Low Rank Tensor Completion via Partial Sum Minimization of Singular Values

Feng Zhang, Jianjun Wang and Jia Jing*
School of Mathematics and Statistics
Southwest University
Chongqing, China
*Corresponding author: jingj@swu.edu.cn

***Abstract*—In this paper, we investigate the low-rank tensor completion problem, in which we wish to estimate missing values of tensors from incomplete samples its entries. In real world, the low-rank tensor can be seen everywhere and the exact rank of it is often known. Based on the fact that singular values before the target rank does not affect rank minimization of tensors, we propose low rank tensor completion via partial sum minimization of singular values algorithm(PSSV-LRTC). Some experiments are performed on both synthetic data and real applications; all results show that our algorithm has a higher precision and convergence rate than previous work.**

Keywords—tensor completion; matrix completion; nuclear norm minimization; alternating direction method of multipliers

I. INTRODUCTION

Low rank matrix completion [1]-[3] mainly deals with estimating missing information from the incomplete observations of a data matrix. The problem has triggered wide attention and has various applications in the industrial field, such as image inpainting [4], [5], *Netflix* competition [6], High Dynamic Range [7]. We not only need to estimate the missing values, and more important is to recover underlying low rank structure of the interested matrix. The key to solve this problem is how to establish the relationship between the observed value and unknown elements. The traditional methods [8]-[10] assume that the missing values are only related to its neighbors and ignore the correlation of the remaining points, so that it can't capture the global information. Just now, the rank minimization is put forward to overcome this disadvantage. So, the low rank matrix completion problem can be formulated as follows:

$$\begin{aligned} &\min_{\boldsymbol{L}} \ rank(\boldsymbol{L}) \\ &s.t.\ \boldsymbol{L}_{\Omega} = \boldsymbol{O}_{\Omega} \end{aligned}, \tag{1}$$

where $\boldsymbol{L}$, $\boldsymbol{O} \in \mathbf{P}^{m\times n}$ are a low rank matrix to be reconstructed and an observation matrix, respectively. Ω is the support set of the observed entries. However, $rank(\cdot)$ is a nonconvex and nonsmooth function, so the problem (1) is a NP-hard problem that can't be effectively solved in practice. As we known, although, the nuclear norm $\|\cdot\|_*$ is not a best approximation of the rank function; it is a convex relaxation of the optimization (1). Therefore, some scholars [1], [11], [12] have investigated the following convex optimization model:

$$\begin{aligned} &\min_{\boldsymbol{L}} \ \|\boldsymbol{L}\|_* \\ &s.t.\ \boldsymbol{L}_{\Omega} = \boldsymbol{O}_{\Omega} \end{aligned}, \tag{2}$$

where $\|\boldsymbol{L}\|_* = \sum_i \sigma_i(\boldsymbol{L})$ and $\sigma_i(\boldsymbol{L})$ is the ith largest singular value of $\boldsymbol{L}$. In practical application, we often known some prior information about the rank of $\boldsymbol{L}$, for instance, $rank(\boldsymbol{L})=1$ for background subtraction, $rank(\boldsymbol{L})=3$ for photometric stereo. In 2015, Tae-Hyun Oh, Yu-Wing Tai *et al.* [13] noticed the major shortcoming of using $\|\cdot\|_*$ to approximate $rank(\cdot)$ is that the nuclear norm minimizes not only the rank of $\boldsymbol{L}$, but also the variance of $\boldsymbol{L}$ by simultaneously minimizing all the singular values of $\boldsymbol{L}$ including the singular values within the target rank. Therefore, they put forward the partial sum minimization of singular values(**PSSV**), that is using $\|\boldsymbol{L}\|_{p=r}$ to approximate $rank(\cdot)$ to take full advantage of the prior information, where $\|\boldsymbol{L}\|_p = \sum_{i=p+1}^{\min(m,n)} \sigma_i(\boldsymbol{L})$ and r is the target rank of $\boldsymbol{L}$. Innovatively, we apply this technique to the matrix completion and form the following model:

$$\begin{aligned} &\min_{\boldsymbol{L}} \ \|\boldsymbol{L}\|_{p=r} \\ &s.t.\ \boldsymbol{L}_{\Omega} = \boldsymbol{O}_{\Omega} \end{aligned}. \tag{3}$$

When the number of observation in $\boldsymbol{O}_{\Omega}$ is very sufficient, the singular value thresholding algorithm(**SVT**), the fixed-point shrinkage algorithm(**FPC**), and the accelerated proximal gradient algorithm(**APC**) *et al.* [14]-[16] can well reconstruct the unknown matrix based on the model (2). Although the problem (3) is nonconvex, when the number of observation in $\boldsymbol{O}_{\Omega}$ is deficient, the method (3) can lead to a higher success rate than the model (2) [13].

