

Received April 21, 2022, accepted May 4, 2022, date of publication May 16, 2022, date of current version May 31, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3175199

Subspace Segmentation by Low Rank Representation via the Sparse-Prompting Quasi-Rank Function

HAIYANG LI¹, YUSI WANG¹, QIAN ZHANG², ZHIWEI XING³,
SHOUJIN LIN⁴, AND JIGEN PENG¹

¹School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

²School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

³School of Science, Xi'an Polytechnic University, Xi'an 710048, China

⁴Zhongshan MLTOR CNC Technology Company Ltd., Zhongshan 528400, China

Corresponding authors: Haiyang Li (fplihaiyang@126.com) and Jigen Peng (jgpeng@gzhu.edu.cn)

This work was supported in part by the National Science Foundation of China under Grant 12031003 and Grant 11771347, and in part by the Basic Research Joint Funding Project of University and Guangzhou City under Grant 202102010434.

ABSTRACT In this paper, a general optimization formulation is proposed for the subspace segmentation by low rank representation via the sparse-prompting quasi-rank function. We prove that, with the clean data from independent linear subspaces, the optimal solution to our optimization formulation not only is the lowest rank but also forms a block-diagonal matrix, which implies that it is reasonable to use any sparse-prompting quasi-rank function as the measure of the low rank in subspace clustering. With the data contaminated by Gaussian noise and/or gross errors, the alternating direction method of multipliers is applied to solving it and every sub-optimization problem has a closed-form optimal solution when the band restricted thresholding operator induced by its corresponding sparse-prompting function has an analytic expression, in which the gross errors part is replaced with the sparse-prompting matrix function. Finally, taking a specific sparse-prompting function, the fraction function, we conduct a series of simulations on different databases to get the performance of our algorithm tested, and experimental results show that our algorithm can obtain lower clustering error rate and higher value of evaluation indicators ACC, NMI and ARI than other state-of-the-art subspace clustering algorithms on different databases.

INDEX TERMS Band restricted thresholding operator, face clustering, low rank representation, motion segmentation, sparse-prompting quasi-rank function, subspace segmentation.

I. INTRODUCTION

In the past few decades, there has been an explosion of high dimensional databases in many fields, such as machine learning, computer vision and signal processing. With the number of dimensions in a database increasing, distance measures become increasingly meaningless. In fact, in very high dimensions, the distances between data points are almost equidistant [12], which leads to the performance damage of many traditional clustering algorithms. While the high dimensional databases, which are not uniformly distributed across the circumambient space and their essential dimensions are much lower than that of the circumambient space,

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine¹.

often lie in the union of some low dimensional subspaces [2], [29], [45]. To cluster such data into clusters, in which every cluster corresponds to a underlying subspace, the methods of high-dimensional data clustering, subspace clustering methods [41], [48], are proposed.

In general, the subspace clustering methods roughly consist of four kinds: iterative methods [3], [24], statistical methods [19], [35], [44], algebraic methods [36], [43], [49] and spectral clustering-based methods [8], [28], [34], of which the first three are sensitive to noise, outliers, and initialization [16]. The spectral clustering-based methods, which firstly build an affinity matrix that can accurately describe the similarity of each pair of data points and then apply this affinity matrix to the framework of spectral clustering [39], have performed superiorly in many fields such as computer

vision, signal and image processing and machine learning [1], [25], [33], [54]. We can see that the key to spectral clustering-based methods is to build a good affinity matrix.

Recently, some methods for constructing a good affinity matrix have been proposed based on sparse and low-rank representation, such as the sparse subspace clustering (SSC) [16], the low rank representation (LRR) [31], [32] and the low rank subspace clustering (LRSC) [26], [50]. In SSC schemes, the sparse representation and the ℓ_1 -norm minimization are employed for a desired affinity matrix. If the clean data are drawn from independent linear subspaces, it is proved that the points which have nonzero coefficients in the sparse representation of a point are in the same subspace. Different from SSC, the low rank representation and the nuclear norm minimization are employed for a desired low-rank affinity matrix in the LRR and LRSC schemes. If the clean data are from independent linear subspaces, both LRR and LRSC also show that the optimal solution matrix to the nuclear norm minimization not only coincides with the optimal solution to the rank minimization but also forms a block-diagonal matrix, i.e., its (i, j) th entry is nonzero only if the i th and the j th points are from the same subspaces, which implies that the optimal solution is a good affinity matrix. However, if the data are corrupted by noise, especially by the gross errors, it is not clear whether the obtained affinity matrix is the lowest rank or whether the obtained sparse noise matrix is the sparsest from theoretical viewpoint, which are extremely important for a good affinity matrix. To obtain a better affinity matrix, many non-convex low rank approximation functions, such as ℓ_0 -norm [4], the Schatten p -norm [9], [52], [58], the weighted Schatten p -norm [53], log-determinant rank function [27] and multivariate GMC penalty function [4] are employed. Extensive experiments demonstrate that these approximation functions have a better performance. However, are these non-convex functions used as the measure of the low rank in subspace clustering reasonable in theory and is there any other function that can replace them?

In this paper, the two questions above are mainly discussed. First, we show that the sparse-prompting penalty function can produce a sparse solution for the noiseless signals in sparse signal recovery and based on this result, we demonstrate that the optimal solution to the sparse-prompting quasi-rank function minimization is the lowest rank and forms a block-diagonal matrix if the uncorrupted data are drawn from independent linear subspaces. This conclusion that coincides with the nuclear norm minimization used in LRR and LRSC implies that any sparse-prompting quasi-rank function used as the measure of the low rank in subspace clustering is reasonable. Second, if the data are corrupted by Gaussian noise and/or gross errors, we replace the low rank and gross errors parts with the sparse-prompting quasi-rank function and the sparse-prompting matrix function, respectively. The optimization solution to this problem is different from LRR and LRSC. Subsequently, we apply the alternating direction method of multipliers to solving it and every sub-optimization

problem has a closed-form optimal solution when the band restricted thresholding operator induced by its corresponding sparse-prompting function has an analytic expression. Finally, taking a specific sparse-prompting function, the fraction function, we conduct a series of simulations on different databases including Extended Yale B face database, ORL face database, AR face database and Hopkins 155 motion segmentation database to demonstrate the performance of our algorithm. The results show that our algorithm can obtain lower clustering error rate and higher value of evaluation indicators ACC, NMI and ARI than other state-of-the-art subspace clustering algorithms on different databases.

The paper is organized as follows. In section II, we mainly review some existing approaches on subspace clustering by sparse and low rank representation, including sparse subspace clustering(SSC), low rank representation(LRR) and low rank subspace clustering(LRSC). In section III, we review some existing results on sparse signal recovery and show that the sparse-prompting penalty function can lead to the sparse solution to the constrained minimization problem, which is useful in discussing the subspace segmentation by low rank representation with the clean data. In section IV, we discuss the regularized rank minimization via the sparse-promoting quasi-rank function and show that the optimal solution can be acquired by a matrix band restricted thresholding operator. The subspace segmentation by low rank representation via the sparse-prompting quasi-rank function with the clean data is discussed in section V. We show that any sparse-prompting quasi-rank function used as the measure of the low rank in subspace clustering is reasonable. The subspace segmentation by low rank representation via the sparse-prompting quasi-rank function with the corrupted data is discussed in section VI. A series of simulations are conducted to test the performance of our algorithm by taking the fraction function as a specific sparse-promoting function in section VII. Finally, section VIII gives the conclusions.

II. RELATED WORK

In this section, some classic methods on subspace clustering by sparse and low rank representation, including sparse subspace clustering(SSC) [16], subspace segmentation by low rank representation(LRR) [31], [32] and low rank subspace clustering(LRSC) [26], [50], is reviewed.

Let $X = (x_1, x_2, \dots, x_N) \in \mathbb{R}^{d \times N}$ be a set of d -dimension data points drawn from an unknown union of k linear subspaces $S_j (j = 1, 2, \dots, k)$ with corresponding dimension d_j . Subspace clustering aims to find out the number k of subspaces and cluster data into clusters, in which every cluster corresponds to an underlying subspace. As is described in Section I, the key to spectral clustering-based methods is to construct a good affinity matrix that can accurately depict the similarity of each pair of points. If the set of data points has the self-expressive property, i.e., each data point in a union of subspace can be efficiently represented as a linear combination of other data points, the representation coefficients, considered as a similarity measure, can be regarded as a good

affinity matrix. SSC, LRR and LRSC all construct an affinity matrix based on representation coefficients, which is widely used at present. The major difference between them is that SSC is based on a sparse representation and LRR and LRSC are based on a low rank representation.

