

Research Article

Low-Rank Representation for Incomplete Data

Jiarong Shi,¹ Wei Yang,¹ Longquan Yong,² and Xiuyun Zheng¹

¹School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China

²School of Mathematics and Computer Science, Shaanxi University of Technology, Hanzhong 723001, China

Correspondence should be addressed to Jiarong Shi; jiarongs3@163.com

Received 20 August 2014; Revised 25 November 2014; Accepted 19 December 2014; Published 31 December 2014

Academic Editor: Wanquan Liu

Copyright © 2014 Jiarong Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Low-rank matrix recovery (LRMR) has been becoming an increasingly popular technique for analyzing data with missing entries, gross corruptions, and outliers. As a significant component of LRMR, the model of low-rank representation (LRR) seeks the lowest-rank representation among all samples and it is robust for recovering subspace structures. This paper attempts to solve the problem of LRR with partially observed entries. Firstly, we construct a nonconvex minimization by taking the low rankness, robustness, and incompleteness into consideration. Then we employ the technique of augmented Lagrange multipliers to solve the proposed program. Finally, experimental results on synthetic and real-world datasets validate the feasibility and effectiveness of the proposed method.

1. Introduction

In the community of pattern recognition, machine learning, and computer vision, the investigated datasets usually have intrinsically low-rank structure although they are probably high-dimensional. Low-rank matrix recovery (LRMR) [1–3] is just a type of model which utilizes the crucial low-complexity information to complete missing entries, recover sparse noise, identify outliers, and build an affinity matrix. It also can be regarded as the generalization of compressed sensing from one order to two orders due to the fact that the low rankness of a matrix is equivalent to the sparsity of its singular values. Recently, LRMR has received more and more attentions in the fields of information science and engineering and achieved great success in video background modeling [4, 5], collaborative filtering [6, 7], and subspace clustering [3, 8, 9], to name just a few.

Generally, LRMR is mainly composed of three appealing types, that is, matrix completion (MC) [1], robust principal component analysis (RPCA) [2, 4], and low-rank representation (LRR) [3, 8]. Among them, MC aims to complete the missing entries with the aid of the low-rank property and is initially described as an affine rank minimization problem. In the past few years, the affine rank minimization is convexly relaxed into a nuclear norm minimization [10] and it is proven that if the number of sampled entries and singular

vectors satisfy some conditions, then most low-rank matrices can be perfectly recovered by solving the aforementioned convex program [1].

Classical principal component analysis (PCA) [11] is very effective to small Gaussian noise, but it does not work well in practice when data samples are corrupted by outliers or large sparse noise. For this purpose, several robust variants of PCA have been proposed successively during the past two decades [12, 13]. Since the seminal research work [4], the principal component pursuit (PCP) approach has become a standard for RPCA. This approach minimizes a weighted combination of the nuclear norm and the l_1 norm with linear equality constraints. It is proven that both the low-rank and the sparse components can be recovered exactly with dominant probability under some conditions by solving PCP [4].

In subspace clustering, a commonly used assumption is that the data lie in the union of multiple low-rank subspaces and each subspace has sufficient samples compared with its rank. Liu et al. [3] proposed a robust subspace recovery technique via LRR. Any sample in each subspace can be represented as the linear combination of the bases. The low complexity of the linear representation coefficients is very useful in exploiting the low-rank structure. LRR attempts to seek the lowest-rank representation of all data jointly and it is demonstrated that the data contaminated by outliers can be exactly recovered under certain conditions by solving

a convex program [8]. If the bases are chosen as the columns of an identity matrix and the l_1 norm is employed to measure the sparsity, then LRR is changed into the PCP of RPCA.

For the datasets with missing entries and large sparse corruption, the robust recovery of subspace structures may be a challenging task. The available algorithms to MC are not robust to gross corruption. Moreover, a large quantity of missing values will bring out the degeneration of recovering performance for LRR or RPCA. In this paper, we attempt to address the problem of low-rank subspace recovery in the presence of missing values and sparse noise. Specifically speaking, we present a model of incomplete low-rank representation (ILRR) which is a direct generalization of LRR. The ILRR model can be boiled down to a nonconvex optimization model which minimizes the combination of the nuclear norm and the $l_{2,1}$ -norm. To solve this program, we design an iterative scheme by applying the method of inexact augmented Lagrange multipliers (ALM).

The rest of this paper is organized as follows. Section 2 briefly reviews preliminaries and related works on LRMR. The model and algorithm for ILRR are presented in Section 3. In Section 4, we discuss the extension of ILRR and its relationship with the existing works. We compare the performance of ILRR with the state-of-the-art algorithms on synthetic data and real-world datasets in Section 5. Finally, Section 6 draws some conclusions.

2. Preliminaries and Related Works

This section introduces the relevant preliminary material concerning matrices and representative models of low-rank matrix recovery (LRMR).

The choice of matrix norms plays a significant role in LRMR. In the following, we present four important types of matrix norms. For arbitrary $\mathbf{X} \in \mathbb{R}^{m \times n}$, the Frobenius norm of \mathbf{X} is expressed by $\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n X_{ij}^2}$, the $l_{2,1}$ -norm is $\|\mathbf{X}\|_{2,1} = \sum_{j=1}^n \sqrt{\sum_{i=1}^m X_{ij}^2}$, the l_1 -norm is $\|\mathbf{X}\|_1 = \sum_{i=1}^m \sum_{j=1}^n |X_{ij}|$, and the nuclear norm is $\|\mathbf{X}\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$, where X_{ij} is the (i, j) th entry of \mathbf{X} and σ_i is the i th largest singular value. Among them, the matrix nuclear norm is the tightest convex relaxation of the rank function, and the $l_{2,1}$ -norm and the l_1 -norm are frequently used to measure the sparsity of a noise matrix.

