

Regularized Class-Specific Subspace Classifier

Rui Zhang, Feiping Nie, and Xuelong Li, *Fellow, IEEE*

Abstract—In this paper, we mainly focus on how to achieve the translated subspace representation for each class, which could simultaneously indicate the distribution of the associated class and the differences from its complementary classes. By virtue of the reconstruction problem, the class-specific subspace classifier (CSSC) problem could be represented as a series of biobjective optimization problems, which minimize and maximize the reconstruction errors of the related class and its complementary classes, respectively. Besides, the regularization term is specifically introduced to ensure the whole system's stability. Accordingly, a regularized class-specific subspace classifier (RCSSC) method can be further proposed based on solving a general quadratic ratio problem. The proposed RCSSC method consistently converges to the global optimal subspace and translation under the variations of the regularization parameter. Furthermore, the proposed RCSSC method could be extended to the unregularized case, which is known as unregularized CSSC (UCSSC) method via orthogonal decomposition technique. As a result, the effectiveness and the superiority of both proposed RCSSC and UCSSC methods can be verified analytically and experimentally.

Index Terms—Class-specific subspace, quadratic ratio problem, reconstruction problem, regularization term, translation.

I. INTRODUCTION

THE reconstruction problem is majorly concerned with the low-rank representation of the original data with a high-quality reconstruction. In other words, not only the dimensionality reduction [1]–[5] could be achieved but also the significant statistical properties could be maintained during the reconstruction [6]. The principal component analysis (PCA) [7]–[11] surely serves as an effective reconstruction approach [12]–[16], via which the dimensionality reduction can be achieved with minimizing the mean square error. However, the efficiency of the PCA method could be compromised, since it is sensitive to the outliers. To enhance the robustness of the PCA method, robust PCA [17], [18] is under further investigation by means of the robust subspace learning [19]–[21]. Nonetheless, most of the related PCA methods train a general subspace under the empirical data without taking the relations among the labeled classes into

consideration. Therefore, the classification results could be awkward, while the low-rank reconstruction might be efficient for the associated PCA approaches.

As to the supervised classification [22], the linear discriminant analysis (LDA) [23] is proposed for the efficient dimensionality reduction. Based on LDA, multiple works [24]–[26], e.g., discriminative embedded clustering [25] and least squares discriminant analysis [26] are further derived and proposed correspondingly. Besides, the maximum margin criterion (MMC) [27] method reformulates LDA from the fraction form into the subtraction form to avoid the malfunction, from which LDA may suffer when it subjects to the small sample size problem [28]. In general, LDA and the related works utilize scatter matrices to find a subspace, such that data in the same class are projected to be more close, while data in the different classes are projected to be more segregated. However, some important statistical properties might still be lost for both LDA and MMC approaches, since only the scatter matrices are included without taking the reconstruction issue into account.

To address all the referred defects, regularized class-specific subspace classifier (RCSSC) problem is proposed and illustrated as a series of biobjective optimization problems, which simultaneously minimize and maximize the reconstruction errors for each class and the associated complementary classes, respectively. In other words, both the low-rank reconstruction and the relations among the labeled classes are taken into consideration. Accordingly, the RCSSC method could be further derived based on solving a general quadratic ratio problem, which is connected with the translation variable. Moreover, it can be proved that the proposed RCSSC method converges to the global optimal solution of the related problem. Furthermore, the RCSSC method can be extended to the unregularized case, i.e., unregularized class-specific subspace classifier (UCSSC) method. Eventually, extensive experiments are performed to demonstrate the superiority of both the proposed RCSSC and UCSSC approaches. Framework of RCSSC and UCSSC is illustrated in Fig. 1.

This paper is organized as follows. Section II revisits the reconstruction problem. Section III at first investigates the RCSSC problem. Via the coordinate blocking method, the RCSSC method can be further derived based on solving a general quadratic ratio problem. The convergent analysis on the RCSSC method is also provided. In Section IV, the RCSSC method is further extended to the unregularized case, which is known as UCSSC method. In Section V, the experiments on the recognition data sets are presented to show the superiority of the proposed RCSSC and UCSSC methods. Eventually, Section VI concludes this paper.

Notations: For any matrix M , Frobenius norm is defined as $\|M\|_F^2 = \text{Tr}(M^T M)$, where $\text{Tr}(\cdot)$ is the trace operator. For any

Manuscript received January 25, 2016; revised May 25, 2016 and August 3, 2016; accepted August 4, 2016. Date of publication August 30, 2016; date of current version October 16, 2017. This work was supported by the Fundamental Research Funds for the Central Universities under Grant 3102015BJ(II)JJZ01. (Corresponding author: Feiping Nie.)

R. Zhang and F. Nie are with School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China (e-mail: ruizhang8633@gmail.com; feipingnie@gmail.com).

X. Li is with Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China (e-mail: xuelong_li@opt.ac.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TNNLS.2016.2598744

column vector v , l_2 -norm is defined as $\|v\|_2 = \sqrt{v^T v}$. For any positive integer i , $1_i = (1, 1, \dots, 1)^T \in \mathbb{R}^{i \times 1}$.

II. RECONSTRUCTION PROBLEM REVISITED

Suppose the input data set as $\{x_1, x_2, \dots, x_s\}$, where each data point $x_i \in \mathbb{R}^{d \times 1}$, ($i = 1, \dots, s$). The reconstruction problem can be recapitulated as the issue seeking a low-rank matrix Y to best approximate the original data matrix $X = [x_1, x_2, \dots, x_s] \in \mathbb{R}^{d \times s}$ as

$$\min_{m, \text{rank}(Y)=k} \|X - m 1_s^T - Y\|_F^2 \quad (1)$$

where $Y \in \mathbb{R}^{d \times s}$ with the translation variable $m \in \mathbb{R}^{d \times 1}$.

Based on the rank factorization of Y , problem (1) can be further reformulated into

$$\min_{W^T W = I, V, m} \|X - m 1_s^T - W V^T\|_F^2 \quad (2)$$

where both variables $W \in \mathbb{R}^{d \times k}$ and $V \in \mathbb{R}^{s \times k}$ are full column rank matrices. Apparently, the variable V is free from any constraint in (2). By virtue of the extreme value condition with respect to V , we could derive that

$$\begin{aligned} \frac{\partial \|X - m 1_s^T - W V^T\|_F^2}{\partial V} &= 0 \\ \Rightarrow \frac{\partial \text{Tr}(V V^T - 2(X - m 1_s^T)^T W V^T)}{\partial V} &= 0 \\ \Rightarrow V &= (X - m 1_s^T)^T W. \end{aligned}$$

By substituting the above result as $V = (X - m 1_s^T)^T W$, the reconstruction problem (1) can be eventually rewritten into

$$\min_{W^T W = I, m} \|X - m 1_s^T - W W^T (X - m 1_s^T)\|_F^2 \quad (3)$$

with the subspace W and the translation m .

III. REGULARIZED CLASS-SPECIFIC SUBSPACE CLASSIFIER

The CSSC [30]–[32] aims at obtaining the translated subspace for each class, such that the associated subspace not only can serve as the indicator for the distribution of the related class but could be able to identify the differences from other labeled classes as well.