A. SPARSE SUBSPACE CLUSTERING (SSC)

The goal of SSC is to find out a sparse representation of the data points set X by minimizing the number of nonzero coefficients. If the data are clean, the following model is employed

$$\min_C \|C\|_0 \quad s.t. \quad X = XC \text{ and } \text{diag}(C) = 0, \quad (\text{II.1})$$

where C is a sparse representation matrix and $\|C\|_0$ means the number of non-zero elements in the matrix C . Because this minimization problem in general is NP-hard, ℓ_1 minimization, as the tightest convex relaxation of ℓ_0 minimization, is considered. That is

$$\min_C \|C\|_1 \quad s.t. \quad X = XC \text{ and } \text{diag}(C) = 0, \quad (\text{II.2})$$

where $\|C\|_1 = \sum_{ij} |c_{ij}|$, i.e., the sum of absolute value of elements in the matrix. It is shown in [16] that under some conditions, the solutions to (II.1) and (II.2) coincide and that $c_{ij} \neq 0$ only if points i and j are in the same subspace.

If the data are corrupted by Gaussian noise G and gross errors E , i.e., only a small percentage of the entries of X are corrupted [7], the affinity matrix C can be acquired by solving the following convex optimization model

$$\begin{aligned} \min_C \quad & \|C\|_1 + \frac{\lambda}{2} \|G\|_F^2 + \delta \|E\|_1 \\ s.t. \quad & X = XC + G + E \text{ and } \text{diag}(C) = 0. \end{aligned} \quad (\text{II.3})$$

B. SUBSPACE CLUSTERING BY LOW RANK REPRESENTATION (LRR)

Different from SSC, the goal of LRR is to find out a low rank representation. If the data are clean, the following model is considered

$$\min_C \text{rank}(C) \quad s.t. \quad X = XC. \quad (\text{II.4})$$

Because of the discrete nature of the rank function, this problem is difficult to solve in general. Hence, the following convex optimization model is considered:

$$\min_C \|C\|_* \quad s.t. \quad X = XC, \quad (\text{II.5})$$

in which $\|C\|_*$ is the nuclear norm of C , i.e., the sum of the singular values of C . It is shown that in [31], [32] the solution to model (II.5) is the Costeria and Kanade affinity matrix $C = V_1 V_1^T$, where $X = U_1 \Sigma V_1^T$ is the rank d singular value decomposition of X . In fact, $V_1 V_1^T$ is a block-diagonal matrix if linear subspaces are independent [11], which implies that $C = V_1 V_1^T$ is a good affinity matrix.

With the data corrupted by Gaussian noise and/or gross errors, the following convex optimization problem is considered,

$$\min_C \|C\|_* + \lambda \|E\|_{2,1} \quad s.t. \quad X = XC + E, \quad (\text{II.6})$$

in which $\|E\|_{2,1} = \sum_{i=1}^N \sqrt{\sum_{j=1}^N |e_{ij}|^2}$ is the $\ell_{2,1}$ norm of E .

C. LOW RANK REPRESENTATION SUBSPACE CLUSTERING (LRSC)

Although the low rank representation is used in LRR and LRSC, LRSC finds a symmetric, low rank affinity matrix C . Thus, the LRSC solves the following non-convex optimization problems

$$\min_C \|C\|_* \quad s.t. \quad X = XC, \quad C = C^T, \quad (\text{II.7})$$

if X is clean, and

$$\begin{aligned} \min_{Y, C, G, E} \quad & \|C\|_* + \frac{\lambda}{2} \|Y - YC\|_2^2 + \frac{\alpha}{2} \|G\|_2^2 + \gamma \|E\|_1 \\ s.t. \quad & X = Y + G + E, \quad Y = YC \text{ and } C = C^T, \end{aligned} \quad (\text{II.8})$$

if X is corrupted by the Gaussian noise G and the gross errors E . It is shown that, if the data X are clean, the optimal solution to problem (II.7) is also the Costeria and Kanade affinity matrix $C = V_1 V_1^T$ which is similar to LRR, and hence C can be used as an affinity matrix. Moreover, if X is corrupted only by the Gaussian noise G (i.e., $\gamma = \infty$), then Y and C can be acquired with the singular values of X and Y thresholded, respectively.

III. SPARSE SIGNAL RECOVERY VIA THE SPARSE-PROMPTING PENALTY FUNCTION

We briefly review some definitions and results on sparse signal recovery in this section, especially the optimal solution to the regularized minimization via the sparse-prompting penalty function is the band restricted thresholding operator. In addition, for the constrained minimization problem, we show that the sparse-prompting penalty function can lead to the sparse solution, which will be used in discussing the subspace segmentation by low rank representation if the data are clean.

Sparse signal recovery, solving a high dimension under-determined system for sparse solutions, has attracted much attention in recent years and there is no doubt that it is definitely beneficial and favorable in different fields, such as compressed sensing [13], [15], [21], subspace clustering [16], [31], [32]. To model this problem, the following constrained minimization problem is commonly considered,

$$(P_0) \quad \min_{x \in \mathbb{R}^N} \|x\|_0 \quad \text{subject to } Ax = b, \quad (\text{III.1})$$

which is also known as ℓ_0 minimization problem [38], in which A is an $m \times N$ real matrix with $\text{rank}(A) \ll N$, $x \in \mathbb{R}^N$ and $b \in \mathbb{R}^m$, and $\|x\|_0$ named ℓ_0 norm is the number of nonzero entries in it. If A is a full row rank matrix,

$rank(A) \ll N$ becomes $m \ll N$. However, the ℓ_0 optimization problem is actually NP-hard and sensitive to noise because of the discontinuous nature of the ℓ_0 norm.

An important method to solve ℓ_0 minimization problem is to replace ℓ_0 -norm with a continuous relaxation function, which is called sparsity-promoting penalty function in this paper.

Definition 1: A continuous function $p : \mathbb{R} \rightarrow \mathbb{R}$ is called a (strict) sparse-promoting function if $p(t) = p(-t) (\forall t \in \mathbb{R})$, $p(0) = 0$ and $p(t)$ is a nondecreasing (strict) concave function on $t \in (0, +\infty)$.

Definition 2: Let $p(\cdot)$ be the (strict) sparse-promoting function. Then

(1) For a signal $x \in \mathbb{R}^N$, the function $P : \mathbb{R}^n \rightarrow \mathbb{R}$, $P(x) = \sum_{i=1}^N p(x_i)$ is called the (strict) sparse-promoting penalty function.

(2) Given a matrix $X \in \mathbb{R}^{m \times N}$, the function $\mathcal{P}^* : \mathbb{R}^{m \times N} \rightarrow \mathbb{R}$, $\mathcal{P}^*(X) = \sum_{i=1}^m \sum_{j=1}^N p(x_{ij})$ is called the (strict) sparse-promoting matrix function.

(3) Assume that $X = U \Sigma V^T$ is the singular value decomposition (SVD) of the matrix $X \in \mathbb{R}^{m \times N}$, in which U and V are $m \times m$ and $N \times N$ unitary matrices and $\Sigma = diag(\sigma_i)_{1 \leq i \leq m}$ is the diagonal matrix with the singular values σ_i of X . Then the function $\mathcal{P} : \mathbb{R}^{m \times N} \rightarrow \mathbb{R}$, $\mathcal{P}(X) = \sum_{i=1}^m p(\sigma_i)$ is called the (strict) sparse-promoting quasi-rank function.

There are many different sparse-promoting functions used in literature; some relevant examples are given in Table 1.

TABLE 1. Continuous sparse-promoting function.