Consider a data matrix $\mathbf{D} \in \mathbb{R}^{m \times n}$ stacked by n training samples, where each column of \mathbf{D} indicates a sample with the dimensionality of m . Within the field of LRMR, the following three proximal minimization problems [3, 4] are extensively employed:

$$\begin{aligned} \min_{\mathbf{X}} \lambda \|\mathbf{X}\|_1 + \frac{1}{2} \|\mathbf{X} - \mathbf{D}\|_F^2, \\ \min_{\mathbf{X}} \lambda \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{D}\|_F^2, \\ \min_{\mathbf{X}} \lambda \|\mathbf{X}\|_{2,1} + \frac{1}{2} \|\mathbf{X} - \mathbf{D}\|_F^2, \end{aligned} \quad (1)$$

where λ is a positive constant used to balance the regularization term and the approximation error. For given λ and \mathbf{D} , we define three thresholding operators $\mathcal{S}_\lambda(\mathbf{D})$, $\mathcal{D}_\lambda(\mathbf{D})$, and $\mathcal{W}_\lambda(\mathbf{D})$ as follows:

$$\begin{aligned} (\mathcal{S}_\lambda(\mathbf{D}))_{ij} &= \max(|D_{ij}| - \lambda, 0), \\ \mathcal{D}_\lambda(\mathbf{D}) &= \mathbf{U} \mathcal{S}_\lambda(\boldsymbol{\Sigma}) \mathbf{V}^T, \\ (\mathcal{W}_\lambda(\mathbf{D}))_{ij} &= \frac{\max\left(\sqrt{\sum_{k=1}^m D_{kj}^2} - \lambda, 0\right) D_{ij}}{\sqrt{\sum_{k=1}^m D_{kj}^2}}, \end{aligned} \quad (2)$$

where $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ is the singular value decomposition (SVD) of \mathbf{D} . It is proven that the aforementioned three optimization problems have closed-form solutions denoted by $\mathcal{S}_\lambda(\mathbf{D})$ [4], $\mathcal{D}_\lambda(\mathbf{D})$ [14], and $\mathcal{W}_\lambda(\mathbf{D})$ [8], respectively.

We assume that \mathbf{D} is low-rank. Because the degree of freedom of a low-rank matrix is far less than its number of entries, it is possible to recover exactly all missing entries from partially observed entries as long as the number of sampled entries satisfies certain conditions. Formally, the problem of matrix completion (MC) [1] can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & X_{ij} = D_{ij}, \quad (i, j) \in \Omega, \end{aligned} \quad (3)$$

where $\Omega \subset [m] \times [n]$, $[m] = \{1, 2, \dots, m\}$, $[n] = \{1, 2, \dots, n\}$. We define a linear projection operator $\mathcal{P}_\Omega(\bullet) : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$ as follows:

$$(\mathcal{P}_\Omega(\mathbf{X}))_{ij} = \begin{cases} X_{ij}, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Hence, the constraints in problem (3) can be rewritten as $\mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{D})$.

PCA obtains the optimal estimate for small additive Gaussian noise but breaks down for large sparse contamination. Here, the data matrix \mathbf{D} is assumed to be the superposition of a low-rank matrix \mathbf{A} and a sparse noise matrix \mathbf{E} . In this situation, robust principal component analysis (RPCA) is very effective to recover both the low-rank and the sparse components by solving a convex program. Mathematically, RPCA can be described as the following nuclear norm minimization [2]:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{E}} \quad & \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{A} + \mathbf{E}, \end{aligned} \quad (5)$$

where λ is a positive weighting parameter. Sequentially, RPCA is generalized into a stable version which is simultaneously stable to small perturbations and robust to gross sparse corruption [15].

We further assume that the dataset is self-expressive and the representation coefficients matrix is also low-rank.

Based on the above two assumptions, the model of low-rank representation (LRR) [3] is expressed as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{DZ} + \mathbf{E}, \end{aligned} \quad (6)$$

where $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is the coefficient matrix, \mathbf{E} is the noise matrix, and λ is a positive trade-off parameter. This model is very effective to detect outliers and the optimal \mathbf{Z} is in favor of the robust subspace recovery. In subspace clustering, the affinity matrices can be constructed by the optimal \mathbf{Z} to problem (6).

Problems (3), (5), and (6) belong to the nuclear norm minimizations. The existing algorithms to the preceding optimizations mainly include the iterative thresholding, the accelerated proximal gradient, the dual approach, and the augmented Lagrange multipliers (ALM) [16]. These algorithms are scalable owing to the adoption of first-order information. Among them, ALM, also called alternating direction method of multipliers (ADMM) [17], is a very popular and effective method to solve the nuclear norm minimizations.

3. Model and Algorithm of Incomplete Low-Rank Representation

This section proposes a model of low-rank representation for incomplete data and develops a corresponding iterative scheme for this model.

3.1. Model. We consider an incomplete data matrix $\mathbf{M} \in \mathbb{R}^{m \times n}$ and denote the sampling index set by Ω . The (i, j) th entry M_{ij} of \mathbf{M} is missing if and only if $(i, j) \notin \Omega$. For the sake of convenience, we set all missing entries of \mathbf{M} to zeros. To recover simultaneously the missing entries and the low-rank subspace structure, we construct an incomplete low-rank representation (ILRR) model:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{D}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{DZ} + \mathbf{E} \\ & \mathcal{P}_\Omega(\mathbf{D}) = \mathbf{M}, \end{aligned} \quad (7)$$

where λ is a positive constant and \mathbf{D} corresponds to the completion argument of \mathbf{M} . If there is not any missing entry, that is, $\Omega = [m] \times [n]$, then the above model is equivalent to LRR. In other words, LRR is a special case of ILRR.

In order to solve conveniently the nonconvex nuclear norm minimization (7), we introduce two auxiliary matrix variables $\mathbf{X} \in \mathbb{R}^{m \times n}$ and $\mathbf{J} \in \mathbb{R}^{n \times n}$. Under this circumstance, the above optimization problem is reformulated as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{D}, \mathbf{J}, \mathbf{X}} \quad & \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{XZ} + \mathbf{E} \\ & \mathbf{Z} = \mathbf{J}, \quad \mathbf{X} = \mathbf{D}, \quad \mathcal{P}_\Omega(\mathbf{D}) = \mathbf{M}. \end{aligned} \quad (8)$$

This minimization problem is equivalent to

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{D}, \mathbf{J}, \mathbf{X}} \quad & \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ & + \frac{\mu}{2} (\|\mathbf{D} - \mathbf{XZ} - \mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{X} - \mathbf{D}\|_F^2) \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{XZ} + \mathbf{E} \\ & \mathbf{Z} = \mathbf{J}, \quad \mathbf{X} = \mathbf{D}, \quad \mathcal{P}_\Omega(\mathbf{D}) = \mathbf{M}, \end{aligned} \quad (9)$$

where the factor $\mu > 0$. Without considering the constraint $\mathcal{P}_\Omega(\mathbf{D}) = \mathbf{M}$, we construct the augmented Lagrange function of problem (9) as follows:

$$\begin{aligned} \mathcal{L} = & \|\mathbf{J}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ & + \frac{\mu}{2} (\|\mathbf{D} - \mathbf{XZ} - \mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2 + \|\mathbf{X} - \mathbf{D}\|_F^2) \\ & + (\langle \mathbf{Y}_1, \mathbf{D} - \mathbf{XZ} - \mathbf{E} \rangle + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} \rangle + \langle \mathbf{Y}_3, \mathbf{X} - \mathbf{D} \rangle), \end{aligned} \quad (10)$$

where $\langle \bullet, \bullet \rangle$ is the inner product operator between matrices and \mathbf{Y}_i is a Lagrange multiplier matrix, $i = 1, 2, 3$. In the next part, we will propose an inexact augmented Lagrange multipliers (IALM) method to solve problem (8) or problem (9).