Given the whole data set $\Omega \in \mathbb{R}^{d \times s}$ distributed into l different classes, we assume $X_c \in \mathbb{R}^{d \times n}$ as c -th class data and $X_{\bar{c}} = \Omega \setminus X_c \in \mathbb{R}^{d \times (s-n)}$ as the complementary data of c -th class, where $1 \leq c \leq l$. Thereinafter, we try to achieve both the class-specific subspace $W_c \in \mathbb{R}^{d \times k}$, ($k \leq d$) and the translation $m_c \in \mathbb{R}^{d \times 1}$, under which the c -th class data has the minimal reconstruction error. Meanwhile, we try to identify the difference between the c -th class data X_c and its complementary data $X_{\bar{c}}$ by maximizing the reconstruction error of $X_{\bar{c}}$ under the same subspace W_c and translation m_c . By virtue of the reconstruction problem (3), the RCSSC

problem can be proposed as

$$\begin{cases} \min_{W_c^T W_c = I, m_c} \|(I - W_c W_c^T)(X_c - m_c 1_n^T)\|_F^2 \\ \quad + \gamma \|W_c^T m_c\|_2^2 \\ \max_{W_c^T W_c = I, m_c} \|(I - W_c W_c^T)(X_{\bar{c}} - m_c 1_{s-n}^T)\|_F^2 \\ \quad + \gamma \|W_c^T m_c\|_2^2 \end{cases} \quad (4)$$

where the regularization term $\gamma \|W_c^T m_c\|_2^2$, ($0 < \gamma < 1$) is introduced to ensure the stability of system (4). In other words, the regularization term takes the form as $\gamma \|W_c^T m_c\|_2^2$ to guarantee an invertible matrix $C - \lambda D$ defined in (9). Otherwise, the matrix $C - \lambda D = (n - \lambda(s - n))(I - W_c W_c^T)$ would be singular, such that the proposed RCSSC method becomes invalid. Actually, problem (4) under $\gamma = 0$ is recognized as the unregularized case, whose discussion will be further covered in Section IV.

To solve the proposed problem (4), the discussion is further divided into two cases via the coordinate blocking method, i.e., fixing each variable (W_c and m_c) alternatively. For the convenience of dealing with the biobjective optimization in (4), we utilize its ratio form hereinafter.

A. Case 1 (Fixing m_c)

By fixing the translation variable m_c , the proposed RCSSC problem (4) is equivalent to the following biobjective optimization as:

$$\begin{cases} \min_{W_c^T W_c = I} \text{Tr}(W_c^T (aI - A) W_c) \\ \max_{W_c^T W_c = I} \text{Tr}(W_c^T (bI - B) W_c) \end{cases} \quad (5)$$

where $A = (X_c - m_c 1_n^T)(X_c - m_c 1_n^T)^T - \gamma m_c m_c^T$ and $B = (X_{\bar{c}} - m_c 1_{s-n}^T)(X_{\bar{c}} - m_c 1_{s-n}^T)^T - \gamma m_c m_c^T$ with $a = \text{Tr}((X_c - m_c 1_n^T)^T (X_c - m_c 1_n^T)/k)$ and $b = \text{Tr}((X_{\bar{c}} - m_c 1_{s-n}^T)^T (X_{\bar{c}} - m_c 1_{s-n}^T)/k)$. Thus, we could further reformulate problem (5) into the following fraction form as:

$$\min_{W_c^T W_c = I} \frac{\text{Tr}(W_c^T (aI - A) W_c)}{\text{Tr}(W_c^T (bI - B) W_c)}. \quad (6)$$

Problem (6) is known as the trace ratio (TR) problem, which has been well investigated and proved in [29] to ensure a global optimal solution W_c .

B. Case 2 (Fixing W_c)

By fixing the variable W_c , we try to deal with problem (4). Via the definition of l_2 -norm, problem (4) can be rewritten as

$$\min_{m_c} \frac{\sum_{i=1}^n \|(I - W_c W_c^T)(X_c(i) - m_c)\|_2^2 + \gamma \|W_c^T m_c\|_2^2}{\sum_{i=1}^{s-n} \|(I - W_c W_c^T)(X_{\bar{c}}(i) - m_c)\|_2^2 + \gamma \|W_c^T m_c\|_2^2} \quad (7)$$

where $X_c(i)$ and $X_{\bar{c}}(i)$ are the i -th column vectors for X_c and $X_{\bar{c}}$, respectively. Problem (7) can be further expanded into the following quadratic ratio problem as:

$$\min_{m_c} \frac{m_c^T C m_c + m_c^T e + g}{m_c^T D m_c + m_c^T f + h} \quad (8)$$

Algorithm 1 To Solve the Quadratic Ratio Problem (8) Based on the Proposed $p(\lambda)$ in (10)

Input: $C \in \mathbb{R}^{d \times d}$, $D \in \mathbb{R}^{d \times d}$, $e \in \mathbb{R}^{d \times 1}$, $f \in \mathbb{R}^{d \times 1}$, $g \in \mathbb{R}$ and $h \in \mathbb{R}$ are defined in (9) based on the given data matrices $X_c \in \mathbb{R}^{d \times n}$, $X_{\bar{c}} \in \mathbb{R}^{d \times (s-n)}$ and the given subspace $W_c \in \mathbb{R}^{d \times k}$.

Initialize: λ via the algorithm 2 such that $C - \lambda D$ is positive definite and $\gamma \in (0, 1)$ such that C is positive definite.

While $p(\lambda) \neq 0$ **do**

1: $m_c \leftarrow \frac{(C - \lambda D)^{-1}(\lambda f - e)}{2} = \arg \min_m m^T (C - \lambda D)m + m^T (e - \lambda f) + (g - \lambda h)$.

2: $\lambda \leftarrow \frac{m_c^T C m_c + m_c^T e + g}{m_c^T D m_c + m_c^T f + h}$.

End while

Output: $m_c \in \mathbb{R}^{d \times 1}$ and λ .

where

$$\begin{cases} C = nI - (n - \gamma)W_c W_c^T \\ D = (s - n)I - (s - n - \gamma)W_c W_c^T \\ e = -2(I - W_c W_c^T) \sum_{i=1}^n X_c(i) \\ f = -2(I - W_c W_c^T) \sum_{i=1}^{s-n} X_{\bar{c}}(i) \\ g = \sum_{i=1}^n X_c(i)^T (I - W_c W_c^T) X_c(i) \\ h = \sum_{i=1}^{s-n} X_{\bar{c}}(i)^T (I - W_c W_c^T) X_{\bar{c}}(i). \end{cases} \quad (9)$$

To solve problem (8), we introduce the following characteristic function $p(\lambda)$ as:

$$p(\lambda) = \min_m m^T (C - \lambda D)m + m^T (e - \lambda f) + (g - \lambda h) \quad (10)$$

where $\lambda \leftarrow ((m^T C m + m^T e + g)/(m^T D m + m^T f + h))$, $m^T D m + m^T f + h > 0$ and the positive definite matrix C . Accordingly, we could infer that

$$\begin{aligned} \min_m m^T (C - \lambda D)m + m^T (e - \lambda f) + (g - \lambda h) \\ \Rightarrow \min_m m^T (C - \lambda D)m + m^T (e - \lambda f) \\ \Rightarrow m = \frac{(C - \lambda D)^{-1}(e - \lambda f)}{2}. \end{aligned} \quad (11)$$

Based on $p(\lambda)$ in (10) and the derivation in (11), Algorithm 1 can be outlined to solve problem (8) under Case 2. Besides, Algorithm 1 monotonically converges to the global optimal solution of the quadratic ratio problem (8) due to the following theoretical supports.

Theorem 1: Algorithm 1 decreases the quadratic ratio problem (8) monotonically during each iteration until it converges.

Proof: $\tilde{m}_c$ is assumed to be the updated m_c in Algorithm 1 under each iteration. According to step 1 in Algorithm 1, we have

$$\begin{aligned} \tilde{m}_c^T (C - \lambda D)\tilde{m}_c + \tilde{m}_c^T (e - \lambda f) + (g - \lambda h) \\ \leq m_c^T (C - \lambda D)m_c + m_c^T (e - \lambda f) + (g - \lambda h) \\ = m_c^T C m_c + m_c^T e + g - \lambda(m_c^T D m_c + m_c^T f + h) \\ = 0 \end{aligned} \quad (12)$$

where $\lambda = ((m_c^T C m_c + m_c^T e + g)/(m_c^T D m_c + m_c^T f + h)) \geq 0$. Therefore, (12) can be further rewritten into

$$\frac{\tilde{m}_c^T C \tilde{m}_c + \tilde{m}_c^T e + g}{\tilde{m}_c^T D \tilde{m}_c + \tilde{m}_c^T f + h} \leq \lambda = \frac{m_c^T C m_c + m_c^T e + g}{m_c^T D m_c + m_c^T f + h}$$

which implies that Algorithm 1 decreases the objective value of the problem (8) monotonically with lower bound zero. ■

Theorem 2: If $p(\lambda^*) = 0$, then λ^* is the global minimum of problem (8).

