“Low-rank + Dual” Model Based Dimensionality Reduction

Si-Qi Wang*, Xiang-Chu Feng, Wei-Wei Wang

School of Mathematics and Statistics, Xidian University, Xi'an, China, 710126.

Abstract

We propose a novel “low-rank + dual” model for the matrix decomposition problems. Based on the unitarily invariant property of the Schatten p -norm, we prove that the solution of the proposed model can be obtained by an “ $l_\infty + l_1$ ” minimization problem, thus a simple and fast algorithm can be provide to solve our new model. Furthermore, we find that applying “ $l_\infty + l_1$ ” to any vector can achieve a shifty threshold on the values. Experiments on the simulation data, the real surveillance video database and the Yale B database prove the proposed method to outperform the state-of-the-art techniques.

Keywords: Dimensionality reduction, Background modeling, Singular value decomposition, Thresholding method, l_p -minimization problem

1. Introduction

Dimensionality reduction is an extraordinary important tool for processing the high-dimensional data, such as video, image, audio and bio-medical signals. A well-known technique for dimension reduction is the low-rank approximation method, which is actually solving the rank minimization problems (RMP) [1, 2, 3, 4]. And recently, RMP attract more and more attentions in many fields of science and engineering, such as computer vision [5, 6, 7, 8, 9, 10], machine learning [7, 8, 11, 12, 13], and signal and image processing [2, 5, 8, 9, 14]. Due to the effectiveness of the rank minimization, many researchers have proposed various algorithms to solve the problem. An important classical model is reduced-rank regression, which limits the coefficient matrix to be low-rank [15, 16, 17]. A nuclear norm penalized least squares estimator has been proposed by Yuan [18], and the estimator encourages the singular values sparsity and the simultaneous rank reduction [19, 20]. Rohde & Tsybakov investigate the Schatten- q quasi-norm penalty and non-asymptotic bounds of prediction risk [21].

Reduced-rank is a very effective dimension reduction assumption, which connections with many popular tools including principal component analysis (PCA) and canonical correlation analysis (CCA), and it is

*Corresponding author, tel. (8629)88202860

Email addresses: pearlwangxd@163.com (Si-Qi Wang), xcfeng@mail.xidian.edu.cn (Xiang-Chu Feng), wwwang@mail.xidian.edu.cn (Wei-Wei Wang)

extensively studied in matrix completion problems [5, 17, 22, 23, 24]. Actually, the rank function and nuclear norm penalized methods in robust PCA (RPCA) problems can be respectively viewed as l_0 and l_1 penalized methods in the singular value decomposition (SVD) domain respectively. Therefore, the general strategy is to solve the nuclear norm minimization problem (NNMP) instead of the RMP, but the solution to NNMP suffers from the high computation cost of multiple SVDs.

In this paper, a “low-rank + dual” model is formalized from the viewpoint of dimension complexity reduction, and an efficient iterative update algorithm is designed to solve the model. It is necessary to particularly mentioned that “dual” here refers to the duality between the norms in functional analysis [25], rather than the typical concept of dual problem in computational science. Dual norms to be used as regular items was originally proposed by Meyer in image decomposition problem [26]. A model with total variation (TV) norm and G-norm (the dual of the TV-norm) was introduced by Meyer, which can better capture the cartoon and texture part in the decomposition of image. Thus, we try to apply dual norms to the regularization terms in the background and foreground modeling problems. The novel model can limit the correlation between the components, and the unitarily invariant property of the Schatten p -norm ensure the model can be replaced by a simplified “ $l_\infty + l_1$ ” minimization problem.

In order to test the efficiency and effectiveness of the proposed model, we apply our method both in the simulation data and real database experiments. In the first simulation experiment, randomly generated square matrix, which is actually the sum of the low-rank component and the error interference component, will be used as the test data. We apply our proposed method and other five state-of-the-art methods on the simulation test data to recover the low-rank and the interference component. In the second simulation experiment, we test the performances of IALM, GoDec and our method in matrix completion tasks. We discard a few entries from a randomly generated low-rank matrix, and recover the whole matrix through three different methods. Background modeling and specularly removing are two challenging task for computer vision applications, which are the fundamental of the follow-up work, such as motion tracking [27] and feature extraction [28]. Therefore, in the real database experiments, we apply our model to the background-foreground modeling and specularly removing problems, and compare the performance of the proposed method with two existing widely used methods. Experimental results demonstrate that the proposed algorithm is superior to other algorithms in most cases.

The rest of the paper is organized as follows. First, we briefly review the “low-rank + sparse” model and present our new “low-rank + dual” model in the next section. In Section 3, we simplify the problem into the vector level and propose the relevant algorithm in details. Next in Section 4, we discuss the thresholding essence of our method, and compare the numerical results with other existing methods on some practical experiments. Finally, we give some conclusion remarks in Section 5.

