

## Research Article

# A Novel Robust Principal Component Analysis Algorithm of Nonconvex Rank Approximation

E. Zhu <sup>1,2</sup>, M. Xu,<sup>3</sup> and D. Pi<sup>2</sup>

<sup>1</sup>College of Internet of Things Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China

<sup>2</sup>College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China

<sup>3</sup>College of Electronic and Information Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China

Correspondence should be addressed to E. Zhu; [erxi666@163.com](mailto:erxi666@163.com)

Received 21 June 2020; Revised 4 September 2020; Accepted 17 September 2020; Published 30 September 2020

Academic Editor: Francesc Pozo

Copyright © 2020 E. Zhu 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.

Noise exhibits low rank or no sparsity in the low-rank matrix recovery, and the nuclear norm is not an accurate rank approximation of low-rank matrix. In the present study, to solve the mentioned problem, a novel nonconvex approximation function of the low-rank matrix was proposed. Subsequently, based on the nonconvex rank approximation function, a novel model of robust principal component analysis was proposed. Such model was solved with the alternating direction method, and its convergence was verified theoretically. Subsequently, the background separation experiments were performed on the Wallflower and SBMnet datasets. Furthermore, the effectiveness of the novel model was verified by numerical experiments.

## 1. Introduction

As fuelled by the advancement of big data and artificial intelligence industry, the data in the image processing, semantic analysis, and other application fields exhibit extremely large scale and dimension, leading to a significant difficulty of data processing and analysis. It has been reported from research that the nature of the mentioned data is commonly characterized by low-rank subspace [1]. For this reason, the problems in the fields mentioned can be transformed into the problems such as low-rank matrix representation [2] or recovery [3]. In industrial production, the collected data usually contain outliers or noises for the factors (e.g., environment and equipment). For instance, as impacted by cloud or other objects' covering, the geomorphological map collected by the sensors contains shadows; face image exhibits differences in brightness for the effect of the light; the quality of nuclear magnetic resonance image degrades because of object motion, and the defects of imaging systems, as well as inherent noise and external interference of recording equipment. Under the impact of the mentioned practical problems, image denoising, image

reconstruction, image restoration, and image completion have become research hotspots.

Image reconstruction and image denoising aim at discovering the low-rank part of the image-related data, i.e., the background in the image. Moreover, image recognition of face recognition and motion segmentation can exploit the subspace clustering method to recover the low-rank part of the image data, i.e., seeking the low-rank matrix representation of the data. Seeking the low-rank part and low-rank matrix representation in the data, denoising, and completing can be essentially considered the low-rank matrix recovery from the perspective of mathematics. Accordingly, the low-rank matrix recovery model creates a unified framework for the mentioned problems existing in image processing.

From the mathematic perspective, each piece of data is arranged into a column vector, and then a data matrix  $X \in R^{m \times n}$  is formed. It is assumed that  $X$  can be decomposed as

$$X = Y + E, \quad (1)$$

where  $X, Y, E \in R^{m \times n}$ ;  $Y$  denotes the low-rank matrix, representing the intrinsic low-rank subspace characteristics

of the real data; and  $E$  is the sparse matrix, representing the outliers or noises in the real data.

To solve the low-rank matrix recovery problem, classical principal component analysis and robust principal component analysis can be conducted, whereas numerous problems may exist in the problem-solving process.

*Problem 1.* Classical principal component analysis [4, 5] method is adopted to solve the model below:

$$\begin{cases} \min & \|X - Y\|_F^2 \\ \text{s.t.} & \text{rank}(Y) \leq k, \end{cases} \quad (2)$$

where  $\|\cdot\|_F^2$  denotes the Frobenius norm of the matrix and  $\text{rank}(Y)$  represents the rank of the matrix. First, the rank of low-rank matrix was assessed, which was not larger than  $k$ . Second,  $E$  should be the only small independent identically distributed Gaussian noises. Equation (2) exploits the singular value matrix decomposition to determine the singular value matrix decomposition of data matrix  $X$ ,  $k$  main singular values, and corresponding left singular vectors, which were subsequently employed to indicate the major features of the analysis or recognition pattern. Principal component analysis (PCA), as an effective tool for data analysis and dimensional reduction, has been extensively applied in various fields. However, if the data are seriously polluted (e.g., large Gaussian noise or only sparse non-Gaussian noise), conventional PCA will fail.

*Problem 2.* In conventional subspace learning models, principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis [6] (ICA) can significantly denoise small Gaussian noise contained in data, whereas they are very sensitive to the data containing outliers or sparse large noises. To recover the low-rank structure from the data containing sparse large noises, Chandrasekaran et al. [7] and Wright et al. [8] also built a robust principal component analysis model, which is expressed below:

$$\begin{cases} \min_{Y,E} & \text{rank}(Y) + \eta \|E\|_0 \\ \text{s.t.} & X = Y + E, \end{cases} \quad (3)$$

where  $\|E\|_0$  denotes the  $\ell_0$ -norm of matrix  $E$ , i.e., the number of nonzero elements in  $E$ , and  $\eta$  represents the positive trade-off parameter.

The objective function of equation (3) is a nonlinear nonconvex function, and its solution is NP problem and relatively difficult. The compression sensing theory was introduced in this problem to solve the convex relaxation problem:

$$\begin{cases} \min_{Y,E} & \|Y\|_* + \eta \|E\|_1 \\ \text{s.t.} & X = Y + E, \end{cases} \quad (4)$$

where  $\|Y\|_*$  denotes the kernel norm of matrix  $Y$  (trace norm or Ky-Fan norm), i.e.,  $\|Y\|_* = \sum_{i=1}^r \sigma_i(Y)$ , where  $\gamma$  denotes the rank of  $Y$  and  $\sigma_i(Y)$  is the  $i$ -th singular value of  $Y$ .  $\|E\|_1$

represents the  $\ell_1$ -norm of matrix  $E$ , i.e.,  $\|E\|_1 = \sum_{i,j} |E_{ij}|$  is the sum of the absolute values of all elements of  $E$ .

When the singular vector of the low-rank matrix distributes reasonably and the nonzero elements of the sparse matrix distribute uniformly, equation (4) recovers the low-rank part and the noisy part from the matrix of the real data exhibiting a probability close to 1. However, in the definition of nuclear norm, all singular values are treated unequally. All singular values in the problem should upregulate the rank by 1 regardless of their values. However, in terms of the nuclear norm, the contribution of the nonzero singular values to the nuclear norm is directly determined by its value. Accordingly, the nuclear norm is commonly regulated by some large singular values, even if there are only a few large singular values, whose effect cannot be overlooked. Thus, in the presence of a large singular value, the nuclear norm will deviate from the rank of the low-rank matrix, so it cannot accurately approximate the rank, and the degree of inaccuracy is upregulated with the increase in the singular value.

*Problem 3.* The situation that the nuclear norm cannot well approximate the rank of the matrix can be considered a problem arising from the compression sensing. Since  $\ell_1$ -norm is the minimum convex envelope of  $\ell_0$ -norm [9], conventional compression sensing technology exploits  $\ell_1$ -norm instead of  $\ell_0$ -norm to recover the original data. However, under several conditions,  $Y$  is both a low-rank matrix and a sparse matrix, while  $E$  denotes a sparse matrix exhibiting low-rank properties. It was demonstrated that  $\ell_1$ -norm is not an exact approximation of  $\ell_0$ -norm [10, 11], and in most cases, this approximation bias is not negligible.

Over the past few years, to improve the approximation rank of nuclear norm, numerous scholars have been committed to develop nonconvex rank approximation functions [12–14] to remedy the defect of nuclear norm convex rank approximation functions, and they have been extensively employed. Considerable experiments proved that the nonconvex rank approximation can more accurately and fastly recover the low-rank part of the data than the nuclear norm. For instance, Sun et al. [11] proposed a nonconvex model of robust principal component analysis with capped trace norm and capped  $\ell_1$ -norm; Kang et al. [15] built a more accurate nonconvex rank approximation function, termed as log-determinant function. In literature [16], Kang et al. defined a novel norm as the nonconvex rank approximation, which is termed as  $\gamma$ -norm, and the experimental results revealed that it is a better rank approximation than the nuclear norm. Xie et al. [17] built a nonconvex regularization tool, termed as weighted Schatten  $p$ -norm, which acted as the rank approximation to achieve a better approximation. Lu et al. [12] introduced a class of nonconvex approximation functions of  $\ell_0$ -norm to the singular values of low-rank matrices and built a nonconvex low-rank minimization problem with the nonconvex approximation functions thus obtained. Yu et al. [18] achieved a threshold representation as a recursive function for general  $\ell_p$  regularization problems; subsequently, they designed a filtering algorithm suitable for image reconstruction. Xu et al. [19] introduced an algorithm of background subtraction based on low rank

and structured sparse decomposition. The algorithm process fully complied with Gao's algorithm [20], block-sparse RPCA. The difference between algorithms of Gao and Liu was that the section of first-pass RPCA introduced structure sparse norm considering the foreground structural continuity. It utilized the adaptive regularization parameter based on the significant adjustment of motion detection method to do group-sparsity operation, and thereby better results were obtained. Liu et al. [21] proposed an efficient numerical solution, Grassmannian Online Subspace Updates with Structured-Sparsity (GOSUS), for online subspace learning in the context of sequential observations involving structured perturbations. Xu modelled both homogeneous perturbations of the subspace and structural contiguities of outliers by Grassmannian fluid, and after certain manipulations, they solved via an alternating direction method of multipliers. Javed et al. [22] introduced a spatiotemporal low-rank modelling method to overcome occlusions by foreground objects and redundancy in video data. The proposed method encodes spatiotemporal constraints by regularizing spectral graphs and uses optical flow information to remove redundant data and to create a set of dynamic frames from the input video sequence. Through their experiments, it can be confirmed that the algorithm has a high efficiency in processing background reconstruction. However, the 11 parameters required by the two models proposed by them need to be optimized, which greatly affects the robustness of the algorithm.

