

RESEARCH

Open Access



# Regularized linear discriminant analysis via a new difference-of-convex algorithm with extrapolation

Chunyan Wang<sup>1</sup>, Wenjie Wang<sup>1\*</sup> and Mengzhen Li<sup>1</sup>

\*Correspondence:

2582097913@qq.com

<sup>1</sup>School of Management Science,  
Qufu Normal University, 276826,  
Rizhao, Shandong, P.R. China

## Abstract

In this paper, we transform the classical linear discriminant analysis (LDA) into a smooth difference-of-convex optimization problem. Then, a new difference-of-convex algorithm with extrapolation is introduced and the convergence of the algorithm is established. Finally, for a face recognition problem, the proposed algorithm achieves better classification performance compared with several current algorithms in the literature.

**Mathematics Subject Classification:** 15A18; 15A69

**Keywords:** Regularized linear discriminant analysis; Difference-of-convex; Face recognition

## 1 Introduction

Dimensionality reduction (DR) plays an important role in pattern recognition, and it has been studied extensively. Several kinds of DR methods are widely studied, such as principal component analysis (PCA) [6], linear discriminant analysis (LDA) [1], etc. Among them, LDA is a powerful tool for feature extraction and has been extensively studied, including multimodal DR [12], audiovisual speech recognition [8], and tensor extension on image representation [3, 11, 13]. In LDA the dimensionality is reduced from a  $d$ -dimensional space to an  $h$ -dimensional space (where  $h < d$ ). LDA tries to find the optimal projection direction by maximizing the between-class variance while simultaneously minimizing the within-class variance in the projected space.

So far, there have been two major concerns for traditional LDA. First, the within-class scatter matrix is singular and cannot be inverted. Although one can use the generalized inverse instead, the estimate will be very unstable due to lack of observations. Second, high dimensionality makes direct matrix operation formidable, hence hindering the applicability of this method. To resolve the singularity problem, the authors [2] gave a regularized linear discriminant analysis (RLDA), i.e., added a multiple of identity matrix  $\gamma I$  to the within-class scatter matrix  $S_w$ .

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

By introducing a slightly biased covariance estimate, not only the singularity problem, but also the stabilization of the sample covariance estimate is solved. However, the difficulties caused by high dimensional matrix direct operations have not been solved.

In this paper, motivated by [2] and [5], we transform the LDA into a smooth difference-of-convex optimization problem, a new difference-of-convex algorithm with extrapolation is introduced and the convergence of the algorithm is established. Furthermore, the proposed new RLDA can resolve the singularity problem. More importantly, the algorithm has great advantages in computation time and the number of iterations. Finally, we prove the convergence of this algorithm, and the new RLDA achieves better classification performance compared with some other algorithms for face recognition.

The article is organized as follows. In Sect. 2, we recall some useful notations and definitions. In Sect. 3, the new RLDA is given, and a new difference-of-convex algorithm with extrapolation is introduced to solve the new RLDA. Then the convergence of the generated subsequence is given. Numerical results are given in Sect. 4. Finally, Sect. 5 concludes this paper.

## 2 Notation and preliminaries

We now define the notation used in this paper. All vectors are column vectors. Given a training data set  $T = \{(\mathbf{x}_1, l_1), \dots, (\mathbf{x}_m, l_m)\}$ , where  $\mathbf{x}_t \in \mathbb{R}^n$  is the input and  $l_t \in \{1, \dots, c\}$  is the corresponding label,  $t = 1, \dots, m$ , we organize the  $m$  inputs by a matrix  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m) \in \mathbb{R}^{n \times m}$ . Assume that the  $i$ th class contains  $m_i$  samples. Then  $\sum_{i=1}^c m_i = m$ . Denote  $\bar{\mathbf{x}}_i$  as the mean of samples in the  $i$ th class and  $\bar{\mathbf{x}}$  as the center of the whole set of samples, that is,  $\bar{\mathbf{x}}_i = (1/m_i) \sum_{j=1}^{m_i} \mathbf{x}_{ij}$  and  $\bar{\mathbf{x}} = (1/m) \sum_{l=1}^m \mathbf{x}_l$ , where  $x_{ij}$  is the  $j$ th element in the  $i$ th class. Based on this, the following matrices are useful in the following analysis:

