Reply to Editor and Reviewers for the manuscript " Generalized Linear Discriminant Analysis based on Trace Ratio Criterion for Feature Extraction in Classification"

We would like to express our sincere appreciation to the **Associate Editor** and the two **Reviewers** for their valuable comments and professional handling of the manuscript. We have carefully gone through the paper and made significant changes based on the comments, as explained in our replies below.

1. To Associate Editor

We sincerely appreciate your professional handling of our paper. We would like to highlight the major changes as follows:

1. Based on the comments from Reviewer 1 and Reviewer 2, we have rewritten the Introduction and overview more LDA based methods and discussed the differences of our proposed method. To this end, a lot of related references have been included in the newly revised manuscript.
2. More simulations examples and results has been illustrated to compare our method with existing methods as suggested by Reviewer 1 and

Reviewer 2. Specifically, we added one more example (Example 2), 4 figures (Figures 2-6) and two tables (Tables 1 and 2) in the revised manuscript.

3. As suggested by Reviewer 2, the relationships between the formulas and the feature extraction problem are clarified. Our main theoretical conclusions are given in Theorems 3.6-3.8, which are also verified in the Simulation Section. Lemmas 3.1-3.5 are presented and proven as they are needed in proving the Theorems.

4. The spelling mistakes have also been carefully checked.

2. Reply to Reviewer 1

We sincerely appreciate your comments and recommendation. In the following, we would like to answer how your comments and suggestions are taken into account in revising our paper.

Comment 1: In this paper, a generalized linear discriminant analysis based on trace ratio criterion (GLDA-TRA) algorithm has been proposed. This is to overcome the problem that linear discriminant analysis (LDA) can only extract limited features in classification. It is shown that, in GLDA-TRA, a set of orthogonal features can be extracted one by one.

My detail comments are as follows: In the introduction section, the authors should overview the LDA Cbased methods in the pass and recent years, and update the new references. Please refer to the following two papers and its references:

Zhihui Lai et al, Sparse tensor discriminant analysis. IEEE Transactions on Image Processing, 2013, 22(10): 3904-3915.

Xiaoshuang Shi et al., Face Recognition by Sparse Discriminant Analysis via Joint L2,1-norm Minimization, Pattern Recognition.

Answer 1: Thanks for your suggestion. We have now updated the references list and rewritten the Introduction section. The major changes can be seen in Paragraphs 2-5, Introduction section.

For you convenience, these three Paragraphs 2-5 are also given in this reply:

In our opinion, there are three main problems for LDA methods. The ?rst problem is its difficulty of dealing with the high-dimensional data, where the number of observed samples is much lower than the samples feature dimension [5]. Many methods have been studied and proposed to address this problem, see for examples, the Linear Discriminant Feature Selection (LDFS) [6], Sparse Discriminant Analysis [5] [7], and Sparse Tensor Discriminant Analysis [8].

The second problem is the well known under-sampled problem [9] in LDA mehtod, in which scatter matrices may become singular due to the insufficient samples. The solutions of this problem have also been well investigated in the following works such as the Regularized LDA [10][11] using regularization techniques [12] [13] and the Penalized LDA [14], the Pseudo Fisher Linear Discriminant [15], the Generalized Singular Value Decomposition [16], and the Uncorrelated LDA [17] and the Orthogonal LDA [17].

Basically the first two problems are quite similar and they can be considered as the same one problem, which has been extensively investigated in above schemes. However, the third problem is tricky since LDA method can only extract quite limited features for classi?cation problems [4]. For example, applying LDA in two-class classi?cation problem, one can only ?nd one nonzero eigenvalue (extracted feature), as the between-class scatter matrix is a rank-one matrix. To our best knowledge, currently there is no good way to deal with this problem yet.