Since the rank approximation of the low-rank matrix refers to the critical problem in the low-rank matrix recovery, how to build a more accurate approximation function should be studied [12, 13, 23], and on that basis, the corresponding low-rank matrix recovery model is built, which is of great practical significance to solving image processing problems (e.g., image reconstruction, image denoising, and image recognition).

Based on the mentioned inspirations, starting from the construction of nonconvex rank approximation function of low-rank matrix, the present study built a low-rank matrix recovery model suitable for nonconvex approximation in different image processing problems and then designed the corresponding algorithm to obtain more accurate low-rank matrix part from the original data, as an attempt to solve the specific problems (e.g., image reconstruction and image denoising).

## 2. Relevant Work

The so-called low-rank matrix is presented as follows: assuming  $X = (x_{ij})_{m \times n} \in R^{m \times n}$ , its rank is  $r$ , and  $r \ll \min\{m, n\}$ . Subsequently,  $X$  is the low rank. Each row and column of low-rank matrix covers considerable redundant information, which can be exploited for image recovery and image feature extraction.

*2.1. Evolution of Low-Rank Matrix Recovery Model.* The present study involved the relevant application of low-rank representation; its model evolution and solving algorithm

were first presented. Starting from the simplest case, assuming that the data were clean, i.e., the data were not polluted or not covered with any noise and outliers, considering the following rank minimization problems

$$\begin{cases} \min_Y \text{rank}(Y) \\ \text{s.t. } X = PY, \end{cases} \quad (5)$$

where  $P$  is termed as the dictionary matrix and can also be considered the basis matrix of the data space;  $Y$  denotes the low-rank matrix; and the lowest rank of the data matrix was expressed by solving equation (5). The general solution of equation (5) is to replace the rank with the nuclear norm, which is transformed into the following problem:

$$\begin{cases} \min_Y \|Y\|_* \\ \text{s.t. } X = PY. \end{cases} \quad (6)$$

Liu et al. proved the properties of the solution of equation (6) in literature [24].

*2.1.1. Optimal Solution Is Unique.* Since the kernel norm is not a strictly convex function, equation (6) may achieve multiple optimal solutions, whereas the theorem below can still prove that equation (6) achieves a unique optimal solution.

**Theorem 1** (see [24]). *Assuming  $P \neq 0$  and  $X = PY$  shows a feasible solution, i.e.,  $X \in \text{span}(P)$ , then*

$$Y^* = P^\dagger X. \quad (7)$$

$Y^* = P^\dagger X$  denotes the unique solution to equation (6), where  $P^\dagger$  expresses the generalized inverse of  $P$ .

*2.1.2. Optimal Solution Is Block Diagonal.* In the selection of a right dictionary, the lowest rank representation will reveal the true segmentation result. In other words, when the columns of  $P$  and  $X$  are indeed from independent subspaces, the optimal solution of equation (6) can reveal the subspace relationship between sample points. The mentioned conclusion can be expressed by the theorem below.

**Theorem 2.** (see [24]). *Set  $\{S_1, S_2, \dots, S_k\}$  as the set of  $k$  subspaces, and the rank (or dimension) of  $S_i$  is  $r_i$ ,  $P = [P_1, P_2, \dots, P_k]$  and  $X = [X_1, X_2, \dots, X_k]$ . Assuming that  $P_i$  denotes the set of  $m_i$  sample points of subspace  $S_i$ ,  $X_i$  represents the set of  $n_i$  sample points of subspace  $S_i$ , and  $P_i$  satisfies  $\text{rank}(P_i) = r_i$ . If the subspace  $\{S_1, S_2, \dots, S_k\}$  is independent, the optimal solution  $Y^*$  of problem (6) is block diagonal:*

$$Y^* = \begin{bmatrix} Y_1^* & 0 & \cdots & 0 \\ 0 & Y_2^* & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_k^* \end{bmatrix}, \quad (8)$$

where  $Y_i^* \in \mathbb{R}^{m_i \times n_i}$  is the coefficient matrix and  $\text{rank}(Y_i^*) = \text{rank}(X_i)$ .

Clean data building is mentioned above. Besides, a general situation is given below. For data with noise, the following rank minimization model should be built:

$$\begin{cases} \min_{Y,E} & \text{rank}(Y) + \eta \|E\|_\ell \\ \text{s.t.} & X = PY + E, \end{cases} \quad (9)$$

where  $\eta$  denotes a positive compromise factor, i.e., the equilibrium parameter.  $\|\cdot\|_\ell$  represents  $\ell$ -norm, where  $\ell$  denotes a parameter of norm, i.e.,  $\ell$  can be taken as (1,2), (2,1) etc., and  $E$  denotes the noise matrix. Obviously, when  $P = I$  ( $I$  denotes the unit matrix), equation (9) is converted into the form of equation (4). Accordingly, the low-rank representation can be considered as the generalization of robust principal component analysis with orthonormal basis as a dictionary. By taking a proper basis, the data from different subspaces can be extracted with low-rank representation, thereby remedying the defect that the robust principal component analysis can only process the data from the identical subspace.

For equation (9), the kernel norm acts as the convex approximation of the rank function; since it is more robust to noises and outliers, it is converted into the following model:

$$\begin{cases} \min_{Y,E} & \|Y\|_* + \eta \|E\|_{2,1} \\ \text{s.t.} & X = PY + E. \end{cases} \quad (10)$$

Equation (10) can be solved by the augmented Lagrange multiplier method. For the separation of variables, a novel variable matrix  $J$  is introduced, and constraints  $Y = J$  are added. Equation (10) has the following equivalent form:

$$\begin{cases} \min_{Y,E} & \|J\|_* + \eta \|E\|_{2,1} \\ & X = PY + E \\ \text{s.t.} & Y = J. \end{cases} \quad (11)$$

The augmented Lagrange function of equation (11) is

$$\begin{aligned} \Psi(Y, J, E, \Lambda_1, \Lambda_2) &= \|J\|_* + \eta \|E\|_{2,1} + \langle \Lambda_1, X - PY - E \rangle \\ &+ \langle \Lambda_2, Y - J \rangle + \frac{\rho}{2} (\|X - PY - E\|_F^2 \\ &+ \|Y - J\|_F^2). \end{aligned} \quad (12)$$

Equation (11) can be solved with precise and imprecise Lagrange multiplier methods. The convergence of the former has been verified in literature [25]. For the latter, convergence is not ensured for cases with more than two pieces for  $Y$ . The primary steps to solve equation (12) with the precise Lagrange multiplier method are presented below:

Step 1: with singular value thresholding operator [26] (SVT operator), the optimal solution  $J_{k+1}$  of the  $(k+1)$ -th iteration is solved by

$$J_{k+1} = \arg \min_J \frac{1}{\rho_k} \|J\|_* + \frac{1}{2} \left\| J - \left( Y_k + \frac{\Lambda_2^k}{\rho_k} \right) \right\|_F^2. \quad (13)$$

Step 2: the optimal solution  $Y_{k+1}$  of the  $(k+1)$ -th iteration is solved:

$$\begin{aligned} Y_{k+1} &= (I_n + P^T P)^{-1} (P^T (X - E_k) + J_{k+1} \\ &+ \frac{(P^T \Lambda_1^k - \Lambda_2^k)}{\rho_k}). \end{aligned} \quad (14)$$

Step 3: based on the conclusion in literature [27], the optimal solution  $E_{k+1}$  of the  $(k+1)$ -th iteration is solved by using the following formula:

$$\begin{aligned} E_{k+1} &= \arg \min_{E} \frac{\eta}{\rho_k} \|E\|_{2,1} \\ &+ \frac{1}{2} \left\| E - \left( X - PY_{k+1} + \frac{\Lambda_1^k}{\rho_k} \right) \right\|_F^2. \end{aligned} \quad (15)$$

Step 4: the multiplier matrices  $\Lambda_1$  and  $\Lambda_2$  and parameter  $\rho$  are updated:

$$\begin{aligned} \Lambda_1^{k+1} &= \Lambda_1^k + \rho_k (X - PY_{k+1} - E_{k+1}), \\ \Lambda_2^{k+1} &= \Lambda_2^k + \rho_k (Y_{k+1} - J_{k+1}), \\ \rho_{k+1} &= \min(\beta \rho_k, \rho_{\max}), \end{aligned} \quad (16)$$

where  $\beta$  and  $\rho_{\max}$  denote the predetermined positive numbers; generally,  $\beta = 1.1$  and  $\rho_{\max} = 106$ .

According to equations (13)–(16), each variable is alternately updated until the termination condition of the algorithm is reached. For instance, a previously selected small positive number  $\varepsilon$  (generally,  $\varepsilon = 10^{-7}$ ,  $10^{-8}$ ) can be taken in the following form [24]:  $\|X - PY - E\|_\infty \leq \varepsilon$ ,  $\|Y - J\|_\infty \leq \varepsilon$ .

