Journal of Mathematical Imaging and Vision (2021) 63:541–554
https://doi.org/10.1007/s10851-020-01008-w

A Novel Discriminant Locality Preserving Projections Method

Ruisheng Ran[1] · Yinshan Ren[2] · Shougui Zhang[3] · Bin Fang[4]

Received: 6 December 2019 / Accepted: 28 November 2020 / Published online: 5 February 2021

Abstract

Locality preserving projections (LPP) is a popular unsupervised dimensionality reduction method based on manifold learning. As a supervised version of the LPP method, discriminant locality preserving projections (DLPP) method has been recently proposed and paid much attention to by researchers. However, the DLPP method has the small-sample-size (SSS) problem. In this paper, in the view of the eigenvalues of scattering matrices of DLPP, they are first mapped to the new values by two polynomial functions, and with the properties of the matrix function of the two polynomial functions, the criterion of the DLPP method is reconstructed; thus, a novel dimensionality reduction method, named polynomial discriminant locality preserving projections (PDLPP) method, is proposed. The proposed PDLPP method has two advantages: one is that it addresses the SSS problem of DLPP, and the other is that, with the nonlinear mapping implied by PDLPP, the distance between inter-class samples is much enlarged and then the better performance of pattern classification is achieved. The experiments are conducted on the COIL-20 database, ORL, Georgia Tech, and AR face datasets, and the results show that the PDLPP is superior to state-of-the-art methods.

Keywords Manifold learning · Dimensionality reduction · Discriminant locality preserving projections · The small-sample-size problem · Matrix function

1 Introduction

Dimensionality reduction has always been one of the research hot spots in the field of machine learning and computer vision. Recently, many nonlinear dimensionality reduction techniques based on manifold learning have been proposed. The well-known methods are local linear embedding (LLE) [1], ISOMAP [2], Laplacian feature map (LE) [3], local linear spatial alignment (LTSA) [4]. These dimensionality reduction methods map the original dataset to the low-dimensional dataset and keep the intrinsic geometry structure of the dataset as far as possible in the low-dimensional space. However, the maps are only defined on the training datasets, and so they are not suitable for pattern classification, for example, face recognition. As a result, many linear methods are developed. Locality preserving projections (LPP) [5] is a representative one, which is the linear version of LE. From the idea of LPP, Laplacianface was proposed for face recognition [6].

When the class label information of samples is considered in the training stage, the discriminant locality preserving projection (DLPP) method is proposed [7–10]. DLPP is more effective than LPP, for example, it achieves better classification performance in the face recognition task. However, in real life, the number of samples is usually much smaller than the dimensionality of the samples, which results in the fact that the scatter matrices of DLPP may be singular and it is

✉ Shougui Zhang
shgzhang@cqnu.edu.cn

Ruisheng Ran
rshran@cqnu.edu.cn

Yinshan Ren
viprys@cqnu.edu.cn

Bin Fang
fb@cqu.edu.cn

1 College of Computer and Information Science, College of Intelligent Science, Chongqing Engineering Research Center of Educational Big Data Intelligent Perception and Application, Chongqing Normal University, Chongqing 401331, China

2 School of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China

3 School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China

4 College of computer science, Chongqing University, Chongqing 400044, China

difficult to solve the generalized eigenvalue problem. This is the small-sample-size (SSS) problem; the DLPP method has such difficulty [11–14].

The common method to address the SSS problem is that it firstly reduces the dimensionality of original samples using the principal component analysis (PCA) method [15]; thus, the scatter matrices of DLPP become non-singular, and then, DLPP is used to extract features in such reduced dimension space. However, when the PCA method is used to reduce the dimension, some important information is also lost at the same time.

To address the SSS problem, Yang et al. [11] proposed the null space discriminant locality preserving projections (NDLPP) method, which solved the eigenvalue problem in null space and then the SSS problem has been avoided. The discriminant locality preserving projections based on maximum margin criterion (DLPP/MMC) [12] was also a method to solve the SSS problem of DLPP. It constructed the objective function with the maximum margin criterion (MMC), i.e., maximizing the difference between inter-class scatter and intra-class scatter. The DLPP/MMC method does need to calculate the inverse of the scatter matrix, thus naturally avoiding the SSS problem of the DLPP method.

Recently, the exponential DLPP (EDLPP) [13,14] based on the matrix exponential is proposed to solve the SSS problem DLPP. It replaced the scatter matrix with the corresponding matrix exponential, therefore avoiding the singularity of the original scatter matrix. The main idea of EDLPP is also used to solve the SSS problem of some dimensionality reduction methods in the manifold learning field, such as the exponential neighborhood preserving embedding (ENPE) [16], the exponential LPP (ELPP) [17], the exponential local discriminant embedding (ELDE) [18] and the exponential semi-supervised discriminant embedding (ESDE) [19]. In [20], a general exponential framework has been presented, and in this framework, the exponential marginal Fisher analysis (EMFA) and the exponential unsupervised discriminant projections (EUDP) are proposed. These methods have all solved the SSS problems of the corresponding methods, and better performances have been gotten.

In this paper, a novel discriminant locality preserving projections method is proposed. The eigenvalues of the scatter matrix play an important role: One is that the eigenvalues of the scatter matrix must be some zero values when the SSS problem of the DLPP method occurs, and the other is that the eigenvalues of the inter-class scatter matrix and the intra-class scatter matrix measure the distance of the inter-class and intra-class samples, respectively. It is well-known that any function can be approximated by a polynomial function. According to this, in view of the eigenvalues of the scatter matrix, we use two polynomial functions to map the eigenvalues into new values. On the one hand, the new values are nonzero. On the other hand, the new values corresponding to the inter-class samples increase, while the new values corresponding to the intra-class samples decrease or maintain simultaneously, which makes the margins between the inter-class samples greatly enlarged after mapping. Then, with the matrix functions of the two polynomial functions, the criterion of DLPP is reconstructed, and a novel discriminant locality preserving projections method is proposed. Because the polynomial functions are used to reconstruct the criterion of DLPP, the method is called as the polynomial DLPP (PDLPP). The proposed PDLPP can avoid the singularity of the scatter matrices, and then, the SSS problem of DLPP is addressed. More interesting, PDLPP can enlarge the distance between the inter-class samples, which is more beneficial to pattern classification compared with the existed methods. Thus, PDLPP can achieve better classification performance than state-of-the-art methods.

The next arrangement of this paper is as follows. In Sect. 2, the LPP and DLPP methods are reviewed. In Sect. 3, the polynomial discriminant locality preserving projections (PDLPP) is presented, and the theoretical analysis is made. In Sect. 4, some experiments are conducted to validate the effectiveness of PDLPP. Finally, Sect. 5 concludes this study.

2 Review of LPP and DLPP

2.1 Locality Preserving Projections (LPP)

Assume that $X = (x_1, x_2, ..., x_m)$ is the original n-dimensional dataset in R^n space. To implement dimensionality reduction, the LPP method seeks to find an optimal transformation matrix U to make a linear projection:

$$y_i = U^T x_i, \tag{1}$$

so that the dimensionality of projected vectors $y_i (i = 1, 2, ..., m)$ is smaller than that of the original vectors $x_i (i = 1, 2, ..., m)$, And, the LPP method tries to make that the reduced vectors $y_i (i = 1, 2, ..., m)$ keep the internal structure of the original vectors as much as possible. The criterion of LPP is designed to minimize the following objective function:

$$\sum_{ij} \| y_i - y_j \|^2 W_{ij} = \sum_{ij} \| U^T x_i - U^T x_j \|^2 W_{ij} = \mathrm{tr}\left(U^T X L X^T U\right), \tag{2}$$

where $\mathrm{tr}(\cdot)$ represents the trace of a matrix, W_{ij} represents the weight of the vertex pairs i and j. Usually, the heat-kernel function is selected to construct the weight. The matrix L is the Laplacian matrix, $L = D - W$, where W is the weight matrix composed of W_{ij} , D is row sum (or column sum), i.e., $D_{ii} = \sum_j W_{ij}$ (or $D_{jj} = \sum_i W_{ij}$) . A constraint condition $U^T XDX^T U = I$ is imposed to Eq. (2); then, the criterion of LPP is formulated as follows:

$$\underset{U^T XDX^T U=I}{\arg\min} \ \operatorname{tr}\left(U^T XLX^T U\right). \tag{3}$$