$ x ^p$	$0 < p \leq 1$	[6], [14], [20], [42]
$\frac{ x }{ x +p}$	$p > 0$	[23], [30], [40]
$\frac{(p+1) x }{p+ x }$	$p > 0$	[59]
$\frac{\log(1+p^{-1} x)}{\log(1+p^{-1})}$	$p > 0$	[22]
$\tanh(\frac{ x }{p})$	$p > 0$	[10]
$1 - e^{-\frac{ x }{p}}$	$p > 0$	[47]
$\begin{cases} 1 & x > p^{-1} \\ 2p x - p^2x^2 & x \leq p^{-1} \end{cases}$	$p > 0$	[59]
$\begin{cases} 1 & x > ap \\ \frac{-x^2+2ap x -p^2}{p^2(a^2-1)} & x \leq 1 \end{cases}$	$p > 0, a > 1$	[17]
$\begin{cases} 1 & x > p \\ \frac{2 x }{p(a+1)} & x \leq p \end{cases}$		

Replacing the ℓ_0 norm with a sparse-promoting penalty functions $P(\cdot)$, the minimization (III.1) is rewritten as

$$\min_{x \in \mathbb{R}^N} P(x) \text{ subject to } Ax = b \quad (III.2)$$

for the constrained problem and

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} \|Ax - b\|_2^2 + \lambda P(x) \quad (III.3)$$

for the regularization problem, where $\lambda > 0$ is the regularized parameter. It is noted that all the sparse-promoting functions are strict except for ℓ_1 norm. ℓ_1 norm is typical and popular

and based on it, considerable excellent theoretical work has been done [6], [14]. However, because ℓ_1 -norm, as a convex surrogate form of the ℓ_0 -norm, often fails not only to produce the sparsest solution but also to select the position of non-zero coefficient [17], [57], many strict sparse-promoting functions are discussed [17], [20], [22], [30], [42], [57], [59].

In the following, we first demonstrate that any sparse-promoting penalty function can reduce the sparse optimal solution in sparse signal recovery.

Lemma 1: (1) Suppose that $P(x)$ is any strict sparse-promoting penalty function and \hat{x} is the optimal solution to problem (III.2), then the columns of matrix $A_{m \times N}$ corresponding with the support of \hat{x} are linear independent, i.e., $\|\hat{x}\|_0 = k \leq rank(A)$.

(2) If $P(x) = \|x\|_1$, the problem (III.2) has at least one optimal solution x^* with at most $rank(A)$ non-zeros.

Proof: (1) Assume that \hat{x} is the optimal solution to problem (III.2) with $\|\hat{x}\|_0 = k > rank(A)$, then the k columns of matrix $A_{m \times N}$ corresponding with the support of \hat{x} are linear dependent. Thus, a non-trivial vector $h \in \mathbb{R}^N$ exists which has the same support as \hat{x} such that $Ah = 0$ and

$$\max_{1 \leq j \leq k} |h_j| < \min_{1 \leq j \leq k} |\hat{x}_j|.$$

Obviously, $\hat{x} - h$, $\hat{x} + h$ are also solutions to $Ax = b$ and for arbitrary j , $\hat{x}_j - h_j$, $\hat{x}_j + h_j$ and \hat{x}_j have the same sign.

Because $p(t)$ is a strictly concave function on $t \in (0, +\infty)$, we have

$$p(|\hat{x}_j - h_j|) + p(|\hat{x}_j + h_j|) < 2p(|\hat{x}_j|),$$

which implies that

$$\sum_{j=1}^n p(|\hat{x}_j - h_j|) + \sum_{j=1}^n p(|\hat{x}_j + h_j|) < 2 \sum_{j=1}^n p(|\hat{x}_j|).$$

Furthermore, we can get

$$\sum_{j=1}^n p(|\hat{x}_j - h_j|) < \sum_{j=1}^n p(|\hat{x}_j|)$$

or

$$\sum_{j=1}^n p(|\hat{x}_j + h_j|) < \sum_{j=1}^n p(|\hat{x}_j|),$$

which are in contradiction to the fact that \hat{x} is the optimal solution to problem (III.2), so $\|\hat{x}\|_0 \leq k$.

(2) See [15]. □

Remark 1: It is necessary to point out that, when $P(x) = \|x\|_1$, the problem (III.2) may have more than one solution and the number of non-zeros of other solutions may be larger than $rank(A)$, which is different from Lemma 1 (1).

Second, we introduce an important result that any sparse-promoting function can induce a band restricted thresholding operator.

Definition 3: A odd function $h_\tau : \mathbb{R} \rightarrow \mathbb{R}$ with parameter $\tau > 0$ is called a band restricted thresholding operator if h_τ in \mathbb{R}^+ satisfies

- (1) for all $u \leq \tau$, $h_\tau(u) = 0$;
- (2) if $u \leq v$, $h_\tau(u) \leq h_\tau(v)$;
- (3) there is a constant $c \in [0, 1]$ such that $u - c\tau \leq h_\tau(u) \leq u$ for all $\tau < u$, where τ is called a thresholding parameter and c is called a band parameter.

Definition 4: Let h_τ be a band restricted thresholding operator. Then

(1) the vector band restricted thresholding operator $H_\tau : \mathbb{R}^N \rightarrow \mathbb{R}^N$ is defined as

$$H_\tau(x) = (h_\tau(x_1), h_\tau(x_2), \dots, h_\tau(x_N))^T \quad (\forall x \in \mathbb{R}^N), \quad (III.4)$$

and

(2) the matrix band restricted thresholding operator $\mathcal{H}_\tau : \mathbb{R}^{m \times N} \rightarrow \mathbb{R}^{m \times N}$ is defined as

$$\mathcal{H}_\tau(X) = U\mathcal{H}_\tau(\Sigma)V^T \quad (\forall X \in \mathbb{R}^{m \times N}) \quad (III.5)$$

where $\mathcal{H}_\tau(\Sigma) = \text{diag}(h_\tau(\sigma_i))$ and $X = U\Sigma V^T$ is the singular value decomposition of the matrix X .

Lemma 2: Let u^* be an optimal solution to the problem

$$\min_{u \in \mathbb{R}} f_\lambda(u) = \frac{1}{2}(u - t)^2 + \lambda p(u), \quad (III.6)$$

where $p(\cdot)$ is a sparse-promoting function and $t \in \mathbb{R}$, $\lambda > 0$ are two given numbers. Then there is a band restricted thresholding operator $h_\tau(t)$ such that $u^* = h_\tau(t)$, where τ and c are defined as

$$\tau = \inf_{t > 0} \left\{ t \mid \frac{t}{2} + \frac{\lambda p(t)}{t} \right\}, \quad c = \lim_{t \rightarrow \tau^+} \frac{t - h_\tau(t)}{t}.$$

Proof: See [60]. □

Lemma 3: Let $x^* = (x_1^*, x_2^*, \dots, x_N^*)^T$ be the optimal solution to (III.3). Then there exists a vector band restricted thresholding operator $H_\tau(x)$ such that $x_i^* = \text{sign}(x_i^*)h_\tau(B_\mu(x_i^*))$, where $B_\mu(x) = x + \mu A^T(b - Ax)$ and $0 < \mu < \|A\|_2^{-2}$.

Proof: See [60]. □

Remark 2: Lemma 3 shows that any sparse-promoting penalty function can lead to a band restricted thresholding operator H_τ , but we cannot make sure whether its analytic expression exists. For example, only when $p = \frac{1}{2}, \frac{2}{3}$ and 1, the band restricted thresholding operators induced by $P(x) = \|x\|^p$ ($0 < p \leq 1$) have analytic expressions.

At the end of this section, we introduce a specific sparse-promoting function, the fraction function $p(x) = \frac{a|x|}{a|x|+1}$ ($a > 0$), in detail, since it will be taken in section VII. In fact, the fraction function is widely used in image restoration [23], [40] and sparse signal recovery [30]. In [30], the equivalence between ℓ_0 minimization and fractional function minimization is discussed and the analytic expression of the optimal solution to model (III.3) is given.

Lemma 4: If $x^* = (x_1^*, x_2^*, \dots, x_N^*)^T$ is an optimal solution to (III.3), where $P(x)$ is the fraction function, a and λ are positive parameter and μ satisfies $0 < \mu < \|A\|_2^{-2}$, then the optimal solution x^* is

$$x_i^* = \begin{cases} g_{\lambda\mu}((B_\mu(x^*))_i), & |(B_\mu(x^*))_i| > t^*, \\ 0, & \text{otherwise.} \end{cases} \quad (III.7)$$

where

$$g_{\lambda\mu}(t) = \text{sign}(t) \left(\frac{\frac{1+at}{3}(1 + 2\cos(\frac{\phi(t)}{3} - \frac{\pi}{3})) - 1}{a} \right), \quad (III.8)$$

$$\phi(t) = \arccos\left(\frac{27\lambda\mu a^2}{2(1+a|t|)^3} - 1\right), \quad B_\mu(x^*) = x^* + \mu A^T(b - Ax^*), \quad (III.9)$$

and

$$t^* = \begin{cases} \lambda\mu a, & \text{if } \lambda \leq \frac{1}{2\mu a^2}, \\ \sqrt{2\lambda\mu} - \frac{1}{2a}, & \text{if } \lambda > \frac{1}{2\mu a^2}. \end{cases} \quad (III.10)$$

Proof: See [30]. □

IV. LOW RANK MINIMIZATION VIA THE SPARSE-PROMOTING QUASI-RANK FUNCTION

In this section we mainly discuss the regularized rank minimization via the sparse-promoting quasi-rank function and show that the optimal solution can be acquired by a matrix band restricted thresholding operator.