**2.2. Principle of Constructing Nonconvex Rank Approximation Function.** Since the kernel norm is the minimum convex envelope of the rank function, the conventional low-rank matrix recovery model employs the kernel norm as the rank approximation function of the low-rank matrix; as a result, the problem is transformed into a convex optimization problem, e.g., equations (4), (6), and (10). However, the kernel norm approximation rank function has some defects, so the conventional low-rank matrix recovery model should be optimized. Since  $\ell_1$ -norm is only loose approximation of  $\ell_0$ -norm, only the suboptimal solution of the original problem can be obtained. Thus, a function better than the  $\ell_1$ -norm should be developed to replace the  $\ell_0$ -norm. Under this context, scholars proposed numerous nonconvex approximations of the  $\ell_0$ -norm, and some nonconvex approximate functions are presented in literature [12] (e.g.,  $\ell_p$ -norm [18] ( $0 < p < 1$ ), Geman and Young [28], smoothly clipped absolute deviation (SCAD) [29], minimax concave penalty

TABLE 1: Eight nonconvex approximate functions of  $\|x\|_0$ .

Name of alternate function	Expression of alternate function $f(x)$ , $x \geq 0$ , $\mu > 0$
$\ell_p$ [18]	$\mu x^p$
Geman and Young [28]	$(\mu x / (x + \gamma))$
SCAD [29]	$\begin{cases} \mu x, & x \leq \mu, \\ (-x^2 + 2\gamma\mu x - \mu^2) / (2(\gamma - 1)), & \mu \leq x \leq \gamma\mu, \\ (\mu^2(\gamma + 1) / 2), & x > \gamma\mu. \end{cases}$
MCP [30]	$\begin{cases} (\mu x - x^2) / (2\gamma), & x < \gamma\mu, \\ (\gamma\mu^2 / 2), & x \geq \gamma\mu. \end{cases}$
Laplace [31]	$\mu(1 - \exp(-x/\gamma))$
Capped $\ell_1$ [32]	$\begin{cases} \mu x, & x < \gamma, \\ \mu\gamma, & x \geq \gamma. \end{cases}$
ETP [33]	$(\mu / (1 - \exp(-\gamma))) (1 - \exp(-\gamma x))$
Logarithm [34]	$(\mu / \log(\gamma + 1)) \log(\gamma x + 1)$

(MCP) [30], Laplace [31], capped  $\ell_1$  [32], exponential-type penalty (ETP) [33], and logarithmic function [34]). The specific function forms are listed in Table 1.

The images of the eight nonconvex approximate functions are shown in Figure 1.

As shown in Figure 1, the graphs of the mentioned functions show a similar trend, which also inspires us to develop novel nonconvex approximate functions. Moreover, by combining with the problem of approximate rank of nuclear norm, to develop the approximate function of the rank function more accurately than the nuclear norm, our principle aims to satisfy the following properties:

- (1) The contribution of a singular value complying with a large value  $\sigma$  the rank approximation function should be bounded; its contribution should be significantly reduced compared with the nuclear norm, i.e., it satisfies

$$0 \leq f(\sigma) \leq M \quad (17)$$

$$\text{or } \lim_{\sigma \rightarrow \infty} \frac{f(\sigma)}{\sigma} = 0,$$

where  $f(\sigma)$  denotes the function of singular value  $\sigma$ , representing the contribution of the singular value, and  $\sigma$  is a positive constant.

- (2)  $f(\sigma)$  denotes the nonsubtractive function of singular value  $\sigma$ .
- (3) When the singular value  $\sigma = 0$ , the contribution is 0, i.e.,  $f(0) = 0$ .

**2.3. A Novel Low-Rank Matrix Recovery Model with Nonconvex Rank Approximation.** According to the definition of nuclear norm, under a too large value of a certain  $\sigma_i(X)$ , the rank estimation will be excessively large, affecting the approximation effect. The present section will give a novel nonconvex approximation function to approximate the rank function of the matrix. Lu et al. [35] summarized the existing

nonconvex approximation model and developed the following general form model:

$$\min_{X \in \mathbb{R}^{m \times n}} F(X) = \sum_{i=1}^m f(\sigma_i(X)) + g(X), \quad (18)$$

where  $f$  denotes a monotonically increasing continuous concave function defined in  $[0, \infty)$ . The gradient function of  $g \nabla g$  Lipschitz is continuous.

In terms of the matrix  $Y \in \mathbb{R}^{m \times n}$ ,  $F(Y) = \sum_{i=1}^{\min\{m,n\}} f(\sigma_i(Y))$ , where  $f(\cdot)$  denotes a certain form of nonconvex approximation function. Subsequently, the rank approximation function in equations (3) and (5) can be approximately replaced with  $F(Y)$ . Thus, a model based on nonconvex rank approximation can be developed.

For the robust principal component analysis model, the nonconvex rank approximation model of equation (3) is expressed as

$$\begin{cases} \min_{Y,E} F(Y) + \eta \|E\|_0 \\ \text{s.t. } X = Y + E, \end{cases} \quad (19)$$

where  $Y, E \in \mathbb{R}^{m \times n}$ ;  $\|E\|_0$  denotes the  $\ell_0$ -norm of matrix  $E$ ; and  $\eta$  is the positive compromise factor.

A general low-rank matrix recovery model based on nonconvex rank approximation is mentioned above. For instance, equation (4) indicates a model with sparse noise only in the original data, while equation (5) builds the model with no noise in the original data. For the low-rank matrix recovery model under other noises, the rank function of the low-rank matrix in the model can also be replaced by several appropriate nonconvex rank approximation functions.

In the nonconvex rank approximation model, by adopting nonconvex rank approximation function  $F(Y)$ , exhibiting higher accuracy than the nuclear norm, the use of the nonconvex rank approximation model can more effectively restore low-rank matrix part in the original data, which can enhance the application effect of the original low-rank matrix recovery model in image processing (e.g., image reconstruction, image denoising, and image recognition).

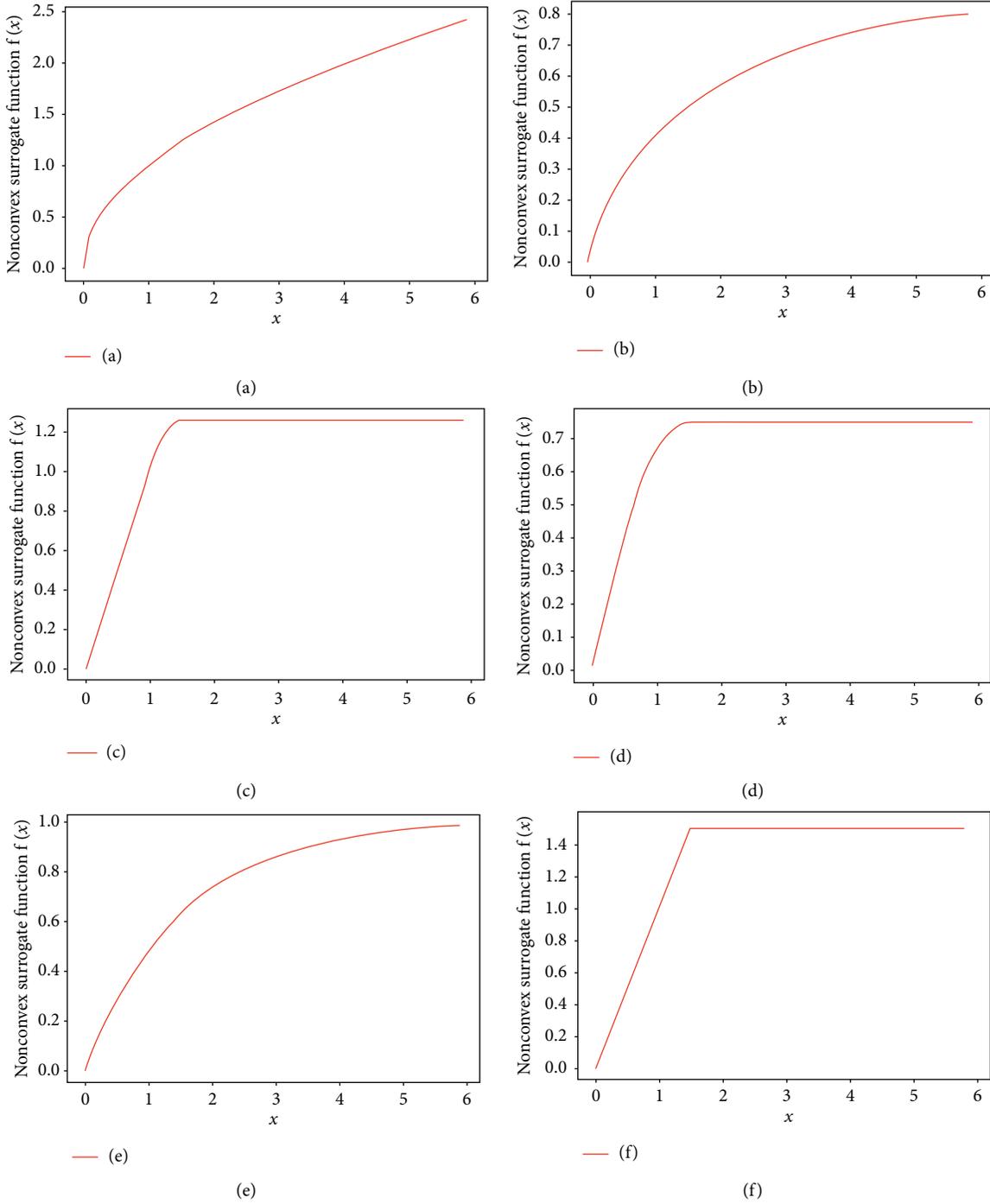


FIGURE 1: Continued.