2.2 Discriminant Locality Preserving Projections (DLPP)

Different from LPP, DLPP is a supervised method. The class label information of the samples is used during the training phase. DLPP firstly constructs an adjacency matrix $W^c = \left(W^c_{ij}\right)$ for the cth intra-class samples and construct a total intra-class adjacency matrix $W = \operatorname{diag}\left(W^1, W^2, \cdots, W^C\right)$, where C is the number of sample categories. For the inter-class samples, the DLPP method finds a mean point in each category, and then, for the C mean points, it constructs an adjacency matrix $B = \left(B_{ij}\right)$. The DLPP method seeks a linear projection $y_i = U^T x_i$ to map the original sample x_i into a point y_i in low-dimensional space, where U is the desired projection matrix. The criterion of DLPP is:

$$\max \frac{\sum_{i,j=1}^{C} \left(\overline{y}_i - \overline{y}_j\right)^2 B_{ij}}{\sum_{c=1}^{C}\sum_{i,j=1}^{n_c} \left(y^c_i - y^c_j\right)^2 W^c_{ij}}, \tag{4}$$

where n_c is the sample number in the cth class and y^c_i is the ith projection vector the cth class. With some algebra transformation, Eq. (4) can be reduced to:

$$J(U) = \max_U \frac{\operatorname{tr}\left(U^T \bar{X} H \bar{X}^T U\right)}{\operatorname{tr}\left(U^T XLX^T U\right)} = \max_U \frac{\operatorname{tr}\left(U^T S_H U\right)}{\operatorname{tr}\left(U^T S_L U\right)}, \tag{5}$$

where $\overline{X} = [\overline{x_1}, \overline{x_2}, ..., \overline{x_C}]$, $H = E - B$, E is the row sum matrix of B, i.e., $E_{ii} = \sum_j B_{ij}$, $L = D - W$, D is the row sum matrix of W , i.e., $D_{ii} = \sum_j W_{ij}$, $S_H = \bar{X} H \bar{X}^T$, $S_L = XLX^T$. The matrix S_H and the matrix S_L may be called as the locality preserving inter-class and the locality preserving intra-class scatter matrix, respectively. Equation (5) can be reduced to the following generalized eigenvector problem:

$$S_H u = \lambda S_L u. \tag{6}$$

For the rank of the matrix S_L, one has

$$rank\,(S_L) = rank\left(XLX^T\right) \le rank\,(X) \le \min\,(m, n)\,,$$

where $rank(\cdot)$ represents the rank of a matrix, m is the number of the training samples, and n is the dimensionality of the samples. In real life, the number of training samples is much smaller than the dimensionality of the samples, i.e., $m \ll n$, then $rank\,(S_L) \ll n$. Note that the matrix $S_L = XLX^T$ is a matrix of $n \times n$, so the matrix S_L is singular. For example, in face recognition, assuming that the size of the face image is 100×100, the corresponding image vector is 10,000 dimensions. Then the matrix S_L is a matrix of $10,000 \times 10,000$. But $rank\,(S_L) \ll 10,000$, so the matrix S_L is singular, which makes it difficult to solve Eq. (6). This is the small-sample-size (SSS) problem, and the DLPP method has such difficulty.

3 Polynomial Discriminant Locality Preserving Projections (PDLPP)

3.1 Matrix Function and its Eigensystem

The matrix function is used to construct the proposed PDLPP method. Thus, the matrix function definition is firstly introduced.

Definition 1 ([21]). Let A be an n-order square matrix. Let $f(x)$ be a scalar function is defined on $\lambda(A)$, where $\lambda(A)$ is the eigenvalue set of the matrix A, then the matrix function $f(A)$ is defined by replacing A with x in the "formula" of $f(x)$.

Theorem 1 *Suppose that A is an n-order real symmetric square matrix, $f(x)$ is a scalar function, $f(A)$ is the matrix function of the matrix A, the eigenvalues of A are λ_i, and the corresponding eigenvectors of the matrix A are v_i, i.e., $Av_i = \lambda_i v_i$ $(i = 1, 2, \cdots, n)$. Then one has*

$$f(A)\,v_i = f(\lambda_i) v_i \ (i = 1, 2, \cdots, n)\,, \tag{7}$$

i.e., $f(\lambda_i)$ is an eigenvalue of $f(A)$ and v_i is the eigenvector of $f(A)$ belong to $f(\lambda_i)$.

Proof Because the matrix A is a real symmetric square matrix, A can be diagonalized. Therefore, there is an n-order orthogonal matrix $V = [v_1, v_2, \cdots, v_n]$, subject to $A = V \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)\, V^{-1}$, where λ_i are the eigenvalues of A and v_i are the corresponding eigenvectors of A. Then, the matrix function $f(A)$ can also be diagonalization [21], and the form is

$$f(A) = V \operatorname{diag}\left(f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)\right) V^{-1}.$$

The above diagonalizable form may be rewritten as:

$$f(A)\,v_i = f(\lambda_i)\,v_i \ (i = 1, 2 \cdots, n)\,.$$

3.2 PDLPP

Recall that the DLPP method is to maximize the criterion function of Eq. (5). For convenience, denote the criterion as

$$J(U) = \max_U \frac{\mathrm{tr}\left(U^T S_H U\right)}{\mathrm{tr}\left(U^T S_L U\right)}. \tag{8}$$

Let λ_{bi} be the eigenvalue of the matrix S_H, v_{bi} be the eigenvector belonging to λ_{bi} $(i = 1, 2, \cdots, n)$, and denote the matrix $\Lambda_H = \mathrm{diag}(\lambda_{b1}, \lambda_{b2}, \cdots, \lambda_{bn})$, the matrix $V_H = (v_{b1}, v_{b2}, \cdots, v_{bn})$, one has $S_H = V_H \Lambda_H V_H^T$. Similarly, let λ_{wi} be the eigenvalue of the matrix S_L, v_{wi} be the eigenvector of the eigenvalue λ_{wi} $(i = 1, 2, \cdots, n)$, and denote the matrix $\Lambda_L = \mathrm{diag}(\lambda_{w1}, \lambda_{w2}, \cdots, \lambda_{wn})$, the matrix $V_L = (v_{w1}, v_{w2}, \cdots, v_{wn})$, one has $S_L = V_L \Lambda_L V_L^T$. Thus, Eq. (8) becomes

$$J(U) = \max_U \frac{\mathrm{tr}\left(U^T V_H \Lambda_H V_H^T U\right)}{\mathrm{tr}\left(U^T V_L \Lambda_L V_L^T U\right)}. \tag{9}$$

When the SSS problem occurs, the matrices S_H and S_L are singular, so some eigenvalues of the two matrices must be zero. If these zero eigenvalues are transformed into nonzero values, then the singularity of the matrices S_H and S_L can be avoided. On the other hand, in pattern classification, we always want to make the distance of inter-class samples as large as possible, and the distance of intra-class samples as close as possible simultaneously. Note that the eigenvalues λ_{bi} and λ_{wi} measure the inter-class and intra-class distance, respectively, which is explained in the next section. If we can select a function $f(x)$ to increase the eigenvalue of the matrix S_H, i.e., $f(\lambda_{bi}) > \lambda_{bi}$, and select a function $g(x)$ to reduce or maintain the eigenvalues of the matrix S_L, i.e., $g(\lambda_{wi}) < \lambda_{wi}$ or $g(\lambda_{wi}) \approx \lambda_{wi}$, the above goal can be achieved.

According to the above analysis, the two functions selected need to meet the following conditions: (1) $g(x) \neq 0$; (2) $f(x) > x$, $g(x) < x$ or $g(x) \approx x$; (3) in order to highlight the pattern classification after mapping, $f(x)$ should be a monotonically increasing function, and as x increases, the function curve of $f(x)$ is as steep as possible.