2. Related works

2.1. “low-rank + sparse” model

Suppose $Y \in \mathbb{R}^{m \times n}$ is a large matrix whose columns are arranged by the data set of the observational images, the size of each image is $m_1 \times n_1 = m$ and the number of the image is n . Then the low-rank matrix approximation problem is to efficiently and accurately obtain the low-rank matrix X from the measurement Y which has been corrupted by the errors E , $X, E \in \mathbb{R}^{m \times n}$. The common “low-rank + sparse” model formulate as follows:

$$\begin{cases} \min_{X, E} \lambda \text{Rank}(X) + \|E\|_0 \\ \text{s.t.} \quad Y = X + E \end{cases}, \quad (1)$$

where $\text{Rank}(\cdot)$ represents the rank function of the desired matrix; $\|\cdot\|_0$ is the regularization term for promoting sparsity; and λ is a positive weighting parameter providing a trade-off between the sparse and low-rank components. Model (1) suggests to seek the lowest-rank matrix X subject to the sparse constraint E , we hope to exactly obtain the pair (X, E) that could have generated the data matrix Y . Nevertheless, due to the discrete of the rank function and the highly non-convex of the l_0 -norm, (1) is NP-hard and no efficient solution is known to its sub-problems.

An tractable approach for solving (1) is to relax the model into the following convex optimization problem:

$$\begin{cases} \min_{X, E} \lambda \|X\|_* + \|E\|_1 \\ \text{s.t.} \quad Y = X + E \end{cases}, \quad (2)$$

where $\|\cdot\|_*$ is the nuclear norm defined by the sum of all singular values, which is the convex hull of the matrix rank; and $\|\cdot\|_1$ is the l_1 -norm defined by the component-wise sum of absolute values of all entries, which is the relaxation of the l_0 -norm. Moreover, recent advances have proposed a variety of different methods to solve (2): the iterative thresholding approach [14], the accelerated proximal gradient approach [29], the exact and inexact augmented lagrange multiplier approaches [3], etc. However, almost all the available algorithms of the “low-rank + sparse” model are plagued by the complex calculation accompany with SVD.

2.2. “low-rank + dual” model

In this paper, we propose a novel “low-rank + dual” model:

$$\begin{cases} \min_{X, E} \lambda \|X\|_{\sigma, 1} + \|E\|_{\sigma, \infty} \\ \text{s.t.} \quad Y = X + E \end{cases}. \quad (3)$$

Here $\|\cdot\|_{\sigma, p}$ is the Schatten p -norm, which is actually defined as the l_p -norm of the singular value vector ¹. To illustrate the advantages of our model, we need to recall the definition of Schatten p -norm first.

¹Notice that this is not an equivalence conception for the singular value vector and the singular vectors - the former, a unique vector, means the vector be composed of the elements of the matrix singular values, while the latter refers to the columns of the unitary matrixes in the singular value decomposition.

Definition 1. If σ_i denote the singular values of the matrix $A \in \mathbb{R}^{m \times n} (m \geq n)$, then the Schatten p -norm can be defined as $\|A\|_{\sigma,p} = (\sum_{i=1}^n \sigma_i^p)^{1/p}$. All Schatten norms are sub-multiplicative and unitarily invariant, which means that $\|A\| = \|UAV^T\|$ holds for any matrix A and unitary matrices U, V .

From the definition, we can know that the most familiar case $p = 1$ of the Schatten norm is the sum of all singular values, which yields the nuclear norm and associates with low-rank. Moreover, $\|\cdot\|_{\sigma,1}$ and $\|\cdot\|_{\sigma,\infty}$ in (3) are dual to each other, then due to the property of dual norms, we should have $|\langle X, E \rangle| \leq \|X\|_{\sigma,1} \|E\|_{\sigma,\infty}$ holds for the matrix X and E . Thus, minimizing the sum of $\|X\|_{\sigma,1}$ and $\|E\|_{\sigma,\infty}$ should achieve a smaller inner product $|\langle X, E \rangle|$, which leads to the less correlated decomposition X, E of the data Y .

The varichange of the regularization term is inspired by duality theory applied to the image decomposition problems [26, 30]. Through such improvement, it has no longer need to limit the non-low-rank part of the matrix to be “sparse”, dual regularization terms in the model can directly lead to the decomposition of the data matrix with two less correlated components. More importantly, the particular norm constraint enables the novel “low-rank + dual” model to be solved in the vector field. A simple and fast algorithm of this model will be demonstrated both in the theoretical and experimental fields. Numerical experiment results show that our “low-rank + dual” model performance well in the matrix decomposition problems, and the time consumption of the corresponding algorithm is much less than other “low-rank + sparse” methods. The theoretical basis and the relevant algorithm will be provided in the following section.

3. Problem simplification and theoretical derivation

Our approach minimize the model (3) as follows, we substitute the equation constrain into the minimization function to get the following form

$$\min_X \|Y - X\|_{\sigma,\infty} + \lambda \|X\|_{\sigma,1}, \quad (4)$$

$Y \in \mathbb{R}^{m \times n}, m \geq n$. We confirm (4) as our original problem in this paper, and all the matrixes involved in the minimization problem now are not square matrix. In the following sections, we will show the simplification of the original problem in details.

