LETTER

A Simple and Effective Generalization of Exponential Matrix Discriminant Analysis and Its Application to Face Recognition

Ruisheng RAN[†a)], **Bin FANG**[††b)], ***Nonmembers***, **Xuegang WU**[††c)], ***Member***, ***and*** **Shougui ZHANG**[†d)], ***Nonmember***

SUMMARY As an effective method, exponential discriminant analysis (EDA) has been proposed and widely used to solve the so-called small-sample-size (SSS) problem. In this paper, a simple and effective generalization of EDA is presented and named as GEDA. In GEDA, a general exponential function, where the base of exponential function is larger than the Euler number, is used. Due to the property of general exponential function, the distance between samples belonging to different classes is larger than that of EDA, and then the discrimination property is largely emphasized. The experiment results on the Extended Yale and CMU-PIE face databases show that, GEDA gets more advantageous recognition performance compared to EDA.

key words: *matrix exponential, linear discriminant analysis, the small sample size problem, face recognition*

1. Introduction

The classical Linear Discriminant Analysis (LDA) [1] is an important and effective approach in pattern recognition. Unfortunately, in most cases, the dimension of the sample is much larger than the number of the samples, which results in the within-class scatter matrix in LDA is in question singular. This is so-called small size sample (SSS) problem and LDA has to suffer from this problem.

As an effective method, Exponential Discriminant Analysis (EDA) [2] is proposed to overcome the SSS problem of classical LDA. The main idea of EDA is that the Euler matrix exponential is introduced. The scatter matrices are firstly mapped into a new space with matrix exponential transformation, and then the LDA criterion is applied in such a space. Due to the property of the matrix exponential, the SSS problem LDA is addressed by EDA. EDA method is improved in Ref. [3]. The author proposed two inexact Krylov subspace algorithms to solve the large matrix exponential eigenproblem efficiently. And the new algorithm showed the superiority over their original counterpart for face recognition.

After the release of EDA, it is widely applied to solve the SSS problem, especially in the manifold learning field. In Ref. [4], a general exponential framework for dimensionality reduction has been proposed. Under this framework, matrix exponential is applied to extend many popular Laplacian embedding algorithms. Many of manifold learning algorithms, such as Local Preserving Projection (LPP) [5], Discriminant Locality Preserving Projection (DLPP) [6], Local Discriminant Embedding (LDE) [7] and Semi-supervised Discriminant Embedding (SDE) [8], have to suffer from the SSS problem. Then, the EDA method is introduced to solve this problem. The exponential LPP (ELPP) [9], the exponential DLPP method (EDLPP) [10], the exponential LDE method (ELDE) [11] and the exponential SDE method (ESDE) [12] are proposed. They are the exponential versions of the corresponding methods. They avoid the SSS problem and show better performance in face recognition.

In this paper, a simple and effective generalization of EDA is made, and called as GEDA. The main idea of GEDA is that the general exponential function $f(x) = a^x$, not Euler exponential function $f(x) = e^x$, is used. Due to the property of exponential function, one has $a^x > e^x$ if $a > e$ and $x > 0$. And then, the distance of GEDA between samples belonging to different classes, which is with the matrix exponential transformation $a(\boldsymbol{A})$, is larger than that of EDA, which is with the Euler matrix exponential $\exp(\boldsymbol{A})$. So, we can believe that GEDA will show advantageous performance over EDA for classification.

2. Generalized EDA (GEDA)

2.1 Theoretical Basis of GEDA

In this section, some theories about the exponential function and matrix function for generalized EDA method are presented.

Theorem 1. Let $f(x) = a^x$ is a general exponential function, if $a > e$ and $x > 0$, the inequality $a^x > e^x$ holds.

The following Fig. 1 shows this property of the exponential function.