Assume that the corrupted data matrix $D = Z + G$, where G is the Gaussian noise and Z is an unknown clean low rank matrix, then the following regularized rank minimization is commonly employed to find a low rank approximation of Z ,

$$\min_Z \frac{1}{2} \|D - Z\|_F^2 + \lambda \text{rank}(Z), \quad (IV.1)$$

where $\lambda > 0$ is a parameter. Because of the discrete nature of the rank function, this problem is NP-hard in general. Hence, the following convex optimization model is considered:

$$\min_Z \frac{1}{2} \|D - Z\|_F^2 + \lambda \|Z\|_*, \quad (IV.2)$$

in which $\|Z\|_*$ is the nuclear norm of Z . The optimal solution to this problem is $Z = U\mathcal{S}_\lambda(\Sigma)V^T$ [5], where $D = U\Sigma V^T$, $\Sigma = \text{diag}(\sigma_i)_{1 \leq i \leq m}$ is the singular value decomposition of the matrix D , and $\mathcal{S}_\lambda(x) = \text{diag}(s_\lambda(\sigma_i))$ is the matrix soft thresholding operator, i.e.,

$$s_\lambda(\sigma_i) = \begin{cases} \sigma_i - \lambda, & \text{if } \lambda \leq \sigma_i, \\ 0, & \text{else.} \end{cases} \quad (IV.3)$$

Before giving the main conclusion of this section, we need the following inequality.

Lemma 5: (Von Neumann's Inequality) Let matrices $X, Y \in \mathbb{R}^{m \times N}$ be given, and let $\sigma_1(X) \geq \sigma_2(X) \geq \dots \geq \sigma_N(X)$ and $\sigma_1(Y) \geq \sigma_2(Y) \geq \dots \geq \sigma_N(Y)$ denote the singular values of X and Y , respectively. Then

$$\text{tr}(X^T Y) \leq \sum_{i=1}^N \sigma_i(X)\sigma_i(Y), \quad (IV.4)$$

The equality achieves if and only if there exist two unitary matrices U and V such that

$$X = U\Sigma_X V^T \quad \text{and} \quad Y = U\Sigma_Y V^T$$

hold simultaneously, where Σ_X and Σ_Y are the diagonal matrices and its diagonal elements are arranged in decreasing singular values of X and Y , respectively.

Proof: See [37]. □

Theorem 1: Let $D \in \mathbb{R}^{m \times N}$ be a given matrix and $\mathcal{P}(Z)$ the sparse-promoting quasi-rank function. Then the optimal solution Z^* to the following minimization

$$\min_Z \frac{1}{2} \|D - Z\|_F^2 + \lambda \mathcal{P}(Z) \quad (IV.5)$$

is the matrix band restricted thresholding operator $\mathcal{H}_\tau(D)$ defined in Definition 4, where the band restricted thresholding operator h_τ is induced by sparse-promoting function $p(\cdot)$.

Proof: Let $D = U_D \Sigma_D V_D^\top$ and $Z = U_Z \Sigma_Z V_Z^\top$ be the SVD of D and Z , where Σ_D and Σ_Z are the $m \times N$ diagonal matrices and its diagonal elements are arranged in decreasing singular values of D and Z , respectively. We have

$$\begin{aligned} & \frac{1}{2} \|D - Z\|_F^2 + \lambda \mathcal{P}(Z) \\ &= \frac{1}{2} \text{tr}(D^\top D) - \text{tr}(D^\top Z) + \frac{1}{2} \text{tr}(Z^\top Z) + \lambda \mathcal{P}(Z) \\ &= \frac{1}{2} \sum_{i=1}^N \sigma_i(D) - \text{tr}(D^\top Z) + \frac{1}{2} \sum_{i=1}^N \sigma_i(Z) + \lambda \sum_{i=1}^N p(\sigma_i(Z)) \\ &\geq \frac{1}{2} \sum_{i=1}^N \sigma_i(D) - \sum_{i=1}^N \sigma_i(D) \sigma_i(Z) + \frac{1}{2} \sum_{i=1}^N \sigma_i(Z) + \lambda \sum_{i=1}^N p(\sigma_i(Z)) \\ &= \frac{1}{2} \sum_{i=1}^N (\sigma_i(D) - \sigma_i(Z))^2 + \lambda \sum_{i=1}^N p(\sigma_i(Z)), \end{aligned} \quad (IV.6)$$

where the inequality holds by (IV.4). Since the optimal problem

$$\min_{\sigma_1(Z), \sigma_2(Z), \dots, \sigma_N(Z)} \sum_{i=1}^N \frac{1}{2} (\sigma_i(D) - \sigma_i(Z))^2 + \lambda \sum_{i=1}^N p(\sigma_i(Z)) \quad (IV.7)$$

is separable, it can be converted to N sub-problems

$$\min_{\sigma_i(Z)} \frac{1}{2} (\sigma_i(D) - \sigma_i(Z))^2 + \lambda p(\sigma_i(Z)). \quad (IV.8)$$

By Lemma 2, its optimal solution $(\sigma_i(Z))^*$ is a band restricted thresholding operator $h_\tau(\sigma_i(D))$.

From Lemma 5, we notice that the equality

$$\text{tr}(D^\top Z) = \sum_{i=1}^N \sigma_i(D) \sigma_i(Z)$$

is achieved if and only if

$$Z^* = U_D \Sigma_Z^* V_D^\top,$$

where Σ_Z^* denotes the $m \times N$ diagonal matrices and its diagonal elements are arranged in decreasing singular values of Z^* , i.e., the optimal solution Y^* to problem (IV.5) is

$$\begin{aligned} Z^* &= U_D \Sigma_Z^* V_D^\top = U_D \text{diag}((\sigma_i(Z))^*) V_D^\top \\ &= U_D \text{diag}(h_\tau(\sigma_i(D))) V_D^\top = \mathcal{H}_\tau(D). \end{aligned}$$

□

V. SUBSPACE SEGMENTATION BY LOW RANK REPRESENTATION VIA THE SPARSE-PROMPTING QUASI-RANK FUNCTION WITH UNCORRUPTED DATA

In this section, we show that the optimal solution to the optimization model based on the sparse-prompting quasi-rank function is the lowest rank and forms a block-diagonal matrix if the uncorrupted data are drawn from independent linear subspaces. This conclusion implies that any sparse-prompting quasi-rank function used as the measure of the low rank in subspace clustering is reasonable.

Given a clean data matrix $Y \in \mathbb{R}^{m \times N}$, whose columns are drawn from a union of k low dimensional linear subspaces of unknown dimensions $\{d_i\}_{i=1}^k$ satisfying $d_i \ll m$ and $\sum_{i=1}^k d_i \ll N$, we consider the following optimization problem

$$\min_C \mathcal{P}(C) \text{ subject to } Y = YC, \quad (V.1)$$

where $\mathcal{P}(C)$ is any sparse-prompting quasi-rank function. The following theorem shows that the optimal solution to problem (V.1) is also the Costeria and Kanade affinity matrix $C = V_1 V_1^\top$, which is similar to the nuclear norm used in LRR and LRSC. In fact, this matrix is called Shape Interaction Matrix (SIM) in [11] and has been widely used for subspace clustering.