3.1. Simplified to vector level

Let $Y = L\Sigma R^T$ be the singular value decomposition of the data matrix Y , where L, R are unitary matrix of size $m \times m, n \times n$ and satisfy $L^T L = L L^T = I, R^T R = R R^T = I$. Then

$$\|Y - X\|_{\sigma,\infty} + \lambda \|X\|_{\sigma,1} = \|L(\Sigma - L^T X R)R^T\|_{\sigma,\infty} + \lambda \|LL^T X R R^T\|_{\sigma,1}. \quad (5)$$

Due to the unitarily invariant of the Schatten norms, we have

$$\|L(\Sigma - L^T X R)R^T\|_{\sigma,\infty} + \lambda \|LL^T X R R^T\|_{\sigma,1} = \|\Sigma - L^T X R\|_{\sigma,\infty} + \lambda \|L^T X R\|_{\sigma,1}. \quad (6)$$

If we denote $\Sigma = \begin{pmatrix} \text{diag}(\sigma_i) \\ O \end{pmatrix}$ and $L^T X R = \tilde{X} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix}$, then (4) is simplified as

$$\min_X \left\| \begin{pmatrix} \text{diag}(\sigma_i) - \tilde{X}_1 \\ -\tilde{X}_2 \end{pmatrix} \right\|_{\sigma, \infty} + \lambda \left\| \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix} \right\|_{\sigma, 1}. \quad (7)$$

Here $\text{diag}(\sigma_i)$ means diagonal matrix with σ_i as the i -th diagonal element of the vector σ , and $\text{size}(\tilde{X}_1) = \text{size}(\text{diag}(\sigma_i))$, $\text{size}(\tilde{X}_2) = \text{size}(O)$.

Proposition 1. *Let $A \in \mathbb{R}^{n \times m}$, and let $B \in \mathbb{R}^{k \times l}$ be a sub-matrix of A . Then, for all $i = 1, 2, \dots, \min(k, l)$, $\sigma_i(B) \leq \sigma_i(A)$.*

Due to Proposition 1 (Fact 9.14.10 in [31]), we known that

$$\left\| \begin{pmatrix} \text{diag}(\sigma_i) - \tilde{X}_1 \\ -\tilde{X}_2 \end{pmatrix} \right\|_{\sigma, \infty} + \lambda \left\| \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{pmatrix} \right\|_{\sigma, 1} \geq \left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} + \lambda \left\| \tilde{X}_1 \right\|_{\sigma, 1}, \quad (8)$$

and the equality is hold if and only if $\tilde{X}_2 = O$. It means that if we want to get the minimization of the function, we should let $\tilde{X}_2 = O$ and rewrite the problem as

$$\min_{\tilde{X}_1} \left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} + \lambda \left\| \tilde{X}_1 \right\|_{\sigma, 1}. \quad (9)$$

Now the matrix involved are all square ones.

Let $\tilde{X}_1 = P D Q^T$, then

$$\left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} + \lambda \left\| \tilde{X}_1 \right\|_{\sigma, 1} = \left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} + \lambda \|D\|_{\sigma, 1}. \quad (10)$$

In order to further simplify, we need to give the following proposition (Theorem 3.3.16 in [32]).

Proposition 2. *Let $A, B \in \mathbb{R}^{n \times n}$, then the inequality $\sigma_{i+j-1}(A+B) \leq \sigma_i(A) + \sigma_j(B)$ hold for all $i, j = 1, 2, \dots, n$ and $i+j \leq n+1$.*

If we set $i = 1$, and $C = A + B$, then $\sigma_j(C) - \sigma_j(B) \leq \sigma_1(C - B)$, $j = 1, \dots, n$. Changing the order of B and C , we have $|\sigma_j(C) - \sigma_j(B)| \leq \sigma_1(C - B)$ holds for $j = 1, \dots, n$, which means that $\max |\sigma_j(C) - \sigma_j(B)| \leq \|C - B\|_{\sigma, \infty}$. Thus we have $\left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} \leq \left\| \text{diag}(\sigma_i) - D \right\|_{\sigma, \infty}$, then

$$\left\| \text{diag}(\sigma_i) - \tilde{X}_1 \right\|_{\sigma, \infty} + \lambda \|D\|_{\sigma, 1} \geq \left\| \text{diag}(\sigma_i) - D \right\|_{\sigma, \infty} + \lambda \|D\|_{\sigma, 1}, \quad (11)$$

and the equality is held if and only if $\tilde{X}_1 = D$ is a diagonal matrix. Therefore we have that $L^T X R = \tilde{X} = \begin{pmatrix} D \\ O \end{pmatrix}$, which means that the solution X to the problem (4) equals to $L \begin{pmatrix} D \\ O \end{pmatrix} R^T$. What's more, the problem (9) now becomes a diagonal one and can be rewritten with a form of vectors

$$\min_d J(d) \triangleq \|\sigma - d\|_{\infty} + \lambda \|d\|_1, \quad (12)$$

where $\sigma_i \geq \sigma_j \geq 0, d_i \geq d_j \geq 0$, if $i > j$.