Theorem 2([13]). For an arbitrarily n-order real symmetric square matrix $\boldsymbol{A}$, $\boldsymbol{A}$ can be diagonalization, i.e., there exists an n-order orthogonal matrix $\boldsymbol{T}$ and a matrix $\boldsymbol{D}$, subject to

$$\boldsymbol{D} = \boldsymbol{T}'\boldsymbol{A}\boldsymbol{T} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$$

is a diagonal matrix. Where $\boldsymbol{T}'$ is the transpose matrix of $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$ is the inverse matrix of $\boldsymbol{T}$ respectively.

For the above diagonalization, let λ_i $(i = 1, 2, \cdots, n)$

Manuscript received September 12, 2017.
Manuscript publicized October 18, 2017.

[†]The authors are with Chongqing Normal University, Chongqing, 401331, China.
[††]The authors are with Chongqing University, Chongqing, 400044, China.
a) E-mail: rshran@163.com
b) E-mail: fb@cqu.edu.cn
c) E-mail: xgwu@cqu.edu.cn
d) E-mail: shgzhang@cqnu.edu.cn

DOI: 10.1587/transinf.2017EDL8198

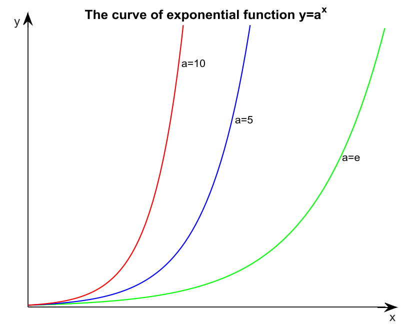

Fig. 1 The curve of exponential function $y = a^x$, $a^x > e^x$ if $a > e$ and $x > 0$

are the eigenvalues of the n-order matrix $\boldsymbol{A}$, then $\boldsymbol{D} = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$. And the ith column vector of $\boldsymbol{T}$ is the eigenvector of the matrix $\boldsymbol{A}$ corresponding to the ith eigenvalue λ_i $(i = 1, 2, \cdots, n)$.

In this paper, the proposed GEDA method is based on the matrix function. In the following, the definition and eigenproblem of the matrix function are introduced.

Definition 1([13]). Let $f(x)$ be a scalar function with x as independent variable and $\boldsymbol{A}$ be a square matrix, if the independent variable x is replaced with the square matrix $\boldsymbol{A}$, the function $f(\boldsymbol{A})$ may be gotten and is called as a matrix function.

Theorem 3([13]). Let $\boldsymbol{A}$ be a n-order square matrix and $\boldsymbol{A}$ can be diagonalization, i.e., there exists a n-order matrix $\boldsymbol{X}$, subject to $\boldsymbol{A} = \boldsymbol{X} diag(\lambda_1, \lambda_2, \cdots, \lambda_n)\boldsymbol{X}^{-1}$. Let $f(\boldsymbol{A})$ is a matrix function, then

$$f(\boldsymbol{A}) = \boldsymbol{X} diag(f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n))\boldsymbol{X}^{-1}.$$

According to Theorem 2 and Theorem 3, the relationship between the eigen-system of a real symmetric square matrix $\boldsymbol{A}$ and the eigen-system of the matrix function $f(\boldsymbol{A})$ may be easily obtained as follows:

Theorem 4. Let $\boldsymbol{A}$ be a n-order real symmetric square matrix and $f(\boldsymbol{A})$ is a matrix function, and let $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n]$ are eigenvectors of the matrix $\boldsymbol{A}$ that correspond to the eigenvalues $[\lambda_1, \lambda_2, \cdots, \lambda_n]$ of $\boldsymbol{A}$, then $\boldsymbol{V}$ are also eigenvectors of the matrix function $f(\boldsymbol{A})$ that correspond to the eigenvalues of $f(\boldsymbol{A})$:

$$[f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)].$$

In this paper, the general matrix exponential function is used. We define it as follows.

Given an arbitrarily $n \times n$ square matrix $\boldsymbol{A}$, its Euler matrix exponential is defined as [13]:

$$\exp(\boldsymbol{A}) = \boldsymbol{I} + \boldsymbol{A} + \frac{\boldsymbol{A}^2}{2!} + \cdots + \frac{\boldsymbol{A}^k}{k!} + \cdots,$$

where $\boldsymbol{I}$ is an identity matrix.

