numerical stability. Notably, our classification experiments with spectral datasets reveal that En-LPPLSDA and LPPLSDAMSD achieve peak performance with only a few components, surpassing other state-of-the-art dimension reduction methods. These results provide compelling evidence of the capabilities of En-LPPLSDA and LPPLSDAMSD in capturing crucial discriminative features within spectral data in a lower-dimensional space, thereby paving the way for more efficient dataset classification.

It is noteworthy that the effectiveness of En-LPPLSDA relies on the choice of the regularization parameters, namely σ_w and σ_b, while the performance of LPPLSDAMSD depends on the scatter-difference parameter θ and the regularization parameter σ_θ. Empirical analysis conducted on the datasets employed in this study proved sufficient for determining optimal values of these parameters,

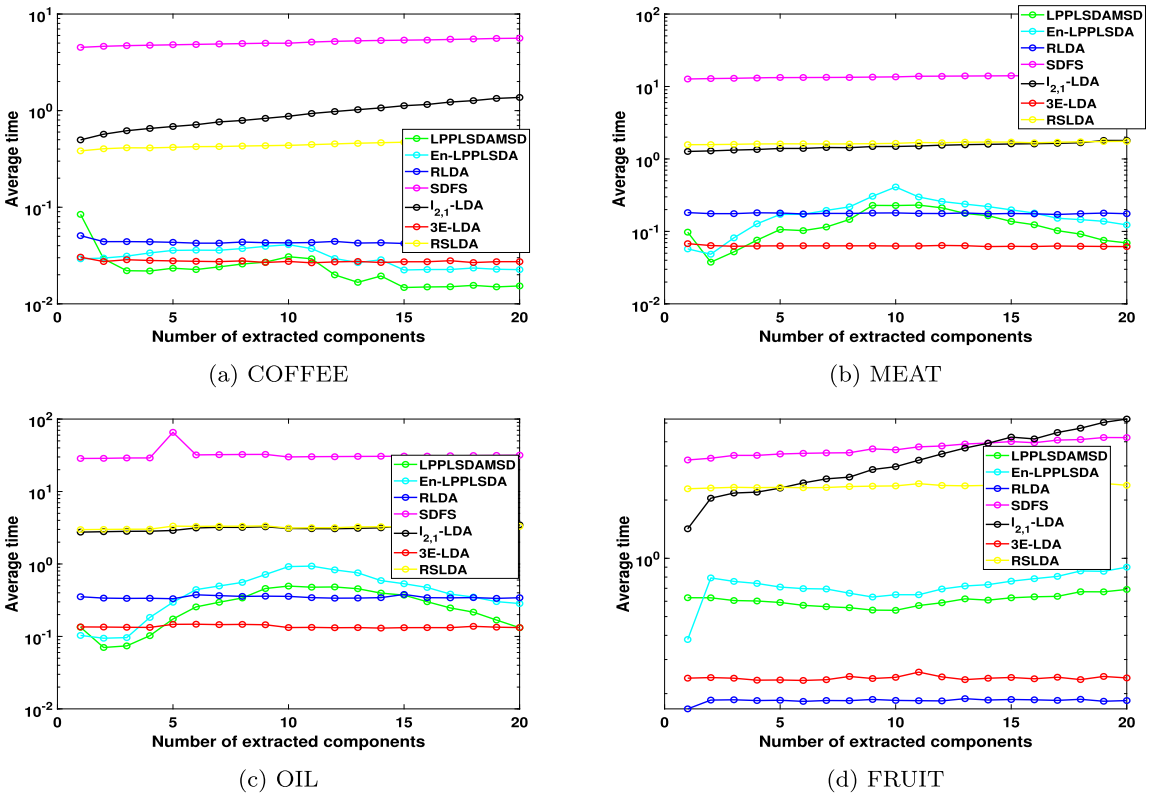


Fig. 17. Comparison of average runtime of LPPLSDAMSD and En-LPPLSDA with state-of-the-art dimension reduction methods.

leading to the highest average classification accuracy. Interestingly, it was observed that the optimal values for each dataset exhibit minimal variance. Nevertheless, it is imperative to highlight that these observations may not universally apply to other datasets. To ensure optimal performance, a comprehensive assessment of these parameters is essential. A more in-depth mathematical analysis is warranted to ascertain the optimal values of regularization parameters from training data. This aspect will be investigated in future studies.