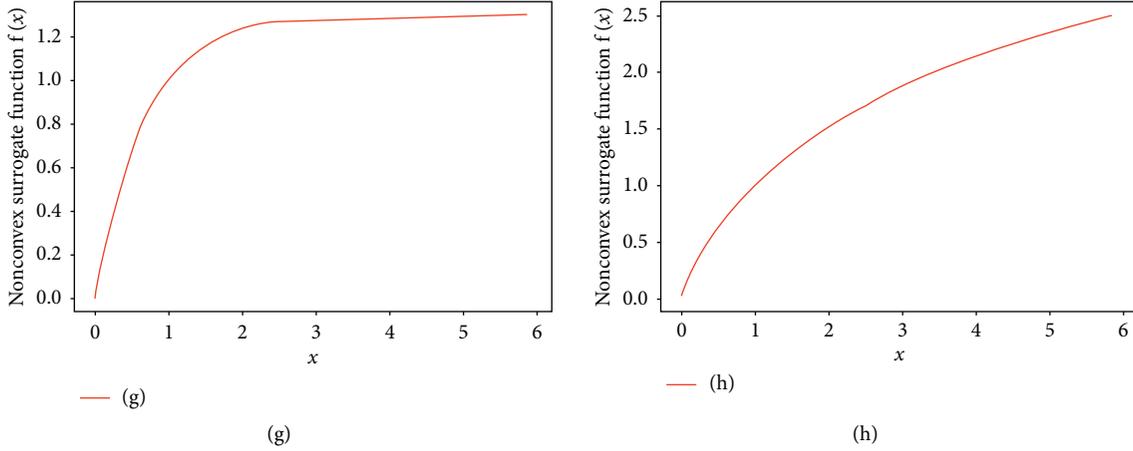


FIGURE 1: Eight nonconvex approximate functions of  $\|x\|_0$ : (a)  $\ell_p$ ; (b) Geman and Young; (c) SCAD; (d) MCP; (e) Laplace; (f) Capped  $\ell_1$ ; (g) ETP; (h) Logarithm  $p = 0.5\mu = 1$ ,  $\gamma = 1.5$ .

Subsequently, a novel low-rank matrix recovery model with nonconvex rank approximation was proposed, and the rank function of low-rank matrix in the existing model was substituted with the nonconvex rank approximation function. Next, a correlation algorithm was developed to address the problem.

**2.4. Nonconvex Rank Approximation Function.** Consistent with the approximation function of  $\ell_0$ -norm given by Geman and Young [28], a novel nonconvex rank approximation function was developed in the present study:

$$f(x) = \frac{\mu x}{x + e^{(-\gamma x)}}, \quad (20)$$

where  $\mu$  denotes the equilibrium factor, and  $\mu > 0$ ;  $\gamma$  represents the convergence rate parameter of  $f(x)$ , and  $\gamma > 0$ . Next, it will be verified that  $f(x)$  satisfies the three properties in the previous section:

- (1)  $\sigma_i$ , ( $i = 1, 2, \dots, l$ ) is assumed as the singular value of matrix  $Y$ , and  $Y \in R^{m \times n}$ , and  $0 \leq \sigma_1 \leq \sigma_2 \leq \dots \leq \sigma_l$ :

$$0 \leq \frac{\mu \sigma_l}{\sigma_l + e^{(-\gamma \sigma_l)}} \leq \frac{\mu \sigma_l}{\sigma_l} = \mu. \quad (21)$$

Moreover,  $0 \leq f(\sigma) \leq M$ ,

$$\begin{aligned} \lim_{\sigma \rightarrow \infty} \frac{f(\sigma)}{\sigma} &= \lim_{\sigma \rightarrow \infty} \frac{(\mu \sigma / (\sigma + te^{(-\gamma \sigma)}))}{\sigma} \\ &= \lim_{\sigma \rightarrow \infty} \frac{\mu}{\sigma + e^{(-\gamma \sigma)}} = 0. \end{aligned} \quad (22)$$

In other words, the contribution of the singular value exhibiting a large value to the rank approximation function should be bounded.

- (2)  $0 \leq \sigma_1 \leq \sigma_2$ :

$$\begin{aligned} f(\sigma_2) - f(\sigma_1) &= \frac{\mu \sigma_2}{\sigma_2 + e^{(-\gamma \sigma_2)}} - \frac{\mu \sigma_1}{\sigma_1 + e^{(-\gamma \sigma_1)}} \\ &= \frac{\mu \sigma_2 (\sigma_1 + e^{(-\gamma \sigma_1)}) - \mu \sigma_1 (\sigma_2 + e^{(-\gamma \sigma_2)})}{(\sigma_2 + e^{(-\gamma \sigma_2)}) (\sigma_1 + e^{(-\gamma \sigma_1)})} \\ &= \frac{\mu \sigma_2 e^{(-\gamma \sigma_1)} - \mu \sigma_1 e^{(-\gamma \sigma_2)}}{(\sigma_2 + e^{(-\gamma \sigma_2)}) (\sigma_1 + e^{(-\gamma \sigma_1)})} \geq 0. \end{aligned} \quad (23)$$

In other words,  $f(\sigma)$  denotes a nondecreasing function of singular value  $\sigma$ .

- (3)  $f(0) = (\mu \times 0 / (0 + e^{(-\gamma \times 0)})) = (0 / (0 + t1)) = 0$ .

That is, under the zero singular value  $\sigma$ , the contribution is zero.

$f(x)$  satisfies the property of nonconvex approximate rank function. Based on the function in equation (20), the following New-norm can be obtained.

**Definition 1.** (New-norm)  $Y \in R^{m \times n}$ ,  $\gamma > 0$  is assumed as a parameter. Subsequently, it is defined as

$$\|Y\|_N = \sum_i \frac{\sigma_i(Y)}{\sigma_i(Y) + e^{(-\gamma \sigma_i(Y))}}, \quad (24)$$

where  $\sigma_i(Y)$  denotes the  $i$ -th singular value of matrix  $Y$ .

When parameter  $\gamma$  takes different values, the true rank of the matrix is approximated by the nonconvex function derived from Definition 1, as presented in Figure 2.

In Figure 2, given the rank and the rank approximation function, the horizontal coordinate  $x$  denotes the singular value, while the true rank is 1 under the nonzero singular

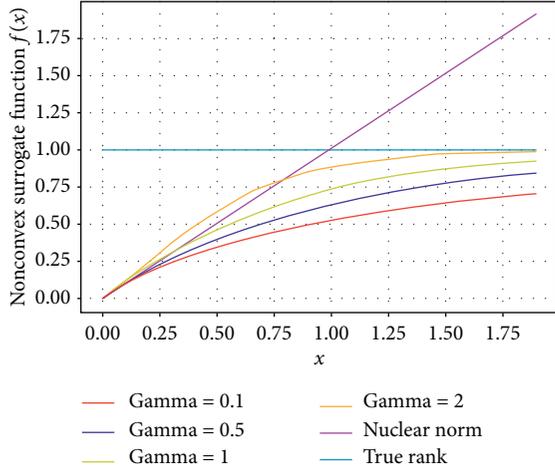


FIGURE 2: Comparisons among New-norm, nuclear norm, and the rank under the different values of  $\gamma$ .

value, and the rank is zero under the zero singular value; thus, it is expressed as a horizontal line with the function value of 1 in the figure. The curve indicates the nonconvex approximation function  $f(x)$ . With the increase in the singular value  $x$ , its function value will not surge, and it is relatively consistent with the real rank. The difference from the rank can indicate the approximation effect. With the increase in the singular value  $x$ , the function value of  $f(x)$  is close to 1, which differs slightly from the real rank. Accordingly, as a rank approximation function, its approximation effect is significantly high. Under the relatively small value of  $\gamma$ , for instance,  $\gamma = 0.1$  or  $\gamma = 0.5$ , a relatively large difference is identified between the curve obtained and the real rank. However, with the increase in the value of  $\gamma$ , the obtained curve approximates the real rank more efficiently. For instance, when  $\gamma = 2$ , the curve's approximation effect has been prominent.

Moreover, Figure 2 presents the simplification in the presence of only one singular value. The nuclear norm complies with a line starting from the origin with a slope of 1. Since the nuclear norm is the sum of all singular values, under only one singular value, the nuclear norm is exactly the singular value. Furthermore, the figure indicates the obvious defect of approximate rank of nuclear norm. The greater the singular value, the greater the degree of deviation of nuclear norm from true rank. Thus, when the nuclear norm is adopted to approximate the rank, even if there are only a few large singular values, the deviation of approximation should not be overlooked.

Next, this study proves theoretically and tests equation (24) as a more accurate nonconvex rank approximation function based on experimental results. It is known that the mentioned norm equation (24) does not satisfy the direct proportionality of the matrix norm (for arbitrary constant  $c$ ,  $\|cY\| = |c|\|Y\|$ ), whereas it is only a virtual norm, or a pseudonorm. However, it can also be verified that they satisfy the properties below.

*Property 1.*  $\|Y\|_N$  defined in equation (24) exhibits the following properties:

- (1)  $\lim_{\gamma \rightarrow 0} \|Y\|_N = \text{rank}(Y)$
- (2)  $\|Y\|_N$  denotes unitary invariant, i.e., any two orthogonal matrices  $U \in R^{m \times m}$  and  $V \in R^{n \times n}$  meet  $\|Y\|_N = \|UYV\|_N$
- (3) Positive definitiveness: for any  $Y \in R^{m \times n}$ ,  $\|Y\|_N \geq 0$  and  $\|Y\|_N = 0$  if and only if  $Y = O$

*Proof*

- (1) For

$$f(\sigma_i(Y)) = \frac{\sigma_i(Y)}{\sigma_i(Y) + e^{(-\gamma\sigma_i(Y))}}, \quad (25)$$

the following is obtained:

$$\lim_{\gamma \rightarrow 0} f(\sigma_i(Y)) = \begin{cases} 1, & \sigma_i(Y) \neq 0, \\ 0, & \sigma_i(Y) = 0. \end{cases} \quad (26)$$

In other words, when  $\gamma \rightarrow 0$ , the function value  $f(\sigma_i(Y))$  is 1 for the nonzero singular value, and 0 for the singular value of zero. On the contrary, given the rank of the matrix  $\text{rank}(Y)$ , it is also the case that each singular value with the value of nonzero upregulates the rank by 1, while each zero singular value does not increase the rank of the matrix.