Note that a general scalar exponential function $f(x) = a^x$ $(a > 0, a \neq 1)$ may be written as:

$$a^x = \exp(x \ln a).$$

According to Definition 1, the general matrix exponential function may be defined as follows:

Definition 2. Let $f(x) = a^x$ $(a > 0, a \neq 1)$ be a scalar exponential function with the x as argument and $\boldsymbol{A}$ be a square matrix. Denote a general matrix exponential function as $a(\boldsymbol{A})$, $a(\boldsymbol{A})$ may be defined as:

$$a(\boldsymbol{A}) = \exp(\boldsymbol{A} \ln a).$$

Obviously, Theorem 3 and Theorem 4 hold for the general matrix exponential function $a(\boldsymbol{A})$.

2.2 GEDA

Let $\boldsymbol{S}_b$ be between-class scatter matrix, $\boldsymbol{S}_w$ be within-class scatter matrix, and let $\boldsymbol{\phi}_b = (\boldsymbol{\varphi}_{b1}, \boldsymbol{\varphi}_{b2}, \cdots, \boldsymbol{\varphi}_{bn})$ is the eigenvector matrix of $\boldsymbol{S}_b$ that corresponds to eigenvalues $\boldsymbol{\Lambda}_b = diag(\lambda_{b1}, \lambda_{b2}, \cdots, \lambda_{bn})$, and let $\boldsymbol{\phi}_w = (\boldsymbol{\varphi}_{w1}, \boldsymbol{\varphi}_{w2}, \cdots, \boldsymbol{\varphi}_{wn})$ is the eigenvector matrix of $\boldsymbol{S}_w$ that corresponds to eigenvalue $\boldsymbol{\Lambda}_w = diag(\lambda_{w1}, \lambda_{w2}, \cdots, \lambda_{wn})$. The objective of LDA is to find an optimal projection matrix $\boldsymbol{W}$ by maximizing the ratio of between-class scatter to within-class scatter as follows:

$$J(\boldsymbol{W}) = \arg\max_{\boldsymbol{W}} \frac{|\boldsymbol{W}^T \boldsymbol{S}_b \boldsymbol{W}|}{|\boldsymbol{W}^T \boldsymbol{S}_w \boldsymbol{W}|} = \arg\max_{\boldsymbol{W}} \frac{|\boldsymbol{W}^T \boldsymbol{\phi}_b^T \boldsymbol{\Lambda}_b \boldsymbol{\phi}_b \boldsymbol{W}|}{|\boldsymbol{W}^T \boldsymbol{\phi}_w^T \boldsymbol{\Lambda}_w \boldsymbol{\phi}_w \boldsymbol{W}|}. \tag{1}$$

Let $f(x) = a^x$ $(a > 0)$ be a general exponential function. In the above LDA criterion (1), we replace the eigenvalue λ_{wj} of $\boldsymbol{S}_w$ with $a^{\lambda_{wj}}$, and replace the eigenvalue λ_{bj} of $\boldsymbol{S}_b$ with $a^{\lambda_{bj}}$, and denote

$$a(\boldsymbol{\Lambda}_w) = diag(a^{\lambda_{w1}}, a^{\lambda_{w2}}, \cdots, a^{\lambda_{wn}}),$$
$$a(\boldsymbol{\Lambda}_b) = diag(a^{\lambda_{b1}}, a^{\lambda_{b2}}, \cdots, a^{\lambda_{bn}}).$$

Then, the criterion of LDA can be transformed to:

$$J(\boldsymbol{W}) = \arg\max_{\boldsymbol{W}} \frac{|\boldsymbol{W}^T \boldsymbol{\phi}_b^T a(\boldsymbol{\Lambda}_b) \boldsymbol{\phi}_b \boldsymbol{W}|}{|\boldsymbol{W}^T \boldsymbol{\phi}_w^T a(\boldsymbol{\Lambda}_w) \boldsymbol{\phi}_w \boldsymbol{W}|}. \tag{2}$$