$$\mathbf{S}_b = \frac{1}{m} \sum_{i=1}^c m_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^\top, \quad (1)$$

$$\mathbf{S}_w = \frac{1}{m} \sum_{i=1}^c \sum_{j=1}^{m_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^\top. \quad (2)$$

For  $\mathbf{w} \in \mathbb{R}^n$  and a sample  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{w}^\top \mathbf{x}$  maps  $\mathbf{x}$  into a 1-D vector. Generally, if  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_d) \in \mathbb{R}^{n \times d}$  with  $d \leq n$ , then  $\mathbf{W}^\top \mathbf{x}$  maps each  $\mathbf{x} \in \mathbb{R}^n$  into a  $d$ -dimensional space.

As a supervised dimensionality reduction method, LDA aims at finding the optimal transformation vector  $\mathbf{w}_1, \dots, \mathbf{w}_d$ ,  $d \leq c - 1$  that maximizes the Rayleigh coefficient

$$J_{\text{LDA}}(\mathbf{w}_h) = \frac{\mathbf{w}_h^\top \mathbf{S}_b \mathbf{w}_h}{\mathbf{w}_h^\top \mathbf{S}_w \mathbf{w}_h} \quad (3)$$

such that  $\mathbf{w}_h^\top \mathbf{S}_w \mathbf{w}_l = 0$ ,  $1 \leq l < h \leq d$ . It is proved that (3) can be reduced to the following eigenvalue decomposition problem:

$$\mathbf{S}_b \mathbf{w}_h = \lambda \mathbf{S}_w \mathbf{w}_h, \quad (4)$$

where  $\mathbf{S}_w$  is nonsingular, and  $\lambda \neq 0$ . Since the rank of  $\mathbf{S}_b$  is at most  $c - 1$ , the number of extracted features is less than or equal to  $c - 1$ .

### 3 New DC algorithm for the new form of RLDA

As mentioned before, the classical LDA requires that  $S_w$  is nonsingular. In addition, high dimensionality makes direct matrix operation formidable, hence hindering the applicability of the method proposed in [2].

To solve such a problem, RLDA adds a multiple of identity matrix  $\gamma I$  to the within-class scatter matrix  $S_w$ . It is noted that the regularization parameter  $\gamma$  is larger than zero. The corresponding objective function and eigenvalue decomposition problem become

$$J_{RLDA}(\mathbf{w}_h) = \frac{\mathbf{w}_h^T S_b \mathbf{w}_h}{\mathbf{w}_h^T (S_w + \gamma I) \mathbf{w}_h} \tag{5}$$

and

$$(S_w + \gamma I)^{-1} S_b \mathbf{w}_h = \lambda \mathbf{w}_h. \tag{6}$$

For (5), we transform a fraction problem into a difference-of-convex problem, and a new difference-of-convex algorithm with extrapolation is proposed to solve the problem efficiently. We construct a new form of RLDA with a minus operator as follows. In this paper, we only consider the binary classification case, i.e.,  $d = 1$ . The formulation is given as follows:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} f(\mathbf{w}) = -\mathbf{w}^T S_b \mathbf{w} + \lambda \mathbf{w}^T S_w \mathbf{w} + \gamma \mathbf{w}^T \mathbf{w}, \tag{7}$$

where  $\lambda$  is a positive tuning parameter and  $\gamma$  is a positive regularization parameter.

The geometric interpretation of problem (7) is clear. Optimizing the first term of (7) means maximizing the scatter between class scatter, which forces the data points from different classes to be as far as possible; whereas minimizing the second term of (7) makes the within-class scatter as small as possible. The third term is the regularization item and it can be avoided if  $S_w$  is nonsingular.