Theorem 2: Let $\text{rank}(Y) = k$ and $Y = U \Sigma V^\top$ be the SVD of Y , where $\Sigma = \text{diag}(\sigma_i)$ is a diagonal matrix and its diagonal elements are arranged in decreasing singular values of Y . Then the optimal solution to problem (V.1) is

$$C = V_1 V_1^\top, \quad (V.2)$$

where $V = [V_1, V_2]$ is partitioned according to the sets $\mathbf{I}_1 = \{i : \sigma_i > 0\}$ and $\mathbf{I}_2 = \{i : \sigma_i = 0\}$. Moreover, the optimal value is

$$\mathcal{P}(C) = \text{rank}(Y) = k. \quad (V.3)$$

Proof: Let $C = P \Delta Q^\top$ be the SVD of C , where $\Delta = \text{diag}(\delta_i)$ is the diagonal matrix and its diagonal elements are arranged in decreasing singular values of C . Then $Y = YC$ can be rewritten as $U \Sigma V^\top = U \Sigma V^\top P \Delta Q^\top$, which implies that

$$\Sigma V^\top Q = \Sigma V^\top P \Delta. \quad (V.4)$$

Let $W = \Sigma V^\top Q = (w_{ij})$ and $Z = \Sigma V^\top P = (z_{ij})$. Then, $\text{rank}(Z) = k$ and the result

$$\left(\sum_{j=1}^N w_{1j}, \sum_{j=1}^N w_{2j}, \dots, \sum_{j=1}^N w_{kj} \right)^\top = Z \delta$$

holds, where $\delta = (\delta_1, \delta_2, \dots, \delta_N)^\top$, which reduces to the following optimization problem

$$\min_{\delta} P(\delta) \text{ s.t. } Z \delta = \left(\sum_{j=1}^N w_{1j}, \sum_{j=1}^N w_{2j}, \dots, \sum_{j=1}^N w_{kj} \right)^\top. \quad (V.5)$$

Hence, the optimal solution δ^* with $\|\delta^*\|_0 \leq k$ is equal to $\text{rank}(Y)$ by Lemma 1. On the other hand, $Y = YC$ implies that $\text{rank}(Y) \leq \text{rank}(C)$ and hence $\text{rank}(Y) = \text{rank}(C) = k$.

It follows from $\text{rank}(Y) = \text{rank}(C) = k$ that equation (V.4) can be rewritten as

$$\begin{aligned} & \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{11} & V_{21} \\ V_{12} & V_{22} \end{pmatrix}^\top \begin{pmatrix} Q_{11} & Q_{21} \\ Q_{12} & Q_{22} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{11} & V_{21} \\ V_{12} & V_{22} \end{pmatrix}^\top \begin{pmatrix} P_{11} & P_{21} \\ P_{12} & P_{22} \end{pmatrix} \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

where $\Sigma_1 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ and $\Delta_1 = \text{diag}(\delta_1, \delta_2, \dots, \delta_k)$, which implies that

$$V_1^\top Q_2 = 0 \tag{V.6}$$

and

$$V_1^\top Q_1 = V_1^\top P_1 \Delta_1 \tag{V.7}$$

Equation (V.6) means that the columns of Q_2 must be orthogonal to the columns of V_1 , and hence the columns of Q_1 must be in the range of V_1 . Thus, $Q_1 = V_1 R$ for a rotation matrix R , which reduces to $V_1^\top Q_1 = V_1^\top V_1 R = R$. Together with Equation (V.7), we obtain that $R = V_1^\top P_1 \Delta_1$, which leads to $\Delta_1 = I_k$ and $P_1 = V_1 R_1$. Hence,

$$C = (P_1 \ P_2) \begin{pmatrix} \Delta_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Q_1^\top \\ Q_2^\top \end{pmatrix} = V_1 R R^\top V_1^\top = V_1 V_1^\top.$$

The proof is completed. \square

By the properties of Shape Interaction Matrix (SIM) [11], the following Corollary holds, which implies that any sparse-promoting quasi-rank function as the measure of the low rank in subspace clustering is reasonable.

Corollary 1: When linear subspaces are independent, the optimal solution, $V_1 V_1^\top$, to problem (V.1) forms a block-diagonal matrix: The (i, j) th entry of $V_1 V_1^\top$ can be nonzero only if the i th entry and j th entry samples are from the same subspace.

VI. SUBSPACE SEGMENTATION BY LOW RANK REPRESENTATION VIA THE SPARSE-PROMPTING QUASI-RANK FUNCTION WITH CORRUPTED DATA

In this section, we discuss subspace segmentation by low rank representation via the sparse-prompting quasi-rank function if the data are corrupted by Gaussian noise and/or gross errors. We replace the low rank and gross errors parts with the sparse-prompting quasi-rank function and the sparse-prompting matrix function, respectively. Subsequently, we apply the alternating direction method of multipliers to solving it and every sub-optimization problem has a closed-form optimal solution when the band restricted thresholding operator induced by its corresponding sparse-prompting function has an analytic expression.

Let a corrupted data matrix $X = Y + G + E$ be given, where Y is an unknown clean matrix, G is a Gaussian noise and E is the gross errors, we consider the following

optimization problem

$$\begin{aligned} & \min_{Y, C, G, E} \frac{1}{2} \|G\|_F^2 + \lambda_1 \mathcal{P}(C) + \lambda_2 \mathcal{P}^*(E) \\ & \text{s.t. } Y = YC \text{ and } X = Y + G + E, \end{aligned} \tag{VI.1}$$

where $\mathcal{P}(C)$ and $\mathcal{P}^*(E)$ are the sparse-promoting quasi-rank function and sparse-promoting matrix function, respectively. Clearly, this problem is equivalent to the following problem

$$\begin{aligned} & \min_{C, G, E} \frac{1}{2} \|G\|_F^2 + \lambda_1 \mathcal{P}(C) + \lambda_2 \mathcal{P}^*(E) \\ & \text{s.t. } X = (X - G - E)C + G + E. \end{aligned} \tag{VI.2}$$

We introduce an auxiliary variable S , and the problem (VI.2) is rewritten as

$$\begin{aligned} & \min_{C, G, E} \frac{1}{2} \|G\|_F^2 + \lambda_1 \mathcal{P}(C) + \lambda_2 \mathcal{P}^*(E) \\ & \text{s.t. } X = (X - G - E)S + G + E, S = C. \end{aligned} \tag{VI.3}$$

This problem can be solved by the alternating direction method of multipliers (ADMM), which minimizes the following augmented Lagrangian function

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \|G\|_F^2 + \lambda_1 \mathcal{P}(C) + \lambda_2 \mathcal{P}^*(E) \\ &+ \text{tr}(\Lambda_1^\top (X - (X - G - E)S - G - E)) + \text{tr}(\Lambda_2^\top (S - C)) \\ &+ \frac{\nu}{2} (\|X - (X - G - E)S - G - E\|_F^2 + \|S - C\|_F^2), \end{aligned} \tag{VI.4}$$

where Λ_1 and Λ_2 are Lagrange multipliers and ν is a penalty parameter.

Given variables S_k, C_k, E_k, G_k and $\Lambda_{1,k}, \Lambda_{2,k}$ and ν_k , the updated rules are as follows:

$$C_{k+1} = \arg \min_{C \in \mathbb{R}^{N \times N}} \frac{1}{2} \|C - S_k - \frac{\Lambda_{2,k}}{\nu_k}\|_F^2 + \frac{\lambda_1}{\nu_k} \mathcal{P}(C), \tag{VI.5}$$

$$\begin{aligned} S_{k+1} &= \arg \min_{S \in \mathbb{R}^{N \times N}} \frac{1}{2} \|S - C_{k+1} + \frac{\Lambda_{2,k}}{\nu_k}\|_F^2 \\ &+ \frac{1}{2} \|(X - G_k - E_k) - (X - G_k - E_k)S + \frac{\Lambda_{1,k}}{\nu_k}\|_F^2, \end{aligned} \tag{VI.6}$$

$$\begin{aligned} G_{k+1} &= \arg \min_{G \in \mathbb{R}^{m \times N}} \frac{1}{2} \|G\|_F^2 \\ &+ \frac{\nu_k}{2} \|G(S_{k+1} - I) + (E_k - X)(S_{k+1} - I) + \frac{\Lambda_{1,k}}{\nu_k}\|_F^2, \end{aligned} \tag{VI.7}$$

$$\begin{aligned} E_{k+1} &= \arg \min_{E \in \mathbb{R}^{m \times N}} \frac{1}{2} \|ES_{k+1} - (E_k + (X - G_{k+1})(S_{k+1} - I) \\ &- \frac{\Lambda_{1,k}}{\nu_k})\|_F^2 + \frac{\lambda_2}{\nu_k} \mathcal{P}^*(E). \end{aligned} \tag{VI.8}$$

Update the Lagrange multipliers as follows:

$$\begin{aligned} \Lambda_{1,k+1} &= \Lambda_{1,k} + \nu_k (X - (X - G_{k+1} - E_{k+1})S_{k+1} \\ &- G_{k+1} - E_{k+1}), \end{aligned} \tag{VI.9}$$

$$\Lambda_{2,k+1} = \Lambda_{2,k} + \nu_k (S_{k+1} - C_{k+1}). \tag{VI.10}$$

TABLE 2. Clustering error rate (%) on Extended Yale B face database.