Proof: The prerequisite as $p(\lambda^*) = 0$ can be specifically illustrated as

$$\min_m m^T (C - \lambda^* D)m + m^T (e - \lambda^* f) + (g - \lambda^* h) = 0. \quad (13)$$

Based on (13), we could conduct the following derivation as:

$$\begin{aligned} m_c^T (C - \lambda^* D)m_c + m_c^T (e - \lambda^* f) + (g - \lambda^* h) &\geq 0 \\ \Rightarrow \frac{m_c^T C m_c + m_c^T e + g}{m_c^T D m_c + m_c^T f + h} &\geq \lambda^* \end{aligned} \quad (14)$$

which represents that λ^* is the global minimum of problem (8) if $p(\lambda^*) = 0$. ■

Theorem 3: Algorithm 1 consistently converges to the global minimum of the problem (8).

Proof: λ^* and m_c^* are assumed to be the final solutions for the proposed Algorithm 1. From step 2 in Algorithm 1, we have

$$\lambda^* = \frac{m_c^{*T} C m_c^* + m_c^{*T} e + g}{m_c^{*T} D m_c^* + m_c^{*T} f + h}. \quad (15)$$

Based on step 1 in Algorithm 1, we have

$$\begin{aligned} m_c^* = \arg \min_m m^T (C - \lambda^* D)m + m^T (e - \lambda^* f) \\ + (g - \lambda^* h). \end{aligned} \quad (16)$$

According to (15) and (16), we have $p(\lambda^*) = 0$. Hence, Algorithm 1 converges to the global minimum of the problem (8) due to Theorem 2. ■

Theorem 4: If the matrix C is positive definite, then there exists some parameter λ , such that $C - \lambda D$ is also positive definite.

Proof: The symmetric matrix D can be decomposed as $D = P \Lambda P^T$, where P is the eigenvector matrix and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ is the eigenvalue matrix of D .

Without loss of generality, we assume $\lambda_i > 0$, ($i = 1, \dots, q$) while $\lambda_j < 0$, ($j = q+1, \dots, r$). Accordingly, two diagonal matrices can be constructed as

$$\begin{cases} \Lambda_1 = \text{diag}(2\lambda_1, \dots, 2\lambda_q, -\lambda_{q+1}, \dots, -\lambda_r, 1, \dots, 1) \\ \Lambda_2 = \text{diag}(\lambda_1, \dots, \lambda_q, -2\lambda_{q+1}, \dots, -2\lambda_r, 1, \dots, 1) \end{cases}$$

where $\Lambda = \Lambda_1 - \Lambda_2$. With two positive definite matrices as $E = P \Lambda_1 P^T$ and $F = P \Lambda_2 P^T$, D can be further disintegrated into

$$D = P \Lambda P^T = P(\Lambda_1 - \Lambda_2)P^T = E - F.$$

If $\lambda < \min_{x^T x=1} ((x^T C x)/(x^T E x))$, $C - \lambda E$ is positive definite, which leads to the positive definite matrix $C - \lambda D$, since

$C - \lambda D = (C - \lambda E) + \lambda F$. Therefore, $C - \lambda D$ is positive definite when $\lambda < \min_{x^T C x = 1} ((x^T C x)/(x^T E x))$. ■

Theorem 5: If $p(\lambda) > 0$, then $\lambda < \lambda^*$. If $p(\lambda) < 0$, then $\lambda > \lambda^*$, where λ^* is the global minimum of the quadratic ratio problem (8).

Proof: According to Theorem 2, we have $p(\lambda^*) = 0$.

1) If $\lambda \geq \lambda^*$, we have

$$\begin{aligned} p(\lambda) &= \min_m m^T C m + m^T e + g \\ &\quad - \lambda(m^T D m + m^T f + h) \\ &\leq m^{*T} C m^* + m^{*T} e + g \\ &\quad - \lambda(m^{*T} D m^* + m^{*T} f + h) \\ &= (\lambda^* - \lambda)(m^{*T} D m^* + m^{*T} f + h) + p(\lambda^*) \\ &\leq 0 \end{aligned} \quad (17)$$

where m^* is the solution of $p(\lambda^*)$. Thus, $p(\lambda) \leq 0$ if $\lambda \geq \lambda^*$.

2) If $\lambda \leq \lambda^*$, we have

$$\begin{aligned} p(\lambda) &= \min_m m^T C m + m^T e + g \\ &\quad - \lambda(m^T D m + m^T f + h) \\ &= \tilde{m}^T C \tilde{m} + \tilde{m}^T e + g - (\lambda^* - \lambda + \lambda) \\ &\quad \times (\tilde{m}^T D \tilde{m} + \tilde{m}^T f + h) \\ &= \tilde{m}^T C \tilde{m} + \tilde{m}^T e + g - \lambda^*(\tilde{m}^T D \tilde{m} + \tilde{m}^T f + h) \\ &\quad + (\lambda^* - \lambda)(\tilde{m}^T D \tilde{m} + \tilde{m}^T f + h) \\ &\geq p(\lambda^*) + (\lambda^* - \lambda)(\tilde{m}^T D \tilde{m} + \tilde{m}^T f + h) \\ &\geq 0 \end{aligned} \quad (18)$$

where $\tilde{m}$ is the solution of $p(\lambda)$. Hence, $p(\lambda) \geq 0$ if $\lambda \leq \lambda^*$.

Based on inverse negative propositions of the results in (17) and (18), Theorem 5 can be proved. ■

Theorem 6: As for the positive definite matrix $C - \lambda^* D$ in $p(\lambda^*)$, we have

$$\lambda < \lambda^* < \min \left(\frac{n}{s-n}, \frac{n-\gamma}{s-n-\gamma}, 1 \right)$$

where $p(\lambda^*) = 0$ and $p(\lambda) > 0$.

Proof: Based on Theorem 4, the existence of λ^* can be guaranteed, such that $C - \lambda^* D$ is positive definite. According to Theorem 5, it is obvious that $\lambda < \lambda^*$. Besides, we have

$$\begin{aligned} C - \lambda^* D &= (n - \lambda^*(s - n))I - (n - \gamma - \lambda^*(s - n - \gamma))W_c W_c^T. \end{aligned}$$

Since $C - \lambda^* D$ is positive definite, we can infer that

$$\begin{cases} n - \lambda^*(s - n) > 0 \\ n - \gamma - \lambda^*(s - n - \gamma) > 0 \\ \frac{n - \gamma - \lambda^*(s - n - \gamma)}{n - \lambda^*(s - n)} < 1. \end{cases}$$

Therefore, we could obtain $\lambda^* < \min((n/(s-n)), ((n-\gamma)/(s-n-\gamma)), 1)$. In sum, we have $\lambda < \lambda^* < \min((n/(s-n)), ((n-\gamma)/(s-n-\gamma)), 1)$. ■

Theorem 7: Algorithm 1 has the quadratic convergence rate.

Algorithm 2 To Initialize λ

Initialize: $p = 1$, $\lambda_1 = 0$ and $\lambda_2 = \min(\frac{n}{s-n}, \frac{n-\gamma}{s-n-\gamma}, 1)$.

While $p(\lambda) > 0$ **do**

```

1:  $\lambda \leftarrow \frac{\lambda_1 + \lambda_2}{2}$ .
2:  $m_c \leftarrow \frac{(C - \lambda D)^{-1}(\lambda f - e)}{2}$ .
3:  $p \leftarrow m_c^T (C - \lambda D) m_c + m_c^T (e - \lambda f) + (g - \lambda h)$ .
   if  $p > 0$  do
4:  $\lambda_1 \leftarrow \lambda$ .
5: else break.
   end if

```

End while

Output: λ .