In mathematics, there are many functions that satisfy the above conditions, and their forms are different. For the generality and simplicity of the proposed method, we can use two polynomial functions to achieve the above goal. It is from the fact that any function can be approximated by a polynomial function. In this paper, we select an n-order polynomial function $f(x) = a_0 + a_1 x + \cdots + a_n x^n (a_k > 0 (k = 0, 1, \cdots, n))$ to map the eigenvalues of the matrix S_H, an one-order polynomial function, i.e., simple linear function, $g(x) = b + x (b > 0)$ to map the eigenvalues of the matrix S_L. For convenience, denote $f(x) = \sum_{k=0}^{n} a_k x^k$. The detailed theoretical analysis about the selection of two functions will be given in the next section.

Thus, the eigenvalues of the matrices S_H and S_L are mapped into the corresponding form, respectively:

$$\lambda_{bi} \to \sum_{k=0}^{n} a_k \lambda_{bi}^k,\ \lambda_{wi} \to b + \lambda_{wi}. \tag{10}$$

By Eq. (10), the matrix function of the matrix Λ_H is $f(\Lambda_H) = \mathrm{diag}\left(\sum_{k=0}^{n} a_k \lambda_{b1}^k, \cdots, \sum_{k=0}^{n} a_k \lambda_{bn}^k\right)$, and the matrix function of the matrix Λ_L is $g(\Lambda_L) = \mathrm{diag}(b + \lambda_{w1}, b + \lambda_{w2}, \cdots, b + \lambda_{wn})$. Then, with the mapping (10), Eq. (9) becomes:

$$J_p(U) = \max_U \frac{\mathrm{tr}\left(U^T V_H f(\Lambda_H) V_H^T U\right)}{\mathrm{tr}\left(U^T V_L g(\Lambda_L) V_L^T U\right)}. \tag{11}$$

Let $f(S_H)$ be the matrix function of the function $f(x)$, $f(S_H)$ is a matrix polynomial, i.e., $f(S_H) = a_0 I + a_1 S_H + \cdots + a_n S_H^n = \sum_{k=0}^{n} a_k S_H^k$, and $g(S_L)$ is the matrix function of the function $g(x)$, i.e., $g(S_L) = bI + S_L$. Note that the matrices S_H and S_L are symmetric, Theorem 1 holds in this case. That is, $\left(\sum_{k=0}^{n} a_k \lambda_{bi}^k\right)$ is the eigenvalue of the matrix function $f(S_H)$ and v_{bi} is still the eigenvector of $f(S_H)$ belonging to the eigenvalue $\left(\sum_{k=0}^{n} a_k \lambda_{bi}^k\right)$; $b + \lambda_{wi}$ is the eigenvalue of the matrix function $g(S_L)$ and v_{wi} is still the eigenvector of $g(S_L)$ belonging to the eigenvalue $b + \lambda_{wi}$. Then, one has:

$$f(S_H) = V_H f(\Lambda_H) V_H{}^T,$$
$$g(S_L) = V_L g(\Lambda_L) V_L{}^T.$$

Thus, Eq. (11) becomes:

$$J_p(U) = \max_U \frac{\mathrm{tr}\left(U^T f(S_H) U\right)}{\mathrm{tr}\left(U^T g(S_L) U\right)}. \tag{12}$$

Equation (12) can be reduced to the generalized eigenvector problem by linear algebraic processing:

$$f(S_H)\, u = \lambda g(S_L)\, u,$$

i.e.,

$$\left(\sum_{k=0}^{n} a_k S_H^k\right) u = \lambda (bI + S_L)\, u. \tag{13}$$

The desired projection matrix U can be composed of the d largest generalized eigenvectors corresponding with the largest d eigenvalues of Eq. (13) ordered by descending.

Because the polynomial functions are used to reconstruct the criterion of DLPP, the method is called as the polynomial

DLPP (PDLPP). For matrix polynomial $f(S_H) = a_0 I + a_1 S_H + \cdots + a_n S_H^n = \sum_{k=0}^{n} a_k S_H^k$, there is the calculation of matrix power. To reduce the calculation cost of matrix power, Ref. [21] gives a skill. Inspired by [21], we use a small skill to calculate the matrix power, see Appendix.

Actually, for the proposed PDLPP method, by Eq. (10), the matrix S_H and the matrix S_L are mapped into the corresponding matrix function $f(S_H) = \sum_{k=0}^{n} a_k S_H^k$ and $g(S_L) = bI + S_L$, respectively. Note that the two matrices are composed of the original samples; we may think that the PDLPP method implies a nonlinear mapping from the original sample space to another space:

$$\begin{aligned}
&\Theta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},\\
&S_H \to \Theta(S_H) = \sum_{k=0}^{n} a_k S_H^k,\\
&\quad S_L \to \Theta(S_L) = bI + S_L.
\end{aligned} \tag{14}$$

And then, based on the new criterion function Eq. (12), DLPP method works on such a new space.

3.3 Theoretical Analysis

(1) Avoiding the SSS problem

From Sect. 3.2, the PDLPP method is based on the new criterion Eq. (12) and is reduced to the generalized eigenvector problem Eq. (13). Note that the eigenvalues of the matrix function $f(S_H) = \sum_{k=0}^{n} a_k S_H^k$ and $g(S_L) = rI + S_L$ are the forms $\sum_{k=0}^{n} a_k \lambda_{bi}^k (a_k > 0)$ and $b + \lambda_{wi}\ (b > 0)$, where λ_{bi} and λ_{wi} are the eigenvalues of the matrices S_H and S_L respectively. Obviously, $\sum_{k=0}^{n} a_k \lambda_{bi}^k > 0 (a_k > 0)$ and $b+\lambda_{wi} > 0 (b > 0)$, so the matrix functions $\sum_{i=0}^{n} a_i S_H^i$ and $bI + S_L$ are non-singular, and then, Eq. (13) is always solvable. Therefore, the SSS problem of DLPP is naturally avoided.

(2) The design of the polynomial function and distance diffusion effect

For the DLPP method, W_{ij}^c and B_{ij} are weight coefficients, W_{ij}^c is the similarity between the homogeneous samples i and j, B_{ij} is the similarity between the ith and jth class mean value vector. The inter-class distance d_b and the intra-class distance d_w in the sample space can be formulated as:

$$d_b = \frac{1}{2}\sum_{i,j} \left\| \bar{x}_i - \bar{x}_j \right\|^2 B_{ij} = \operatorname{tr}\left(\bar{X} H \bar{X}^T\right) = \operatorname{tr}(S_H). \tag{15}$$

$$d_w = \frac{1}{2}\sum_{i,j} \left\| x_i - x_j \right\|^2 W_{ij}^C = \operatorname{tr}\left(X L X^T\right) = \operatorname{tr}(S_L), \tag{16}$$

Note that the trace of a matrix equals the sum of all eigenvalues of the matrix; then, the above two distances may be written as:

$$d_b = \operatorname{tr}(S_H) = \lambda_{b1} + \lambda_{b2} + \cdots + \lambda_{bn}, \tag{17}$$

$$d_w = \operatorname{tr}(S_L) = \lambda_{w1} + \lambda_{w2} + \cdots + \lambda_{wn}. \tag{18}$$

For the PDLPP method, with the mapping Eq. (14), the matrix S_H is mapped into $f(S_H) = \sum_{k=0}^{n} a_k S_H^k$, and the matrix S_L is mapped into $g(S_L) = bI + S_L$, and by Theorem 1, the two distances d_b and d_w in Eqs. (17) (18) are replaced by d_b^p and d_w^p:

$$\begin{aligned}
d_b^p &= \operatorname{tr}(f(S_H)) = \left(\sum_{k=0}^{n} a_k \lambda_{b1}^k\right) + \left(\sum_{k=0}^{n} a_k \lambda_{b2}^k\right) + \cdots\\
&\quad + \left(\sum_{k=0}^{n} a_k \lambda_{bn}^k\right),
\end{aligned} \tag{19}$$

$$\begin{aligned}
d_w^p &= \operatorname{tr}(g(S_L)) = (b + \lambda_{w1}) + (b + \lambda_{w2}) + \cdots\\
&\quad + (b + \lambda_{wn}).
\end{aligned} \tag{20}$$