3.2. Algorithm. Inexact ALM (IALM) method employs an alternating update strategy and it minimizes or maximizes the function \mathcal{L} with respect to each block variable at each iteration. Let $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3\}$.

Computing J. When \mathbf{J} is unknown and other variables are fixed, the calculation procedure of \mathbf{J} is as follows:

$$\begin{aligned} \mathbf{J} & := \arg \min_{\mathbf{J}} \mathcal{L} \\ & = \arg \min_{\mathbf{J}} \|\mathbf{J}\|_* + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} \rangle + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J}\|_F^2 \\ & = \arg \min_{\mathbf{J}} \frac{1}{\mu} \|\mathbf{J}\|_* + \frac{1}{2} \left\| \left(\mathbf{Z} + \frac{\mathbf{Y}_2}{\mu} \right) - \mathbf{J} \right\|_F^2 \\ & = \mathcal{D}_{1/\mu} \left(\mathbf{Z} + \frac{\mathbf{Y}_2}{\mu} \right). \end{aligned} \quad (11)$$

Computing Z. If matrix \mathbf{Z} is unknown and other variables are given, \mathbf{Z} is updated by minimizing \mathcal{L} :

$$\begin{aligned} \mathbf{Z} &:= \arg \min_{\mathbf{Z}} \mathcal{L} \\ &= \arg \min_{\mathbf{Z}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{XZ} - \mathbf{E} \rangle + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{J} \rangle \\ &\quad + \frac{\mu}{2} (\|\mathbf{D} - \mathbf{XZ} - \mathbf{E}\|_F^2 + \|\mathbf{Z} - \mathbf{J}\|_F^2) \\ &= \arg \min_{\mathbf{Z}} \left\| \left(\mathbf{D} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right) - \mathbf{XZ} \right\|_F^2 \\ &\quad + \left\| \mathbf{Z} - \left(\mathbf{J} - \frac{\mathbf{Y}_2}{\mu} \right) \right\|_F^2. \end{aligned} \quad (12)$$

Let $g(\mathbf{Z}) = \|\mathbf{D} - \mathbf{E} + \mathbf{Y}_1/\mu - \mathbf{XZ}\|_F^2 + \|\mathbf{Z} - (\mathbf{J} - \mathbf{Y}_2/\mu)\|_F^2$. By setting the derivative of $g(\mathbf{Z})$ to zero, we have

$$\begin{aligned} \nabla g(\mathbf{Z}) &= 2\mathbf{X}^T \mathbf{XZ} - 2\mathbf{X}^T \left(\mathbf{D} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right) \\ &\quad + 2\mathbf{Z} - 2 \left(\mathbf{J} - \frac{\mathbf{Y}_2}{\mu} \right) = 0 \end{aligned} \quad (13)$$

or, equivalently,

$$\mathbf{Z} := (\mathbf{X}^T \mathbf{X} + \mathbf{I})^{-1} \left(\mathbf{X}^T \left(\mathbf{D} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right) + \left(\mathbf{J} - \frac{\mathbf{Y}_2}{\mu} \right) \right), \quad (14)$$

where \mathbf{I} is an n -order identity matrix.

Computing D. The update formulation of matrix \mathbf{D} is calculated as follows:

$$\begin{aligned} \mathbf{D} &:= \arg \min_{\mathbf{D}} \mathcal{L} \\ &= \arg \min_{\mathbf{D}} \langle \mathbf{Y}_1, \mathbf{D} - \mathbf{XZ} - \mathbf{E} \rangle + \langle \mathbf{Y}_3, \mathbf{X} - \mathbf{D} \rangle \\ &\quad + \frac{\mu}{2} (\|\mathbf{D} - \mathbf{XZ} - \mathbf{E}\|_F^2 + \|\mathbf{X} - \mathbf{D}\|_F^2) \\ &= \arg \min_{\mathbf{D}} \left\| \mathbf{D} - \left(\mathbf{XZ} + \mathbf{E} - \frac{\mathbf{Y}_1}{\mu} \right) \right\|_F^2 + \left\| \left(\mathbf{X} + \frac{\mathbf{Y}_3}{\mu} \right) - \mathbf{D} \right\|_F^2 \\ &= \frac{1}{2} \left(\mathbf{XZ} + \mathbf{E} - \frac{\mathbf{Y}_1}{\mu} + \mathbf{X} + \frac{\mathbf{Y}_3}{\mu} \right). \end{aligned} \quad (15)$$

Considering the constraint $\mathcal{P}_{\Omega}(\mathbf{D}) = \mathbf{M}$, we further obtain the iteration formulation of \mathbf{D}

$$\mathbf{D} := \mathbf{M} + \frac{1}{2} \mathcal{P}_{\bar{\Omega}} \left(\mathbf{XZ} + \mathbf{E} - \frac{\mathbf{Y}_1}{\mu} + \mathbf{X} + \frac{\mathbf{Y}_3}{\mu} \right), \quad (16)$$

where $\bar{\Omega}$ is the complementary set of Ω .