It is obvious that (7) is a smooth difference-of-convex optimization problem. For this problem, we proposed a new DC algorithm with extrapolation.

To move on, let

$$g(\mathbf{w}) = \lambda \mathbf{w}^T S_w \mathbf{w} + \gamma \mathbf{w}^T \mathbf{w}, \quad h(\mathbf{w}) = \mathbf{w}^T S_b \mathbf{w}.$$

Then  $f(\mathbf{w}) = g(\mathbf{w}) - h(\mathbf{w})$ .

Obviously,  $g(\mathbf{w})$  and  $h(\mathbf{w})$  are smooth convex functions. Motivated by [4, 7], we introduce a new DC algorithm with extrapolation to find stationary points of the smooth problem  $f(\mathbf{w})$ .

We note from

$$\mathbf{w}^{t+1} = \arg \min_{\bar{\mathbf{w}}} \left\{ \lambda \bar{\mathbf{w}}^T S_w \bar{\mathbf{w}} + \gamma \bar{\mathbf{w}}^T \bar{\mathbf{w}} - \langle 2S_b \bar{\mathbf{w}}^t, \bar{\mathbf{w}} \rangle + \frac{L}{2} \|\bar{\mathbf{w}} - \bar{\mathbf{w}}^t\|^2 \right\} \tag{8}$$

that  $\mathbf{w}^{t+1}$  is the global minimizer of a strongly convex function.

In this algorithm, motivated by [7], we set

$$\beta_t = \frac{\theta_{t-1} - 1}{2\theta_t} \quad \text{with } \theta_{t+1} = \frac{1 + \sqrt{1 + 4\theta_t^2}}{2}, \theta_{-1} = \theta_0 = 1.$$

---

**Algorithm 1** A new DC algorithm with extrapolation ( $DC_e$ ) for (7)

---

**Input:**  $\mathbf{w}^0 \in \text{dom } P_1, \{\beta_t\} \subseteq [0, \frac{1}{2})$  with  $\sup \beta_t < \frac{1}{2}$ . Set  $\mathbf{w}^{-1} = \mathbf{w}^0$

**Process:**

for  $t = 0, 1, 2, \dots$

Set

$$\bar{\mathbf{w}}^t = \mathbf{w}^t + \beta_t(\mathbf{w}^t - \mathbf{w}^{t-1}),$$

$$\begin{aligned} \mathbf{w}^{t+1} &= \arg \min_{\bar{\mathbf{w}}} \langle \nabla h(\bar{\mathbf{w}}^t), \bar{\mathbf{w}} \rangle + \frac{L}{2} \|\bar{\mathbf{w}} - \bar{\mathbf{w}}^t\|^2 \\ &= \arg \min_{\bar{\mathbf{w}}} \{\lambda \bar{\mathbf{w}}^\top \mathbf{S}_w \bar{\mathbf{w}} + \gamma \bar{\mathbf{w}}^\top \bar{\mathbf{w}} - \langle 2\mathbf{S}_b \bar{\mathbf{w}}^t, \bar{\mathbf{w}} \rangle + \frac{L}{2} \|\bar{\mathbf{w}} - \bar{\mathbf{w}}^t\|^2\}. \end{aligned}$$

end for

---

In what follows, we prove a global subsequential convergence result of Algorithm 1, which is applied to solving (7).

**Theorem 1** (*Global subsequential convergence*) Let  $\{\mathbf{w}^t\}$  be a sequence generated by the DC algorithm with extrapolation for solving (7). Then the following statements hold.

- (i) The sequence  $\{\mathbf{w}^t\}$  is bounded.
- (ii)  $\lim_{t \rightarrow \infty} \|\mathbf{w}^{t+1} - \mathbf{w}^t\| = 0$ .
- (iii) The accumulation point of  $\{\mathbf{w}^t\}$  is a stationary point of (7).