Because the matrices $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$ are symmetric and semi-positive definite matrices [2], Theorem 4 holds in this case. And so, the Eq. (2) can be rewritten as:

$$J(\boldsymbol{W}) = \arg\max_{\boldsymbol{W}} \frac{|\boldsymbol{W}^T a(\boldsymbol{S}_b) \boldsymbol{W}|}{|\boldsymbol{W}^T a(\boldsymbol{S}_w) \boldsymbol{W}|}. \tag{3}$$

The columns vector of optimal transformation matrix $\boldsymbol{W}$ can be obtained by solving the following generalized eigenvectors problem:

$$a(\boldsymbol{S}_b)\boldsymbol{x} = \lambda a(\boldsymbol{S}_w)\boldsymbol{x}. \tag{4}$$

Because $f(x) = a^x$ is a general exponential function, the proposed method is called as generalized EDA (GEDA). If the base $a = e$, where e is Euler number 2.71828, i.e. $f(x) = e^x$, the GEAD becomes EDA.

3. Distance Diffusion Mapping

LDA finds an optimal projection by simultaneously maximizing the between-class distance and minimizing the within-class distance. The two distances can be measured by the traces of two scatter matrices as:

$$d_b = trace(\boldsymbol{S}_b) = \lambda_{b1} + \lambda_{b2} + \cdots + \lambda_{bn}, \tag{5}$$

$$d_w = trace(\boldsymbol{S}_w) = \lambda_{w1} + \lambda_{w2} + \cdots + \lambda_{wn}. \tag{6}$$

Note that the between-class scatter matrix $\boldsymbol{S}_b$ and within-class scatter matrix $\boldsymbol{S}_w$ are symmetric and semi-positive definite matrices [2], so for all the eigenvalues in Eqs. (5) and (6), one has $\lambda_{bi} \geq 0$, $\lambda_{wi} \geq 0$ $(i = 1, 2, \cdots, n)$.

In fact, for GEDA, there is an implicit non-linear mapping of samples:

$$\Theta\colon\ \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},$$
$$\boldsymbol{S}_b \to \Theta(\boldsymbol{S}_b) = a(\boldsymbol{S}_b),\ \boldsymbol{S}_w \to \Theta(\boldsymbol{S}_w) = a(\boldsymbol{S}_w). \tag{7}$$

With the mapping (7) and Theorem 4, in the same input space, such that the distance d_b and d_w are replaced by d_b^a and d_w^a:

$$d_b^a = trace(a(\boldsymbol{S}_b)) = a^{\lambda_{b1}} + a^{\lambda_{b2}} + \cdots + a^{\lambda_{bn}}, \tag{8}$$

$$d_w^a = trace(a(\boldsymbol{S}_w)) = a^{\lambda_{w1}} + a^{\lambda_{w2}} + \cdots + a^{\lambda_{wn}}. \tag{9}$$

In general, the distance between samples in different classes is bigger than the related distance between samples in the same class, one has $d_b > d_w$. So, for most of the eigenvalues in Eqs. (5) and (6), one has $\lambda_{bi} > \lambda_{wi}$. These eigenvalues may be viewed as a set. In the set, we have $\lambda_{bi} > \lambda_{wi} \geq 0$, according to the value of λ_{wi}, we make the discussion as follows:

1) In the case of $\lambda_{wi} > 0$.

Because of $\lambda_{bi} > \lambda_{wi}$, one has $a^{\lambda_{bi}} > a^{\lambda_{wi}}$, and so

$$\frac{a^{\lambda_{bi}}}{a^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \tag{10}$$

Especially, if $a = e$, the above inequality (10) also holds.

In this paper, the base a of the exponential function is chosen to meet the condition: $a > e$. According to Theorem 1, if $a > e$ and $x > 0$, one has $a^x > e^x$. In this case, $\lambda_{bi} > \lambda_{wi} > 0$, so one has:

$$\frac{a^{\lambda_{bi}}}{a^{\lambda_{wi}}} > \frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}\ (a > e). \tag{11}$$

2) In the case of $\lambda_{wi} = 0$.