For the eigenvalues in Eq. (19), one can always design a polynomial function $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ $(a_k > 0 (k = 0, 1, \cdots, n))$ such that $a_0 + a_1\lambda_{bi} + \cdots + a_n\lambda_{bi}^n \gg \lambda_{bi}$. For the eigenvalues in Eq. (20), if b takes a little value, for example $b = 0.01$, one has $(b + \lambda_{wi}) \approx \lambda_{wi}$. In general, the distances of inter-classes are larger than that of the intra-classes, i.e., $d_b > d_w$, so for the most of λ_{bi}, λ_{wi} in Eq. (17) and (18), one has $\lambda_{bi} > \lambda_{wi}$. So the following inequality holds:

$$\frac{a_0 + a_1\lambda_{bi} + \cdots + a_n\lambda_{bi}^n}{b + \lambda_{wi}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \tag{21}$$

And then, by Eqs. (17), (18), (19), (20) and (21), one has

$$\frac{d_b^p}{d_w^p} > \frac{d_b}{d_w}. \tag{22}$$

Recently, the exponential DLPP (EDLPP) method based on the matrix exponential is proposed to solve the SSS problem of the DLPP method [13,14]. Although it does not express that the EDLPP method has the distance diffusion effect in [13,14], it is not difficult to know that this method also has a distance diffusion effect from similar research about the matrix exponential method, for example [18,19]. The distance diffusion effect of the EDLPP method can be described as:

$$\frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}.$$

For eigenvalue items in Eq. (20), if the parameter b takes a little value, one has $(b+\lambda_{wi}) \ll e^{\lambda_{wi}}$. To make the proposed PDLPP method also exceeds the distance diffusion ability of

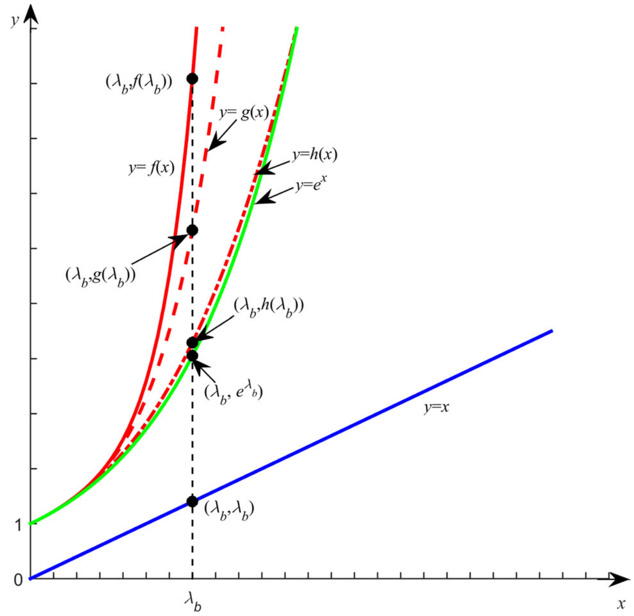


Fig. 1 The geometric meanings of the nonlinear mapping

the EDLPP method, we need to design the polynomial function $f(x) = \sum_{k=0}^{n} a_k x^k (a_k > 0)$. Inspired by the definition of the function $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, we can design a polynomial function $f(x) = 1 + \sum_{k=1}^{n} \frac{x^k}{k}$ such that $f(x) = 1 + \sum_{k=1}^{n} \frac{x^k}{k} > \sum_{k=0}^{\infty} \frac{x^k}{k!} = \exp(x)$. The motivation for this design is to make $f(\lambda_{bi}) = 1 + \sum_{k=1}^{n} \frac{\lambda_{bi}{}^k}{k} > \sum_{k=0}^{\infty} \frac{\lambda_{bi}^k}{k!} = e^{\lambda_{bi}}$. But, how to set the order n of the polynomial function? For the polynomial function $f(x) = 1+\sum_{k=1}^{n} \frac{x^k}{k}$, the larger the n value is, the steeper the curve is, the larger the eigenvalue after mapping is. But the larger the n value, the more calculation cost is needed. Especially in calculating n-order matrix polynomials $f(S_H) = 1 + \sum_{k=1}^{n} \frac{S_H{}^k}{k}$, the calculation cost is much large. On the other hand, the smaller the order n of the polynomial function $f(x)$, the small effect on eigenvalue, even lower than that of the EDLPP method.

In order to illustrate how we determine the order n of the polynomial function $f(x)$ and the geometric meanings of the eigenvalues before and after the nonlinear mapping, we have presented the following Fig. 1. In this figure, the blue line represents the function $y = x$, which expresses the eigenvalue before mapping, and the green line represents the function $y = e^x$, which expresses the eigenvalue after the mapping $\lambda_b \to e^{\lambda_b}$. The red short dotted line represents the function $y = h(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3}$, and the red long dotted line respresents the function $y = g(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5}$, the red solid line represents the function $y = f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + \frac{x^6}{6} + \frac{x^7}{7}$, which expresses the eigenvalue after the mapping $\lambda_b \to 1 + \sum_{k=1}^{3} \frac{\lambda_b^k}{k}$, $\lambda_b \to 1 + \sum_{k=1}^{5} \frac{\lambda_b^k}{k}$, and $\lambda_b \to 1 + \sum_{k=1}^{7} \frac{\lambda_b^k}{k}$, respectively. From Fig. 1, it can be seen that:

(1) Let λ_b be eigenvalue of the locality preserving inter-class scatter matrix S_H. For the polynomial function $y = 1 + \sum_{k=1}^{n} \frac{x^k}{k}$, when $n = 7$, the mapped value $f(\lambda_b)$ of the eigenvalue λ_b is the largest, followed by $g(\lambda_b)$ when $n = 5$, and the worst is $h(\lambda_b)$ when $n = 3$ (note that the polynomial function curve when $n = 3$ almost coincides with the exponential function curve). If the order n is larger, mathematically, the steeper the curve is, the larger the eigenvalue after mapping is. However, if the larger the n value is, the larger the computational complexity of computing n-order matrix polynomials $f(S_H) = 1 + \sum_{k=1}^{n} \frac{S_H{}^k}{k}$ is. On the other hand, for the EDLPP method, the calculation of the matrix $\exp(S_H)$ and $\exp(S_L)$ will result in large values, so the matrices S_H and S_L are usually normalized with their norms in the real calculation. Thus, the eigenvalues of the matrices S_H and S_L are shrinked to the interval [0, 1]. We know that the polynomial function $y = 1 + \sum_{k=1}^{n} \frac{x^k}{k}$, if the orders p and q, if $p > q$, compared with the polynomial $1 + \sum_{k=1}^{q} \frac{\lambda_b^k}{k}$, the increased items of the polynomial $1 + \sum_{k=1}^{p} \frac{\lambda_b^k}{k}$ are $\sum_{k=q+1}^{p} \frac{\lambda_b^k}{k}$. when the eigenvalue $0 < \lambda_b < 1$, the value of the above increased items is little.

So, we set a relative compromise value, namely $n = 7$.

(2) For the inter-class eigenvalue λ_b, from Fig. 1, one has: $f(\lambda_b) > e^{\lambda_b} > \lambda_b$, i.e., the effect of the polynomial function on eigenvalue is much larger than that of the exponential function $y = e^x$ and the linear function $y = x$.