Computing E. Fix $\mathbf{Z}, \mathbf{D}, \mathbf{J}, \mathbf{X}$, and \mathbf{Y} and minimize \mathcal{L} with respect to \mathbf{E} :

$$\begin{aligned} \mathbf{E} &:= \arg \min_{\mathbf{E}} \mathcal{L} \\ &= \arg \min_{\mathbf{E}} \frac{\lambda}{\mu} \|\mathbf{E}\|_{2,1} + \frac{1}{2} \left\| \mathbf{D} - \mathbf{XZ} - \mathbf{E} + \frac{1}{\mu} \mathbf{Y}_1 \right\|_F^2 \\ &= \mathcal{W}_{\lambda/\mu} \left(\mathbf{D} - \mathbf{XZ} + \frac{\mathbf{Y}_1}{\mu} \right). \end{aligned} \quad (17)$$

Computing X. Fix $\mathbf{Z}, \mathbf{D}, \mathbf{J}, \mathbf{E}$, and \mathbf{Y} to calculate \mathbf{X} as follows:

$$\begin{aligned} \mathbf{X} &:= \arg \min_{\mathbf{X}} \mathcal{L} \\ &= \arg \min_{\mathbf{X}} h(\mathbf{X}), \end{aligned} \quad (18)$$

where $h(\mathbf{X}) = \|\mathbf{D} - \mathbf{XZ} - \mathbf{E} + \mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{X} - \mathbf{D} + \mathbf{Y}_3/\mu\|_F^2$. The derivative of $h(\mathbf{X})$ is $\nabla h(\mathbf{X}) = 2\mathbf{XZZ}^T - 2(\mathbf{D} - \mathbf{E} + \mathbf{Y}_1/\mu)\mathbf{Z}^T + 2\mathbf{X} - 2(\mathbf{D} - \mathbf{Y}_3/\mu)$. Hence, we obtain the update of \mathbf{X} by setting $\nabla h(\mathbf{X}) = 0$:

$$\mathbf{X} := \left(\left(\mathbf{D} - \mathbf{E} + \frac{\mathbf{Y}_1}{\mu} \right) \mathbf{Z}^T + \left(\mathbf{D} - \frac{\mathbf{Y}_3}{\mu} \right) \right) (\mathbf{ZZ}^T + \mathbf{I})^{-1}. \quad (19)$$

Computing Y. Given $\mathbf{Z}, \mathbf{E}, \mathbf{D}, \mathbf{J}$, and \mathbf{X} , we calculate \mathbf{Y} as follows

$$\mathbf{Y} := \arg \max_{\mathbf{Y}} \mathcal{L}. \quad (20)$$

In the detailed implementation, \mathbf{Y} is updated according to the following formulations:

$$\begin{aligned} \mathbf{Y}_1 &:= \mathbf{Y}_1 + \mu (\mathbf{D} - \mathbf{XZ} - \mathbf{E}), \\ \mathbf{Y}_2 &:= \mathbf{Y}_2 + \mu (\mathbf{Z} - \mathbf{J}), \\ \mathbf{Y}_3 &:= \mathbf{Y}_3 + \mu (\mathbf{X} - \mathbf{D}). \end{aligned} \quad (21)$$

We denote by $\mathbf{O}_{m \times n}$ the $m \times n$ zeros matrix. The whole iterative procedure is outlined in Algorithm 1. The stopping condition of Algorithm 1 can be set as

$$\max \{ \|\mathbf{D} - \mathbf{XZ} - \mathbf{E}\|_F, \|\mathbf{Z} - \mathbf{J}\|_F, \|\mathbf{X} - \mathbf{D}\|_F \} < \varepsilon, \quad (22)$$

where ε is a sufficiently small positive number.

3.3. Convergence. When solving ILRR via the IALM method, the block variables $\mathbf{J}, \mathbf{Z}, \mathbf{D}, \mathbf{E}, \mathbf{X}$ are updated alternatively. Now, we update simultaneously these five block variables; namely,

$$(\mathbf{J}, \mathbf{Z}, \mathbf{D}, \mathbf{E}, \mathbf{X}) := \arg \min_{\mathbf{J}, \mathbf{Z}, \mathbf{D}, \mathbf{E}, \mathbf{X}} \mathcal{L}. \quad (23)$$

The modified method is called the exact ALM method. Since the objective function in problem (8) is continuous, the exact ALM method is convergent [18]. However, it is still difficult to

Input: Data matrix \mathbf{M} , sampling index set Ω , compromising parameter λ .
Initialize: $\mathbf{D} = \mathbf{M}$, $\mathbf{X} = \mathbf{M}$, $\mathbf{Z} = \mathbf{O}_{n \times n}$, $\mathbf{J} = \mathbf{O}_{n \times n}$, $\mathbf{E} = \mathbf{O}_{m \times n}$, $\mathbf{Y}_1 = \mathbf{O}_{m \times n}$,
 $\mathbf{Y}_2 = \mathbf{O}_{n \times n}$, $\mathbf{Y}_3 = \mathbf{O}_{m \times n}$, $\mu = 10^{-6}$, $\bar{\mu} = 10^8$, $\rho = 1.1$.
Output: \mathbf{Z} and \mathbf{D} .
While not converged **do**
(1) Update \mathbf{J} according to (11).
(2) Update \mathbf{Z} according to (14).
(3) Update \mathbf{D} according to (16).
(4) Update \mathbf{E} according to (17).
(5) Update \mathbf{X} according to (19).
(6) Update \mathbf{Y} according to (21).
(7) Update μ as $\mu := \min(\rho\mu, \bar{\mu})$.
End while

ALGORITHM 1: Solving ILRR by IALM.

prove the convergence of the IALM. There are two reasons for the difficulty: one is the existence of nonconvex constraints in (8) and the other is that the number of block variables is more than two. Nevertheless, the experimental results of Section 5 demonstrate the validity and effectiveness of Algorithm 1 in practice.

4. Model Extensions

In ILRR model, the main aim of the term $\|\mathbf{E}\|_{2,1}$ is to enhance the robustness to noise and outliers. If we do not consider outliers, then $\|\mathbf{E}\|_{2,1}$ should be replaced with $\|\mathbf{E}\|_1$. For the new ILRR model, we can design an algorithm by only revising Step 4 of Algorithm 1 as follows:

$$\begin{aligned} \mathbf{E} &:= \arg \min_{\mathbf{E}} \frac{\lambda}{\mu} \|\mathbf{E}\|_1 + \frac{1}{2} \left\| \mathbf{D} - \mathbf{XZ} - \mathbf{E} + \frac{1}{\mu} \mathbf{Y}_1 \right\|_F^2 \\ &= \mathcal{S}_{\lambda/\mu} \left(\mathbf{D} - \mathbf{XZ} + \frac{\mathbf{Y}_1}{\mu} \right). \end{aligned} \quad (24)$$

If we substitute $\|\mathbf{E}\|_F^2$ for $\|\mathbf{E}\|_{2,1}$ and set $\mathbf{Z} = \mathbf{Z}^T$, then problem (7) is the incomplete version of low-rank subspace clustering (LRSC) with uncorrupted data [9]. If we replace $\|\mathbf{Z}\|_*$ and $\|\mathbf{E}\|_{2,1}$ by $\|\mathbf{Z}\|_1$ and $\|\mathbf{E}\|_1$, respectively, and incorporate $\text{diag}(\mathbf{Z}) = 0$ into the constraints, then problem (7) is the incomplete version of sparse subspace clustering (SSC) without dense errors [19].