$\|Y\|_N$  is derived by accumulating the function value  $f(\sigma_i(Y))$  of all singular values, while the number of nonzero singular values of all singular values is expressed as  $\text{rank}(Y)$ . Thus, the formula  $\lim_{\gamma \rightarrow 0} \|Y\|_N = \text{rank}(Y)$  is proved to be true.

- (2) It is assumed that  $(\cdot)^H$ ,  $(\cdot)^{-1}$ , and  $\lambda$ , respectively, denote the conjugate transpose, inverse, and eigenvalues of the matrix and  $I$  represents the unit matrix of the appropriate order. Given the characteristic polynomial,

$$\begin{aligned} |(UYV)^H UYV - \lambda I| &= |V^H Y^H U^H UYV - \lambda I| \\ &= |V^H Y^H YV - \lambda I| = |V^{-1} Y^H YV - \lambda I| = |Y^H Y - \lambda I|. \end{aligned} \quad (27)$$

Based on the derivation, it is suggested that the matrices  $(UYV)^H UYV$  and  $Y^H Y$  show the identical characteristic polynomial, so they achieve the same eigenvalue. Subsequently, the singular value of  $UYV$  and  $Y$  refers to the nonnegative square root of the eigenvalues  $(UYV)^H UYV$  and  $Y^H Y$ , respectively, and the singular value of  $UYV$  is equal to that of  $Y$ .

Thus,  $\|Y\|_N = \|UYV\|_N$  is true, and  $\|Y\|_N$  is unitarily invariant.

- (3) First, the eigenvalue of singular value is the non-negative square root, so  $\sigma_i(Y) \geq 0$ . Next,  $f(\sigma_i(Y)) \geq 0$ . Accordingly,

$$\|Y\|_N \geq 0. \quad (28)$$

Then,

$$f(\sigma_i(Y)) = 0 \iff Y = O. \quad (29)$$

So,

$$\|Y\|_N = 0 \implies Y = O. \quad (30)$$

End. □

**2.5. RPCA of New Nonconvex Rank Approximation.** As impacted by the limitations of PCA and RPCA, in the present study, a novel nonconvex rank approximation robust principal component analysis (NNCRPCA) was developed, capable of effectively dealing with the overall corrosion and exhibiting a good generalization ability. For any novel data damaged by gross error, the learning projection can effectively eliminate the possible damage by projecting the data into its low-rank subspace. As revealed from the results of considerable experiments performed on three static continuous image datasets and one dynamic video dataset, NNCRPCA is robust to serious corruption and can effectively process novel data.

In the low-rank matrix recovery model, the rank function  $\text{rank}(Y)$  is approximately replaced by New-norm of  $Y$  to obtain the following model:

$$\begin{cases} \min_{Y,E} & \|Y\|_N + \eta \|E\|_1 \\ \text{s.t.} & X = PY + E, \end{cases} \quad (31)$$

where  $X, Y, E \in R^{m \times n}$ , and  $\eta > 0$  is a positive compromise factor.

For equation (31), we can change it into an unconstrained optimization form by convex relaxation:

$$\min_{Y,E} \frac{1}{2} \|X - PY - E\|_F^2 + \eta_1 \|Y\|_N + \eta_2 \|E\|_1, \quad (32)$$

where  $\eta_1$  and  $\eta_2$  are nonnegative parameters used to balance data consistency.

For the unconstrained optimization problem (32), we can use the alternating direction method to solve large-scale optimization problems that are commonly encountered and dealt with in the field of machine learning.

Using the alternating direction method, in the  $(k+1)$ -th iteration, equation (32) requires solving the following two subproblems:

$$Y_{k+1} = \arg \min_Y \left\{ \eta_1 \|Y\|_N + \frac{1}{2} \|X - PY - E_k\|_F^2 \right\}, \quad (33)$$

$$E_{k+1} = \arg \min_E \left\{ \eta_2 \|E\|_1 + \frac{1}{2} \|X - PY_k - E\|_F^2 \right\}. \quad (34)$$

To solve equation (33), first of all, its objective function is the combination of nonconvex functions and convex functions is needed to be noticed: the first term  $\|Y\|_N$  is the nonconvex function, and the second term  $\|X - PY - E_k\|_F^2$  is the convex function. Thus, the solution can be obtained with the help of the difference of convex function (DC) [36]. The convex function algorithm proposed by Tao in Literature [37] refers to an algorithm based on duality theory and local optimality conditions of convex function optimization, and the convergence proof is presented in Literature [36].

**Lemma 1.** *The subdifferential of  $\|Y\|_N$ :*

$$\partial \|Y\|_N = \left\{ U \text{diag}(y) V^T : y_i = \frac{df(\sigma_i)}{d(\sigma_i)}, \quad i = 1, 2, \dots, \text{rank}(Y) \right\}, \quad (35)$$

where the columns of  $U$  and  $V$ , respectively, denote left and right singular vectors of  $Y$ , and the following formula is yielded from equation (25):

$$y_i = \frac{2\sigma_i(Y) + (1 - \gamma\sigma_i(Y))e^{-\gamma\sigma_i(Y)}}{(\sigma_i(Y) + e^{-\gamma\sigma_i(Y)})^2}. \quad (36)$$

According to Lemma 1, the  $(k+1)$ -th iteration of  $Y$  in equation (33) is written as

$$Y_{k+1} = \arg \min_Y \left\{ \mu_1 \langle \partial \|Y_k\|_N, Y \rangle + \frac{1}{2} \|X - PY - E_k\|_F^2 \right\}. \quad (37)$$

In equation (37), the partial derivative of the objective function is with respect to  $Y$ , and then it is set as zero, so the following formula can be obtained:

$$Y_{k+1} = (P^H P)^{-1} (P^H X - P^H P E_k - \mu_1 \partial \|Y_k\|_N). \quad (38)$$

Equation (38) refers to the iterative formula for solving  $Y$ .

Next, the solution of  $E$  is to be discussed. Soft thresholding of singular values of  $E$  can be adopted iteratively. Accordingly, the  $(k+1)$ -th iteration in equation (34) can be expressed as

$$E_{k+1} = \mu_2 (X_k - Y_k), \quad (39)$$

where  $X_0 = P^H X$  and matrix  $X_k = Y_k + E_k$ .

The iterative formula of matrix  $X$  is as follows:

$$X_{k+1} = Y_{k+1} + E_{k+1} - P^H (X - PY_{k+1} - E_{k+1}). \quad (40)$$

In equation (40), to keep the data consistency, the novel  $X_{k+1}$  is derived by subtracting the residuals  $P^H (X - PY_{k+1} - E_{k+1})$  from  $Y_{k+1} + E_{k+1}$ .

Input data: initial data  $X$ ;  
 $P$  dictionary matrix;  
 $\mu_1, \mu_2$  thresholds.  
Initialization:  $Y_0 = X_0 = P^H X, E_0 = O, k = 0$ .  
Output data:  $Y_{k+1}, E_{k+1}$   
Iteration steps:  
(1) Update variables  $Y_{k+1}$ : according to the following formula  $Y_{k+1} = (P^H P)^{-1} (P^H X - P^H P E_k - \mu_1 \partial \|Y_k\|_N)$   
(2) Update variables  $E_{k+1}$ : according to the following formula  $E_{k+1} = \mu_2 (X_k - Y_k)$   
(3) Update variables  $X_{k+1}$ : according to the following formula  $X_{k+1} = Y_{k+1} + E_{k+1} - P^H (X - A Y_{k+1} - E_{k+1})$   
(4) Stopping condition:  $(\|Y_{k+1} + E_{k+1} - (Y_k + E_k)\|_F) / \|(Y_k + E_k)\|_F \leq 2.5 \times 10^{-3}$   
Repeat Steps 1–3; otherwise, stop iteration and return to  $Y_{k+1}, E_{k+1}$ .

ALGORITHM 1: Solve the alternating direction method of equation (32).

The mentioned solution process is summarized in Algorithm 1.

**2.6. Algorithm Convergence.** In the present section, the convergence proof of the algorithm is provided.

**Lemma 2.** Set  $\{Y_k\}$  and  $\{E_k\}$  as the sequence generated by the algorithm, then  $\{Y_k\}$  and  $\{E_k\}$  are bounded.

*Proof.* First, the objective function of equation (32) is denoted as the following form:

$$\Psi(Y, E) = \frac{1}{2} \|X - PY - E\|_F^2 + \eta_1 \|Y\|_N + \eta_2 \|E\|_1. \quad (41)$$

Based on equation (33) and (34),

$$\begin{aligned} \Psi(Y_{k+1}, E_k) &\leq \Psi(Y_k, E_k), \\ \Psi(Y_{k+1}, E_{k+1}) &\leq \Psi(Y_{k+1}, E_k). \end{aligned} \quad (42)$$

Then,

$$\Psi(Y_{k+1}, E_{k+1}) \leq \Psi(Y_k, E_k). \quad (43)$$

Thus, the sequence  $\{\Psi(Y_k, E_k)\}$  is nonincreasing.

Next, it is verified that the sequences  $\{Y_k\}$  and  $\{E_k\}$  are bounded, and the boundedness of  $\{Y_k\}$  is subsequently proved by the proof by contradiction.