3.2. $l_\infty + l_1$ minimization problem

Inspired by the split method and the alternating direction method in [33, 34, 35], we introduce a splitting algorithm to solve the problem (12). By adding an intermediate vector $\tilde{\sigma} = \sigma - d$, the problem (12) has been converted into the following constrained optimization problem:

$$\min_{\tilde{\sigma}, d} \|\tilde{\sigma}\|_\infty + \lambda \|d\|_1 \quad s.t. \tilde{\sigma} = \sigma - d. \quad (13)$$

Moreover, we can get an unconstrained form to (13) with the l_2 penalty term as follows

$$\arg \min_{\tilde{\sigma}, d} \|\tilde{\sigma}\|_\infty + \lambda \|d\|_1 + \frac{\mu}{2} \|\tilde{\sigma} - (\sigma - d)\|_2^2, \quad (14)$$

where μ is the positive trade-off parameters. It is very important to note that $\|\cdot\|_2^2$ is the l_2 -norm of the vector, which is different from the spectral norm for the matrix. In order to satisfy the constraint condition, μ must take a very large value that may cause the stability problems in the calculation. Thus we need to define the Bregman distance of the convex functional (14) as:

$$D_E^p(\tilde{\sigma}, d, \tilde{\sigma}_k, d_k) = E(\tilde{\sigma}, d) - E(\tilde{\sigma}^k, d^k) - \langle p_{\tilde{\sigma}}^k, \tilde{\sigma} - \tilde{\sigma}_k \rangle - \langle p_d^k, d - d_k \rangle, \quad (15)$$

where $E(\tilde{\sigma}, d) = \|\tilde{\sigma}\|_\infty + \lambda \|d\|_1$, $p_{\tilde{\sigma}}^k$ and p_d^k are the sub-gradient of E with respect to the variables $\tilde{\sigma}$ and d . Rather than solving the problem (14) we recursively solve

$$\arg \min_{\tilde{\sigma}, d_1} \|\tilde{\sigma}\|_\infty + \lambda \|d\|_1 - \langle p_{\tilde{\sigma}}^k, \tilde{\sigma} - \tilde{\sigma}_k \rangle + \langle p_d^k, d - d_k \rangle + \frac{\mu}{2} \|\tilde{\sigma} + d - \sigma\|_2^2. \quad (16)$$

Using the explicit formulas of the sub-gradients $p_{\tilde{\sigma}}^k$ and p_d^k can help us to get the simplified iteration

$$\begin{cases} (\tilde{\sigma}^k, d^k) = \arg \min_{\tilde{\sigma}, d_1} \|\tilde{\sigma}\|_\infty + \lambda \|d\|_1 + \frac{\mu}{2} \|\tilde{\sigma} + d - \sigma - b^{k-1}\|_2^2 \\ b^k = b^{k-1} + \sigma - \tilde{\sigma}^k - d^k \end{cases}. \quad (17)$$

Furthermore, we can break (17) into the following three steps to be solved alternatively in each iteration:

$$Step1 : \tilde{\sigma}^k = \arg \min_{\tilde{\sigma}} \frac{\mu}{2} \|\tilde{\sigma} + d^{k-1} - \sigma - b^{k-1}\|_2^2 + \|\tilde{\sigma}\|_\infty, \quad (18a)$$

$$Step2 : d^k = \arg \min_d \frac{\mu}{2} \|d + \tilde{\sigma}^k - \sigma - b^{k-1}\|_2^2 + \lambda \|d\|_1, \quad (18b)$$

$$Step3 : b^k = b^{k-1} + \sigma - \tilde{\sigma}^k - d^k. \quad (18c)$$

In this algorithm, the sum of l_1 -norm and l_∞ -norm is a non-differentiable but convex functional, and l_2 -norm is differentiable and convex functional, which satisfy the conditions in the l_1 regularized problems as [33], so the iterative convergence is guaranteed, for more details please refer to literature.

3.3. Algorithm

The $l_\infty + l_1$ minimization problem can be split into two sub-problems, the l_∞ and l_1 minimization problem, and the closed-form solutions to these two sub-problems are provided in the following two propositions.

Proposition 3. [36] *For any given data vector $g \in \mathbb{R}^N$ and $\tau > 0$, the unique closed-form solution to the following problem*

$$x^* = \arg \min_x \frac{1}{2} \|x - g\|_2^2 + \tau \|x\|_\infty, \quad (19)$$

takes the form

$$x^* = \begin{cases} 0, & \text{if } \|g\|_1 \leq \tau \\ g, & \text{if } \|g\|_\infty \geq \tau \\ g - S_\tau(g), & \text{otherwise} \end{cases} \quad (20)$$

where

$$S_\tau(g(i)) = \max(1 - \frac{v}{|g(i)|}, 0) \cdot g(i) \quad (21)$$

and $v > 0$ is chosen by

$$v = \frac{\tau - G(t)}{G(t+1) - G(t)} \cdot g_\alpha(N-t) + \frac{G(t+1) - \tau}{G(t+1) - G(t)} \cdot g_\alpha(N-t+1) \quad (22)$$

where t satisfies $G(t) \leq \tau < G(t+1)$. Here g_α is such a sequence that sorting the elements of the vector g in an increasing order ($g_\alpha(1) \leq g_\alpha(2) \leq \dots \leq g_\alpha(N)$), and G is the cumulated ordered values which has been defined as $G(j) = \|g_\alpha\|_1 - \sum_{i=1}^{N-j} g_\alpha(i)$.