*Proof* First we prove (i). We note from (8) that  $\mathbf{w}^{t+1}$  is the global minimizer of a strongly convex function. Using this and comparing the objective values of this strongly convex function at  $\mathbf{w}^{t+1}$  and  $\mathbf{w}^t$ , we see immediately that

$$\begin{aligned} g(\mathbf{w}^{t+1}) - \langle \nabla h(\bar{\mathbf{w}}^t), \mathbf{w}^{t+1} \rangle + \frac{L}{2} \|\mathbf{w}^{t+1} - \bar{\mathbf{w}}^t\|^2 \\ \leq g(\mathbf{w}^t) - \langle \nabla h(\bar{\mathbf{w}}^t), \mathbf{w}^t \rangle + \frac{L}{2} \|\mathbf{w}^t - \bar{\mathbf{w}}^t\|^2 - \frac{L}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2. \end{aligned} \tag{9}$$

Then we have

$$\begin{aligned} f(\mathbf{w}^{t+1}) &\leq g(\mathbf{w}^{t+1}) - h(\bar{\mathbf{w}}^t) - \langle \nabla h(\bar{\mathbf{w}}^t), \mathbf{w}^{t+1} - \bar{\mathbf{w}}^t \rangle \\ &\leq g(\mathbf{w}^t) - h(\bar{\mathbf{w}}^t) - \langle \nabla h(\bar{\mathbf{w}}^t), \mathbf{w}^t - \bar{\mathbf{w}}^t \rangle + \frac{L}{2} \|\mathbf{w}^t - \bar{\mathbf{w}}^t\|^2 - \frac{L}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 \\ &\leq g(\mathbf{w}^t) - h(\mathbf{w}^t) + L \|\mathbf{w}^t - \bar{\mathbf{w}}^t\|^2 - \frac{L}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 \\ &\leq f(\mathbf{w}^t) + L \|\mathbf{w}^t - \bar{\mathbf{w}}^t\|^2 - \frac{L}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2, \end{aligned} \tag{10}$$

where the first inequality follows from the convexity of  $h(\mathbf{w})$ , the second inequality follows from (9), the third inequality follows from the fact that  $\nabla h$  is Lipschitz continuous with a modulus of  $L > 0$ . Now, invoking the definition of  $\bar{\mathbf{w}}$ , we obtain further from (10) that

$$f(\mathbf{w}^{t+1}) \leq f(\mathbf{w}^t) + L\beta_t^2 \|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 - \frac{L}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2. \tag{11}$$

Consequently, we have upon rearranging terms that

$$L\left(\frac{1}{2} - \beta_t^2\right)\|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 \leq \left[ f(\mathbf{w}^t) + \frac{L}{2}\|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 \right] - \left[ f(\mathbf{w}^{t+1}) + \frac{L}{2}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 \right]. \tag{12}$$

Since  $\{\beta_t\} \subset [0, \frac{1}{2})$ , we deduce from (12) that the sequence

$$\left\{ f(\mathbf{w}^t) + \frac{L}{2}\|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 \right\}$$

is nonincreasing. This together with the fact that  $\mathbf{w}^0 = \mathbf{w}^{-1}$  gives

$$f(\mathbf{w}^t) \leq f(\mathbf{w}^t) + \frac{L}{2}\|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 \leq f(\mathbf{w}^0)$$

for all  $t \geq 0$ , which shows that  $\{\mathbf{w}^t\}$  is bounded. This proves (i).

Next we prove (ii). Summing both sides of (12) from  $t = 0$  to  $\infty$ , we obtain that

$$\begin{aligned} & L \sum_{t=0}^{\infty} \left(\frac{1}{2} - \beta_t^2\right)\|\mathbf{w}^t - \mathbf{w}^{t-1}\|^2 \\ & \leq f(\mathbf{w}^0) - \liminf_{t \rightarrow \infty} \left[ f(\mathbf{w}^{t+1}) + \frac{L}{2}\|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 \right] \\ & \leq f(\mathbf{w}^0) - \nu < \infty. \end{aligned}$$

Since  $\sup \beta_t < \frac{1}{2}$ , we deduce immediately from the above relation that

$$\lim_{t \rightarrow \infty} \|\mathbf{w}^{t+1} - \mathbf{w}^t\| = 0.$$

This proves (ii).