ILRR uses the data \mathbf{D} itself as the dictionary. Now, we extend the dictionary and noise sparsity to the more general cases. As a result, we obtain a comprehensive form of ILRR:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}, \mathbf{A}, \mathbf{D}} \quad & \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_l \\ \text{s.t.} \quad & \mathbf{D} = \mathbf{AZ} + \mathbf{E}, \quad \mathcal{P}_\Omega(\mathbf{D}) = \mathbf{M}, \end{aligned} \quad (25)$$

where $\mathbf{A} \in \mathbb{R}^{m \times r}$ represents the dictionary, $\mathbf{Z} \in \mathbb{R}^{r \times n}$ is the coefficients matrix, and $\|\mathbf{E}\|_l$ indicates a certain norm of \mathbf{E} . If \mathbf{A} is an m -order identity matrix and $\|\mathbf{E}\|_l$ is chosen as $\|\mathbf{E}\|_1$, then problem (25) corresponds to the incomplete version of RPCA [20]. If we further reinforce $\lambda = 0$, then problem (25)

becomes the equivalent formulation of MC [16]. Moreover, if \mathbf{A} is an unknown orthogonal matrix and $\lambda = 0$, then problem (25) is equivalent to matrix decomposition method for MC [21]. Finally, if we let \mathbf{D} and \mathbf{A} be stacked by the testing samples and the training samples, respectively, let $\|\mathbf{Z}\|_*$ be replaced by $\|\mathbf{Z}\|_1$, and let $\lambda = 0$, then problem (25) is changed into the incomplete version of robust pattern recognition via sparse representation [22].

5. Experiments

In this section, we validate the effectiveness and efficiency of the proposed method by carrying out experiments on synthetic data and real-world datasets. The experimental results of incomplete low-rank representation (ILRR) are compared with that of other state-of-the-art methods: sparse subspace clustering (SSC), low-rank subspace clustering (LRSC), and low-rank representation (LRR). For the latter three methods, the missing values are replaced by zeros. For SSC and LRSC, their parameters are tuned to achieve the best performance. The tolerant error ε is set to 10^{-8} in all experiments.

5.1. Synthetic Data. We generate randomly an orthogonal matrix $\mathbf{U}_1 \in \mathbb{R}^{200 \times 4}$ and a rotation matrix $\mathbf{T} \in \mathbb{R}^{200 \times 200}$. Four other orthogonal matrices are constructed as $\mathbf{U}_{i+1} = \mathbf{T}\mathbf{U}_i$, $i = 1, 2, 3, 4$. Thus, five independent low-rank subspaces $\{\mathcal{S}_i\}_{i=1}^5 \subset \mathbb{R}^{200}$ are spanned by the columns of \mathbf{U}_i , respectively. We draw 40 data vectors from each subspace by $\mathbf{M}_i = \mathbf{U}_i \mathbf{Q}_i$, $i = 1, \dots, 5$, where the entries of $\mathbf{Q}_i \in \mathbb{R}^{4 \times 40}$ are independent of each other and they obey standard normal distributions. We set $\mathbf{M}' = (\mathbf{M}_1, \dots, \mathbf{M}_5)$ and choose randomly its 20 column vectors to be corrupted. In this part, the chosen column vector \mathbf{m} is contaminated by Gaussian noise with zero mean and standard deviation $0.1 \|\mathbf{m}\|_2$. The resulting matrix is expressed as \mathbf{M}'' .

We draw samples on \mathbf{M}'' according to a uniform distribution and denote by Ω the sampling index set. The sampling ratio (SR) is defined as $\text{SR} = |\Omega|/200^2$, where $|\Omega|$ is the cardinality of Ω and $\text{SR} = 1$ means that no entry is missing. Hence, an incomplete matrix is generated by $\mathbf{M} = \mathcal{P}_\Omega(\mathbf{M}'')$. The trade-off parameter λ in both LRR and ILRR is set to 0.1.

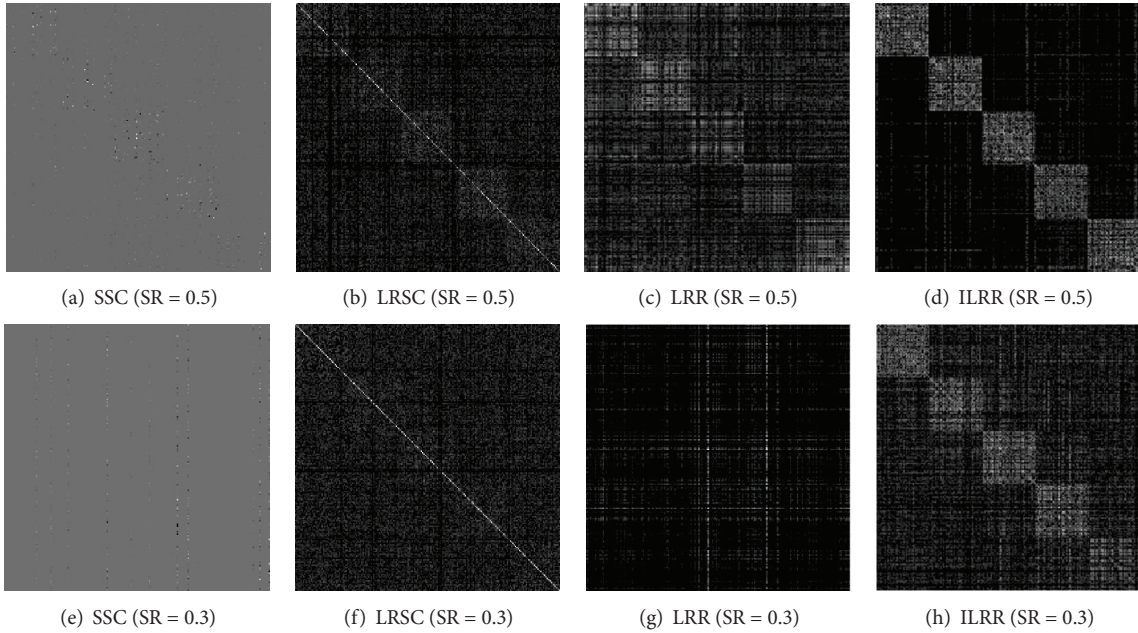


FIGURE 1: Affinity matrices comparison on synthetic data.