If some $\lambda_{wi} = 0$, obviously $a^{\lambda_{wi}} = e^{\lambda_{wi}} = 1$. Because of $\lambda_{bi} > \lambda_{wi} = 0$, in this case, one always has

$$a^{\lambda_{bi}} > a^{\lambda_{wi}}, \quad e^{\lambda_{bi}} > e^{\lambda_{wi}}. \tag{12}$$

$$a^{\lambda_{bi}} > e^{\lambda_{bi}} > \lambda_{bi}\ (a > e). \tag{13}$$

Combining the two cases, i.e., by the inequality (11)~(13), the non-linear mapping function Θ of GEDA has the effect of distance diffusion. As a result, there is a difference in diffusion scale between the between- and within-class distances. The diffusion scale to the between-class distance is larger than that to the within-class distance. According to the inequality (11) and (13), the diffusion scale of GEDA is much larger than that of EDA. Hence, the distances between different class samples of GEDA are larger than that of EDA. This is the superiority of GEDA over EDA. And so, GEDA will show better discrimination power than EDA.

4. Experiment Results

In this section, we evaluate the face recognition performance of the proposed GEDA method. The experiments are made on the two public face image databases: the Extended Yale and CMU-PIE. The proposed GEDA method is compared with the classical PCA [14], LDA+PCA [1], EDA [2]. In order to show the performance of GEDA, the base a of the exponential function is taken as 10,100 and 1000 respectively.

A random subset with p images for each individual is taken to form the training set, and the remaining images are used as the testing set. Note that LDA method may get the maximal $C - 1$ subspace dimension, where C is the number of classes. So we choose the $C - 1$ subspace dimension for all methods to compare. For each given p, the 30 times random sample splits are made to get a stable recognition result. Then the average value of the 30 recognition accuracies, from the 30 times random sample splits, is regarded as the recognition ratio of the corresponding method. The results are illustrated in Table 1 and Table 2 respectively.

In general, the recognition performance varies with the subspace dimension. Figure 2 and Fig. 3 show the recognition rates versus subspace dimension of the above methods on the Extended Yale and CMU-PIE face database respectively. These obtained plots are the average values over 30 random splits.

From the experiment results, we can observe that: compared with the other methods, GEDA has better performance among all subspace dimensions in all the database. And if the base of the exponential function is larger, the discriminant performance is better.

Table 1 Average recognition accuracy of the different methods on Extended Yale database over 30 random splits.

Method	5 trains(%)	7 trains(%)	9 trains(%)
PCA	38.30	40.62	41.06
LDA+PCA	53.18	59.90	62.30
EDA	54.72	59.64	64.01
GEDA(a=10)	57.79	62.63	68.13
GEDA(a=100)	60.47	64.30	71.02
GEDA(a=1000)	62.17	64.95	71.74

Table 2 Average recognition accuracy of the different methods on PIE database over 30 random splits.

Method	8 trains(%)	10 trains(%)	15 trains(%)
PCA	71.11	73.59	86.98
LDA+PCA	75.68	77.48	88.47
EDA	77.85	79.57	90.11
GEDA(a=10)	82.19	81.39	91.73
GEDA(a=100)	85.76	84.96	93.20
GEDA(a=1000)	87.26	86.46	93.17

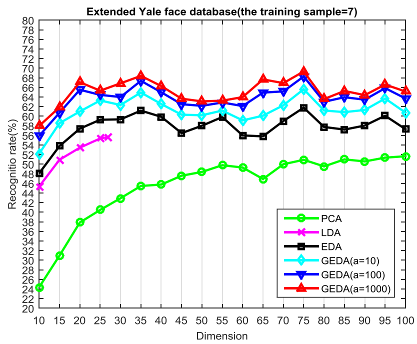

Fig. 2 Recognition accuracy versus the projected dimensions on the Extended Yale face database (seven training images)