Finally, let  $\mathbf{w}^*$  be an accumulation point of  $\{\mathbf{w}^t\}$  and let  $\{\mathbf{w}^{t_i}\}$  be a subsequence such that  $\lim_{i \rightarrow \infty} \mathbf{w}^{t_i} = \mathbf{w}^*$ . Then, from the first-order optimality condition of subproblem (8), we have

$$-L(\mathbf{w}^{t_i+1} - \bar{\mathbf{w}}^{t_i}) = \nabla g(\mathbf{w}^{t_i+1}) - \nabla h(\bar{\mathbf{w}}^{t_i}).$$

Using this together with the fact that  $\bar{\mathbf{w}}^{t_i} = \mathbf{w}^{t_i} + \beta_{t_i}(\mathbf{w}^{t_i} - \mathbf{w}^{t_i-1})$ , we obtain further that

$$-L[(\mathbf{w}^{t_i+1} - \mathbf{w}^{t_i}) - \beta_{t_i}(\mathbf{w}^{t_i} - \mathbf{w}^{t_i-1})] = \nabla g(\mathbf{w}^{t_i+1}) - \nabla h(\bar{\mathbf{w}}^{t_i}). \tag{13}$$

In addition,  $\|\mathbf{w}^{t_i+1} - \mathbf{w}^{t_i}\| \rightarrow 0$  from (ii) together with the continuity of  $\nabla g$  and of  $\nabla h$ , we have upon passing to the limit in (13) that

$$0 = \nabla g(\mathbf{w}^*) - \nabla h(\mathbf{w}^*).$$

This completes the proof. □

#### 4 Numerical examples

In this section, experimental results are given to evaluate the performance of the proposed method. Several related DR methods, including RLDA, RSLDA [5], and PDCA [7], are recalled for comparison. For RLDA and RSLDA [5], parameters  $\rho$  and  $\lambda$  are selected from  $\{0.1, 0.5, 1, 5, 10\}$  and  $\{0.1, 0.5, 1, 5, 10\}$ , respectively. We choose  $\delta$  for RSLDA from  $\{0.01, 0.05, 0.1, 0.5, 1, 5\}$ . For our method, parameters  $\lambda$  and  $\gamma$  are selected from  $\{0.1, 0.5, 1, 5, 10, 50, 100\}$  and  $\{0.01, 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9\}$ , respectively. Note that the parameters for all the methods are optimally selected from their own sets. Numerical experiments are done in Matlab R2018b on a laptop computer with Intel(R), Core(TM), CPU i7-8550U @ 1.80GHz, and 16 GB memory running Microsoft Windows 10.

To show the effectiveness of the proposed method, we focus on testing the proposed algorithm based on human face datasets that are included in FERET and ORL. The FERET dataset includes 200 persons' images, and each one has 7 different images. Each image is  $80 \times 80$  with 256 grayscale levels per pixel. The ORL dataset includes 40 individuals' face images, and each face has 10 different images. Each image is  $112 \times 92$  with 256 grayscale levels per pixel. Figure 1 and Fig. 2 show partial sample faces of FERET and ORL databases. A random subset with  $p(2, 3, \dots, 10)$  images per subject is taken to form the training set, while the rest of the data comprise the test set. For each given  $p$ , the average result over



Figure 1 FERET



Figure 2 ORL

ten random splits is considered. Figure 1 and Fig. 2 show partial sample faces from FERET and ORL databases as follows.

The classification accuracy is used as an indicator to test the performance of the methods. The corresponding numerical results of the algorithms are listed in Tables 1 and 2 respectively. Here, “iter” denotes the number of iterations, time is measured in seconds, and “tnr” denotes the classification accuracy.