Proof: Algorithm 1 is proposed based on solving the following characteristic function as:

$$p(\lambda) = \min_m m^T (C - \lambda D) m + m^T (e - \lambda f) + (g - \lambda h).$$

Suppose $\tilde{m}$ as the solution of $p(\lambda)$, then we have

$$\begin{aligned} p(\lambda) &= \tilde{m}^T (C - \lambda D) \tilde{m} + \tilde{m}^T (e - \lambda f) + (g - \lambda h) \\ \Rightarrow p'(\lambda) &= \frac{dp(\lambda)}{d\lambda} = -(\tilde{m}^T D \tilde{m} + \tilde{m}^T f + h) < 0. \end{aligned} \quad (19)$$

According to Algorithm 1 and (19), we have

$$\begin{aligned} \lambda_{t+1} &= \frac{m_t^T C m_t + m_t^T e + g}{m_t^T D m_t + m_t^T f + h} \\ &= \lambda_t - \frac{m_t^T (C - \lambda_t D) m_t + m_t^T (e - \lambda_t f) + (g - \lambda_t h)}{-(m_t^T D m_t + m_t^T f + h)} \\ &= \lambda_t - \frac{p(\lambda_t)}{p'(\lambda_t)} \end{aligned} \quad (20)$$

which is known as Newton method with m_t and λ_t being the solutions of Algorithm 1 for the t -th iteration.

Therefore, Algorithm 1 has the quadratic convergence rate as Newton method does. ■

According to Theorem 1 and the result in (19), we have the following decreasing series as:

$$\lambda_t \geq \lambda_{t+1} \geq \lambda^*, \quad (t = 1, 2, \dots)$$

where λ_t is the solution of Algorithm 1 for the t -th iteration and λ^* is the global optimal solution of Algorithm 1. Therefore, we have $p(\lambda_t) \leq 0, \forall t$ due to Theorem 5. Accordingly, the initialization for λ can be outlined in Algorithm 2 based on Theorem 6. Owing to the discussion and the analysis in both Cases 1 and 2, the RCSSC can be summarized in Algorithm 3 to converge to the global optimal solution of the RCSSC problem (4) correspondingly.

IV. UNREGULARIZED CLASS-SPECIFIC SUBSPACE CLASSIFIER METHOD ($\gamma = 0$)

When $\gamma = 0$, problem (4) degenerates to the unregularized case as

$$\begin{cases} \min_{W_c^T W_c = I, m_c} \|(I - W_c W_c^T)(X_c - m_c 1_n^T)\|_F^2 \\ \max_{W_c^T W_c = I, m_c} \|(I - W_c W_c^T)(X_c - m_c 1_{s-n}^T)\|_F^2 \end{cases} \quad (21)$$

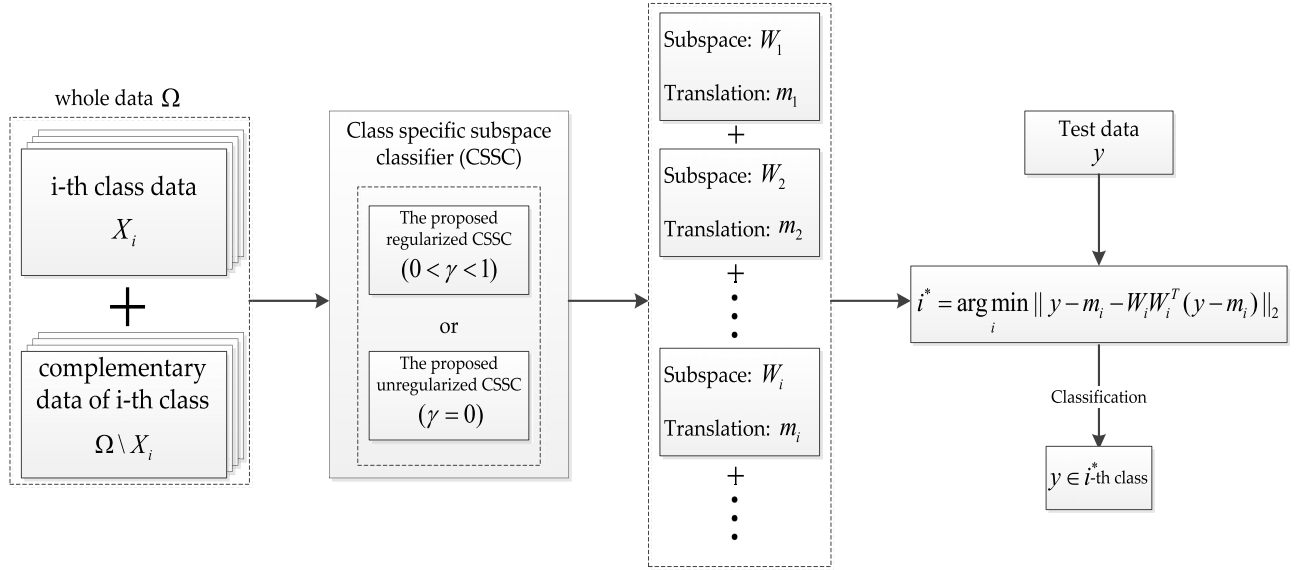


Fig. 1. Framework of the proposed CSSC methods is outlined.

Algorithm 3 Regularized Class-Specific Subspace Classifier Method

Input: The data matrices $X_c \in \mathbb{R}^{d \times n}$ and $X_{\bar{c}} \in \mathbb{R}^{d \times (s-n)}$.

Initialize: Random $m_c \in \mathbb{R}^{d \times 1}$ and $\gamma \in (0, 1)$ such that C is positive definite.

For $c = 1 : l$

While not converge **do**

- 1: Update $W_c \in \mathbb{R}^{d \times k}$ by solving the problem (6) via the trace ratio method in [29].
- 2: Update $m_c \in \mathbb{R}^{d \times 1}$ via the algorithm 1.

End while

End for

Output: $c^* = \arg \min_{1 \leq c \leq l} \|y - m_c - W_c W_c^T (y - m_c)\|_2$, which denotes that y belongs to c^* -th class for any given test data $y \in \mathbb{R}^{d \times 1}$.

where the regularization term $\gamma \|W_c^T m_c\|_2^2$ vanishes. Accordingly, Algorithm 1 becomes invalid, since the matrix $C - \lambda D$ would be singular. In other words, the difficulty is to solve the following extreme value condition of $p(\lambda)$ with respect to m_c as:

$$\begin{aligned} \nabla_{m_c} p &= 0, \quad \left(\nabla_{m_c} = \frac{\partial}{\partial m_c} \right) \\ \Rightarrow \nabla_{m_c} (m_c^T (C - \lambda D) m_c + m_c^T (e - \lambda f) + (g - \lambda h)) &= 0 \\ \Rightarrow 2(n - \lambda(s - n))(I - W_c W_c^T) m_c + e - \lambda f &= 0 \\ \Rightarrow (I - W_c W_c^T) (2(n - \lambda(s - n)) m_c + e - \lambda f) &= 0 \end{aligned} \quad (22)$$

where C , D , e , and f are defined in (9) with $e - \lambda f = (I - W_c W_c^T)(e - \lambda f)$ due to the following derivation as:

$$\begin{aligned} e - \lambda f &= (I - W_c W_c^T) \left(-2 \sum_{i=1}^n X_c(i) + 2\lambda \sum_{i=1}^{s-n} X_{\bar{c}}(i) \right) \\ \Rightarrow (I - W_c W_c^T)(e - \lambda f) &= e - \lambda f \end{aligned}$$

where $I - W_c W_c^T$ is idempotent, i.e., $(I - W_c W_c^T)^2 = I - W_c W_c^T$.