Assuming  $\{Y_k\}$  is unbounded, then for any real number  $M > 0$ , there is always a positive integer  $N$  that satisfies  $\|Y_k\| \geq M$  for  $k \geq N$  and any norm  $\|\cdot\|$ , i.e.,  $\|Y_k\| \rightarrow \infty$ . Accordingly, the first term of equation (41) is  $(1/2) \|X - PY - E\|_F^2 \rightarrow \infty$ .

As the second term  $\|Y\|_N$  in equation (41) is a pseudonorm, next, we will discuss the boundedness of this term under the following two conditions:

- (1) Under unbounded  $\{Y_k\}$ , if  $\sigma_i(Y_k)$  is unbounded, according to equation (24), the following formula can be obtained:

$$\lim_{\sigma_i(Y_k) \rightarrow \infty} \frac{\sigma_i(Y_k)}{\sigma_i(Y_k) + e^{-\gamma \sigma_i(Y_k)}} = 1. \quad (44)$$

In a review of the nonconvex rank approximation, to better approximate the rank function, the value of  $\gamma$  approaches  $\infty$ , so it is suggested that it is bounded in this case.

- (2) When  $\{Y_k\}$  is unbounded, if  $\sigma_i(Y_k)$  is bounded, then according to equation (24), it is obvious that, in this case,  $\|Y\|_N$  is also bounded.

In conclusion,  $\|Y\|_N$  is always bounded.

Consistent with the proof method in  $\{Y_k\}$ , assuming  $\{E_k\}$  is unbounded, the first and third terms in equation (41) are then unbounded as follows:

$\|E\|_1$  and  $(1/2) \|X - PY - E\|_F^2$  are unbounded.

In brief, both  $\{Y_k\}$  and  $\{E_k\}$  are unbounded, so  $\{\Psi(Y_k, E_k)\}$  is unbounded. Combined with its nonnegative property, it is suggested that  $\Psi(Y_k, E_k) \rightarrow \infty$ , which is not consistent with the conclusion in equation (43) that the sequence  $\{\Psi(Y_k, E_k)\}$  is nonincremental. Accordingly, the sequences  $\{Y_k\}$  and  $\{E_k\}$  are bounded.

Based on the mentioned lemma, the below theorem can be developed.  $\square$

**Theorem 3.** Assuming that  $\{Y_k\}$  and  $\{E_k\}$  are the sequences generated by the algorithm, then any clustering point of sequence  $\{(Y_k, E_k)\}$  is a local minimum point of equation (32).

*Proof.* Based on Lemma 2, it can be known that  $\{(Y_k, E_k)\}$  is bounded. So  $\{(Y_k, E_k)\}$  has at least one gathering point, i.e.,  $(Y^*, E^*)$ . Combining with the conclusion that  $\{\Psi(Y_k, E_k)\}$  is a nonincremental sequence in equation (43), it can be known that there is  $\lim_{k \rightarrow \infty} \Psi(Y_k, E_k)$ , and the following formula is true:

$$\lim_{k \rightarrow \infty} \Psi(Y_k, E_k) = \Psi(Y^*, E^*). \quad (45)$$

Thus,  $\{(Y^*, E^*)\}$  refers to the local minimum point of problem (32).  $\square$

### 3. Results and Discussion

In the present study, using Wallflower (<https://www.microsoft.com/en-us/research/people/jckrumm/downloads/>) and SBM-net (<http://www.scenebackgroundmodeling.net/>), the novel model was adopted to solve background separation problem;

subsequently, it was compared with existing GoDec [38], ALM, RegL1-ALM [24], NCRPCA, GOSUS [19], and SLM [22] methods. Video background separation, a process to remove foreground and separate background from video shot with a fixed camera, has been extensively employed in computer moving target detection. In the present study, the experimental operating environment is Intel(R) Core™ i7-8565U CPU @1.80ghz 2.00ghz and memory 16.00 GB. All algorithms are written by MATLAB R2018b.

**3.1. Parameter Analysis.** The novel algorithm was tested and analysed on Wallflower and SBMnet, respectively. By referencing the stop criterion of the four comparison algorithms, the relative error value  $t$  of this algorithm in the Wallflower background separation experiment reaches  $2.5 \times 10^{-3}$  and  $4 \times 10^{-4}$  in the SBMnet background separation experiment, which is expressed as

$$RE = \frac{\|Y_{k+1} + E_{k+1} - (Y_k + E_k)\|_F}{\|(Y_k + E_k)\|_F} \leq t. \quad (46)$$

For Wallflower, the effect of the parameter values on the experimental effect was analysed. To be specific, a common quantitative index, i.e., root-mean-square error (RMSE), was adopted as the measurement standard, which is defined as follows:

$$RMSE = \frac{\|X - Y - E\|_F}{\|X\|_F}, \quad (47)$$

where  $X$ ,  $Y$ , and  $E$ , respectively, denote the original observation data matrix, low-rank matrix, and sparse matrix. The value of  $P$  encoding operator refers to the unit matrix.

Next, with Wallflower as an example, the relationship between the value of parameter  $\gamma$  and RMSE is illustrated. The initial value  $\gamma = 0.1$ , step size = 0.01, and then stops when  $\gamma = 10$ .

- (1) When  $\gamma \in [4, 10]$ , the value of RMSE approaches 1, suggesting that the algorithm cannot recover low-rank matrix  $Y$  from data  $X$  at this time, i.e., the experimental effect is not ideal
- (2) When  $\gamma \in [4, 10]$ , the value of RMSE maintains stably between 0.5 and 0.9, and meantime, the experimental effect of the algorithm is ideal

Accordingly, a conclusion was also drawn that the larger the  $\gamma$ , the better the approximation degree of the approximation rank, whereas the image reconstruction effect of the algorithm in the actual implementation is unsatisfactory. Among the experiments on Wallflower, when the RMSE value is the smallest, and the rank is approximately the optimal,  $\gamma = 2$ .

The statistical information of Wallflower exploited in the experiment is listed in Table 2 (e.g., image dimension, image frame number, and dimension of stretch matrix).

The statistical information of SBMnet exploited in the experiment is listed in Table 3 (e.g., image dimension, image frame number, and dimension of stretch matrix).

TABLE 2: Statistical information of Wallflower applied in the present tests.

Datasets	Dimensions	Frames	Dimension of stretch matrix
Bootstrap	160 × 120	3055	19200 × 3055
Camouflage	160 × 120	353	19200 × 353
Lightswitch	160 × 120	2715	19200 × 2715
Waving trees	160 × 120	287	19200 × 287
Time of day	160 × 120	5890	19200 × 5890
Moved object	160 × 120	1745	19200 × 1745
Foreground aperture	160 × 120	2113	19200 × 2113

**3.2. Experimental Results.** Besides the experimental renderings, the quantitative indexes to test the algorithm consist of calculation time, iteration times, and other common indexes. Moreover, since the premise of the low-rank matrix recovery model adopted for image reconstruction is that the matrix corresponded by the background part is low rank, the rank of matrix  $Y$  recovered by the algorithm should be maximally small, and the most ideal result is that the rank is 1. By equation (33), the RMSE was calculated after applying various algorithms to Wallflower and SBMnet, as an attempt to indicate the background separation effect. Besides the mentioned indexes, the RE is another vital index to reflect the accuracy of the algorithm.

Let us start with Wallflower first. Many challenges are placed in the dataset to evaluate the performance of the algorithms. The Bootstrap refers to a scene of taking food from a restaurant. The characters in the scene are constantly updated with frequent activities, and there exist considerable characters. The Camouflage refers to a scene of electronic computer screen. When recording the Camouflage dataset, the computer screen refreshes at a certain frequency, and someone flashes by in front of the computer screen, so the background is difficult to separate. In LightSwitch, both the electronic computer refreshing, and the behaviour process of characters are identified; meantime, the transformation of lights is added in the video shooting process. The Waving Trees refers to a scene in which a man passes through the shaking trees. The Foreground Aperture refers to a scene of one person sleeping on the table in a low-light environment. The Move Object refers to a scene in which a person enters the classroom, picks up the phone to talk, and leaves. The Time of Day refers to a scene in which a living room passes from night to day and then from day to night. We used many algorithms (i. e. GoDec, ALM, RegL1-ALM, NCRPCA, GOSUS, SLM, and NNCRPCA) to do background separation experiments on these datasets and evaluated their performance. Table 4 lists the algorithms' experimental quantitative index for Wallflower, and Figure 3 illustrates the algorithms' rendering for the background separation experiment.

Table 4 indicates that the performance of the algorithms mentioned above is quite different in the background separation experiment. The earlier algorithms, ALM and RegL1-ALM, have higher iteration times and more running time, but they own better relative error and RMSE than

TABLE 3: Statistical information of SBMnet applied in the present tests.

Category name	Challenges	Frames	Dimension of stretch matrix
Highway	Cars moving	1700	$320 \times 240 \times 1700$
Bus station	Background objects moving away, objects stopping for a short while, and then moving away	617	$360 \times 240 \times 617$
Group campus	Clutter videos	300	$320 \times 240 \times 300$
Camera parameter	Videos with strong illumination changes	451	$320 \times 240 \times 451$
Advertisement board	Videos exhibiting dynamic background motion	598	$504 \times 336 \times 598$
Bus stop morning	Videos containing more than 3,500 frames	3766	$320 \times 240 \times 3766$
Toscana	Videos containing a limited number of frames (less than 20) with a very low framerate	6	$800 \times 600 \times 6$



FIGURE 3: Decomposition of  $Y + E$  in Wallflower.

TABLE 4: The algorithm's quantity indexes for Wallflower experiment.