We focused on the third problem in this paper. A generalized LDA based on trace ratio criterion [18] [19]- [22] is provided to overcome such a problem and an algorithm called generalized linear discriminant analysis based on trace ratio criterion algorithm (GLDA-TRA) is derived to extract features from the input feature space. The algorithm first extracts a feature which maximizes the trace ratio criterion by solving a generalized eigenvalue problem. Actually, it is shown that such a generalized eigenvalue problem is the same as the generalized eigenvalue problem of LDA. Then, the learning data is projected to a subspace orthogonal to the space spanned by the extracted features. In that orthogonal subspace, the algorithm continues to extract a feature which maximizes the proposed trace ratio criterion. This precess continues and in this way, a set of orthogonal features is obtained iteratively. It is proven that each newly extracted feature is the optimal feature that maximizes the trace ratio criterion in the sub-space orthogonal to the space spanned by the previous extracted features. Finally the extracted features are shown to give a sequence of trace ratio criteria corresponding to these features is monotonically decreasing in magnitude.

Comment 2: What is the difference of the proposed method and the most related method should be discussed.

Answer 2: Thanks for your comment. We have discussed this in the Introduction section. As seen in Paragraph 4, it is stated that:

Basically the first two problems are quite similar and they can be considered as the same one problem, which has been extensively investigated in above schemes. However, the third problem is tricky since LDA method can only extract quite limited features for classi?cation problems [4]. For example, applying LDA in two-class classi?cation problem, one can only ?nd one nonzero eigenvalue (extracted feature), as the between-class scatter matrix is a rank-one matrix. To our best knowledge, currently there is no good way to deal with this problem yet.

Comment 3: In the experimental parts, the authors can fix the S_b and S_w to compute the different discriminant subspace by using different computing methods, and compare it similar to Figure 1, and also list the first projection of different methods in one table. Particularly to compare the two methods:

Trance ratio vs ration trace for dimensionality reduction,YE, J. 2005.

Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. J. Mach. Learn. Resear. 6, 483C502.

It is expected to see the numeric solutions difference between the proposed method and the OLDA in YE, J. 2005. (I guess, the final solution of yours and OLDA(in YE, J. 2005, page 495 the last page) are very similar or even the same)

Answer 3: Thanks you very much for this comments. As mentioned in the Introduction section, in our opinion, there are three main problems in LDA. The main difference of our proposed method and the above methods is as follows: the above methods aim to deal with the under-sample problem but our proposed GLDA-TRA actually assumes that both S_b and S_w are full rank, which implies that the samples are sufficient enough. This objective of this paper is to address the problem how to extract more features such that $m > C - 1$, which is a unsolvable problem in LAD.

Comment 4: Recognition rates in recognition task should also be given.

Answer 4: Thanks you very much for this comments. We have now added 4 Figures and 2 Tables to show the performance of method and recognition rates and the comparisons.

3. Reply to Reviewer 2

We sincerely appreciate your helpful questions and suggestions to improve the quality of this paper. In the following, we would like to answer how your comments and suggestions are taken into account in revising our paper.

Comment 1: In this paper, the author proposed a generalized linear discriminant analysis based on trace ratio criterion algorithm (GLDA-TRA) to extract features for classification.This paper cannot be accepted for the following reasons:

In the INTRODUCTION part, the author should clearly describe some related works on feature extraction using LDA.

Answer 1: Thanks for this comment. We have now updated the references list and rewritten the Introduction section. The major changes can be seen in Paragraphs 2-4 in the Introduction section.

Comment 2: In section3, the author has given many mathematic formulas. However, the relationship between those formulas and feature extraction problem is not clarified.

Answer 3: Thanks for this comment in improving the quality of this paper. We would like to clarify the relationships between the feature extraction problem as follows.