Suppose $W_c^\perp$ is the orthogonal complement space of W_c , then the vector $2(n - \lambda(s - n))m_c + e - \lambda f$ can be further disintegrated into

$$2(n - \lambda(s - n))m_c + e - \lambda f = W_c \alpha + W_c^\perp \beta. \quad (23)$$

Based on (22) and (23), we can infer that

$$\begin{aligned} (I - W_c W_c^T) (2(n - \lambda(s - n))m_c + e - \lambda f) &= 0 \\ \Rightarrow (I - W_c W_c^T) (W_c \alpha + W_c^\perp \beta) &= 0 \\ \Rightarrow W_c^\perp \beta &= 0 \\ \Rightarrow m_c &= \frac{\lambda f - e}{2(n - \lambda(s - n))} + W_c \theta \end{aligned} \quad (24)$$

where $\theta \in \mathbb{R}^{k \times 1}$ is an arbitrary column vector. By means of the result in (24), problem (21) can be reformulated into

$$\begin{cases} \min_{W_c^T W_c = I, \lambda} \text{Tr}(\mathbb{X}_c^T (I - W_c W_c^T) \mathbb{X}_c) \\ \max_{W_c^T W_c = I, \lambda} \text{Tr}(\mathbb{X}_{\bar{c}}^T (I - W_c W_c^T) \mathbb{X}_{\bar{c}}) \end{cases} \quad (25)$$

where $\mathbb{X}_c = X_c - ((\lambda f - e)/(2(n - \lambda(s - n))))1_n^T$ and $\mathbb{X}_{\bar{c}} = X_{\bar{c}} - ((\lambda f - e)/(2(n - \lambda(s - n))))1_{s-n}^T$. Obviously, (25) does not depend on θ . Thus, we can choose θ as zero vector for the convenience, such that the translation m_c can be delivered as

$$m_c = \frac{\lambda f - e}{2(n - \lambda(s - n))} \in \mathbb{R}^{d \times 1}. \quad (26)$$

Based on (26), the proposed RCSSC method can be further extended to the unregularized case, i.e., $\gamma = 0$. Accordingly, the UCSSC method can be summarized in Algorithm 4.

Algorithm 4 Unregularized Class-Specific Subspace Classifier Method ($\gamma = 0$)

Input: The data matrices $X_c \in \mathbb{R}^{d \times n}$ and $X_{\bar{c}} \in \mathbb{R}^{d \times (s-n)}$.

Initialize: Random $m_c \in \mathbb{R}^{d \times 1}$, $\lambda_2 = \min(\frac{n}{s-n}, 1)$ and $p = 1$.

For $c = 1 : l$

While not converge **do**

1: $\lambda_1 \leftarrow 0$.

2: Update the subspace $W_c \in \mathbb{R}^{d \times k}$ by solving $\min_{W_c^T W_c = I} \frac{\text{Tr}(W_c^T (aI - A_1) W_c)}{\text{Tr}(W_c^T (bI - B_1) W_c)}$ via the trace ratio method in [29], where $A_1 = (X_c - m_c 1_n^T)(X_c - m_c 1_n^T)^T$ and $B_1 = (X_{\bar{c}} - m_c 1_{s-n}^T)(X_{\bar{c}} - m_c 1_{s-n}^T)^T$.

while $p(\lambda) > 0$ **do**

3: $\lambda \leftarrow \frac{\lambda_1 + \lambda_2}{2}$.

4: $m_c \leftarrow \frac{\lambda f - e}{2(n - \lambda(s-n))}$.

5: $p \leftarrow m_c^T (C - \lambda D) m_c + m_c^T (e - \lambda f) + (g - \lambda h)$.

if $p > 0$ **do**

6: $\lambda_1 \leftarrow \lambda$.

7: **else break.**

end if

end while

while $p(\lambda) \neq 0$ **do**

8: $m_c \leftarrow \frac{\lambda f - e}{2(n - \lambda(s-n))}$.

9: $\lambda \leftarrow \frac{m_c^T C m_c + m_c^T e + g}{m_c^T D m_c + m_c^T f + h}$.

end while

End while

End for

Output: $W_c \in \mathbb{R}^{d \times k}$ and $m_c \in \mathbb{R}^{d \times 1}$, where $1 \leq c \leq l$.

V. EXPERIMENT

In this section, extensive experiments are performed to show the effectiveness of the proposed RCSSC and UCSSC methods. Besides, the data sets are downloaded on the Web site separately and some of them are preprocessed such as resizing or cropping, but it will not affect the experimental results and the analysis.

A. Complexity Analysis With How to Accelerate the RCSSC Method

As for both RCSSC and UCSSC methods, the TR method in the loop takes the same complexity for each iteration as $O(d^2n + d^2(s-n) + d^2kt_1)$ to converge to the class-specific subspace W_c , where t_1 stands for the iteration number within the TR method. Apart from the complexity of the TR method, the proposed RCSSC and UCSSC methods achieve the translation m_c at the expense of $O((d^3 + d^2n + d^2(s-n))t_2)$ and $O((d^2k + d^2n + d^2(s-n))t_2)$, respectively, where t_2 stands for the iteration number in the loop to converge to the translation m_c . In sum, the proposed RCSSC and UCSSC methods perform with the orders of the complexity as $O(d^3t_2t + d^2kt_1t + (d^2n + d^2(s-n))t_2t)$ and $O(d^2k(t_1 + t_2)t + (d^2n + d^2(s-n))t_2t)$, respectively, where t stands for the iteration number for the convergence.

1) *How to Accelerate the RCSSC Method:* The major distinction between the proposed RCSSC and UCSSC algorithms

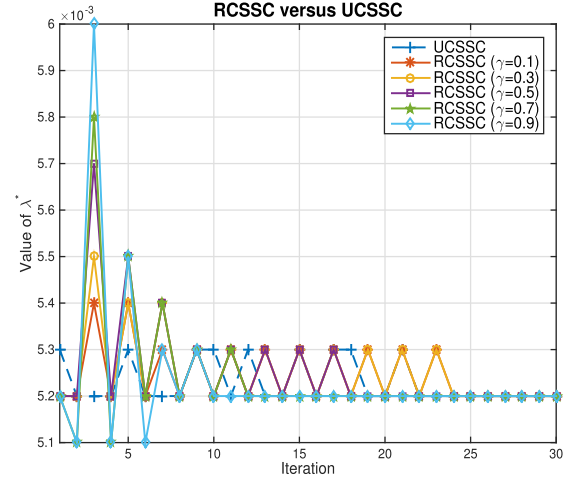


Fig. 2. Comparison of the objective value in (8) is performed for the proposed RCSSC and UCSSC methods over the data set AT&T.

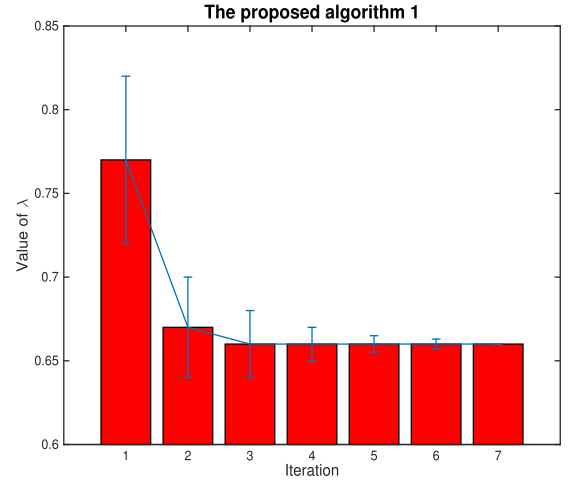


Fig. 3. Convergence of the proposed Algorithm 1 is performed under 100 random initial guesses.

reflects on the translation variable m_c . In other words, the UCSSC method only deals with matrix multiplication due to the derivation in (26), while the RCSSC method needs to cope with the term $(C - \lambda D)^{-1}$, i.e., inverse matrix computation [$O(d^3)$], which enlarges the time consumption for the RCSSC method in Algorithm 3. To accelerate the RCSSC method, we try to transform inverse matrix computation into matrix multiplication to get m_c , which leads to Theorem 8.

Theorem 8: For an invertible matrix $H = UV$, we have

$$(H - UV)^{-1} = H^{-1} + H^{-1}U(I - VH^{-1}U)^{-1}VH^{-1} \quad (27)$$

where $H \in \mathbb{R}^{d \times d}$ is invertible with $U \in \mathbb{R}^{d \times k}$ and $V \in \mathbb{R}^{k \times d}$.