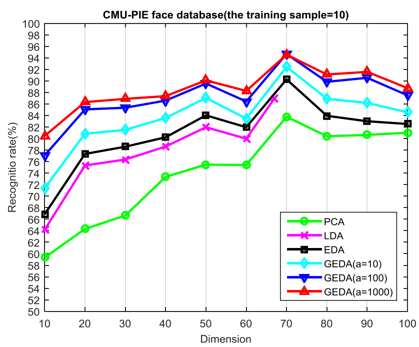

Fig. 3 Recognition accuracy versus the projected dimensions on the CMU-PIE face database (ten training images)

5. Conclusions

In this paper, a generalized exponential matrix discriminant analysis (GEDA) is proposed to improve the EDA method. The proposed GEDA method is equivalent to transforming the scatter matrices to a new space by distance diffusion mapping, and then, the LDA criterion is applied in such a space. The main idea of the improvement is that the general exponential function is used, not the Euler exponential function. It is a simple generalization of the EDA method, but it shows advantageous classification performance over EDA.

Acknowledgments

This research was supported by National Nature Science Foundation of China (61472053, 91420102), the Project of Chongqing Foundation and Frontier Research (cstc2016jcyjA0419) and the School Fund Project of CQNU (16XLB006).

References

[1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.711–720, 1997.

[2] T.P. Zhang, B. Fang, Y.Y. Tang, Z. Shang, and B. Xu, "Generalized discriminant analysis: A matrix exponential approach," IEEE Trans. Syst. Man, Cybern. B, Cybern., vol.40, no.1, pp.186–197, 2010.

[3] G. Wu, T.-T. Feng, L.-J. Zhang, and M. Yang, "Inexact implementation using Krylov subspace methods for large scale exponential discriminant analysis with applications to high dimensionality reduction problems," Patt. Recog., vol.66, pp.328–341, 2017.

[4] S.-J. Wang, S.C. Yan, J. Yang, C.-G. Zhou, and X. Fu, "A General Exponential Framework for Dimensionality Reduction," IEEE Trans. Image Process., vol.23, no.2, pp.920–930, 2014.

[5] X.F. He and P. Niyogi, "Locality preserving projections," Advances in Neural Information Processing Systems (NIPS), vol.16, no.1, pp.186–197, 2003.

[6] W.W. Yu, X.L. Teng, and C.Q. Liu, "Face recognition using discriminant locality preserving projections," Image Vis. Comput., vol.24, no.3, pp.239–248, 2006.

[7] H. Chen, H. Chang, and T. Liu, "Local discriminant embedding and its variants," Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., San Diego, pp.846–853, 2005.

[8] H. Huang, J. Liu, and Y. Pan, "Semi-supervised marginal fisher analysis for hyperspectral image classification," ISPRS Ann. Photogrammetry-Remote Sens. Spatial Inf. Sci., vol.I-3, pp.377–382, 2012.

[9] S.-J. Wang, H.-L. Chen, X.-J. Peng, and C.-G. Zhou, "Exponential locality preserving projections for small sample size problem," Neurocomputing, vol.74, no.17, pp.3654–3662, 2011.

[10] S.C. Huang and L. Zhuang, "Exponential Discriminant Locality Preserving Projection for face recognition," Neurocomputing, vol.208, pp.373–377, 2016.

[11] F. Dornaika and A. Bosaghzadeh, "Exponential Local Discriminant Embedding and Its Application to Face Recognition," IEEE Trans. Cybern., vol.43, no.3, pp.921–934, 2013.

[12] F. Dornaika and Y.E. Traboulsi, "Matrix exponential based semi-supervised discriminant embedding for image classification," Pattern Recognition, vol.6, pp.192–103, 2017.

[13] G.H. Golub and C.F.V. Loan, Matrix Computation, The Johns Hopkins Univ. Press, Baltimore, 1983.

[14] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol.3, no.1, pp.71–86, 1991.