We added a new Remark (Remark 3) in the beginning of the Section 3.2:

Remark 3. Our main theoretical conclusions of this paper are shown in Theorems 3.6-3.8. The implication of Theorem 3.6 is summarized in Remark 4 (which is the Remark 3 in the old version). In Theorem 3.7, the decreasing of the trace ratio sequence $\{F(w_i)\}$ means that the separability of the data set on the each newly extracted features decreases compared with the previous extracted features. This is the main result and contribution of this paper. Note that the trace ratio cost function is to evaluate whether a feature is good or not. By employing our proposed method, basically one can obtain m ($m <= d$ and m can be greater than $C - 1$) new extracted orthogonal features in which the w_i is better than or at least equal to w_{i+1}. Theorem 3.8 gives us more concrete conclusions: if we already have i features denoted as $w_1, ..., w_i$, the $i + i$th feature w_{i+1} extracted by GLDA-TRA is the optimal feature in the space orthogonal to the space spanned by $w_1, ..., w_i$. We present and prove Lemmas 3.1-3.5 as they are needed in proving the Theorems 3.6-3.8. The Theorems will be illustrated and verified the Simulation Section.

Comment 3: In part 4, the simulation work in this paper is not convincing. In example 1, the result of the simulation did not proof that GLDA-TRA is better in classification. Then in example 2, the author did not provide any simulation results so that there is no evidence that GLDA- TRA is better than LDA.

Answer 3: Thank for this comment. More detailed simulation results can be seen in the newly added Figures 2-6 (i.e., the Figures 1-4 shown in this reply later) and Tables (Tables 1 and 2). These results shows that the performance of the proposed GLDA- TRA is better than that of LDA.

Comment 4: There are some spelling mistakes in this paper. For example, in the fourth line of first paragraph, the author wrote LDA as LAD.

Answer 4: Thanks for your comment. In the newly revised manuscript, the spelling mistakes have been carefully checked.

Comment 5: Although this paper has proposed a generalized linear discriminant analysis based on trace ratio criterion algorithm, its reliability is uncertain due to its lack of substantial experiments.

Answer 5: Thanks for your comment. More simulations examples and results has been illustrated to compare our method with LDA. Specifically, we added one more example (Example 2), 4 figures (Figures 2-6) and two tables (Tables 1 and 2) in the newly revised manuscript.

For your convenience, the Tables and Figures are also shown in this reply. More detailed information can be seen in revised manuscript.

Table 1: Comparison of LDA and GLDA-TRA in the Iris data set

Name	Num. extracted features	Accuracy rate	Name	Accuracy rate
LDA	1	98%	GLDA-TRA	98%
LDA	2	98%	GLDA-TRA	98%
LDA	3	Not Available	GLDA-TRA	98.6667%
LDA	4	Not Available	GLDA-TRA	98%

Table 2: Comparison of LDA and GLDA-TRA in the Example 3

Name	Num. extracted features	Accuracy rate	Name	Accuracy rate
LDA	1	69%	GLDA-TRA	69%
LDA	2	Not Available	GLDA-TRA	77.5%
LDA	3	Not Available	GLDA-TRA	88%
LDA	4	Not Available	GLDA-TRA	92.5%
LDA	5	Not Available	GLDA-TRA	96%
LDA	6	Not Available	GLDA-TRA	97%