Proof: The derivation can be conducted as

$$\begin{aligned} I &= I - H^{-1}UV + H^{-1}UV \\ &= I - H^{-1}UV + H^{-1}U(I - VH^{-1}U)^{-1}(I - VH^{-1}U)V \\ &= I + H^{-1}U(I - VH^{-1}U)^{-1}V - H^{-1}UV \\ &\quad - H^{-1}U(I - VH^{-1}U)^{-1}VH^{-1}UV \\ &= (I + H^{-1}U(I - VH^{-1}U)^{-1}V)(I - H^{-1}UV) \\ &= (H^{-1} + H^{-1}U(I - VH^{-1}U)^{-1}VH^{-1})(H - UV) \end{aligned}$$

which leads to (27) as a result. ■

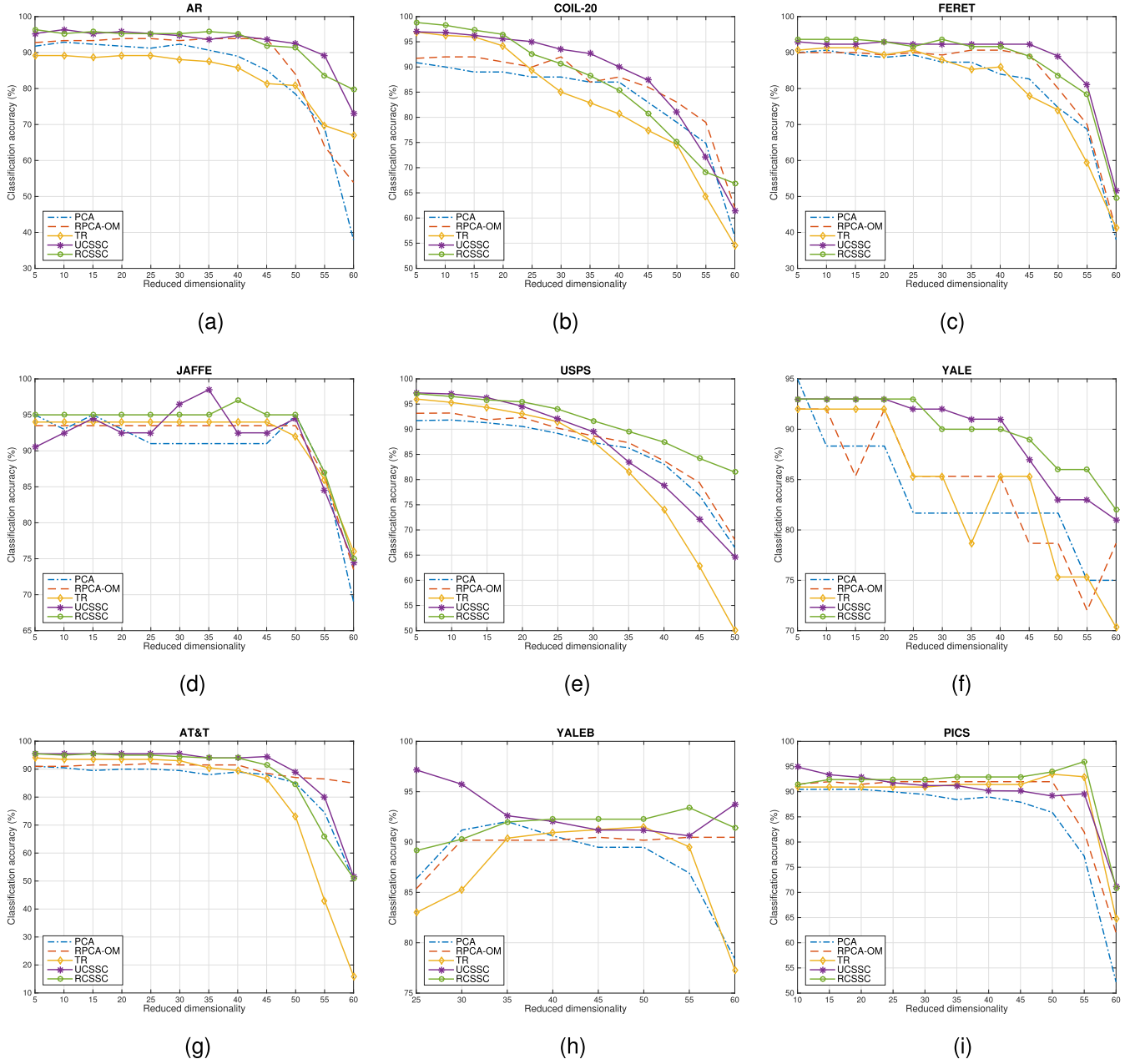


Fig. 4. Classification comparisons of reduced dimensionality are performed for five methods, including PCA [7], Robust PCA-Optimal Mean [18], TR [29], UCSSC, and RCSSC under nine data sets. (a) AR. (b) COIL-20. (c) FERET. (d) JAFFE. (e) USPS. (f) YALE. (g) AT&T. (h) YALEB. (i) PICS.

Based on Theorem 8, we could reformulate $(C - \lambda D)^{-1}$ into

$$\begin{aligned}
 (C - \lambda D)^{-1} &= ((n - \lambda(s - n))(I - aW_c W_c^T))^{-1} \\
 &= \frac{(I - aW_c W_c^T)^{-1}}{(n - \lambda(s - n))} \\
 &= \frac{I + aW_c (I - aW_c^T W_c)^{-1} W_c^T}{(n - \lambda(s - n))} \\
 &= \frac{I + \frac{a}{1-a} W_c W_c^T}{(n - \lambda(s - n))} \\
 &= \frac{(1 - a)I + aW_c W_c^T}{(1 - a)(n - \lambda(s - n))}
 \end{aligned} \tag{28}$$

where $a = ((n - \gamma - \lambda(s - n - \gamma))/(n - \lambda(s - n)))$ with treating $H = I$, $U = aW_c$, and $V = W_c^T$ in (27). By virtue of the result in (28), the RCSSC method can be further expedited. To be specific, the related complexity can be reduced from $O(d^3)$ to $O(d^2k)$ by replacing $(C - \lambda D)^{-1}$ with $((1 - a)I + aW_c W_c^T)/((1 - a)(n - \lambda(s - n)))$ for the RCSSC method. Apparently, the time consumption could be largely reduced if $k \ll d$, i.e., efficient dimensionality reduction under the subspace $W_c \in \mathbb{R}^{d \times k}$.

Consequently, the order of the complexity for the proposed RCSSC method can be reduced to $O(d^2k(t_1 + t_2)t + (d^2n + d^2(s - n)t_2t))$, which is same as the order of the complexity for the UCSSC method.

TABLE I
UCSSC VERSUS RCSSC FOR THE CLASSIFICATION ACCURACY
VIA VARYING THE REGULARIZATION PARAMETER
UNDER THE DATA SET AT&T

Classification accuracy (%)					
Class number	20	25	30	35	40
Reduced dimensionality	300	330	240	280	230
UCSSC ($\gamma = 0$)	91.00	89.12	90.30	89.57	88.00
RCSSC	($\gamma = 0.1$)	91.00	89.00	90.33	89.57
	($\gamma = 0.2$)	91.00	89.00	90.33	89.57
	($\gamma = 0.3$)	91.00	89.00	90.33	89.57
	($\gamma = 0.4$)	91.00	89.00	90.33	89.57
	($\gamma = 0.5$)	91.00	89.00	90.33	89.57
	($\gamma = 0.6$)	91.00	89.00	90.33	89.57
	($\gamma = 0.7$)	91.00	89.00	90.33	89.57
	($\gamma = 0.8$)	91.00	89.00	90.33	89.57
	($\gamma = 0.9$)	91.00	89.00	90.33	89.57

B. RCSSC Versus UCSSC

First, we utilize the data set AT&T to illustrate the performance of both proposed RCSSC and UCSSC methods. In Fig. 3, the consistent convergence of the proposed Algorithm 1 is performed for 100 trials, i.e., 100 different initial guesses to achieve the global minimum of problem (8) under the same input defined in (9). Both the classification accuracy and the objective value for the proposed RCSSC and UCSSC methods are recorded via varying the regularization parameter in Table I and Fig. 2, respectively.

From Table I, Figs. 2 and 3, we can draw the following conclusions.