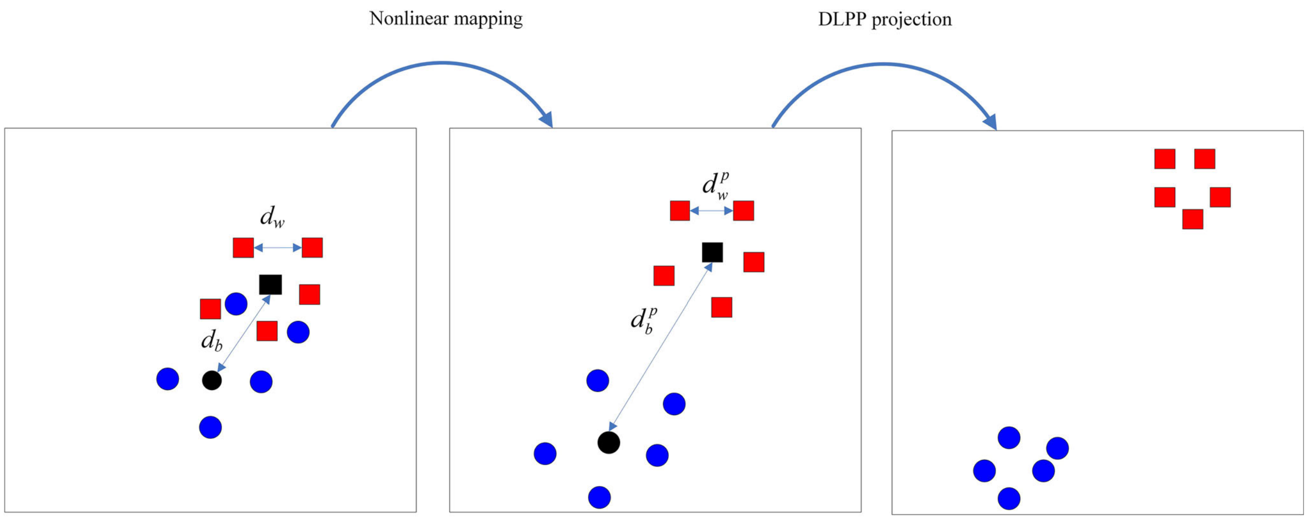
Therefore, from the perspective of exceeding the distance diffusion ability of the EDLPP method and the low calculation cost, we design a seven-order polynomial function $f(x) = 1+x+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\frac{x^5}{5}+\frac{x^6}{6}+\frac{x^7}{7}$. Where n=7 may be viewed as an empirical parameter, and one can also try to set other order or design another polynomial function. Because the eigenvalue λ_{bi} of the inter-class scatter matrix S_H usually is a large value, one always has $1 + \sum_{k=1}^{7} \frac{\lambda_{bi}{}^k}{k} \gg e^{\lambda_{bi}}$. Thus, for the eigenvalues in Eqs. (19) and (20), the following inequality holds:

$$\frac{1 + \sum_{k=1}^{7} \frac{\lambda_{bi}^k}{k}}{b + \lambda_{wi}} > \frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \tag{23}$$

And then one has:

$$\frac{d_b^p}{d_w^p} > \frac{d_b^e}{d_w^e} > \frac{d_b}{d_w}. \tag{24}$$

With the nonlinear mapping function Θ, the inter-class distance d_b and the intra-class distance d_w have been transformed into d_b^p and d_w^p, respectively. From the design of d_b^p



(a) The original sample space

(b) The sample space mapped by the polynomial function

(c) The projected space

Fig. 2 The geometrical interpretation of PDLPP using an illustrative example

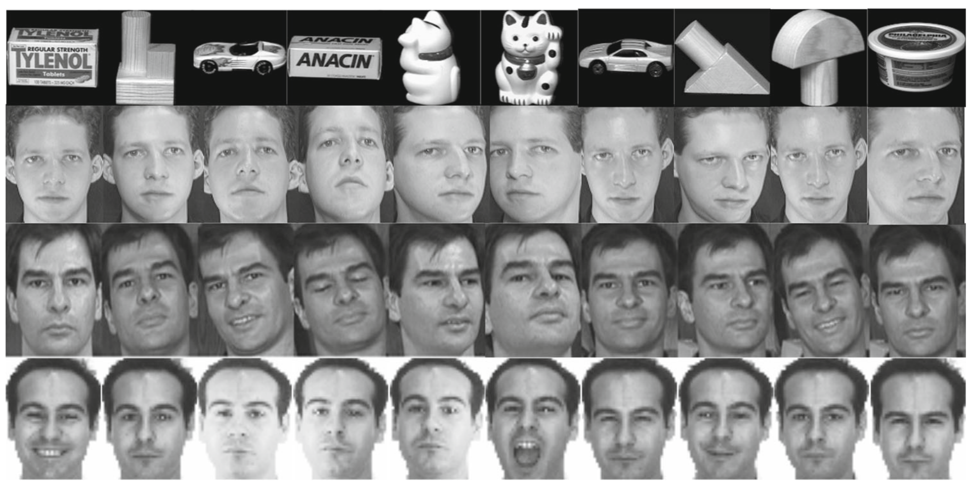


Fig. 3 Some sample images from the four datasets. The first row is from the Coil-20 object dataset, the second is from the ORL face dataset, the third row is from the Georgia Tech face dataset, and the fourth row is from the AR face dataset

Table 1 The Details of the experimental dataset

Dataset	Number of classes	Number of samples per class	Image size	Data dimension
Coil-20	20	72	50 × 50	2500
ORL	40	10	64 × 64	4096
Georgia Tech.	50	15	64 × 64	4096
AR	126	14	40 × 50	2000

and d_w^p, the PDLPP method enlarges the distance between inter-class samples, while the distance between the intra-class samples is almost unchanged. So the margins between the inter-class samples are greatly enlarged, which is beneficial for pattern classification. By the inequality (24), the ratio d_b^p/d_w^p is biggest, it means that the distance diffusion ability of the proposed PDLPP method is strongest compared with DLPP and EDLPP, so the better discrimination power will be gotten by PDLPP. Although the EDLPP method also has a distance diffusion effect, its drawback is that it not only enlarges the distance between inter-class samples but also enlarges the distance between the intra-class samples, which makes it unable to get better classification ability to some extent.

To intuitively explain the main idea of the PDLPP method and the distance diffusion effect of the nonlinear mapping,

Table 2 The results on the COIL-20 database (Recognition accuracy (%) ± standard deviation, and the optimal dimension)

p	8	9	10	11	12
PCA	87.59±1.06(50)	88.41±1.40(90)	89.17±1.77(40)	89.95±1.44(20)	91.11±1.63(40)
LPP	88.15±4.13(60)	89.31±4.23(40)	90.59±4.53(50)	90.65±4.25(30)	92.36±4.33(70)
LDA	89.59±2.65(10)	90.30±2.51(20)	91.21±2.43(20)	91.42±2.11(20)	93.65±2.06(10)
DLPP	89.54±2.94(20)	90.26±1.85(20)	91.65±1.97(10)	93.02±0.93(10)	94.17±1.15(20)
NDLPP	90.43±1.53(30)	91.93±1.84(80)	92.61±1.80(90)	92.73±0.61(50)	94.86±1.71(20)
DLPP/MMC	90.14±2.23(30)	91.45±0.63(60)	92.20±0.90(70)	92.86±0.97(70)	95.84±1.35(10)
EDLPP	91.06±1.60(80)	90.85±2.76(90)	91.07±0.64(10)	92.87±1.25(80)	95.94±1.15(10)
PDLPP	92.30±1.86(70)	92.69±2.04(60)	93.19±0.83(70)	94.16±1.11(70)	96.97±1.15(10)

Table 3 The results on the ORL face database (Recognition accuracy (%) ± standard deviation, and the optimal dimension)

p	2	3	4	5	6
PCA	78.62±1.79(80)	79.66±2.47(90)	88.66±2.63(80)	89.57±2.77(70)	92.10±2.82(80)
LPP	80.12±3.87(90)	81.38±1.86(90)	90.55±3.45(70)	91.52±2.36(90)	93.65±2.52(40)
LDA	80.63±1.83(39)	86.17±2.04(39)	91.67±1.21(30)	93.33±1.61(35)	95.23±1.98(39)
DLPP	74.37±4.20(60)	85.36±1.34(20)	91.00±1.29(40)	95.00±1.78(50)	95.21±2.99(60)
NDLPP	79.69±3.74(50)	90.61±2.78(70)	92.42±2.61(40)	95.12±2.57(70)	96.65±1.18(80)
DLPP/MMC	82.71±3.97(80)	91.07±2.67(60)	93.75±2.23(100)	95.67±0.24(30)	96.88±0.89(90)
EDLPP	82.19±0.88(100)	90.36±1.05(60)	94.45±0.39(40)	96.67±0.94(30)	98.54±0.78(100)
PDLPP	82.92±2.81(100)	91.67±1.71(60)	95.56±0.52(100)	97.17±0.62(50)	99.75±0.51(100)