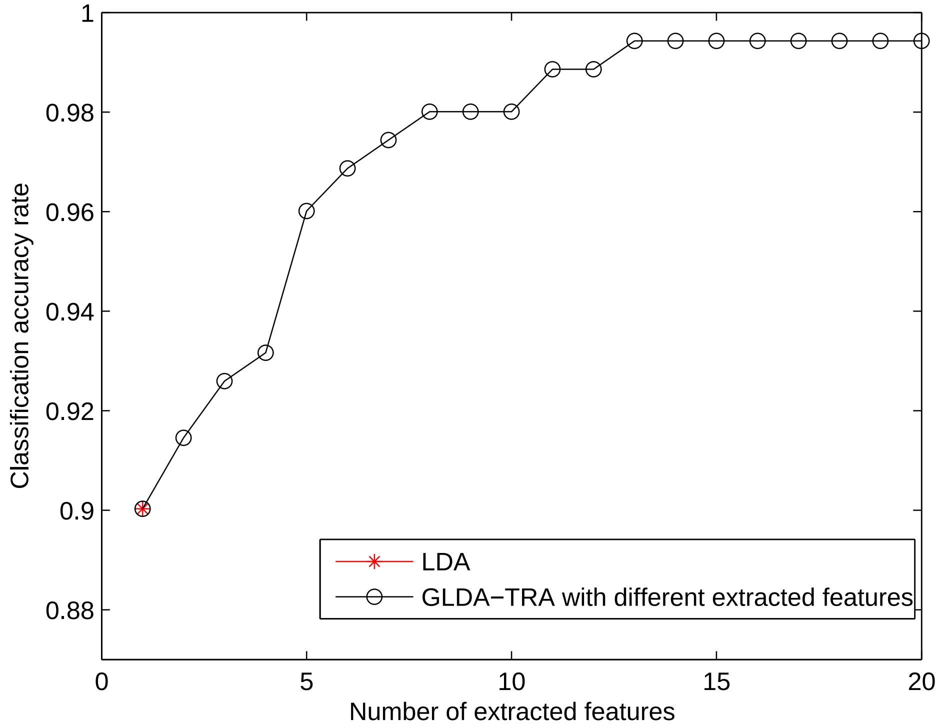

Figure 1: Classification accuracy rate for LDA and GLDA-TRA in example 2

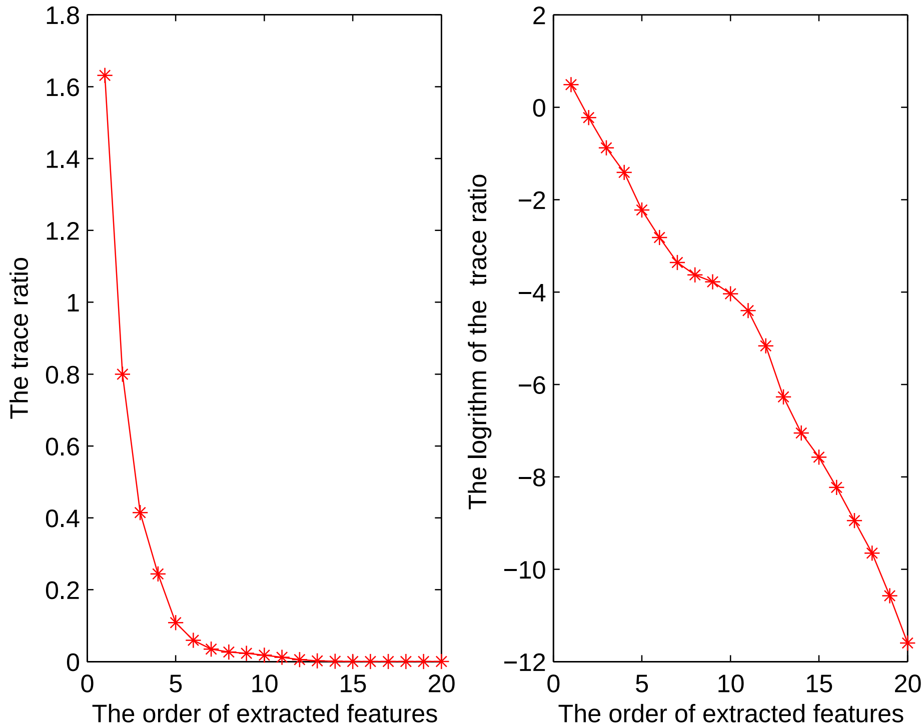

Figure 2: The trace ratio (a) & logrithm of the trace ratio (b) on each extracted feature in order