Proposition 4. [37] *Let $h \in \mathbb{R}^N$ be a given vector, then the optimal solution to*

$$y^* = \arg \min_y \frac{1}{2} \|y - h\|_2^2 + \tau \|y\|_1 \quad (23)$$

can be efficiently obtained by using the shrinkage $y^ = \text{soft}(h, \tau)$.*

We suggest using the following stop criteria for the proposed algorithm to judge whether to end iteration. The first stopping criterion is the feasibility:

$$\|\sigma - \tilde{\sigma}^k - d^k\|_2^2 / \|\sigma\|_2^2 \leq \varepsilon_1 \quad (24)$$

Furthermore, the solution should converge to the fixed point, so the second stopping criterion is

$$\max\{\|\tilde{\sigma}^{k+1} - \tilde{\sigma}^k\|_2^2 / \|\sigma\|_2^2, \lambda \|d^{k+1} - d^k\|_2^2 / \|\sigma\|_2^2\} \leq \varepsilon_2 \quad (25)$$

Therefore, according to the aforementioned analysis, we may safely draw the alternative iterative (AI) algorithm to the minimization problem (3). In the actual experiments, the acceptable solution is obtained even if the stopping criteria ε is not too small.

Algorithm 1 Solving (3) with AI method.

```

1: input:  $Y$ .
2: initialization:  $X = 0, E = 0, \tilde{\sigma}^0 = 0, d^0 = 0, b^0 = 0$ , and  $k = 0$ .
3: Do SVD on the data matrix  $Y = L\Sigma R^T, \Sigma = \begin{pmatrix} \text{diag}(\sigma_i) \\ O \end{pmatrix}$ .
4: repeat
5:    $k = k + 1$ .
6:   update  $\tilde{\sigma}^k$  by formula (18a);
7:   update  $d^k$  by formula (18b);
8:   update  $b^k$  by formula (18c);
9: until the stopping criterion is satisfied.
10:  $X^* = L \cdot \begin{pmatrix} \text{diag}(d_i^*) \\ O \end{pmatrix} \cdot R^T; E^* = Y - X^*$ 
11: output:  $X^*, E^*$ .

```

4. Numerical examples

We confirm the theoretical predictions in this paper with some experimental results. In this section, we examine the effectiveness and efficiency of the proposed method on the simulation data and real database. Since GoDec and our method are related in their motivations, so the numerical experiments in this section are completely imitate [13] as a reference. The simulation experiments are the randomly generated square matrices with different sizes. The practical experiments are apply our model on the real surveillance videos for the applications in object detection and background subtraction, and on the Yale B database to remove shadows and specularities from face images. However, before these experiments, we will present the difference between the thresholding methods in section 4.1 first. All the numerical experiments are performed with Matlab 7.14 on a Intel-Xeon 2.40 GHz PC running Windows 7 with 8 GB main memory.

4.1. Thresholding discussion

Our vectorizational minimization problem (12) can be divided into two sub-problems, and these two minimization problems can be classified as a same model:

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} \|x - y\|_2^2 + \lambda \|x\|_p. \quad (26)$$

When $p = 0$, the solution of (26) can be obtained by applying a hard-thresholding on the vector y , $\text{hard}(y, \lambda) = y \cdot (|y| > \lambda)$; when $p = 1$, the solution of (26) is the soft-thresholding on the vector y , $\text{soft}(y, \lambda) = \text{sign}(y) \cdot (|y| - \lambda)_+$; and when $p = \infty$, the solution is a little complex which can be obtained by the primal and dual minimization problem of (26).

Actually, the solution $new(y, \lambda)$ to the minimization problem (12) is also a threshold essentially, which is completely different from the general hard- and soft-thresholding. The hard-thresholding result can keep the large values of the vector, but it is a discontinuous function; the soft-thresholding result is continuously, but it abandon most of the values of the vector data; and our new-thresholding result can not only be a continuous one, but also retain the data values as much as possible. Besides, applying the $l_\infty + l_1$ minimization problem (12) on any vector is actually doing a shifty threshold on the vector (when the vector value is greater than the setting parameter, the threshold decreases with the increasing of the vector variable).

Finally, we apply the three thresholding functions directly onto an arbitrary vector y , and Fig. 1 shows us the difference between these thresholds, the solid line shows us the value of signals, the dotted line shows us the removing value from the signal via the thresholding methods.

Figure 1: Three types of threshold effect on the signal in range $(0, 1)$, $\lambda = 0.4$. From left to right: Original signal; Hard-thresholded signal; Soft-thresholded signal; New-thresholded signal.

4.2. Simulation experiments

I. Comparison on the Robust PCA Problem. In order to better comparing with other methods, a brief comparison between our proposed method and other five state-of-the-art methods is presented in this section. We consider randomly generated square matrices with different sizes n and different ranks r for our simulations. For a matrix $Y = X + E$, the matrix X is the low-rank component whose values are assigned according to LR' in the interval $[0, 255]$, wherein both L and R are $n \times r$ standard Gaussian matrices with entries independently sampled from a $N(0, 1/n)$ distribution. And the matrix E is the error interference component, whose non-zero entries k are independent and identically distributed uniformly at random. Table 1 lists out the relative errors $\|X - X^*\|_F^2 / \|Y\|_F^2$ and time costs results with $r = 0.05 \times n$ and $k = 0.1 \times n^2$. Notice that in all cases, the relative error is small. The speed improvement of our method is due to the “ $l_\infty + l_1$ ” minimization significantly saves the computation of each iteration round.