- 1) From Fig. 3, the proposed Algorithm 1 monotonically converges to the same objective value of the quadratic ratio problem (8) as proved in Theorem 1.
- 2) From Fig. 2, we can observe that the swing amplitude for the objective value of the RCSSC method is proportional to the regularization parameter γ during the convergence. In other words, the greater value the regularization parameter γ achieves, the larger swing the proposed RCSSC method will obtain. However, the value of the regularization parameter γ will not affect the consistent convergence of the proposed RCSSC method.
- 3) From Table I, the classification accuracy for the proposed RCSSC method is not affected by the regularization parameter γ . Accordingly, it is rational to conjecture that the independence of the regularization parameter γ represented in Table I correlates with the consistent convergence shown in Fig. 2 for the RCSSC method. Therefore, we can choose an arbitrary regularization parameter γ within the open interval (0, 1) to further compare the classification accuracy in Section V-C.

C. Classification Accuracy

1) *Reduced Dimensionality*: Second, we utilize nine data sets, including AR, COIL-20, FERET, JAFFE, USPS, YALE, AT&T, YALEB, and PICS for the classification comparisons. For the convenience, images in each data set are reshaped into the column vectors. Table II provides the comparisons of the CSSC value for five approaches as PCA [7], RPCA-OM [18], TR [29], RCSSC, and UCSSC under the

TABLE II
COMPARISONS OF THE CSSC VALUE FOR FIVE METHODS UNDER
DIFFERENT DIMENSIONALITIES OVER SIX DATA SETS

The value of CSSC is defined as $\frac{\ (I - W_c W_c^T)(X_c - m_c 1_n^T)\ _F^2}{\ (I - W_c W_c^T)(X_c - m_c 1_{s-n}^T)\ _F^2}$.					
Dataset	Reduced dimensionality	30	40	50	60
AR ($\times 10^{-1}$)	PCA [7]	0.2900	0.2914	0.2911	0.2913
	RPCA-OM [18]	0.2751	0.2750	0.4803	0.3212
	TR [29]	0.2503	0.2543	0.2617	0.3158
	UCSSC(our)	0.2262	0.2269	0.2321	0.2620
	RCSSC(our)	0.2229	0.2302	0.2306	0.2493
Dataset	Reduced dimensionality	25	35	45	55
COIL-20 ($\times 10^{-1}$)	PCA [7]	0.2723	0.2682	0.2673	0.2663
	RPCA-OM [18]	0.4096	0.4057	0.3930	0.4300
	TR [29]	0.2794	0.3201	0.3605	0.4316
	UCSSC(our)	0.1900	0.1923	0.2044	0.2689
	RCSSC(our)	0.2002	0.2059	0.2128	0.2214
Dataset	Reduced dimensionality	20	30	40	50
FERET ($\times 10^{-2}$)	PCA [7]	1.5835	1.6033	1.6297	1.6223
	RPCA-OM [18]	1.4882	1.5848	1.2332	1.1701
	TR [29]	1.0303	1.0340	1.1005	1.0944
	UCSSC(our)	0.8481	0.8444	0.8432	0.8323
	RCSSC(our)	0.8406	0.8512	0.8972	0.8852
Dataset	Reduced dimensionality	31	41	51	61
JAFFE ($\times 10^0$)	PCA [7]	0.0313	0.0302	0.0300	0.0299
	RPCA-OM [18]	0.2295	0.1314	0.1790	0.3032
	TR [29]	0.0254	0.0275	0.0280	0.0660
	UCSSC(our)	0.0203	0.0203	0.0250	0.0139
	RCSSC(our)	0.0234	0.0235	0.0252	0.0266
Dataset	Reduced dimensionality	27	37	47	57
USPS ($\times 10^0$)	PCA [7]	0.0714	0.0701	0.0689	0.0687
	RPCA-OM [18]	0.0773	0.0832	0.0587	0.0431
	TR [29]	0.0483	0.0565	0.0698	0.1025
	UCSSC(our)	0.0378	0.0419	0.0554	0.0600
	RCSSC(our)	0.0368	0.0381	0.0401	0.0399
Dataset	Reduced dimensionality	15	25	35	45
YALE ($\times 10^0$)	PCA [7]	0.0615	0.0621	0.0620	0.0621
	RPCA-OM [18]	0.1025	0.0964	0.0643	0.0573
	TR [29]	0.0482	0.0500	0.0567	0.0711
	UCSSC(our)	0.0415	0.0418	0.0426	0.0425
	RCSSC(our)	0.0418	0.0423	0.0440	0.0497

first six data sets mentioned earlier. Besides, the class-specific subspace is trained under the random selected class. In particular, the translation is chosen as the mean vector of the selected class for both PCA and TR methods. In Fig. 4, the classification results of the aforementioned approaches are further compared in terms of the reconstruction classifier as $\arg \min_c \|y - m_c - W_c W_c^T (y - m_c)\|_2$. All the approaches are utilized to train the class-specific subspaces and the translations. From Table II and Fig. 4, the following conclusions can be drawn.

- 1) From Table II, the proposed RCSSC and UCSSC can achieve more optimal subspaces and translations than other methods do for solving the CSSC problem defined as $((\|(I - W_c W_c^T)(X_c - m_c 1_n^T)\|_F^2) / (\|(I - W_c W_c^T)(X_c - m_c 1_{s-n}^T)\|_F^2))$.
- 2) From Fig. 4, we could observe that the proposed RCSSC and UCSSC methods perform with better classification accuracy than other approaches do with minor exceptions.