Algorithms		SCC	LSA	SSC	LRR	LRR-H	LRSC	LRRSP
2 Subjects	Mean	16.82	32.80	1.86	9.52	2.54	5.32	1.05
	Median	7.82	47.66	0.00	5.47	0.78	4.69	0.00
3 Subjects	Mean	38.16	52.29	3.10	19.52	4.21	8.47	1.83
	Median	39.06	50.00	1.04	14.58	2.60	7.81	0.59
5 Subjects	Mean	58.90	58.02	4.31	34.16	6.90	12.24	2.91
	Median	59.38	56.87	2.50	35.00	5.63	11.25	1.32
8 Subjects	Mean	66.11	59.19	5.85	41.19	14.34	23.72	3.75
	Median	64.65	58.59	4.49	43.75	10.06	28.03	2.80
10 Subjects	Mean	73.02	60.42	10.94	38.85	22.92	30.36	7.00
	Median	75.78	57.50	5.63	41.09	23.59	28.75	3.92
All	Mean	50.56	52.54	5.21	28.65	10.18	16.02	3.09
	Median	49.34	54.12	2.73	27.98	8.53	16.10	1.24

Update parameter ν_k as follows:

$$\nu_{k+1} = \min\{\rho\nu_k, \nu_{max}\}. \quad (\text{VI.11})$$

Termination condition:

$$\|X - (X - G_k - E_k)S_k - E_k - G_k\|_\infty < \varepsilon$$

and $\|S - C\|_\infty < \varepsilon. \quad (\text{VI.12})$

The sub-optimization problems above are solved separately, and the specific methods are as follows:

The sub-optimization problem (VI.5) can be solved by Theorem 2 and the optimal solution is

$$C_{k+1} = \mathcal{H}_\tau(S_k + \frac{\Lambda_{2,k}}{\nu_k}) = U_k \text{diag}(\mathcal{H}_\tau(\Sigma_k))V_k^\top, \quad (\text{VI.13})$$

where $\mathcal{H}_\tau(\cdot)$ is the matrix band restricted thresholding operator and $U_k \Sigma_k V_k^\top$ is the SVD of matrix $S_k + \frac{\Lambda_{2,k}}{\nu_k}$.

The sub-optimization problem (VI.6) and (VI.7) have the closed-form solutions, which can be obtained by the Frobenius norm minimization operator:

$$S_{k+1} = (I + D_k^\top D_k)^{-1}(C_{k+1} + D_k^\top D_k + \frac{D_k^\top \Lambda_{1,k} - \Lambda_{2,k}}{\nu_k}), \quad (\text{VI.14})$$

where $D_k = (X - E_k - G_k)$, and

$$G_{k+1} = (\nu_k(X - E_k)Z_{k+1}Z_{k+1}^\top + \Lambda_{1,k}Z_{k+1}^\top)(I + \nu_k Z_{k+1}Z_{k+1}^\top)^{-1}, \quad (\text{VI.15})$$

where $Z_{k+1} = (I - S_{k+1})$.

Since the sub-optimization problem (VI.8) is separable, the i th column vector $e_{i,k+1}^\top$ of the optimal solution E_{k+1}^\top is

$$e_{i,k+1}^\top = \arg \min_{e \in \mathbb{R}^N} \frac{1}{2} \|S_{k+1}^\top e - ((X - G_{k+1})(S_{k+1} - I) + E_k - \frac{\Lambda_{1,k}}{\nu_k})^\top\|_2^2 + \frac{\lambda_2}{\nu_k} P(e), \quad (\text{VI.16})$$

LRRSP Algorithm Formulation

Input: The database $X = (x_1, x_2, \dots, x_N)$, parameters λ_1, λ_2 and ν .

Initialization: $S_1 = C_1 = E_1 = G_1 = 0, \Lambda_{1,1} = \Lambda_{2,1} = 0, \varepsilon = 10^{-8}, \nu_1 = 10^{-6}, \rho = 1.1$ and $\nu_{max} = 10^6$.

(1) Solve the sub-optimization problem (VI.5), (VI.6), (VI.7) and (VI.8), and update variables C_k, S_k, G_k and E_k in turn.

(2) Update the Lagrange multiplier $\Lambda_{1,k}$ and $\Lambda_{2,k}$ by (VI.9) and (VI.10).

(3) Update the parameter ν_k by (VI.11).

(4) Terminate by (VI.12).

(5) Calculate the similarity matrix by $W = C + C^\top$.

(6) Use the spectral clustering algorithm.

Output: The clustering result of each data point.

where $P(\cdot)$ is the corresponding sparse-promoting penalty function of $\mathcal{P}^*(\cdot)$. By Lemma 3, we have

$$e_{i,k+1} = (\text{sign}(e_{i1})h_\tau(B_\mu(e_{i1})), \dots, \text{sign}(e_{iN})h_\tau(B_\mu(e_{iN}))), \quad (\text{VI.17})$$

where h_τ is the induced band restricted thresholding operator by sparse-promoting function $p(\cdot)$. Hence, the algorithm formulation, called LRRSP, for the subspace segmentation by low rank representation via the sparse-prompting quasi-rank function is as follows.

VII. EXPERIMENTS

We conduct a series of simulations to test the performance of the algorithm formulation above in this section by taking the fraction function $p_a(x) = \frac{a|x|}{1+a|x|}$ as a specific sparse-promoting function. We evaluate the performance of the LRRSP algorithm by subspace clustering error rate and cluster evaluation indicators ACC, NMI and ARI. The performance of LRRSP algorithm is compared with other

TABLE 3. Evaluation indicators of algorithms on the Extended Yale B face database (%).

Algorithms	SCC	LSA	SSC	LRR	LRR-H	LRSC	LRRSP
ACC	41.10	41.98	95.69	65.84	93.10	87.76	97.09
NMI	38.53	23.30	86.68	53.92	79.90	75.53	88.90
ARI	16.22	9.01	62.36	29.54	60.48	59.60	70.85



FIGURE 1. Extended Yale B face database.



FIGURE 2. ORL face database.

TABLE 4. Clustering error rate (%) on ORL face database.

Algorithms	LSR	LSA	SSC	LRR	LRSC	LRRSP
5 Subjects	13.65	27.68	13.98	13.83	15.00	10.00
10 Subjects	19.83	35.18	25.50	26.20	22.47	15.95
15 Subjects	23.68	37.99	27.90	27.33	25.64	20.21
20 Subjects	25.44	40.15	26.24	29.75	26.75	22.18
All	30.48	63.50	31.59	37.07	32.76	26.07

state-of-the-art subspace clustering algorithms, such as SCC [8], LSA [55], LRR, LRR-H and LRSC, in which the implementations is provided by the authors. Experiments are performed on three standard databases: Extended Yale B face database, ORL face database, AR face database and Hopkins155 motion segmentation database. The experimental environment is Microsoft Windows10 operating system, Lenovo G500 notebook computer with Inter(R) Core(TM) i5-3230M CPU @2.60GHz processor and 4G memory.

A. EXPERIMENTS WITH THE FACE DATABASE

In this subsection, the face databases including the Extended Yale B database, the ORL face database and the AR face database are used to evaluate the performance of our algorithm. Face clustering refers to the problem of clustering a set of face images from multiple individuals according to the identity of each individual. Here, the data matrix X is of dimension $m \times N$, where m is the number of pixels and

TABLE 5. Evaluation indicators of algorithms on the ORL face database (%).

Algorithms	LSR	LSA	SSC	LRR	LRSC	LRRSP
ACC	80.17	62.85	76.58	78.83	80.17	86.82
NMI	89.14	78.42	87.90	88.53	89.10	94.05
ARI	71.25	58.35	66.01	68.97	70.97	78.43

TABLE 6. Clustering error rate (%) on AR face database.

Algorithms	LSR	LSA	SSC	LRR	LRSC	LRRSP
5 Subjects	11.25	29.97	16.24	10.03	9.81	6.09
10 Subjects	17.09	39.86	24.85	15.76	15.49	11.81
15 Subjects	24.55	46.32	29.96	23.01	22.55	17.64
20 Subjects	25.10	48.02	30.24	23.98	23.12	19.52
All	30.57	55.29	35.62	29.05	28.41	24.86

TABLE 7. Evaluation indicators of algorithms on the AR face database (%).