Table 4 The results on the Georgia Tech database (Recognition accuracy (%) ± standard deviation, and the optimal dimension)

p	7	8	9	10	11
PCA	69.83±1.79(60)	70.57±0.47(90)	73.78±0.63(80)	74.27±3.81(60)	76.00±2.32(70)
LPP	70.12±2.47(90)	71.19±1.65(90)	72.12±1.67(80)	75.20±0.36(90)	77.17±0.52(40)
LDA	70.99±1.13(70)	73.87±2.04(60)	75.62±1.21(90)	76.97±1.61(80)	78.30±2.17(49)
DLPP	70.83±1.53(50)	74.86±1.02(40)	78.22±2.44(40)	80.83±2.54(30)	79.83±2.66(20)
NDLPP	71.69±3.14(50)	77.61±2.34(60)	80.42±2.61(40)	82.62±2.57(50)	82.75±2.18(70)
DLPP/MMC	69.50±3.14(50)	69.71±1.53(50)	70.67±2.84(50)	72.80±0.86(70)	72.50±0.71(70)
EDLPP	79.75±0.94(100)	81.81±0.49(80)	84.89±1.97(40)	84.93±0.50(40)	84.83±1.65(50)
PDLPP	82.83±1.94(100)	83.90±1.52(100)	87.66±0.47(40)	87.87±2.17(30)	89.00±1.22(100)

Table 5 The results on the AR face database (Recognition accuracy (%) ± standard deviation, and the optimal dimension)

p	2	3	4	5	6
PCA	80.79±1.29(70)	86.64±1.82(80)	90.67±0.51(100)	92.53±0.31(90)	91.35±0.23(90)
LPP	83.12±1.67(60)	88.29±2.46(70)	93.63±0.87(80)	93.20±1.03(80)	91.98±0.82(40)
LDA	89.77±1.29(60)	90.15±2.25(80)	95.03±0.69(100)	96.79±0.49(100)	93.89±1.73(90)
DLPP	92.11±2.28(100)	92.49±2.49(50)	94.33±0.54(30)	96.00±0.68(50)	97.23±0.40(70)
NDLPP	92.39±3.14(70)	93.61±2.38(70)	95.42±2.61(40)	96.69±1.57(50)	97.75±1.15(60)
DLPP/MMC	93.03±1.43(100)	93.39±2.21(50)	94.25±0.29(30)	97.04±0.62(70)	98.33±0.23(70)
EDLPP	91.64±1.70(100)	92.96±3.02(50)	94.56±3.44(100)	97.08±0.66(70)	98.10±0.22(70)
PDLPP	94.98±1.37(100)	95.25±2.49(50)	96.39±0.68(30)	98.49±0.49(70)	99.34±0.40(80)

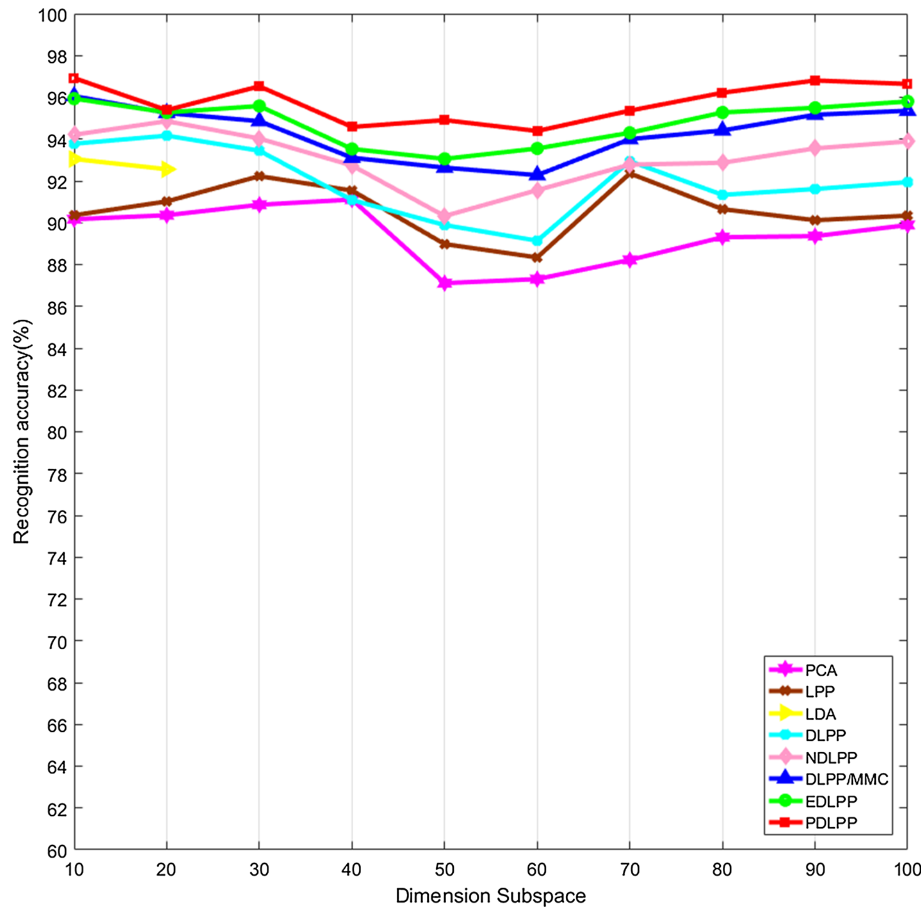


Fig. 4 The comparison of performances versus subspace dimension on COIL-20 dataset (The training sample $p = 12$)

the geometric interpretation of PDLPP is given in the following Fig. 2. Where Figure (a) represents the original sample space, Figure (b) represents the sample space transformed by the nonlinear mapping, and Figure (c) represents the projected space by DLPP. For convenience, two-class samples are used to illustrate. One class is represented by red squares, and the center of the class is represented by a black square. Another class is represented by a blue circle, and the center of the class is represented by a black circle. As shown in Fig.2, with a nonlinear mapping, the PDLPP method maps the original samples into a new space. On such a new space, the distance between the inter-class samples is enlarged and the distance between the intra-class samples is almost unchanged, so the margins between the inter-class samples are greatly enlarged. And then, the DLPP method makes projection and extraction on the new space.

4 Experimental Results

4.1 Experimental Dataset

In this section, the experiments are conducted on the COIL-20 database and three face databases: ORL, Georgia Tech, and AR face databases.

(1) The COIL-20 dataset

Columbia University Image Library (COIL-20) contains 20 objects, each of which rotates 360° horizontally, taking a picture every 5°, and each object has a total of 72 images. Some sample images of this dataset are shown in the first row of Fig. 3. In the experiment, the image size is resized to 50×50 pixels.

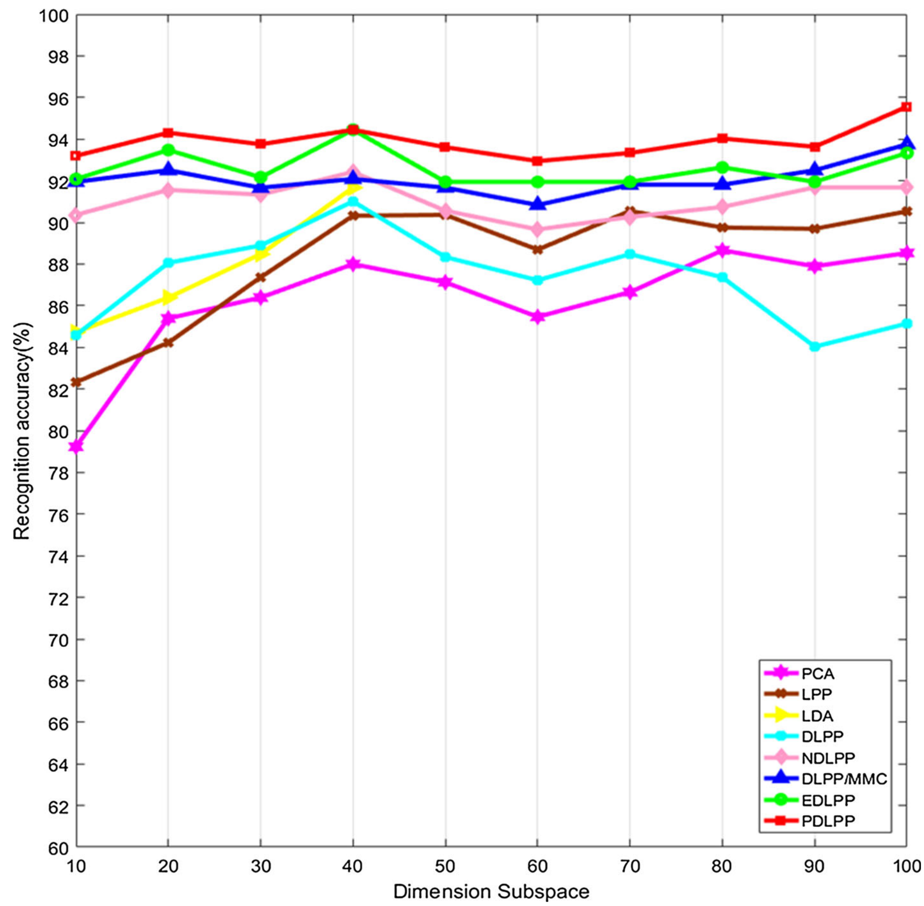


Fig. 5 The comparison of performances versus subspace dimension on the ORL face dataset (The training sample p = 4)