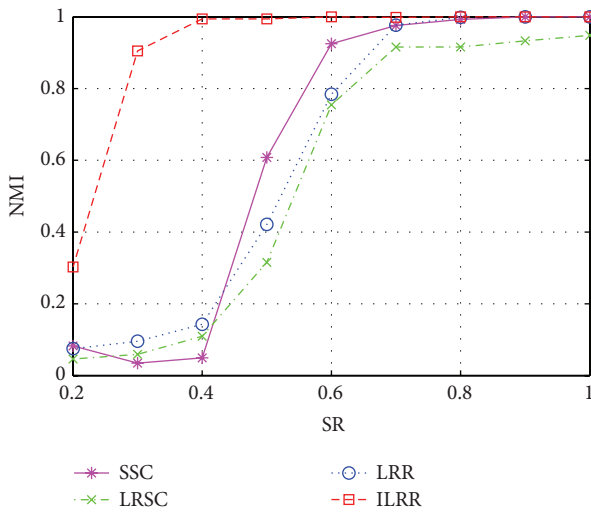


FIGURE 2: NMI comparison on synthetic data.

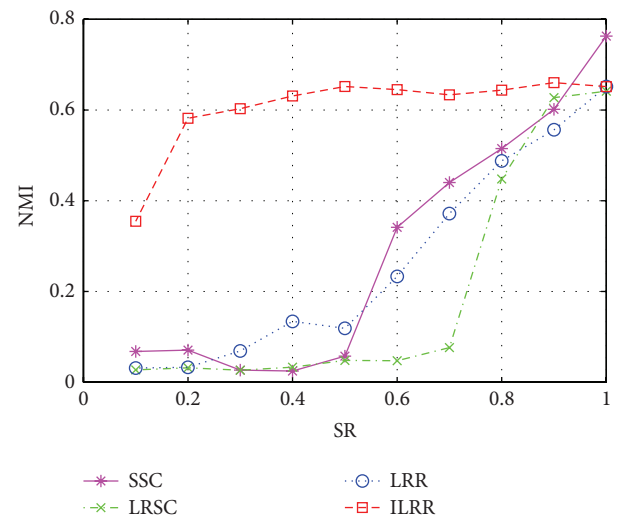


FIGURE 3: NMI comparison on Extended Yale B.

After running Algorithm 1, we obtain the optimal low-rank representation matrix \mathbf{Z} . On the basis of \mathbf{Z} , we construct an affinity matrix $\tilde{\mathbf{Z}} = (|\mathbf{Z}| + |\mathbf{Z}^T|)/2$, where $|\cdot|$ means the absolute value operator. Then we choose spectral clustering [23] as the clustering algorithm and evaluate the clustering performance by normalized mutual information (NMI). Let C be a set of true cluster labels and let C' be a set of clusters obtained from the spectral clustering algorithm. NMI is calculated as

$$\text{NMI} = \frac{\text{MI}(C, C')}{\sqrt{H(C)H(C')}}, \quad (26)$$

where $\text{MI}(C, C')$ is the mutual information metric and $H(C)$ is the entropy of C . NMI takes values in $[0, 1]$ and a larger NMI value indicates a better clustering performance.

In the experimental implementation, we vary the SR from 0.2 to 1 with an interval of 0.1. For each fixed SR, we repeat the experiments 10 times and report the average NMI. We first compare the affinity matrices $\tilde{\mathbf{Z}}$ produced by SSC, LRSC, LRR, and ILRR, as shown partially in Figure 1. From Figure 1, we observe that our method exhibits obvious block structures compared to other three methods, whereas SSC, LRSC, and LRR show no block-diagonal structure with $\text{SR} = 0.3$. This observation means that only ILRR can keep compact representation for the same subspace and divergent