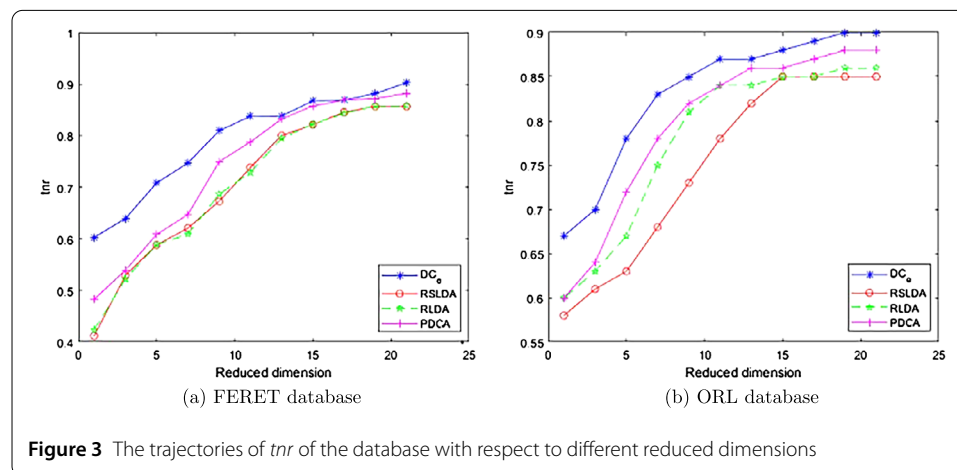
It can be easily seen from Table 1 and Table 2 that the proposed Algorithm 1 requires fewer iterations and less computing time to achieve higher accuracy than both RSLDA and RLDA. Furthermore, we show the relationship between reduced dimension and classification accuracies in Fig. 3. From the figure, we can see that for all of the methods, the accuracies all have ascending trends in terms of the increase of reduced dimensions in general. Algorithm 1 is more superior.

**Table 1** Numerical results for the data of FERET

$\rho$	RLDA	RSLDA	PDCA	DC <sub>e</sub>
	iter/time/tnr	iter/time/tnr	iter/time/tnr	iter/time/tnr
2	12/0.1874/70.74	13/0.1631/64.71	2/0.0123/72.73	2/0.0024/76.67
3	10/0.2114/76.08	10/0.1812/70.59	2/0.0138/76.92	2/0.0062/80.59
4	9/0.2329/77.70	9/0.1912/72.22	2/0.0146/78.47	4/0.0105/85.00
5	11/0.2532/83.34	9/0.1948/83.23	4/0.0171/84.62	4/0.0113/86.92
6	10/0.2793/86.3	9/0.2450/85.75	4/0.030/87.14	4/0.0226/87.25
7	9/0.3258/93.91	9/0.2858/93.33	4/0.0388/94.86	5/0.0394/95.31

**Table 2** Numerical results for the data of ORL

$\rho$	RLDA	RSLDA	PDCA	DC <sub>e</sub>
	iter/time/tnr	iter/time/tnr	iter/time/tnr	iter/time/tnr
2	12/0.1341/54.21	11/0.1166/52.94	2/0.0129/58.33	2/0.0033/66.67
3	9/0.1774/63.84	14/0.1524/66.67	3/0.0158/69.23	3/0.0081/70.59
4	11/0.2142/78.87	12/0.1849/76.47	3/0.0146/80.00	3/0.0142/80.47
5	12/0.2509/79.66	13/0.2229/80.65	4/0.0230/81.25	4/0.0233/83.33
6	8/0.3047/84.04	8/0.2850/82.35	4/0.0379/83.33	4/0.0406/85.00
7	8/0.3563/85.49	8/0.3270/85.29	5/0.0882/87.50	4/0.0610/87.89
8	10/0.4387/87.00	9/0.3995/86.24	6/0.1007/86.67	4/0.1449/88.24
9	11/0.4849/90.91	10/0.4581/90.18	7/0.1710/92.12	4/0.1788/90.91
10	9/0.6017/92.95	9/0.5774/92.12	9/0.2906/93.75	5/0.2864/95.95