(2) ORL face dataset

The ORL face image database consists of 40 persons; each person has 10 different images, in all 400. These images were taken in different facial features and scenes. Some sample images of one individual on this dataset are shown in the second row of Fig. 3. In our experiments, each image is adjusted to a gray-scale image of 64×64 pixels.

(3) The Georgia Tech face Dataset

The Georgia Tech Face Dataset contains photos of 50 individuals taken during two or three meetings between Januanry and November 2007. The image includes front and slanted faces under variations in facial expressions, illumination and proportions, and a cluttered background. Some samples of one individual on the dataset are shown in the third row of Fig. 3. Each image is aligned and scaled to 64×64 pixels in the experiments.

(4) The AR face Dataset

This face database was created by A. Martinez and R. Benavente in the computer vision center (CVC) at the U.A.B. It contains over 4000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions and occlusions (sun glasses and scarf). Some samples of one individual on the dataset are shown in the fourth row of Fig. 3. The size of each image is 40×50 pixels in the experiments.

The following Table 1 summarizes the details of the datasets used in the experiments.

4.2 Experiment Setup and Results

In the experiments, we compare the proposed PDLPP with two global dimension reduction methods: PCA [15] and LDA [22], as well as dimension reduction methods based on manifold learning: LPP [5], DLPP [7], NDLPP [11], DLPP/MMC [12] and EDLPP [13,14] methods, where NDLPP, DLPP/MMC and EDLPP are methods to solve the SSS problem of DLPP. For the DLPP method, PCA is firstly used to reduce the dimensionality of the original samples and then avoids the SSS problem of DLPP. For the proposed PDLPP method, the polynomial function $f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + \frac{x^6}{6} + \frac{x^7}{7}$, and the function $g(x) = 0.01 + x$.

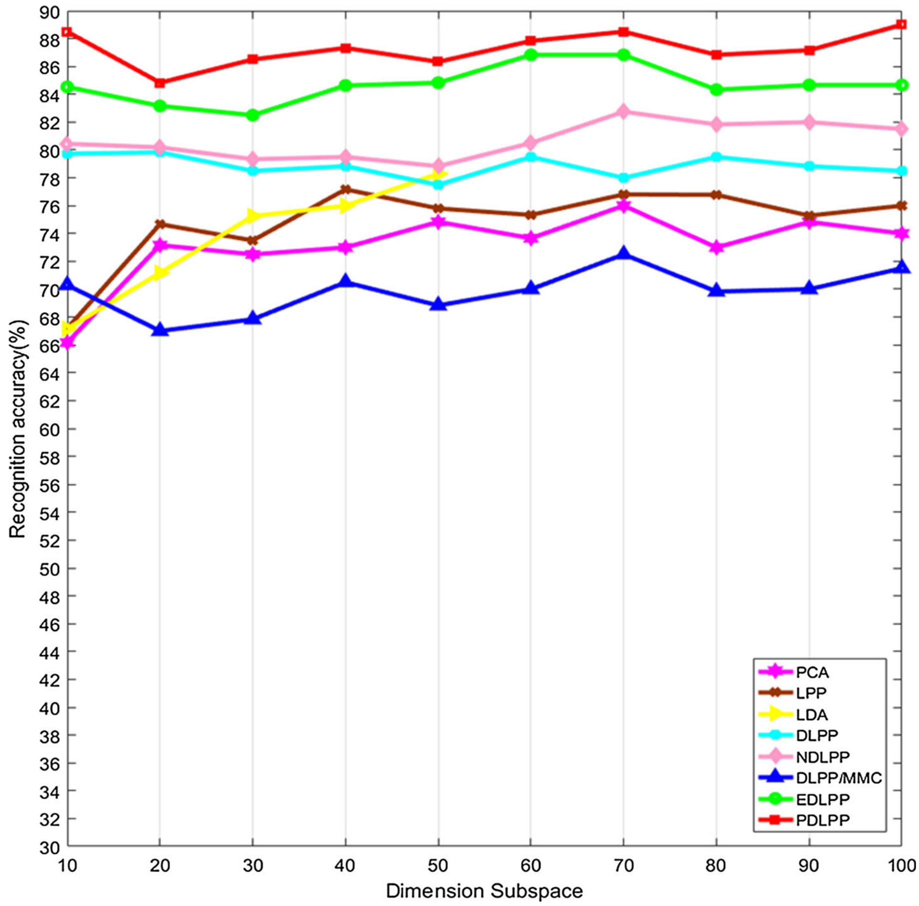


Fig. 6 The comparison of performances versus subspace dimension on the Georgia Tech dataset (The training sample p = 11)

In the experiment, for each subject in each dataset, p samples are selected as training samples, and the rest is used as test samples. To get stable experiment results, the experiment is repeated 20 times. For each experiment, p training samples are randomly selected. In each experiment, for a given p, the subspace dimension after dimensionality reduction is taken from 10 to 100 with step 10.

We evaluate the performance of these methods for pattern classification using the average of the best recognition rates. In each experiment, for p training samples, the recognition rate corresponding to the best subspace dimension is regarded as the best recognition rate. Therefore, for 20 experiments, there are 20 best recognition rates. Finally, their average value is regarded as the recognition rate when the training sample is p. The recognition results (Recognition accuracy (%) $\pm$ standard deviation, and the optimal subspace dimension) of these methods are reported in Tables 2, 3, 4, and 5.

We also evaluated the performance of these methods concerning the subspace dimension. In each experiment, there is a recognition rate for given p training samples and given subspace dimension. In this way, there are 20 recognition rates for 20 experiments. Their average value is taken as the recognition rate for the current subspace dimension. When the subspace dimension is taken from 10 to 100 with step 10, we can get the recognition rate for each dimension. The recognition rates versus the variations of the subspace dimension are shown in Figs. 4, 5, 6, and 7.

From the experimental results, it can be observed that: (1) compared with the other methods, the PDLPP method achieves the best recognition accuracy on all databases. (2) When the original sample is reduced to different dimensions, the PDLPP method has better performance among all subspace dimensions in all the databases. It shows that PDLPP is an effective dimensionality reduction method.