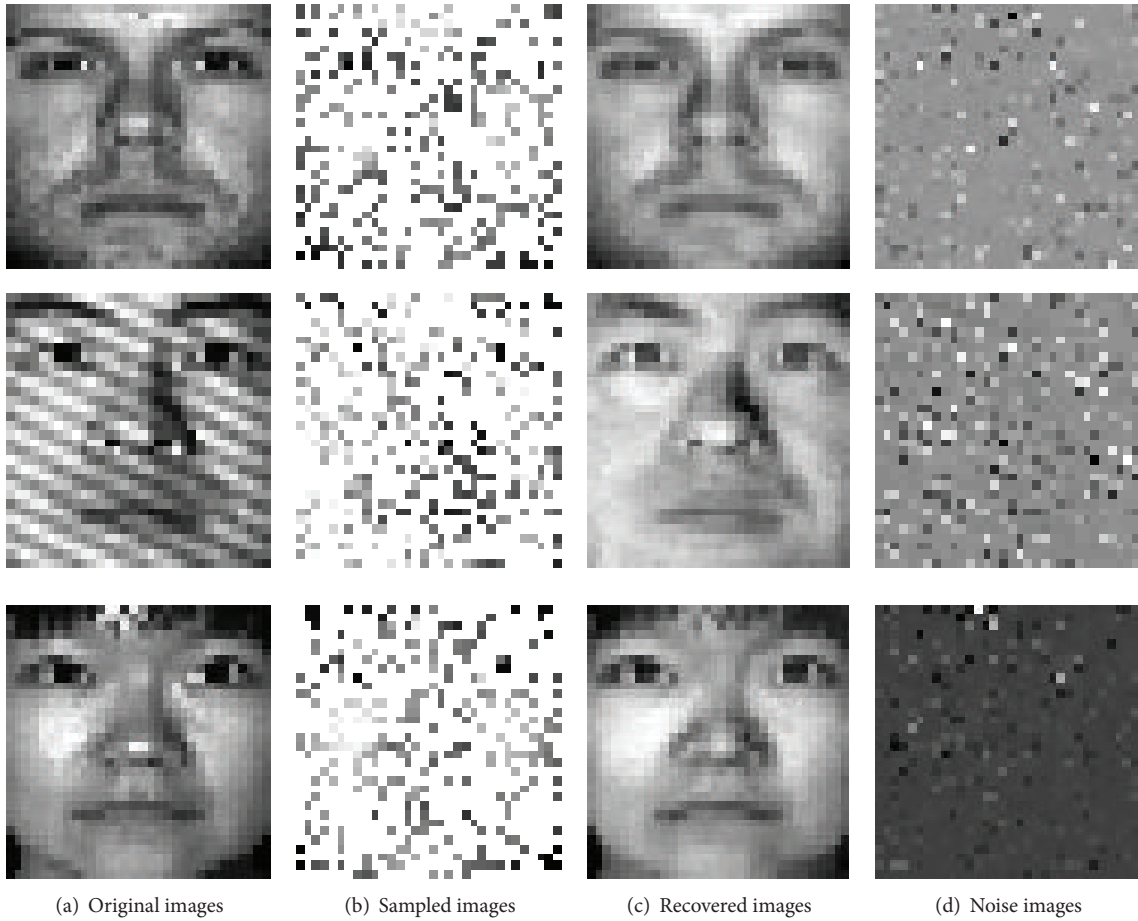


FIGURE 4: Face images recovery of ILRR on Extended Yale B with $SR = 0.3$.

representation for different subspaces when a large number of values are missing.

Secondly, we compare NMI values of ILRR with that of other three methods at different SR, as shown in Figure 2. It can be seen from this figure that NMI values of SSC, LRR, and ILRR are almost one if $SR \geq 0.7$, while ILRR has higher NMI values than other three methods if $SR \leq 0.6$. In other words, the NMI values of SSC, LRSC, and LRR are reduced drastically with the decreasing of SR. These observations verify that ILRR is more robust than other methods in the presence of missing values.

5.2. Face Clustering. We carry out face clustering experiments on a part of Extended Yale Face Database B [24] with large corruptions. This whole database consists of 38 objects and each object has about 64 images. We choose the first 10 objects and each image is resized into 32×32 . Thus, the face images dataset is represented by a matrix with size of 1024×640 . Each column of the data matrix is normalized to an l_2 unit length in consideration of variable illumination conditions and poses.

The generation manner of sampling set Ω is similar to that of the previous part. We vary the values of SR from 0.1 to 1

with an interval of 0.1 and set $\lambda = 1.2$ in LRR and ILRR. The comparison of NMI values is compared among SSC, LRSC, LRR, and ILRR, as shown in Figure 3, where each NMI value is the average result of ten repeated experiments. It can be seen from Figure 3 that ILRR achieves relatively stable NMI values if $SR \geq 0.2$, while other three methods obtain much worse NMI values than ILRR if $SR \leq 0.9$. This observation shows that SSC, LRSC, and LRR are more sensitive than ILRR on SR.

Furthermore, ILRR has an advantage in recovering the low-rank components over other three methods. In the following, we give the recovery performance contrast of ILRR. Here, we only consider two different SR; that is, $SR = 0.3$ and $SR = 0.6$. For these two given SR, the original images, sampled images, recovered images, and noise images are shown partially in Figures 4 and 5, respectively. Among these sampled images, the unsampled entries are shown in white. From these two figures, we can see that ILRR not only corrects automatically the corruptions (shadow and noise) but also recovers efficiently the low-rank components.

5.3. Motion Segmentation. In this part, we test the proposed ILRR method for the task of motion segmentation. We consider eight sequences of outdoor traffic scenes, a subset of

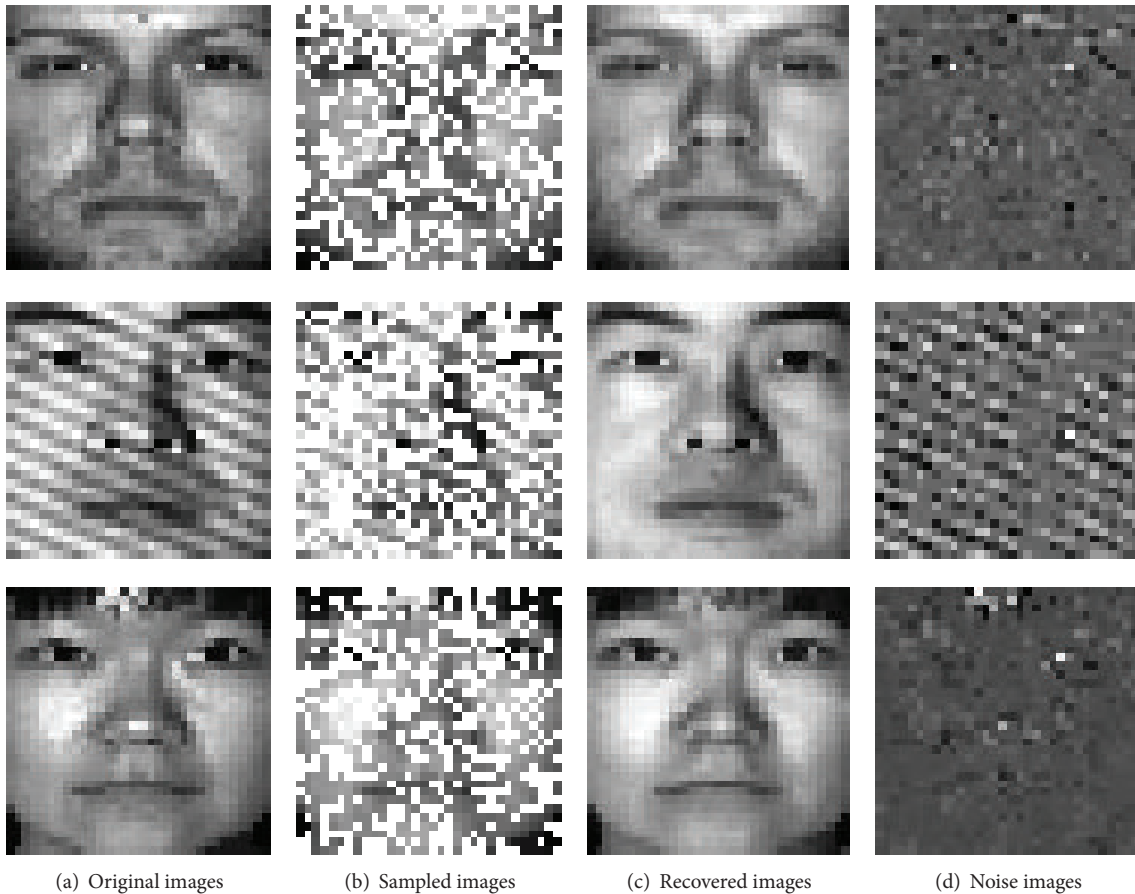


FIGURE 5: Face images recovery of ILRR on Extended Yale B with $SR = 0.6$.

the Hopkins 155 dataset [25]. These sequences were taken by a moving handheld camera and they tracked two cars translating and rotating on a street, as shown partly in Figure 6. These sequences are drawn from two or three motions, where one motion corresponds to one subspace. There are 8 clustering tasks in total since motion segmentation of each sequence is a sole clustering task. The tasks of motion segmentation are carried out according to the given features extracted and tracked in multiple frames of the above video sequences.