CRediT authorship contribution statement

Noor Atinah Ahmad: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Noor Atinah Ahmad reports financial support was provided by The Ministry of Higher Education, Malaysia.

Data availability

Data used in this research is publicly available on https://csr.quadram.ac.uk/example-datasets-for-download/.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work the author(s) used CHATGPT in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgement

The author sincerely thanks all valuable comments and suggestions from anonymous reviewers and editors, which helped to improve the quality of the manuscript. This work was supported by the Ministry of Higher Education, Malaysia Fundamental Research Grant Scheme with Project Code: FRGS/1/2020/ICT06/USM/02/1.

References

[1] W. Sun, K. Liu, G. Ren, W. Liu, G. Yang, X. Meng, J. Peng, A simple and effective spectral-spatial method for mapping large-scale coastal wetlands using China zy1-02d satellite hyperspectral images, Int. J. Appl. Earth Obs. Geoinf. 104 (2021) 102572, https://doi.org/10.1016/j.jag.2021.102572.

[2] S. Deegalla, H. Boström, Classification of microarrays with knn: comparison of dimensionality reduction methods, in: H. Yin, P. Tino, E. Corchado, W. Byrne, X. Yao (Eds.), Intelligent Data Engineering and Automated Learning - IDEAL 2007, Springer, Berlin, Heidelberg, 2007, pp. 800–809.

[3] A.-L. Boulesteix, Pls dimension reduction for classification with microarray data, Stat. Appl. Genet. Mol. Biol. 3 (2004), https://doi.org/10.2202/1544-6115.1075.

[4] R. Huang, Q. Liu, H. Lu, S. Ma, Solving the small sample size problem of lda, in: 2002 International Conference on Pattern Recognition, vol. 3, 2002, pp. 29–32.

[5] W. Wong, H. Zhao, Supervised optimal locality preserving projection, Pattern Recognit. 45 (2012) 186–197, https://doi.org/10.1016/j.patcog.2011.05.014.

[6] F. Nie, Z. Wang, R. Wang, Z. Wang, X. Li, Towards robust discriminative projections learning via non-greedy $\ell_{2,1}$ -norm minmax, IEEE Trans. Pattern Anal. Mach. Intell. 43 (2021) 2086–2100, https://doi.org/10.1109/TPAMI.2019.2961877.

[7] H. Zhao, Z. Wang, F. Nie, A new formulation of linear discriminant analysis for robust dimensionality reduction, IEEE Trans. Knowl. Data Eng. 31 (2019) 629–640, https://doi.org/10.1109/TKDE.2018.2842023.

[8] Z. Wang, F. Nie, L. Tian, R. Wang, X. Li, Discriminative feature selection via a structured sparse subspace learning module, in: IJCAI, 2020, pp. 3009–3015.

[9] Y. Li, B. Liu, Y. Yu, H. Li, J. Sun, J. Cui, 3e-lda: three enhancements to linear discriminant analysis, ACM Trans. Knowl. Discov. Data 15 (2021), https://doi.org/10.1145/3442347.

[10] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, Y. Xu, Robust sparse linear discriminant analysis, IEEE Trans. Circuits Syst. Video Technol. 29 (2019) 390–403, https://doi.org/10.1109/TCSVT.2018.2799214.

[11] J. Park, S. Kumar, S. Han, V. Singh, S. Nam, Y. Lee, Two-step partial least squares-discriminant analysis modeling for accurate classification of edible sea salt products using laser-induced breakdown spectroscopy, Appl. Spectrosc. 76 (2022) 1042–1050, https://doi.org/10.1177/00037028221091581.