II. Comparison on the Matrix Completion Problem. We test the performance of the proposed method in matrix completion task, and compare the result with IALM, GoDec. Each test low-rank matrix is generated

Table 1: Correct recovery for random RPCA problems with varying size.

Size	Method	RPCA[5]	APG[29]	IALM[3]	EALM[3]	GoDec[13]	OUR
250	error	2.06e-4	1.23e-5	4.07e-8	9.03e-8	6.72e-8	3.07e-8
	time	5.41	8.35	0.83	1.88	1.68	0.18
500	error	1.94e-4	1.02e-5	2.29e-8	1.94e-8	5.15e-8	2.85e-8
	time	11.58	11.07	2.32	5.79	5.80	0.69
750	error	1.70e-4	9.49e-6	1.81e-8	1.01e-8	3.60e-8	1.93e-8
	time	23.27	19.14	4.80	12.44	13.86	2.03
1000	error	1.58e-4	8.31e-6	2.26e-8	1.21e-8	3.21e-8	1.69e-8
	time	48.18	32.93	9.89	27.51	26.49	4.61
1500	error	2.05e-4	6.79e-6	2.89e-8	1.53e-8	2.47e-8	1.33e-8
	time	71.57	86.21	25.09	87.86	66.47	11.36
2000	error	1.80e-4	6.02e-6	3.73e-8	1.82e-8	1.83e-8	1.13e-8
	time	192.57	196.46	51.18	193.48	122.31	39.96
3000	error	2.01e-4	4.93e-6	2.09e-8	2.64e-8	1.84e-8	1.35e-8
	time	797.69	597.97	156.08	573.89	310.60	99.04
5000	error	5.61e-4	3.96e-6	3.04e-8	3.87e-8	1.52e-8	1.60e-8
	time	7180.30	2216.98	680.03	2457.63	965.89	353.72

by $X = AB'$, wherein both A and B are $n \times r$ standard Gaussian matrices. We randomly sample a few entries from X and recover the whole matrix through three different methods. Table 2 lists out the relative errors $\|X - X^*\|_F^2 / \|X^*\|_F^2$ and time costs (seconds).

Table 2: Comparison between IALM[3], GoDec[13] and OURs on the matrix completion problem.

Size(X)	Rank(X)	Sampling rate(%)	IALM[3]		GoDec[13]		OUR	
			error	time	error	time	error	time
1000	10	0.1194	9.62e-5	8.45	3.17e-5	22.94	9.74e-5	7.52
	50	0.3900	9.32e-5	12.85	1.30e-5	29.22	7.09e-5	10.83
	100	0.5700	1.49e-4	24.31	7.84e-5	36.20	2.22e-5	11.63
3000	10	0.0399	8.16e-5	61.42	6.46e-5	213.51	2.20e-5	87.59
	50	0.1653	6.07e-5	115.07	5.24e-5	253.94	9.72e-6	88.49
	100	0.2622	7.32e-5	209.54	3.76e-5	290.00	8.60e-6	89.08
5000	10	0.0240	8.19e-5	229.62	1.76e-5	402.78	4.31e-5	151.10
	50	0.0995	7.72e-5	388.33	2.62e-5	655.74	1.40e-5	170.83
	100	0.1584	1.04e-4	666.03	4.69e-5	721.71	5.90e-5	171.42

4.3. Background modeling

In this experiment, we perform our algorithm on a real video dataset which can be downloaded from the website ². Background modeling is a crucial task for motion segmentation in surveillance video sequences,

²http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html

Figure 2: Background modeling results of four 200-frame surveillance video sequences. From up to down: Lobby in an office building (resolution 128×160); Hall of a business building (resolution 144×176); Restaurant (resolution 120×160); Shopping center (resolution 256×320). From left to right: (a) Original sample from the video frames; (b)-(c) Background and Foreground by IALM; (d)-(e) Background and Foreground by GoDec; (f)-(g) Background and Foreground by OURs.

the background of any static frames is controlled by few factors which satisfies the low-rank structure, and the foreground is detected by the indentifying spatially localized residuals which being limited as the most irrelevant part with the background. We compare our method against two state-of-the-art methods: IALM [3] and GoDec [13] on four surveillance videos: Lobby, Hall, Restaurant, and Mall databases, respectively. The pending data matrix consists of 200 frames of each video. For example, the Lobby video is composed of the first 200 frames with the resolution 128×160 , and then we convert these color images into the gray-scale ones and reshape each frame of the gray images into a long vector as one column of the collected data matrix with a size of 20480×200 .