Algorithms	LSR	LSA	SSC	LRR	LRSC	LRRSP
ACC	82.91	60.14	75.15	84.24	84.51	88.19
NMI	92.51	75.22	88.62	92.47	92.33	96.05
ARI	75.62	53.61	66.08	75.25	75.07	78.98

N is the number of images. According to the Lambertian assessment, the set of all images of each individual with a fixed pose and varying illumination forms a cone in the image space and lies close to a linear subspace of dimension 9 [2]. In practice, a few pixels deviate from Lambertian model due to cast shadows and specularities, which can be modeled as sparse outlying entries. Therefore, the face clustering problem reduces to clustering a set of images according to multiple subspaces and corrupted by gross errors.

The Extended Yale B database samples 38 categories of faces with a grayscale size of 192×168 pixels, and each type includes 64 face images taken under different lighting conditions. In order to reduce storage space and computing cost of all algorithms, we down-sample each image to 48×42 pixels, and then vectorize it to a 2016-dimensional vector as a data sample. Figure 1 shows some sample images of the Extended Yale B face database. Following the experimental settings in [16], the 38 subjects of target images are divided into 4 groups. They are 1 to 10, 11 to 20, 21 to 30 and

TABLE 8. Clustering error rate (%) on the Hopkins155 motion segmentation database with the 2F-dimensional data points.

Algorithms		SCC	LSA	SSC	LRR	LRR-H	LRSC	LRRSP
2 Motions	Mean	2.89	4.23	1.52	4.10	2.13	3.69	0.85
	Median	0.00	0.56	0.00	0.22	0.00	0.29	0.00
3 Motions	Mean	8.25	7.02	4.40	9.89	4.03	7.69	2.03
	Median	0.24	1.45	0.56	6.22	1.43	3.80	0.28
All	Mean	4.10	4.86	2.18	5.41	2.56	4.59	1.61
	Median	0.00	0.89	0.00	0.53	0.00	0.60	0.02

TABLE 9. Clustering error rate (%) on the Hopkins155 motion segmentation database with the 4n-dimensional data points.

Algorithms		SCC	LSA	SSC	LRR	LRR-H	LRSC	LRRSP
2 Motions	Mean	3.04	3.61	1.83	4.83	3.41	3.87	1.02
	Median	0.00	0.51	0.00	0.26	0.00	0.26	0.01
3 Motions	Mean	7.91	7.65	4.40	9.89	4.86	7.72	2.40
	Median	1.14	1.27	0.56	6.22	1.47	3.80	0.32
All	Mean	4.14	4.52	2.41	5.98	3.74	4.74	1.86
	Median	0.00	0.57	0.00	0.59	0.00	0.58	0.01

31 to 38, respectively. Then we apply clustering algorithms for each trial, i.e., each set of n subjects. Finally, we take the mean and median of the clustering results of all trials for each n subjects. All choices of $n \in \{2, 3, 5, 8, 10\}$ categories will be considered in the first three groups and all choices of $n \in \{2, 3, 5, 8\}$ in the fourth group. Table 2 shows the mean and median subspace clustering error rates of different algorithms, which include SCC, LSA, LRR, LRR-H, SSC and LRSC. Under the number of categories 5, the three indicators ACC, NMI and ARI are shown in Table 3.

The ORL database contains 400 frontal face images of different ages, genders and races from 40 categories. Each category has 10 face images of size 92×112 pixels and black background, of which 40 subjects are from different ages, genders and races. These images change in facial expressions, facial accessories and details, such as smiling or not smiling, eyes open or closed, wearing or not wearing glasses, etc. The posture of the face also changes, whose depth rotation and plane rotation can reach at 20 degrees, and the face size also changes up to 10%. Some sample images of the ORL face database are shown in Figure 2. Following the experimental settings in [16], we resize all the images in the ORL database to 56×46 and vectorize them to 2576-dimensional vectors as the data samples of 9D subspace. We divide the 40 types of target images into the following two groups: 1 to 20 and 21 to 40. For each cluster, we use all 5, 10, 15, 20 categories that can be selected for the experiments and the clustering errors are shown in Table 4, and the three indicators ACC, NMI and ARI are shown in Table 5 under the number of categories 10.

The AR face database contains frontal face images from 50 men and 50 women. Each sample has 26 images with different shooting angles and facial expressions. There are

**FIGURE 3.** AR face database.

2600 face images in the database, and the image size is 120×165 . These sample images have a front view, some facial expressions change and some faces have some occlusion (sunglasses and scarves). In order to save the computational cost, we down sample all the images in the AR database, and then vectorize them. According to Lambert estimation, we cluster the vectorized 2200 dimensional vector as the data sample of 9D subspace. Figure 3 shows some samples in the AR database. We divided the 100 subjects of target images into the following five groups for repeated experiments: 1 to 20, 21 to 40, 41 to 60, 61 to 80 and 81 to 100. For each cluster, we use all 5, 10, 15, 20 categories that can be selected for the experiments and the clustering errors are shown in Table 6, and the three indicators ACC, NMI and ARI are shown in Table 7 under the number of categories 10.

B. EXPERIMENTS ON MOTION SEGMENTATION

In this subsection, the Hopkins155 motion segmentation database [46] is used to evaluate the performance of our algorithm. Motion segmentation is the problem of dividing the video sequence of multiple rigid moving objects into multiple spatio-temporal regions, which correspond to different

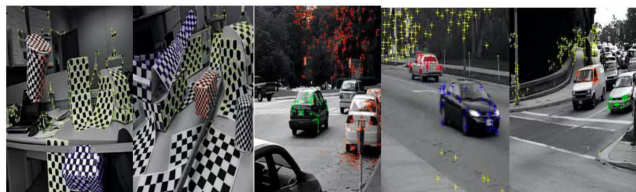


FIGURE 4. Hopkins155 motion segmentation database.

motions in the scene. Specifically, it extracts and tracks a set of N feature points in each frame of the video, where a $2F$ -dimensional vector (F is the total number of frames of the video) obtained by stacking the feature data points corresponds to the feature trajectory. The goal of motion segmentation is separating these feature trajectories according to objects underlying motion. Under the affine projection model, all characteristic trajectories related to the motion of a single rigid lie in an affine subspace of \mathbb{R}^{2F} of dimension $d = 1, 2$ or 3 . Alternatively, they lie in a subspace of equivalent to \mathbb{R}^{2F} , which dimension at most 4 [45]. Therefore, the features of moving objects can be used as data points in subspace, so segmenting the feature trajectory of moving objects is to cluster the samples in subspace.

The Hopkins155 motion segmentation database is made up of 155 video sequences, of which 120 video sequences contain two motions and each sequence contains 30 frames with 266 characteristic trajectories; 35 of the video sequences contain three motions and each sequence contains 29 frames with 398 characteristic trajectories. Some sample images of the Hopkins155 motion segmentation database are shown in Figure 4. Following the experimental settings in [18], we first apply the principal component analysis algorithm to reducing the trajectory from $2F$ (F is the total number of frames of the video) to $4n$ (n is the number of subspaces), and then use $2F$ and $4n$ data to execute our algorithm. The clustering results under $2F$ -dimensional data points are shown in Table 8 and under $4n$ -dimensional data points are shown in Table 9.

The above experimental results show that the clustering error rate of our algorithm is lower and the value of evaluation indicators ACC, NMI and ARI are higher than other algorithms on different databases.

VIII. CONCLUSION

In this paper, we propose a general optimization formulation for subspace segmentation by low rank representation via the sparse-prompting quasi-rank function. Our key contribution is to show that any sparse-prompting quasi-rank function as the measure of the low rank in subspace clustering is reasonable, i.e., the optimal solution to our optimization formulation forms a block-diagonal matrix if the clean data are drawn from independent linear subspaces. In addition, the alternating direction method of multipliers (ADMM) is applied to solving the optimization problem if the data are corrupted by Gaussian noise and gross errors. Finally, a series of simulations on different databases are conducted.