The extracted feature points of each sequence can be reshaped into an approximately low-rank matrix with n columns, where n equals the number of feature points. In the detailed experimental implementation, we consider different SR varying from 0.2 to 1 with an interval 0.2. The regularization parameter set λ in LRR and ILRR is set to 2. For fixed sequence and fixed SR , each method is repeated 10 times and the mean NMI values are recorded. We report the mean and the standard derivation of NMI values among 8 sequences, as shown in Table 1.

From Table 1, we can see that ILRR obtains very high NMI values especially for $SR \geq 0.2$, while NMI values of other three methods are unacceptable for $SR \leq 0.8$. Although SSC segments exactly each subspace when $SR = 1.0$, small SR has a fatal influence on SSC's segmentation performance.

TABLE 1: NMI (%) comparison of different methods on traffic sequences.

SR	SSC	LRSC	LRR	ILRR
0.2	5.15 ± 4.97	8.76 ± 4.66	9.97 ± 5.26	66.18 ± 15.40
0.4	13.74 ± 7.00	15.61 ± 10.50	15.07 ± 11.45	89.98 ± 13.24
0.6	21.45 ± 15.88	18.48 ± 11.84	17.78 ± 13.71	90.60 ± 13.72
0.8	50.97 ± 25.52	18.85 ± 14.06	17.46 ± 15.71	96.71 ± 6.84
1.0	100 ± 0.00	96.38 ± 7.60	96.83 ± 6.41	96.83 ± 6.41

In summary, ILRR is more robust to a large number of missing values than other three methods.

6. Conclusions

In this paper, we investigate the model of low-rank representation for incomplete data, which can be regarded as the generalization of low-rank representation and matrix completion. For the model of incomplete low-rank representation, we propose an iterative scheme based on augmented Lagrange multipliers method. Experimental results show that the proposed method is feasible and efficient in recovering

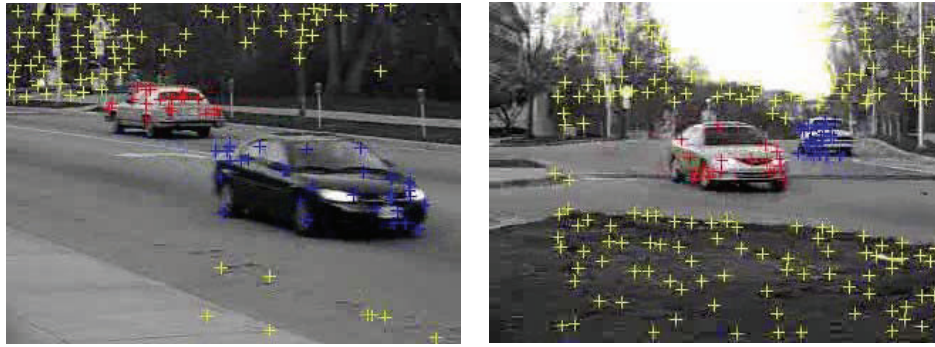


FIGURE 6: Images from traffic sequences on the Hopkins 155 dataset.

low-rank structure, completing missing entries, and removing noise. It still needs further research on how to construct a general model of low-rank matrix recovery and design the corresponding algorithm.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (no. 61403298, no. 11326204, no. 11401457, and no. 11401357), by the Natural Science Basic Research Plan in Shaanxi Province of China (no. 2014JQ8323 and no. 2014JQ1019), by the Shaanxi Provincial Education Department (no. 2013JK0587 and no. 2013JK0588), and by the Hanzhong Administration of Science & Technology (no. 2013hzzx-39).

References

- [1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [2] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao, "Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization," in *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS '09)*, vol. 22, pp. 2080–2088, December 2009.
- [3] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27th International Conference on Machine Learning*, pp. 663–670, 2010.
- [4] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, article 11, 2011.
- [5] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding*, vol. 122, pp. 22–34, 2014.
- [6] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)*, pp. 713–719, August 2005.
- [7] F. Nie, H. Wang, X. Cai et al., "Robust matrix completion via joint Schatten p-norm and lp-norm minimization," in *Proceedings of the 12th IEEE International Conference on Data Mining*, pp. 566–574, 2012.
- [8] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [9] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recognition Letters*, vol. 43, no. 1, pp. 47–61, 2014.
- [10] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.
- [11] I. T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, 2nd edition, 2002.
- [12] F. De La Torre and M. J. Black, "Robust principal component analysis for computer vision," in *Proceedings of the 8th International Conference on Computer Vision*, pp. 362–369, July 2001.
- [13] C. Ding, D. Zhou, X. He et al., "R1-PCA: rotational invariant L1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 281–288, 2006.
- [14] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [15] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, "Stable principal component pursuit," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT '10)*, pp. 1518–1522, Austin, Tex, USA, June 2010.
- [16] Z. Lin, M. Chen, L. Wu et al., "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," Tech. Rep. UILU-ENG-09-2215, UIUC, 2010.
- [17] X. Yuan and J. Yang, "Sparse and low-rank matrix decomposition via alternating direction methods," Technical Report of Hong Kong Baptist University, 2009.
- [18] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, New York, NY, USA, 1982.
- [19] E. Elhamifar and R. Vidal, "Sparse subspace clustering: algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

- [20] J. Shi, X. Zheng, and L. Yong, "Incomplete robust principal component analysis," *ICIC Express Letters Part B: Applications*, vol. 5, no. 6, pp. 1531–1538, 2014.
- [21] Y. Liu and F. Shang, "An efficient matrix factorization method for tensor completion," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 307–310, 2013.
- [22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [23] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, pp. 849–856, 2002.
- [24] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
- [25] R. Vidal and R. Hartley, "Three-view multibody structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 214–227, 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