Fig. 2 illustrates the decomposition results of one frame in each video sequence using IALM, GoDec, and the proposed method. It is clear that the background and moving objects are precisely separated by these three methods. Due to that there only has a small moving object in the foreground, the decomposition results of the three methods all performance well in the Lobby video. But once the moving foreground become complex, such as other three scenes in Fig. 2, the advantages of the proposed algorithm will emerge. The background be modeled by IALM always contains some part of the foreground, which leads to the extraction of the moving objects have partially missing. The background and foreground, which is decomposed by GoDec, both contains some artificial artifacts in the results. By comparison, the proposed algorithm can always get more clean background and more complete foreground. For our method, we set

the parameters as $\lambda = c_1/\sqrt{m}, \mu = c_2/\|Y\|_2^2$, where c_1, c_2 are the constants to ensure the algorithm adapt to different video. For IALM and GoDec, the parameters are set as in [3, 13].

Furthermore, we provide the CPU time consumption of IALM, GoDec, and the proposed method in Table 3, from which we can see that the proposed method is much faster than IALM and GoDec. The improvement of speed is due to the “low-rank + dual” model based low-rank approximation significantly saves the computation of each iteration.

Table 3: Comparison of the CPU time consumption (seconds) on four videos for three methods, and the correlation between the background image and the moving foreground.

Method	IALM		GoDec		OUR	
	Time/sec	correlation	Time/sec	correlation	Time/sec	correlation
Lobby	24.9068	0.1273	39.5113	0.1180	1.0188	0.0725
Hall	34.1413	0.3015	63.5167	0.3469	1.2327	0.2627
Restaurant	25.2194	0.1911	62.6013	0.2459	1.0187	0.1310
Mall	115.6207	0.1934	209.5952	0.2755	4.6415	0.1035

4.4. Specularities removing

In the final experiment, we apply our methods on face images of each individual from the yale B database³. Face recognition is an important field in computer vision applications, but the environment under different illuminations will affect the collection of the face information during the acquisition process. Shadow and specularities are the most common effects in computer vision applications, and our algorithm can remove those effects and only keep the useful information of the human faces. We assume that the useful parts of the face images are low-rank and the rest parts of the images are the shadow and specularities. Yale B database has 64 face images for each individual and the size of each image is 192×168 , thus the data matrix for each individual is of size 32256×64 .

Fig. 3 shows the specularities removing performance of our proposed method, the sub-images (a)-(d) respectively show the results of four different individuals, the Eastern and Western faces of men and women, each sub-image contains original face (left), main information face (middle) and illumination effected shadow (right). From the figure we can see that all the shadows and specularities on each face are successfully removed by the proposed method. The low-rank part contains clean face image with complete information, and the removing part contains all the interference section.

Table 4 shows us the CPU time consumption of IALM, GoDec, and the proposed method, and the average correlation between the informational face and the removal residuals for each method. The learning time of the proposed method for each individual is less than 1s, while IALM and GoDec take more than

³<http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html>

Figure 3: Shadow, specularities and saturations removal from face images of four individuals in Yale B database, Each individual has 64 images with resolution 192×168 . From left to right: (a) Eastern man face; (b) Eastern woman face; (c) Western man face; (d) Western woman face.

60s. For our proposed method, we set the parameters as $\lambda = 7, \mu = 0.02$. And the parameter sets of IALM and GoDec are set as in [3, 13].

Table 4: Comparison of the average CPU time consumption (seconds) on three methods, and the average correlation between the informational face and the removal residuals.

Method	IALM	GoDec	OUR
Time	68.9843	63.6077	0.2018
Correlation	0.7019	0.6750	0.3288

5. Conclusion

In this work, we introduced a special “low-rank + dual” model with dual regularization terms, which enable our novel model to calculate in a simple “ $l_\infty + l_1$ ” minimization problem. In addition, the relative algorithm is easy to implement and surprisingly effective in both computational cost and storage requirement. Compare with the widely used method IALM and GoDec on background modeling and facial effects remove problems, our method, which enables faster recovery for large-scale data matrix, is better at handling the global structures and correcting the corruptions in data automatically. Both theoretical and experimental results illustrate the effectiveness of the proposed method.

Acknowledgement

The authors would like to thank the anonymous reviewers for their considerations and kindly comments. The National Science Foundation of China (No. 61271294, 61472303) and the Fundamental Research Funds

for the Central Universities (NSIY21) support this work.