REFERENCES

- [1] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," *ACM SIGMOD Rec.*, vol. 27, no. 2, pp. 94–105, Jun. 1998.
- [2] R. Basri and D. Jacobs, "Lambertian reflection and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 3, pp. 218–233, Feb. 2003.
- [3] P. Bradley and O. Mangasarian, " k -plane clustering," *J. Global Optim.*, vol. 16, no. 1, pp. 23–32, 2000.
- [4] M. Brbić and I. Kopriva, " ℓ_0 -motivated low-rank sparse subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1711–1725, Apr. 2020.
- [5] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [6] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Nov. 2005.
- [7] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.
- [8] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, Mar. 2009.
- [9] W. Cheng, M. Zhao, and N. Xiong, "Non-convex sparse and low-rank based robust subspace segmentation for data mining," *Sensors*, vol. 17, no. 7, pp. 1–15, 2017.
- [10] E. Chouzenoux, A. Jezierska, and J. Pesquet, "A majorizeminimize subspace approach for ℓ_2 - ℓ_0 image regularization," *SIAM J. Imag. Sci.*, vol. 6, no. 1, pp. 563–591, 2013.
- [11] J. P. Costeira and T. Kanade, "A multibody factorization method for independently moving objects," *Int. J. Comput. Vis.*, vol. 29, pp. 159–179, Sep. 1998.
- [12] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *Amer. Math. Soc. Math Challenges Lect.*, vol. 1, pp. 1–32, Aug. 2000.
- [13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Jan. 2006.
- [14] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 27, pp. 9446–9451, Jul. 2005.
- [15] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.
- [16] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.
- [17] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its Oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [18] P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *Proc. CVPR*, Jun. 2011, pp. 1801–1807.
- [19] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [20] S. Foucart and M. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 395–407, 2009.
- [21] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer, 2013.
- [22] J. H. Friedman, "Fast sparse regression and classification," *Int. J. Forecasting*, vol. 28, no. 3, pp. 722–738, Jul./Sep. 2012.
- [23] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [24] J. Ho, M.-H. Yang, J. Lim, K.-C. Lee, and D. Kriegman, "Clustering appearances of objects under varying illumination conditions," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 11–18.
- [25] W. Hong, J. Wright, K. Huang, and Y. Ma, "Multiscale hybrid linear models for lossy image representation," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3655–3671, Dec. 2006.

- [26] H. Jiang, D. P. Robinson, R. Vidal, and C. You, "A nonconvex formulation for low rank subspace clustering: Algorithms and convergence analysis," *Comput. Optim. Appl.*, vol. 70, no. 2, pp. 395–418, Jun. 2018.
- [27] Z. Kang, C. Peng, J. Cheng, and Q. Cheng, "LogDet rank minimization with application to subspace clustering," *Comput. Intell. Neurosci.*, vol. 2015, pp. 1–10, Jan. 2015.
- [28] F. Lauer and C. Schnorr, "Spectral clustering of linear subspaces for motion segmentation," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 678–685.
- [29] A. J. Lee and M. Verleysen, "Nonlinear dimensionality reduction," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 5, no. 5500, 2009, pp. 1959–1966.
- [30] H. Li, Q. Zhang, A. Cui, and J. Peng, "Minimization of fraction function penalty in compressed sensing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1626–1637, May 2020.
- [31] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [32] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, 2010, pp. 663–670.
- [33] L. Lu and R. Vidal, "Combined central and subspace clustering for computer vision applications," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 593–600.
- [34] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
- [35] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1546–1562, Sep. 2007.
- [36] Y. Ma, A. Y. Yang, H. Derksen, and R. Fossom, "Estimation of subspace arrangements with applications in modeling and segmenting mixed data," *SIAM Rev.*, vol. 50, no. 3, pp. 413–458, Jan. 2008.
- [37] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte Für Math.*, vol. 79, no. 4, pp. 303–306, 1975.
- [38] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Feb. 1995.
- [39] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst., Natural Synthetic*. Cambridge, MA, USA: MIT Press, 2001, pp. 849–856.
- [40] M. Nikolova, "Local strong homogeneity of a regularized estimator," *SIAM J. Appl. Math.*, vol. 61, pp. 633–658, Feb. 2000.
- [41] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: A review," *ACM SIGKDD*, vol. 6, no. 1, pp. 90–105, 2004.
- [42] J. Peng, S. Yue, and H. Li, "NP/CMP equivalence: A phenomenon hidden among sparsity models ℓ_0 minimization and ℓ_p minimization for information processing," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 4028–4033, Jul. 2015.
- [43] S. R. Rao, A. Y. Yang, S. S. Sastry, and Y. Ma, "Robust algebraic segmentation of mixed rigid-body and planar motions from two views," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 425–446, 2010.
- [44] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, Feb. 1999.
- [45] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.
- [46] R. Tron and R. Vidal, "A benchmark for the comparison of 3-D motion segmentation algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [47] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0 -minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.
- [48] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Feb. 2011.
- [49] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component analysis," *IEEE Trans. Pattern Anal. Mach.*, vol. 27, no. 12, pp. 1945–1959, Dec. 2005.
- [50] R. Vidal and P. Favaro, "Low rank subspace clustering," *Pattern Recognit. Lett.*, vol. 43, no. 1, pp. 47–61, 2014.
- [51] J. Woodworth and R. Chartrand, "Compressed sensing recovery via non-convex shrinkage penalties," *Inverse Problems*, vol. 32, no. 7, Jul. 2016, Art. no. 075004.
- [52] K. Xie, W. Liu, Y. Lai, and W. Li, "Discriminative low-rank subspace learning with nonconvex penalty," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 10, Sep. 2019, Art. no. 1951006, doi: 10.1142/S0218001419510066.
- [53] X. Xue, X. Zhang, X. Feng, H. Sun, W. Chen, and Z. Liu, "Robust subspace clustering based on non-convex low-rank approximation and adaptive kernel," *Inf. Sci.*, vol. 513, pp. 190–205, Mar. 2020.
- [54] A. Y. Yang, J. Wright, Y. Ma, and S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, May 2008.
- [55] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 94–106.
- [56] Y. Yu, J. Peng, and S. Yue, "A new nonconvex approach to low-rank matrix completion with application to image inpainting," *Multidimensional Syst. Signal Process.*, vol. 30, no. 1, pp. 145–174, Jan. 2019.
- [57] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
- [58] H. Zhang, J. Yang, F. Shang, C. Gong, and Z. Zhang, "LRR for subspace segmentation via tractable Schatten- p norm minimization and factorization," *IEEE Trans. Cybern.*, vol. 49, no. 5, pp. 1722–1734, May 2019.
- [59] S. Zhang and J. Xin, "Minimization of transformed ℓ_1 penalty: Closed form representation and iterative thresholding algorithms," *Math. Program.*, vol. 169, pp. 307–336, Jan. 2018.
- [60] F. Zhao, "Research on the sparse optimization theory and algorithms based on band restricted thresholding operator," Ph.D. dissertation, School Math. Statist., Xi'an Jiaotong Univ., Xi'an, China, 2021.



HAIYANG LI received the B.S. degree in mathematics from Yan'an University, Shaanxi, China, in 1996, and the M.Sc. and Ph.D. degrees in fundamental mathematics from Shaanxi Normal University, Xi'an, China, in 2005 and 2008, respectively. He is currently a Professor with the School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. His current research interests include quantum logic theory, the machine learning theory, and sparse information processing.



YUSI WANG received the B.S. and M.Sc. degrees from the School of Science, Xi'an Polytechnic University, Xi'an, China, in 2016 and 2020, respectively. She is currently pursuing the Ph.D. degree with the School of Mathematics and Information Sciences, Guangzhou University, Guangzhou, China. Her current research interests include machine learning collision detection computer vision and neural networks.



QIAN ZHANG received the B.S. and M.Sc. degrees from the School of Science, Xi'an Polytechnic University, Xi'an, China, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include graph-based semi-supervised learning, Gaussian process, and Bayesian inference.



SHOUJIN LIN received the M.S. degree from the School of Science, Xi'an Jiaotong University, Xi'an, China, in 2003. He is currently working at Zhongshan MLTOR Intelligent Equipment Company Ltd., Zhongshan, China. His current research interests include machine learning theory, advanced manufacturing, and sparse information processing.



ZHIWEI XING received the B.S. degree in applied mathematics from Luoyang Normal University, China, in 2004, the M.S. degree in basic mathematics from the School of Mathematics and Statistics, Yunnan University, China, in 2007, and the Ph.D. degree in applied mathematics from the School of Mathematics and Statistics, Xi'an Jiaotong University, China, 2012. His research interests include machine learning, image processing, and complex networks.



JIGEN PENG received the B.S. degree in mathematics from Jiangxi University, Jiangxi, China, in 1989, and the M.Sc. and Ph.D. degrees in applied mathematics and computing mathematics from Xi'an Jiaotong University, Xi'an, China, in 1992 and 1998, respectively. He is currently a Professor with the School of Mathematics and Information Science, Guangzhou University, Guangzhou, China. His current research interests include nonlinear functional analysis and applications, data set matching theory, the machine learning theory, and sparse information processing.

...