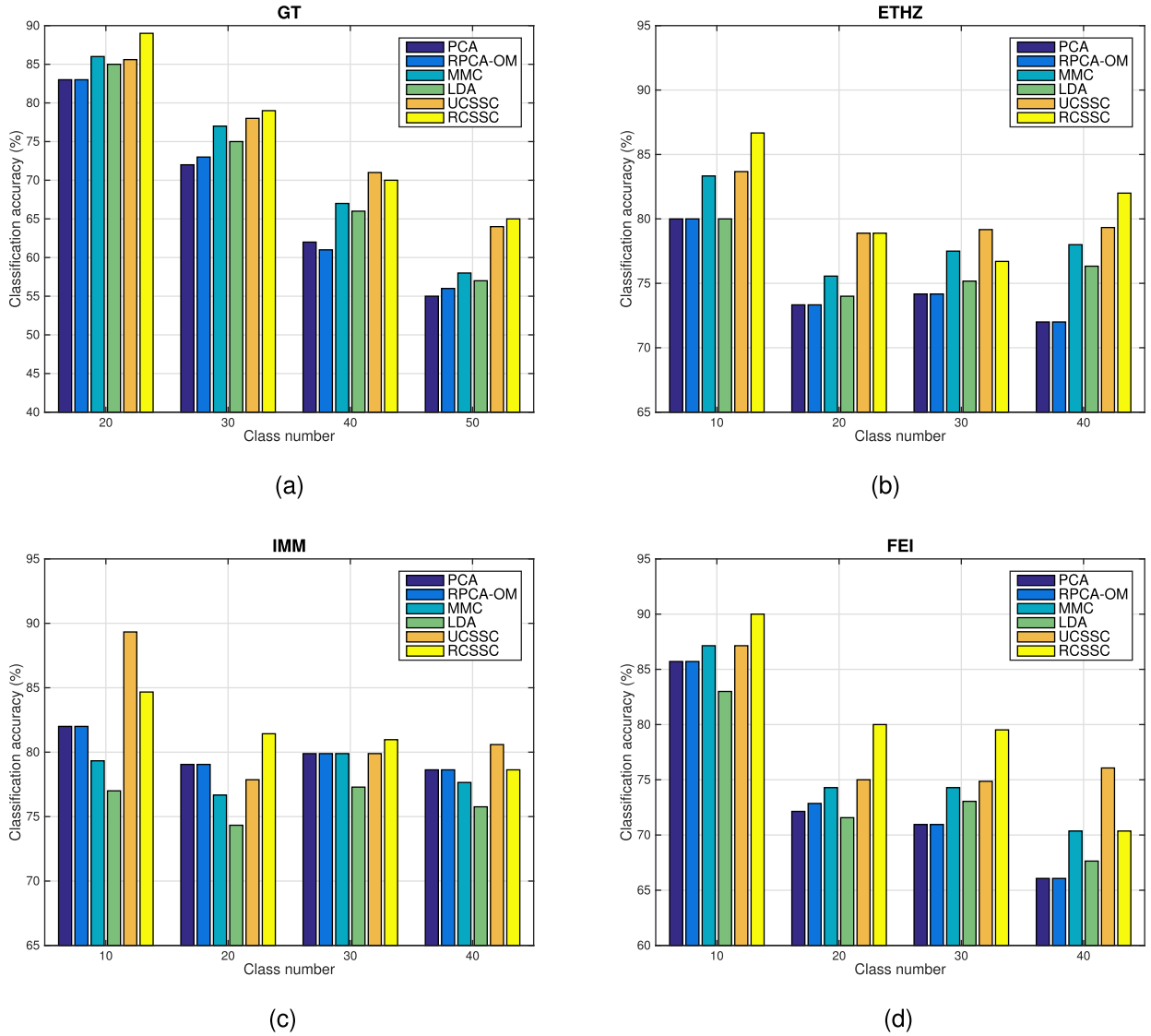


Fig. 5. Classification comparisons of class number for six methods, including PCA [7], RPCA-OM [18], LDA [23], MMC [27], UCSSC, and RCSSC are performed under four data sets. (a) GT. (b) ETHZ. (c) IMM. (d) FEI.

2) *Class Number*: Finally, four recognition data sets as GT, ETHZ, IMM, and FEI are utilized for the classification comparisons in Fig. 5. Besides, the shortest distance classifier is utilized for PCA [7], RPCA-OM [18], LDA [23], and MMC [27] methods, via which general subspaces are obtained for the classification. The reduced dimensionality k is randomly chosen as positive integer within the interval $[1, c - 1]$, where c is the class number. For the convenience, the input data are reshaped into the column vectors. From Fig. 5, we have the following conclusions.

- 1) From Fig. 5, the proposed RCSSC and UCSSC methods are more efficient for the classification under most cases.
- 2) From Figs. 4 and 5, the proposed RCSSC and UCSSC methods perform better than both the reconstruction methods and the supervised classification methods, including PCA [7], RPCA-OM [18], LDA [23], and MMC [27], since both the low-rank reconstruction and the relations among the different classes are taken into consideration for the proposed RCSSC and UCSSC approaches.

VI. CONCLUSION

In this paper, we propose and investigate the CSSC problem. By introducing the regularization term, the RCSSC method is proposed to obtain the global optimal subspace and translation for each class, where both the high-quality reconstruction and the relations among the labeled classes are taken into consideration. Besides, a general framework for solving the quadratic ratio problem is proposed and investigated. By means of the orthogonal decomposition technique, the RCSSC method can be further extended to the unregularized case, i.e., UCSSC method. Moreover, an expedited RCSSC algorithm is also derived and provided. As a result, the effectiveness and the superiority for both RCSSC and UCSSC methods are analytically and empirically verified.

REFERENCES

- [1] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2531–2544, Dec. 2015.

- [2] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.
- [3] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.
- [4] F. Nie, X. Cai, and H. Huang, "Flexible shift-invariant locality and globality preserving projections," in *Machine Learning and Knowledge Discovery in Databases*, T. Calders *et al.*, Eds. Berlin, Germany: Springer-Verlag, 2014, pp. 485–500.
- [5] J. Shi, Q. Jiang, Q. Zhang, Q. Huang, and X. Li, "Sparse kernel entropy component analysis for dimensionality reduction of biomedical data," *Neurocomputing*, vol. 168, pp. 930–940, Nov. 2015.
- [6] L. Yang, H. Cheng, J. Su, and X. Li, "Pixel-to-model distance for robust background reconstruction," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 903–916, May 2016.
- [7] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.
- [8] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.
- [9] X. Li, Y. Pang, and Y. Yuan, " L_1 -norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 4, pp. 1170–1175, Aug. 2010.
- [10] Y. Pang, D. Tao, Y. Yuan, and X. Li, "Binary two-dimensional PCA," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 1176–1180, Aug. 2008.
- [11] J. Ye, "Generalized low rank approximations of matrices," *Mach. Learn.*, vol. 61, nos. 1–3, pp. 167–191, 2005.
- [12] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling*. London, U.K.: Chapman & Hall, 1994.
- [13] J. A. Lee and M. Verleysen, *Nonlinear Dimensionality Reduction*. Berlin, Germany: Springer, 2007.
- [14] F. Nie, S. Xiang, Y. Jia, and C. Zhang, "Semi-supervised orthogonal discriminant analysis via label propagation," *Pattern Recognit.*, vol. 42, no. 11, pp. 2615–2627, 2009.
- [15] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint $l_{2,1}$ -norms minimization," in *Proc. NIPS*, 2010, pp. 1813–1821.
- [16] F. Nie, H. Huang, and C. Ding, "Low-rank matrix recovery via efficient Schatten p -norm minimization," in *Proc. AAAI*, 2012, pp. 655–661.
- [17] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy l_1 -norm maximization," in *Proc. IJCAI*, 2011, pp. 1433–1438.
- [18] F. Nie, J. Yuan, and H. Huang, "Optimal mean robust principal component analysis," in *Proc. ICML*, 2014, pp. 1062–1070.
- [19] S. Li and Y. Fu, "Robust subspace discovery through supervised low-rank constraints," in *Proc. SIAM Int. Conf. Data Mining*, 2014, pp. 163–171.
- [20] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE ICDM*, Dec. 2014, pp. 110–119.
- [21] E. Kim, M. Lee, and S. Oh, "Elastic-net regularization of singular values for robust subspace learning," in *Proc. IEEE CVPR*, Jun. 2015, pp. 915–923.
- [22] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 447–461, Mar. 2016.
- [23] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. Boston, MA, USA: Academic, 1990.
- [24] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [25] C. Hou, F. Nie, D. Yi, and D. Tao, "Discriminative embedded clustering: A framework for grouping high-dimensional data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1287–1299, Jun. 2015.
- [26] F. Nie, S. Xiang, Y. Liu, C. Hou, and C. Zhang, "Orthogonal vs. uncorrelated least squares discriminant analysis for feature extraction," *Pattern Recognit. Lett.*, vol. 33, no. 5, pp. 485–491, 2012.
- [27] H. R. Li, T. Jiang, and K. Zhang, "Efficient and robust feature extraction by maximum margin criterion," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 157–165, Feb. 2006.
- [28] L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.-W. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.
- [29] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 729–735, Apr. 2009.
- [30] G. Goudelis, S. Zafeiriou, A. Tefas, and I. Pitas, "Class-specific kernel-discriminant analysis for face verification," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 570–587, Sep. 2007.
- [31] X. Jing, S. Li, Y. Yao, L. Bian, and J. Yang, "Kernel uncorrelated adjacent-class discriminant analysis," in *Proc. IEEE ICPR*, Aug. 2010, pp. 706–709.
- [32] A. Iosifidis, A. Tefas, and I. Pitas, "Class-specific reference discriminant analysis with application in human behavior analysis," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 3, pp. 315–326, Jun. 2015.



Rui Zhang received the M.S. degree in applied mathematics from Northwest University, Xi'an, China, in 2011, and the M.S. degree in applied mathematics from Georgia Southern University, Statesboro, GA, USA, in 2013. He is currently pursuing the Ph.D. degree in computer science with Northwestern Polytechnical University, Xi'an, under the supervision of Prof. X. Li and Prof. F. Nie.



Feiping Nie received the Ph.D. degree in computer science from Tsinghua University, Beijing, China, in 2009.

He is currently a Full Professor with Northwestern Polytechnical University, Xi'an, China. He has authored over 100 papers in the prestigious journals and conferences, such as the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the International Conference on Machine Learning, the Neural Information Processing Systems, and the Knowledge Discovery and Data Mining. His current research interests include machine learning and its application fields, such as pattern recognition, data mining, computer vision, image processing, and information retrieval.

Dr. Nie is currently serving as an Associate Editor or a Program Committee Member for several prestigious journals and conferences in the related fields.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, P. R. China.