References

- [1] J. Cai, E. Candes, Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4), (2010) 1956–1982.
- [2] E. Candes, B. Recht. Exact low-rank matrix completion via convex optimization. In: *Proc. of 46th Annual Allerton Conference on Communication, Control, and Computing*, (2008) 806–812.
- [3] Z. Lin, M. Chen, L. Wu, Y. Ma. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. *UIUC Technical Report UILU-ENG-09-2215* (2009).
- [4] S. Ma, D. Goldfarb, L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization. Preprint (2009).
- [5] E. Candes, X. Li, Y. Ma, J. Wright. Robust principal component analysis? *Journal of ACM* 58(3), (2011) 1–37.
- [6] X. Ding, L. He, L. Carin. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing* 20(12), (2011) 3419–3430.
- [7] Z. Lin, R. Liu, Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems (NIPS)*, (2011) 612–620.
- [8] G. Liu, Z. Lin, Y. Yu. Robust subspace segmentation by low-rank representation. In *Proceedings of the 25th international conference on machine learning, ICML*, (2010) 663–670.
- [9] Y. Peng, A. Ganesh, J. Wright, Y. Ma. RASL: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR*, (2010) 763–770.
- [10] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2), (2011) 52–68.
- [11] S. Ji, J. Ye. An accelerated gradient method for trace norm minimization. In *Proceedings of the 26th international conference on machine learning, ICML*, (2009) 457–464.
- [12] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), (2013) 171–184.
- [13] T. Zhou, D. Tao. GoDec: randomized low-rank & sparse matrix decomposition in noisy case. In *Proceedings of the 26th international conference on machine learning, ICML*, (2011) 33–40.
- [14] J. Wright, A. Ganesh, S. Rao, Y. Peng, Y. Ma. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In: *Proc. of Neural Information Processing Systems (NIPS)*, (2009) 2080–2088.
- [15] T. W. Anderson. Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, (1951) 327–351.
- [16] A. J. Izenman. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis* 5, (1975) 248–264.
- [17] G. C. Reinsel, P. Velu. *Multivariate reduced-rank regression: theory and applications*. New York: Springer, (1998).
- [18] M. Yuan, A. Ekici, Z. Lu, R. Monteiro. Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society Series B* 69, (2007) 329–346.
- [19] F. Bunea, Y. She, M. Wegkamp. Optimal selection of reduced rank estimators of high-dimensional matrices. *The Annals of Statistics* 39, (2011) 1282–1309.
- [20] S. Negahban, M. J. Wainwright. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics* 39, (2011) 1069–1097.
- [21] A. Rohde, A. Tsybakov. Estimation of High-Dimensional Low-Rank Matrices. *The Annals of Statistics* 39, (2011) 887–930.
- [22] E. Candes, B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.* 9, (2009) 717–772.

- [23] V. Koltchinskii, K. Lounici, A. Tsybakov. Nuclear norm penalization and optimal rates for noisy low rank matrix completion. *The Annals of Statistics* 39, (2011)2302–2329.
- [24] C. Eckart, G. Young. The approximation of one matrix by another of lower rank. *Psychometrika* 1, (1936) 211–218.
- [25] A. N. Kolmogorov, S. V. Fomin. *Elements of the theory of Functions and Functional Analysis*. Courier Dover, (1999).
- [26] Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations. the fifteenth Dean, in: B. Jacqueline (Ed.), *Lewis Memorial Lectures*, AMS, Boston, MA, USA, (2001).
- [27] C. Stauffer, W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), (2000) 747–757.
- [28] W. Yang, Z. Wang, C. Sun. A collaborative representation based projections method for feature extraction. *Pattern Recognition*, 48(1), (2015) 20–27.
- [29] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. submitted to *SIAM Journal on Optimization*, (2008).
- [30] J. L. Xu, X. C. Feng, Y. Hao, et al. Adaptive variational models for image decomposition. *Science China Information Sciences*, 57(2), (2014) 1–8.
- [31] S. Dennis. *Bernstein, Matrix Mathematics - Theory, Facts and Formulas*. Princeton University Press, Princeton and Oxford, (2008).
- [32] R.A. Horn, C.R. Johanson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, (1997).
- [33] T. Goldstein, S. Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2), (2009) 323–343.
- [34] S. Setzer. Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage. In: X.-C. Tai, K. Morken, M. Lysaker, K.-A. Lie (Eds.), *Scale Space and Variational Methods in Computer Vision*. LNCS, Springer, Berlin, vol. 5567, (2009) 464–476.
- [35] B. He, X. Yuan. On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM Journal on Numerical Analysis*. 50(2), (2012) 700–709.
- [36] J. M. Fadili, G. Peyre. Total variation projection with first order schemes. *IEEE Transactions on Image Processing* 20(3), (2011) 657–669.
- [37] A. Chambolle, R. D. Vore, N. Lee, B. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Proc.* 7(3), (1998) 319–335.

Si-Qi WANG received the B.S. degree in computational mathematics from Xidian University in 2009, and earned the M.S. degree and pursued doctoral degree in advance from the Department of Mathematics, Xidian University, Xian, China, in 2011. She is now working towards the Ph.D. degree in applied mathematical at the school of mathematics and statistics, Xidian University. Her research interest covers mathematical problem in image processing.

Xiang-Chu FENG received the B.S. degree in computational mathematics from the Xian JiaoTong University, Xian, China, in 1984, and the M.S. and Ph.D. degrees in applied mathematics from Xidian University, Xian, in 1989 and 1999, respectively. He is currently a Professor of applied mathematical with the school of mathematics and statistics at Xidian University. His research interests include numerical analysis, wavelets, and partial differential equations for image processing.

Wei-Wei Wang received the B.S., M.S. and Ph.D. degrees from Xidian University, Xian, China, in 1993, 1998 and 2001, respectively. She is currently a Professor of applied mathematical with the school of mathematics and statistics, Xidian University. Her research is focused on variational and partial differential equations, multiresolution analysis, sparse representation and their applications in image processing.