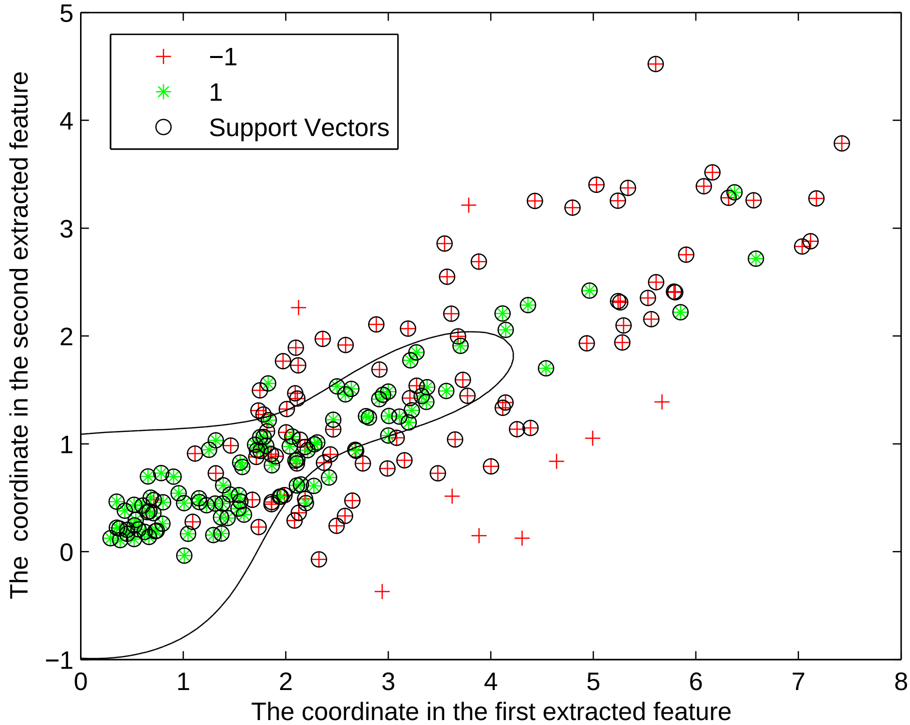

Figure 3: The data points projected onto the first two extracted features in Example 3.

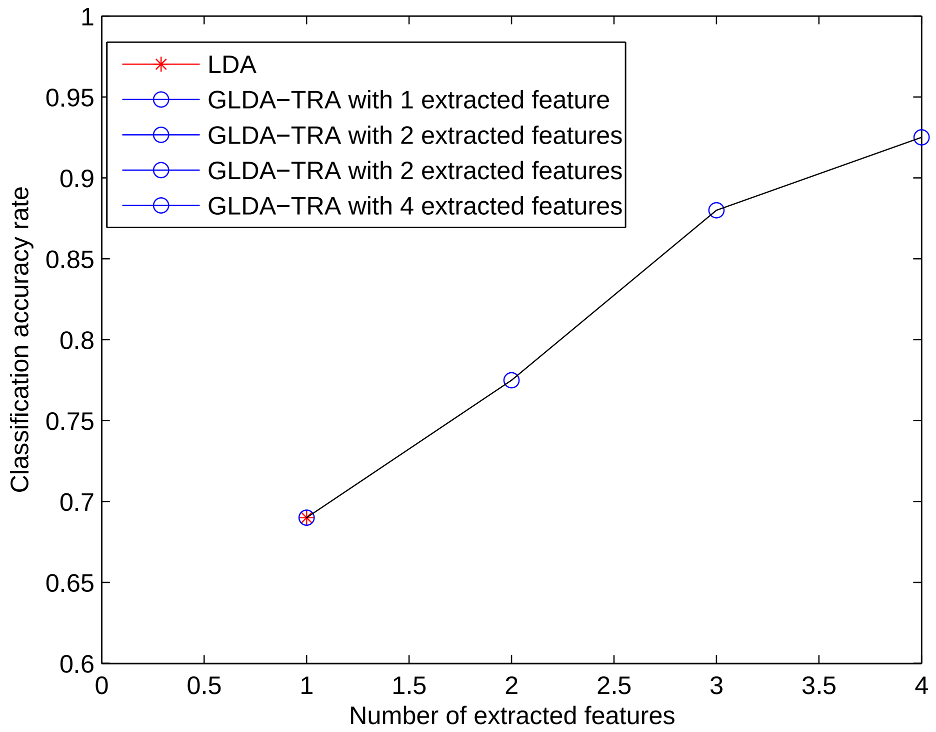

Figure 4: Classification accuracy rate for LDA and GLDA-TRA in example 3