[12] B. Duarte, R. Mamede, J. Carreiras, I.A. Duarte, I. Caçador, P. Reis-Santos, R.P. Vasconcelos, C. Gameiro, P. Ré, S.E. Tanner, V.F. Fonseca, Harnessing the full power of chemometric-based analysis of total reflection X-ray fluorescence spectral data to boost the identification of seafood provenance and fishing areas, Foods 11 (2022), https://doi.org/10.3390/foods11172699.

[13] D. Chung, S. Keles, Sparse partial least squares classification for high dimensional data, Stat. Appl. Genet. Mol. Biol. 9 (2010), https://doi.org/10.2202/1544-6115.1492.

[14] J. Kuligowski, G. Quintás, C. Herwig, B. Lendl, A rapid method for the differentiation of yeast cells grown under carbon and nitrogen-limited conditions by means of partial least squares discriminant analysis employing infrared micro-spectroscopic data of entire yeast cells, Talanta 99 (2012) 566–573, https://doi.org/10.1016/j.talanta.2012.06.036.

[15] W. Song, H. Wang, P. Maguire, O. Nibouche, Nearest clusters based partial least squares discriminant analysis for the classification of spectral data, Anal. Chim. Acta 1009 (2018) 27–38, https://doi.org/10.1016/j.aca.2018.01.023.

[16] F. Tsopelas, D. Konstantopoulos, A.T. Kakoulidou, Voltammetric fingerprinting of oils and its combination with chemometrics for the detection of extra virgin olive oil adulteration, Anal. Chim. Acta 1015 (2018) 8–19, https://doi.org/10.1016/j.aca.2018.02.042.

[17] D.V. Nguyen, D.M. Rocke, Partial least squares discriminant analysis: taking the magic away, Bioinformatics 18 (2002) 39–50, https://doi.org/10.1002/cem.2609.

[18] M. Aminu, N.A. Ahmad, New variants of global-local partial least squares discriminant analysis for appearance-based face recognition, IEEE Access 8 (2020) 166703–166720, https://doi.org/10.1109/ACCESS.2020.3022784.

[19] M. Aminu, N.A. Ahmad, Locality preserving partial least squares discriminant analysis for face recognition, J. King Saud Univ, Comput. Inf. Sci. 34 (2022) 153–164, https://doi.org/10.1016/j.jksuci.2019.10.007.

[20] T. Luo, C. Hou, F. Nie, D. Yi, Dimension reduction for non-gaussian data by adaptive discriminative analysis, IEEE Trans. Cybern. 49 (2019) 933–946, https://doi.org/10.1109/TCYB.2018.2789524.

[21] J. Zhu, J. Chen, B. Xu, H. Yang, F. Nie, Fast orthogonal locality-preserving projections for unsupervised feature selection, Neurocomputing 531 (2023) 100–113, https://doi.org/10.1016/j.neucom.2023.02.021.

[22] A. Martinez, A. Kak, Pca versus lda, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 228–233, https://doi.org/10.1109/34.908974.

[23] M. Aminu, N.A. Ahmad, Complex chemical data classification and discrimination using locality preserving partial least squares discriminant analysis, ACS Omega 5 (2020) 26601–26610, https://doi.org/10.1021/acsomega.0c03362.

[24] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, translation editor, Fritz John in Andrey N. Tikhonov, Vasiliy Y. Arsenin (Eds.), Scripta Series in Mathematics, 1977, Winston; New York: distributed solely by Halsted Press, Washington.

[25] W.-K. Ching, D. Chu, L.-Z. Liao, X. Wang, Regularized orthogonal linear discriminant analysis, Pattern Recognit. 45 (2012) 2719–2732, https://doi.org/10.1016/j.patcog.2012.01.007.

[26] X. Zhang, L. Cheng, D. Chu, L.-Z. Liao, M.K. Ng, R.C.E. Tan, Incremental regularized least squares for dimensionality reduction of large-scale data, SIAM J. Sci. Comput. 38 (2016) B414–B439, https://doi.org/10.1137/15M1035653.