Algorithms	Evaluation indexes	BS	CF	LS	WT	TOD	MO	FA
GoDec [38]	RE	$2.325e-03$	$3.625e-03$	$2.402e-05$	$1.0996e-04$	$2.5267e-03$	$6.7650e-04$	$8.2345e-03$
	Iteration	20	22	43	12	89	30	21
	Rank (Y)	13	3	4	8	16	7	5
	RMSE	0.6213	0.7112	0.6322	0.5546	0.7032	0.6505	0.5890
	CPU time(s)	28.6702	3.5344	19.2457	<b>0.66918</b>	15.4668	<b>4.3642</b>	6.2168
ALM	RE	$3.5567e-04$	$3.5567e-06$	$1.4453e-05$	$2.2310e-06$	$2.2121e-04$	$2.2223e-05$	$2.3512e-05$
	Iteration	<u>178</u>	<u>149</u>	<u>154</u>	23	204	<u>56</u>	<u>104</u>
	Rank (Y)	<u>663</u>	42	<u>26</u>	<u>8</u>	32	13	56
	RMSE	<u>0.6534</u>	<u>0.7243</u>	0.6231	0.5043	0.6439	<u>0.7230</u>	0.5546
	CPU time(s)	<u>241.5925</u>	<u>34.5612</u>	<u>46.7471</u>	18.5634	568.3478	167.2456	356.2323
RegL1-ALM [24]	RE	$2.426e-05$	$1.390e-05$	$1.255e-05$	$3.5624e-07$	$3.4011e-04$	$2.5144e-07$	$2.8995e-07$
	Iteration	144	147	154	<u>55</u>	<u>240</u>	40	48
	Rank (Y)	589	<u>237</u>	16	4	<u>98</u>	<u>983</u>	<u>1153</u>
	RMSE	0.5937	0.6894	0.6025	0.5523	0.5722	0.6332	0.6534
	CPU time(s)	231.5002	28.2874	24.64	<u>20.5867</u>	430.5623	<u>205.8264</u>	284.3467
NCRPCA	RE	$1.0996e-07$	$1.2994e-07$	$2.1757e-07$	$3.6489e-08$	$5.3841e-07$	$2.3394e-07$	$2.2697e-07$
	Iteration	36	23	30	20	48	36	40
	Rank (Y)	8	3	4	1	2	3	4
	RMSE	0.5546	0.6943	0.6001	0.5014	0.5421	0.6522	0.5867
	CPU time(s)	83.8376	1.4279	39.7935	1.0232	128.5847	22.719	54.980
GOSUS [19]	RE	<u><math>7.2312e-02</math></u>	<u><math>5.2562e-03</math></u>	<u><math>4.2231e-02</math></u>	<u><math>3.231e-01</math></u>	<u><math>3.2548e-00</math></u>	<u><math>4.4423e-03</math></u>	<u><math>2.4323e-02</math></u>
	Iteration	<b>10</b>	<b>10</b>	10	7	12	10	<b>10</b>
	Rank (Y)	6	5	6	3	18	5	6
	RMSE	0.5134	0.7356	<u>0.6654</u>	<u>0.6732</u>	<u>0.7622</u>	0.5821	0.6521
	CPU time(s)	<b>2.2312</b>	2.8270	<b>1.2432</b>	1.4058	13.7578	4.6285	<b>5.1747</b>
SLM [22]	RE	$9.6320e-08$	<b><math>6.2156e-08</math></b>	$7.5632e-09$	$6.5432e-08$	$4.5621e-08$	$7.3230e-08$	$2.3120e-08$
	Iteration	27	16	38	14	45	17	32
	Rank (Y)	14	5	8	2	3	2	<b>2</b>
	RMSE	<b>0.5015</b>	0.7283	<b>0.5213</b>	0.5523	<b>0.5018</b>	0.5689	0.5590
	CPU time(s)	29.2312	<b>2.3210</b>	10.2354	3.3421	38.2654	13.6218	24.6654
NNCRPCA	RE	<b><math>9.990e-08</math></b>	$7.6623e-07$	<b><math>3.2243e-09</math></b>	$2.5643e-08$	<b><math>3.5478e-09</math></b>	$2.7865e-08$	$3.6542e-09$
	Iteration	24	16	<b>28</b>	6	<b>22</b>	<b>10</b>	14
	Rank (Y)	<b>6</b>	<b>4</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>2</b>	3
	RMSE	0.5032	<b>0.6650</b>	<b>0.5976</b>	<b>0.5000</b>	<b>0.5032</b>	<b>0.5034</b>	<b>0.5530</b>
	CPU time(s)	51.4532	2.5426	14.6745	4.2214	48.5523	23.2246	35.6240

BS: Bootstrap; CF: Camouflage; LS: LightSwitch; WT: Waving Trees; TOD: Time of Day; MO: Moved Object; FA: Foreground Aperture. The bold represents the minimum value of the same index in the same subject experiment, and the underline represents the maximum value.

GoDec and GOSUS. GoDec and GOSUS are superior in execution time and have fewer iterations. SLM has better performance in all aspects of background separation experiment, with less iterations and less execution time. NCRPCA is inferior to SLM in execution time and iteration. Although NNCRPCA is worse than SLM in execution time, it is better than SLM in other aspects. RMSE and relative error are kept with a good level; especially the rank of tensor matrix of dataset is kept at a low level, which is lower than other algorithms. Figure 3 shows the processing effect of background separation by different algorithms, and it is suggested that the background calculated by the NNCRPCA algorithm is clearer.

In real life, influenced by the variations in quantity, lighting, weather, and other conditions, video background

changes constantly. The rank of the background matrix reaches over 1, and the separation process is relatively complex, which is beyond the reach of ordinary algorithms. In the following section, the approximate model proposed in the present study was applied in SBMnet, and the feasibility of the model was verified by numerical experiments.

In the SBMnet dataset, many challenging scenarios which were illustrated in Table 3 were set up and imposed great challenges on the calculation of the algorithm. Through many experiments, we can observe how the backgrounds of scenarios could be separated by the above algorithms and the performance of the algorithms verified again. The algorithms' background separation experimental quantitative index for SBMnet is listed in Table 5, and Figure 4 shows the

TABLE 5: The algorithm's background separation experiment number index of SBMnet.

Algorithms	Evaluation indexes	AB	BS	CP	GC	HH	To
GoDec [38]	RE	$3.6720e-05$	$3.6753e-05$	$7.6534e-04$	$9.6543e-05$	$0.5632e-04$	$0.5634e-03$
	Iteration	38	86	62	18	99	6
	Rank (Y)	57	68	52	26	89	4
	RMSE	0.6621	0.5793	0.6735	0.6756	0.5987	0.7624
	CPU time(s)	<b>13.5822</b>	7.146	<b>4.3845</b>	<b>2.7945</b>	17.9202	0.32478
ALM	RE	$3.8976e-06$	$9.6735e-06$	$3.7865e-06$	$3.8732e-06$	$4.7650e-06$	$8.6723e-06$
	Iteration	<u>183</u>	<u>267</u>	<u>206</u>	<u>49</u>	<u>667</u>	<u>32</u>
	Rank (Y)	<u>335</u>	<u>438</u>	<u>310</u>	<u>169</u>	<u>1080</u>	<u>6</u>
	RMSE	0.6723	0.6889	<u>0.7623</u>	0.6724	0.6521	0.6617
	CPU time(s)	<u>2819.9870</u>	<u>5816.2657</u>	<u>4890.2370</u>	<u>872.6000</u>	<u>7623.8876</u>	<u>520.2010</u>
RegL1-ALM [24]	RE	$6.582e-06$	$7.4981e-06$	$4.159e-06$	$3.4045e-06$	$7.5781e-08$	$4.7509e-07$
	Iteration	36	38	35	31	42	31
	Rank (Y)	296	321	244	164	953	3
	RMSE	<u>0.7689</u>	0.6229	0.6281	0.5225	0.6790	0.6623
	CPU time(s)	240.8317	111.1153	63.863	39.3734	458.8552	2.4931
NCRPCA	RE	$6.9508e-10$	$2.0525e-09$	$2.128e-09$	$1.5526e-10$	$7.9364e-10$	$1.8266e-11$
	Iteration	<b>33</b>	<b>32</b>	28	26	46	20
	Rank (Y)	2	5	7	13	22	3
	RMSE	0.6022	0.5187	0.5965	0.5562	0.7217	<u>0.7893</u>
	CPU time(s)	30.0727	15.4769	8.8945	5.0345	82.1707	<u>0.28524</u>
GOSUS [19]	RE	<u><math>3.45e-01</math></u>	<u><math>1.69e-01</math></u>	<u><math>2.32e+00</math></u>	<u><math>3.05e-01</math></u>	<u><math>2.71e-01</math></u>	NaN
	Iteration	<b>12</b>	<b>16</b>	<b>15</b>	<b>10</b>	<b>16</b>	NaN
	Rank (Y)	237	308	209	153	389	NaN
	RMSE	0.6832	<u>0.7390</u>	0.6899	<u>0.7023</u>	<u>0.7621</u>	NaN
	CPU time(s)	18.298	<b>6.915</b>	4.6814	3.7101	<b>14.7244</b>	NaN
SLM [22]	RE	$3.8908e-08$	$7.6754e-09$	$8.6543e-08$	$8.3276e-07$	$8.3326e-08$	$0.6732e-07$
	Iteration	37	49	42	36	42	17
	Rank (Y)	48	59	43	32	69	4
	RMSE	0.5622	<b>0.5016</b>	0.6523	<b>0.5286</b>	0.6317	0.7701
	CPU time(s)	23.3612	48.3233	40.6398	42.1232	46.2219	3.2134
NNCRPCA	RE	<b><math>5.7823e-10</math></b>	<b><math>7.6765e-11</math></b>	<b><math>3.6745e-11</math></b>	<b><math>5.7628e-12</math></b>	<b><math>6.3224e-10</math></b>	<b><math>9.7862e-12</math></b>
	Iteration	38	42	28	20	36	13
	Rank (Y)	<b>2</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>18</b>	<b>3</b>
	RMSE	<b>0.5521</b>	0.5032	<b>0.5066</b>	0.5489	<b>0.6213</b>	<b>0.7230</b>
	CPU time(s)	28.3789	56.2231	47.3897	49.2238	52.5487	<b>6.1020</b>

AB: Advertisement Board; BS: Bus Station; CP: Camera Parameter; GC: Group Campus; HH: highway; To: Toscana.

algorithm's rendering for the background separation. The Toscana dataset has only 6 frames, which makes GOSUS unable to run.