5 Conclusions

In this paper, a novel dimensionality reduction method named polynomial discriminant locality preserving projections method (PDLPP) is proposed. The PDLPP method starts from the eigenvalue of the scatter matrices, transforms the eigenvalues into new values by two simple polynomial functions, and then reconstructs the criterion of the DLPP method based on the matrix function of the polynomial function. The PDLPP method solves the SSS problem of DLPP, and it can enlarge the distance of inter-class samples, and so it has better pattern classification power. Experimental results

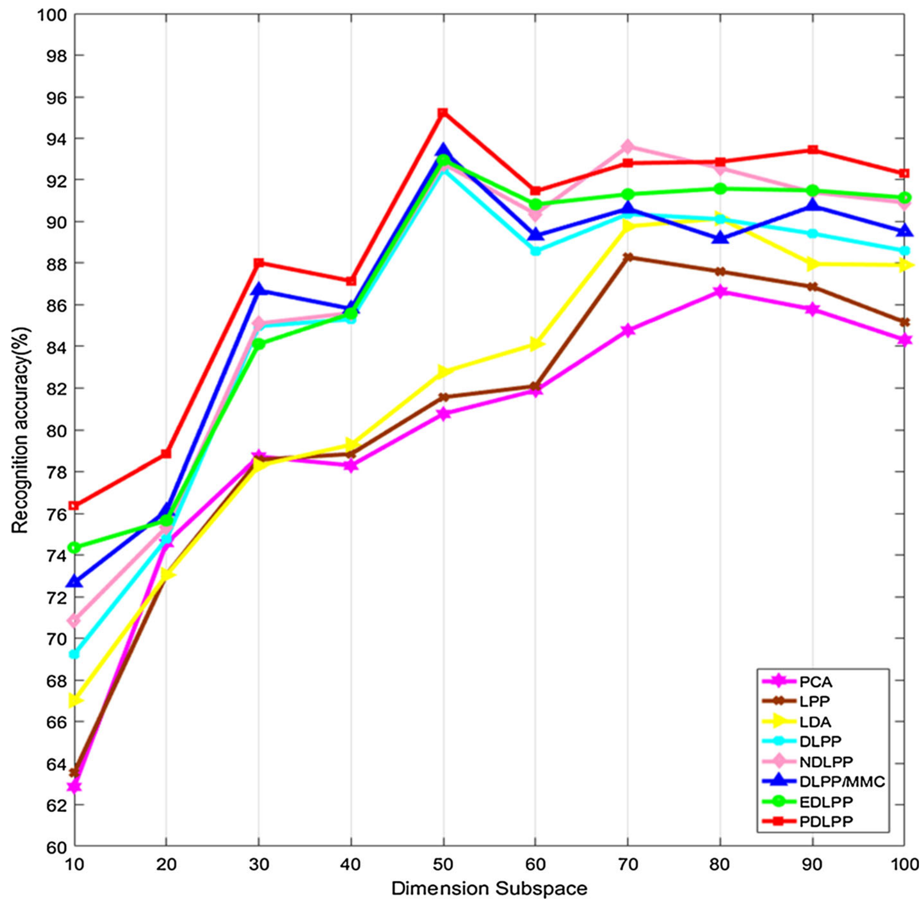


Fig. 7 The comparison of performances versus subspace dimension on the AR face dataset (The training sample p = 3)

on the COIL-20 database and three face databases have validated the effectiveness of the PDLPP method and shown that the PDLPP method has advantageous performance over state-of-the-art methods. Besides, the PDLPP method may be regarded as a general method and extended to manifold learning methods, such as MFA, LDE, SDE, etc.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant 61876026, Humanities and Social Sciences Project of the Ministry of Education of China under Grant 20YJAZH084, Chongqing Technology Innovation and Application Development Project under Grant cstc2020jscx-msxmX0190,cstc2019jscx-mbdxX0061, Project of Natural Science Foundation Project of CQ CSTC of China under grant cstc2016jcyjA0419 and cstc2017jcyjAX0316, and School Fund Project of CQNU under grant 16XLB006 and 16XZH07.

Appendix

Let A be a square matrix, and $f(A) = a_0 I + a_1 A + a_2 A^2 \cdots + a_n A^n$ be a matrix polynomial. The calculation of each power of $f(A)$ can be divided into the calculation of even power and the calculation of odd power.

Let $B = A^2 = A \times A$, one can get each even power recursively as follows:

$$A^4 = B \times B, A^6 = B \times A^4, A^8 = B \times A^6, \cdots$$

For odd power, based on the above even powers, one has:

$$A^3 = A \times B, A^5 = A \times A^4, A^7 = A \times A^6, \cdots$$

Thus, using recursion, each matrix power can be obtained by simple multiplication of two matrices.

References

1. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
2. Tenenbaum, J.B.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
3. Belkin, M., Niyogi, P.: Laplacian Eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**, 1373–1396 (2003)
4. Zhang, Z., Zha, H.: Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment. SIAM J. Sci. Comput. **26**, 313–338 (2005)
5. He, X.F., Niyogi, P.: Locality preserving projections. Adv. Neur. Inf. Process. Syst. **16**, 153–160 (2004)

6. He, X.F., Yan, S., Hu, Y., et al.: Face Recognition Using Laplacian Faces. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 328–340 (2005)
7. Yu, W., Teng, X., Liu, C.: Face recognition using discriminant locality preserving projections. Image Vis. Comput. **24**, 239–248 (2006)
8. Zhu, L., Zhu, S.: Face recognition based on orthogonal discriminant locality preserving projections. Neurocomputing **70**, 1543–1546 (2007)
9. Jie, G., Chao, W., Ling, Z.: Locality preserving discriminant projections. Emerg. Intell. Comput. Technol. Appl. **566–572**, (2009)
10. Zhao, X., Tian, X.X.: Locality preserving fisher discriminant analysis for face recognition. Emerg. Intell. Comput. Technol. Appl. **261–269**, (2009)
11. Yang, L., Gong, W., Gu, X., et al.: Null space discriminant locality preserving projections for face recognition. Neurocomputing **71**, 3644–3649 (2008)
12. Lu, G.F., Lin, Z., Jin, Z.: Face recognition using discriminant locality preserving projections based on maximum margin criterion. Pattern Recogn. **43**(10), 3572–3579 (2010)
13. Huang, S.C., Zhuang, L.: Exponential discriminant locality preserving projection for face recognition. Neurocomputing **208**, 373–377 (2016)
14. Lu, G.F., Wang, Y., Zou, J., et al.: Matrix exponential based discriminant locality preserving projections for feature extraction. Neural Netw. **97**, 127–136 (2018)
15. Turk, M.A., Pentland, A.P.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
16. Ran, R.S., Fang, B., Wu, X.G.: Exponential neighborhood preserving embedding for face recognition, IEICE Trans. Info. Syst, E101-D. **5**, 1410–1420 (2018)
17. Wang, S.J., Chen, H.L., Peng, X.J., et al.: Exponential locality preserving projections for small sample size problem. Neurocomputing **74**, 3654–3662 (2011)
18. Dornaika, F., Bosaghzadeh, A.: Exponential local discriminant embedding and its application to face recognition. IEEE Trans. Cybern. **43**, 921–934 (2013)
19. Dornaika, F., Traboulsi, Y.E.: Matrix exponential based semi-supervised discriminant embedding for image classification. Pattern Recogn. **61**, 92–103 (2017)
20. Wang, S.J., Yan, S., Yang, J., et al.: A general exponential framework for dimensionality reduction. IEEE Trans. Image Process. **23**, 920–930 (2014)
21. Golub, G.H., Loan, C.F.V.: Matrix Computation, The Johns Hopkins Univ. Press, Baltimore, Md, USA, pp. 978-990 (1983)
22. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Hum. Genet. **7**, 179–188 (1936)

Ruisheng Ran received the B.S. degree in mathematics from Chongqing Normal University, Chongqing, China, in 1998. He received the M.S. degree in computational mathematics from University of Electronic Science and Technology of China, Chengdu, China, in 2004. He received the Ph.D. degree in computer application technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2006. He is currently a Professor with the College of computer and information science, Chongqing Normal University, Chongqing, China. He is doing machine learning, computer vision research.

Yinshan Ren is a Master candidate in Chongqing Normal University, his major is computer technology and main interests include computer vision and machine learning.

Shougui Zhang received his B.S. degree from Chongqing Normal University, Chongqing, China. He received his M.S. and Ph.D. degrees from Chongqing University, Chongqing, China, in 2005 and 2012, respectively. He is currently an Associate Professor with the College of Mathematical Sciences, Chongqing Normal University, Chongqing, China. His research interests include numerical solution of PDEs, boundary element methods. Right now, he is particularly interested in computational methods and optimization application in variational inequalities.

Bin Fang received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, the M.S. degree in electrical engineering from Sichuan University, Chengdu, China, and the Ph.D. degree in electrical engineering from the University of Hong Kong, Hong Kong. He is currently a Professor with the Department of Computer Science, Chongqing University, Chongqing, China. His research interests include computer vision, pattern recognition, medical image processing, biometrics applications, and document analysis. He has published more than 100 technical papers and is an Associate Editor of the International Journal of Pattern Recognition and Artificial Intelligence.