**Figure 3** The trajectories of *tnr* of the database with respect to different reduced dimensions

## 5 Conclusions

In this paper, a new RLDA is proposed. A new DC algorithm with extrapolation is introduced for a smooth DC problem, and the convergence of this algorithm is given. Numerical results show that the proposed algorithm achieves better classification performance compared with current algorithms for face recognition. In the future, we may consider several more practical applications of RLDA in optimal control and so on [9, 10].

### Acknowledgements

The authors are very grateful to the reviewers for several valuable and helpful comments, suggestions, and questions, which helped to improve the paper into its present form.

### Funding

This work was supported by the Natural Science Foundation of China (12071249, 12071250) and Shandong Provincial Natural Science Foundation of Distinguished Young Scholars (ZR2021JQ01).

### Availability of data and materials

The data used to support the findings of this study are available from the corresponding author upon request.

## Declarations

### Competing interests

The authors declare no competing interests.

### Author contributions

All authors contributed equally to this work. All authors read and approved the final manuscript.

Received: 28 September 2022 Accepted: 19 June 2023 Published online: 07 July 2023

## References

1. Fukunaga, K.: Introduction to Statistical Pattern Recognition, 2nd edn. Academic Press, New York (1991)
2. Guo, Y., Hastie, T., Tibshirani, R.: Regularized discriminant analysis and its application in microarrays. *Pattern Recognit.* **8**(1), 86–100 (2003)
3. Huang, R., Liu, C., Zhou, J.: Discriminant analysis via jointly  $L_{2,1}$ -norm sparse tensor preserving embedding for image classification. *J. Vis. Commun. Image Represent.* **47**, 10–22 (2017)
4. Le Thi, H.A., Pham Dinh, T., Muu, L.D.: Numerical solution for optimization over the efficient set by D.C. optimization algorithms. *Oper. Res. Lett.* **19**(3), 117–128 (1996)
5. Li, C.N., Shao, Y.H., Yin, W., Liu, M.Z.: Robust and sparse linear discriminant analysis via an alternating direction method of multipliers. *IEEE Trans. Neural Netw. Learn. Syst.* **31**(3), 915–926 (2019)
6. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
7. Wen, B., Chen, X., Pong, T.K.: A proximal difference-of-convex algorithm with extrapolation. *Comput. Optim. Appl.* **69**(2), 297–324 (2018)
8. Zeiler, S., Nicheli, R., Ma, N., Brown, G.J., Kolossa, D.: Robust audiovisual speech recognition using noise-adaptive linear discriminant analysis. In: *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 2797–2801 (2016)
9. Zhang, X., Wang, T.: Elastic and reliable bandwidth reservation based on distributed traffic monitoring and control. *IEEE Trans. Parallel Distrib. Syst.* **33**(12), 4563–4580 (2022). <https://doi.org/10.1109/TPDS.2022.3196840>
10. Zhang, X., Wang, Y., Geng, G., Yu, J.: Delay-optimized multicast tree packing in software-defined networks. *IEEE Trans. Serv. Comput.* **16**(1), 261–275 (2021). <https://doi.org/10.1109/TSC.2021.3106264>
11. Zhang, Z., Chow, W.S.: Tensor locally linear discriminative analysis. *IEEE Signal Process. Lett.* **18**(11), 643–646 (2011)
12. Zhang, Z., Zhao, M., Chow, T.W.S.: Constrained large margin local projection algorithms and extensions for multimodal dimensionality reduction. *Pattern Recognit.* **45**(12), 4466–4493 (2012)
13. Zhao, J., Shi, L., Zhu, J.: Two-stage regularized linear discriminant analysis for 2-D data. *IEEE Trans. Neural Netw. Learn. Syst.* **26**(8), 1669–1681 (2015)

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.