According to Table 5, the performance of NNCRPCA is better than other algorithms. In terms of relative error and RMSE, the results of NNCRPCA are lower than or close to those of SLM, but lower than those of other algorithms. NNCRPCA is still lower in rank approximation than other algorithms, including SLM, but its CPU time is still inferior to some algorithms. Figure 4 shows the processing effect of background separation by different algorithms, and it is suggested that the background calculated by the NNCRPCA algorithm is clearer.

Tables 4 and 5 suggests that the NNCRPCA algorithm achieves prominent experimental results for RE and RMSE in

the background separation application. The NNCRPCA algorithm modifies the existing convex approximation that employs nuclear norm as the low-rank matrix and adopts a more accurate approximation method of nonconvex rank approximation function. In the application of background separation, its effect is higher than that of nuclear norm, and it is significantly improved in time and rank. Besides, the improvement of the RMSE and relative error is also the result of adopting a better rank approximation. A low rank is achieved by the NNCRPCA algorithm, demonstrating that the low-rank matrix to extract the background part of the image is more accurate, so the sparse part extraction of the experiment is more complete. As revealed from Figures 3 and 4, the low-rank part extracted by our algorithm is clearer, and the information of the sparse part is more complete.



FIGURE 4: Background separation of SBMnet.

### 4. Conclusions

A novel algorithm for robust principal component analysis of nonconvex approximation rank function was proposed. Unlike the previous models, a novel nonconvex approximation rank function acted as the rank approximation of the low-rank matrix. Subsequently, the robust principal component analysis based on this nonconvex rank approximation function was established, and the NNCRPCA algorithm based on the alternating direction method was developed. Next, the numerical experiments were performed on Wallflower as well as SBMnet. From the perspective of image effect, as well as the main quantitative indexes (e.g., running time, iteration times, RMSE, and RE), the algorithm effectively improved the image background separation effect. Our major contributions are revealed in the following three aspects:

- (1) A novel nonconvex rank approximation function was constructed. Different from the conventional

methods in which the nuclear norm acted as convex approximation, the new rank approximation function approximated the rank of the low-rank matrix very well, the construction and solution of the novel algorithm for robust principal component analysis were achieved, and the convergence of the algorithm was proved theoretically.

- (2) It is combined with the unconstrained problem of alternating direction method and convex function after solving the differential of convex relaxation. Since the objective function of the optimization problem to be solved is the combination of convex function and nonconvex function, the differential of convex function refers to an effective technique to solve this type of optimization problems.
- (3) The low-rank matrix recovery model was applied for background separation, and the experimental results assessed the application prospect of applying the nonconvex approximation method in

this problem. Moreover, in terms of data processing mode, the background separation done in the present study requires further studies. With the continuous increase in the data size in the video dataset, the model and algorithm proposed in this study will exhibit a broader application prospect.

## Data Availability

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

## Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Acknowledgments

Some of the authors of this publication are also working on these related projects: (1) Higher Vocational Education Teaching Fusion Production Integration Platform Construction Projects of Jiangsu Province under grant no. 2019(26); (2) Natural Science Fund of Jiangsu Province under grant no. BK20131097; (3) “Qin Lan Project” teaching team in Colleges and Universities of Jiangsu Province under grant no. 2017(15); (4) High Level of Jiangsu Province Key Construction Project funding under grant no. 2017(17). Our research was supported by “Geometry problem geometry” project (the National Natural Science Foundation of China, 61073086).

## References

- [1] D. Xia, “Confidence Region of singular subspaces for Low-Rank matrix regression,” *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7437–7459, 2019.
- [2] H. Yuan, J. Li, and L. L. Lai, “Low-rank matrix regression for image feature extraction and feature selection,” *Information Sciences*, vol. 522, pp. 214–226, 2020.
- [3] P. Wang, C. D. Lin, C. Lin, X. Yang, and S. Xiong, “Low-rank and sparse matrix recovery from noisy observations via 3-block Admm algorithm,” *Journal of Applied Analysis & Computation*, vol. 10, no. 3, pp. 1024–1037, 2020.
- [4] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [5] I. Jolliffe, *Principal Component Analysis*, Springer, Berlin, Germany, 1986.
- [6] D. Efimova, *Encyclopedia of Social Network Analysis and Mining*, Springer, Berlin, Germany, 2014.
- [7] V. Chandrasekaran, S. Sanghavi, P. Parrilo et al., “Sparse and low-rank matrix decompositions,” in *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 962–967, IEEE, Monticello, IL, USA, September 2009.
- [8] J. Wright, A. Ganesh, S. Rao et al., “Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization,” in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, pp. 2080–2088, Vancouver, Canada, December 2009.
- [9] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [10] D. L. Donoho, “For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.
- [11] Q. Sun, S. Xiang, and J. Ye, “Robust principal component analysis via capped norms,” in *Proceeding 19th ACM SIGKDD International Conference on Knowledge Discovery Data Mining (KDD)*, pp. 311–319, Chicago, IL, USA, August 2013.
- [12] C. Lu, J. Tang, S. Yan, and Z. Lin, “Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm,” *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 829–839, 2016.
- [13] C. Gao, N. Wang, Q. Yu et al., “A feasible nonconvex relaxation approach to feature selection,” in *Proceeding 25th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 356–361, San Francisco, CA, USA, August 2011.
- [14] J. Fan and R. Li, “Variable selection via nonconcave penalized likelihood and its oracle properties,” *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [15] Z. Kang, C. Peng, J. Cheng et al., “Lodged rank minimization with application to subspace clustering,” *Computational Intelligence and Neuroscience*, vol. 68, p. 2015, 2015.
- [16] Z. Kang, C. Peng, and Q. Cheng, “Robust PCA via nonconvex rank approximation,” in *Proceeding IEEE International Conference on Data Mining (ICDM)*, IEEE, Atlantic City, NJ, USA, pp. 211–220, November 2015.
- [17] Y. Xie, Y. Qu, D. Tao et al., “Hyperspectral image restoration via iteratively regularized weighted Schatten  $p$ -norm minimization,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4642–4659, 2016.
- [18] H. Yu and C. Miao, “General thresholding representation for  $\ell_p$  ( $0 < p < 1$ ) regularization,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5455–5468, 2015.
- [19] J. Xu, V. Ithapu, L. Mukherjee, J. M. Rehg, and V. Singh, “GOSUS: Grassmannian online subspace updates with structured-sparsity,” in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, pp. 3376–3383, Sydney, Australia, December 2013.
- [20] Z. Gao, L. F. Cheong, and Y. X. Wang, “Block-sparse RPCA for salient motion detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 10, pp. 1975–1987, 2014.
- [21] X. Liu, G. Zhao, J. Yao, and S. K. Jung, “Background subtraction based on low-rank and structured sparse decomposition,” *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2502–2514, 2015.
- [22] S. Javed, L.-F. Cheong, Y.-X. Wang et al., “Spatiotemporal low-rank modeling for complex scene background initialization,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1315–1329, 2018.
- [23] T. Zhang, “Analysis of multi-stage convex relaxation for sparse regularization,” *Journal of Machine Learning Research*, vol. 11, pp. 1081–1107, 2010.
- [24] G. Liu, Z. Lin, S. Yan et al., “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.

- [25] J.-F. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, Cambridge, MA, USA, 1982.
- [26] J. Cai, E. J. Candès, Z. Shen, and Y. Wang, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [27] J. Yang, W. Yin, Y. Zhang, and Z. Lin, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [28] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Transactions on Image Processing*, vol. 4, no. 7, pp. 932–946, 1995.
- [29] C.-H. Zhang, "Hyperspectral image denoising with cubic total variation model," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 7, pp. 95–98, 2012.
- [30] C. H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, pp. 894–942, 2010.
- [31] J. Trzasko and A. Manduca, "Highly under sampled magnetic resonance image reconstruction via homotopic 0-minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.
- [32] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *Journal of Machine Learning Research*, vol. 11, pp. 1081–1107, 2010.
- [33] H. Othman and S. E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 397–408, 2006.
- [34] J. H. Friedman, "Fast sparse regression and classification," *International Journal of Forecasting*, vol. 28, no. 3, pp. 722–738, 2012.
- [35] C. Y. Lu, J. H. Tang, S. C. Yan et al., "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear floNn," *IEEE Transactions on Image Processing*, vol. 2, no. 25, pp. 829–839, 2016.
- [36] P. D. Tao and H. A. Thi, "Convex analysis approach to dc programming: theory, algorithms and applications," *Acta Mathematica Vietnamica*, vol. 22, no. 1, pp. 289–355, 1997.
- [37] P. D. Tao, "Duality in dc (difference of convex functions) optimization. Sub gradient methods," *Trends in Mathematical Optimization*, pp. 277–293, Birkhäuser, Basel, Switzerland, 1998.
- [38] T. Zhou and D. Tao, "Godec: randomized low-rank & sparse matrix decomposition in noisy case," in *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, WA, USA, January 2011.