[27] D. Chu, S.T. Goh, A new and fast orthogonal linear discriminant analysis on undersampled problems, SIAM J. Sci. Comput. 32 (2010) 2274–2297, https://doi.org/10.1137/090766772.

[28] G. Fix, R. Heiberger, An algorithm for the ill-conditioned generalized eigenvalue problem, SIAM J. Numer. Anal. 9 (1972) 78–88, https://doi.org/10.1137/0709009.

[29] M.R. Guarracino, C. Cifarelli, O. Seref, P.M. Pardalos, A classification method based on generalized eigenvalue problems, Optim. Methods Softw. 22 (2007) 73–81, https://doi.org/10.1080/10556780600883874.

[30] M. Fordellone, A. Bellincontro, F. Mencarelli, Partial least squares discriminant analysis: a dimensionality reduction method to classify hyperspectral data, Stat. Appl. Italian J. Appl. Stat. 31 (2020) 181–200, https://doi.org/10.26398/IJAS.0031-010.

[31] F. Song, D. Zhang, D. Mei, Z. Guo, A multiple maximum scatter difference discriminant criterion for facial feature extraction, IEEE Trans. Syst. Man Cybern., Part B, Cybern. 37 (2007) 1599–1606, https://doi.org/10.1109/TSMCB.2007.906579.

[32] K. Fukunaga, Introduction to Statistical Pattern Recognition, second edition, Academic Press, 1990.

[33] S. Prasad, L.M. Bruce, Overcoming the small sample size problem in hyperspectral classification and detection tasks, in: IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, vol. 5, 2008, pp. V–381–V–384.

[34] H. Li, K. Zhang, T. Jiang, Robust and accurate cancer classification with gene expression profiling, in: 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05), 2005, pp. 310–321.

[35] L.-F. Chen, H.-Y.M. Liao, M.-T. Ko, J.-C. Lin, G.-J. Yu, A new lda-based face recognition system which can solve the small sample size problem, Pattern Recognit. 33 (2000) 1713–1726, https://doi.org/10.1016/S0031-3203(99)00139-9.

[36] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (2003) 1373–1396, https://doi.org/10.1162/089976603321780317.

[37] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., The Johns Hopkins University Press, 1996.

[38] H. Li, T. Jiang, K. Zhang, Efficient and robust feature extraction by maximum margin criterion, IEEE Trans. Neural Netw. 17 (2006) 157–165, https://doi.org/10.1109/TNN.2005.860852.

[39] G. Downey, R. Briandet, R.H. Wilson, E.K. Kemsley, Near-and mid-infrared spectroscopies in food authentication: coffee varietal identification, J. Agric. Food Chem. 45 (1997) 4357–4361, https://doi.org/10.1021/jf970337t.
[40] W. Zheng, H. Shu, H. Tang, H. Zhang, Spectra data classification with kernel extreme learning machine, Chemom. Intell. Lab. Syst. 192 (2019) 103815, https://doi.org/10.1016/j.chemolab.2019.103815.
[41] O. Al-Jowder, E. Kemsley, R. Wilson, Mid-infrared spectroscopy and authenticity problems in selected meats: a feasibility study, Food Chem. 59 (1997) 195–201, https://doi.org/10.1016/S0308-8146(96)00289-0.
[42] H.S. Tapp, M. Defernez, E.K. Kemsley, Ftir spectroscopy and multivariate analysis can distinguish the geographic origin of extra virgin olive oils, J. Agric. Food Chem. 51 (2003) 6110–6115, https://doi.org/10.1021/jf030232s.
[43] J. Holland, E. Kemsley, R. Wilson, Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purees, J. Sci. Food Agric. 76 (1998) 263–269, https://doi.org/10.1002/(SICI)1097-0010(199802)76:2<263::AID-JSFA943>3.0.CO;2-F.