In recent years, with the development of computer technology, data collection becomes much easier. But the dimension of the collected data is growing geometrically, such as aerospace data, biology data, image data and so on. Owing to the abundant information from these data named higher-order tensors, the research of higher-order tensors have great significance. How to deal with higher-order tensors effectively is an enormous challenge because of "the curse of dimensionality". One of the most basic ideas is to reduce the dimension, namely, higher-order tensors are processed as a 1D signal or a 2D matrix. This processing method can not capture

the structural information of higher-order tensors, which is not conducive to our further analysis. To overcome this problem, we establish a new model based on the partial sum minimization of singular values (**PSSV**) for tensor completion that is an extension of matrix completion in next section.

The paper is organized as follows. In Section II, we introduce some notations and give our problem formulation and its algorithm. In Section III, we perform a series of experiments based on the proposed algorithm in Section II. Finally, the conclusion is addressed in section IV.

II. TENSOR COMPLETION BASED ON PSSV

In order to facilitate the analysis, we firstly provide a brief notational summary following Kolda and Bader's review on tensor decompositions [17].

A. Notation

Let an N-order tensor as $\mathbf{\Lambda} \in \boldsymbol{P}^{I_1 \times I_2 \times \cdots \times I_N}$ and $l_{i1\cdots in\cdots iN}$ represent each element of $\mathbf{\Lambda}$, where $1 \le in \le I_N$. The order N of a tensor is the number of dimensions, also known as ways or modes. A first-order tensor is a vector and a second-order tensor is a matrix. We define the mode-n unfolding matrix as $\boldsymbol{L}_{(n)} = \text{unfold}_n(\mathbf{\Lambda}) = \boldsymbol{P}^{I_n \times \left(\prod_{i \ne n} I_i\right)}$. It's inverse operation defined as $\mathbf{\Lambda} = \text{fold}_n\left(\boldsymbol{L}_{(n)}\right)$. The inner product of two tensors $\mathbf{\Lambda}$ and ς with the same dimension is defined as $\langle \mathbf{\Lambda}, \varsigma \rangle = \sum_{i1}\sum_{i2}\cdots\sum_{iN} l_{i1\cdots iN} v_{i1\cdots iN}$. The Frobenius norm of $\mathbf{\Lambda}$ is defined as $\|\mathbf{\Lambda}\|_F = \sqrt{\langle \mathbf{\Lambda}, \mathbf{\Lambda} \rangle}$ and $\|\mathbf{\Lambda}\|_F = \|\boldsymbol{L}_{(i)}\|_F$ for any $1 \le i \le N$ is clear.

B. Problem Formulation

A core problem of the research on tensor completion is how to define the trace norm of tensors. We should thank to Liu, Musialski *et al.* [18] who proposed a great definition for the tensor trace norm as follows:

$$\|\mathbf{\Lambda}\|_* := \sum_i^N \alpha_i \|\boldsymbol{L}_{(i)}\|_*, \tag{4}$$

where $\sum_i^N \alpha_i = 1$. As we pointed out in our introduction, we can learn some prior information about the rank of the unfolding matrix $\boldsymbol{L}_{(i)}$ along mode-i in advance. For convenience, we assume that the rank r_i of $\boldsymbol{L}_{(i)}$ is known without loss of generality. Based on the model (3) and the definition (4), we initiatively propose our major research model in this paper as follows:

$$\begin{aligned} &\min_{\mathbf{\Lambda}} \sum_i^N \alpha_i \|\boldsymbol{L}_{(i)}\|_{p=r_i}, \\ &s.t.\ \mathbf{\Lambda}_{\Omega} = \mathbf{O}_{\Omega} \end{aligned} \tag{5}$$

However, when the tensor is represented by the mode-i matrices, these matrices share the same variables. It is precisely because of the interdependent between **PSSV**, so we can't directly solve each norm. Hence, we introduce N auxiliary tensors $\mathbf{\Gamma}_i's$ to remove these interdependencies and to optimize these terms independently. Therefore, the optimization problem (5) becomes

$$\begin{aligned} &\min_{\mathbf{\Lambda}, \mathbf{\Gamma}_1, \cdots, \mathbf{\Gamma}_N} \sum_i^N \alpha_i \|\mathbf{\Gamma}_{i(i)}\|_{p=r_i} \\ &s.t.\ \mathbf{\Lambda}_{\Omega} = \mathbf{O}_{\Omega} \\ &\quad \mathbf{\Lambda} = \mathbf{\Gamma}_{(i)},\ i = 1, \cdots, N. \end{aligned} \tag{6}$$

One remaining challenge is to build an algorithm for solving efficiently the problem (6). We will present it in next subsection.

C. Optimization Algorithm

To solve our partial sum objective function with equality constraints, we utilize the Alternating Direction Method of Multipliers(ADMM) [19] and obtain the following the augmented Lagrangian function

$$\begin{aligned} &F_\theta(\mathbf{\Lambda}, \mathbf{\Gamma}_1, \cdots, \mathbf{\Gamma}_N, \mathbf{\Pi}_1, \cdots, \mathbf{\Pi}_N) \\ &= \sum_i^N \alpha_i \|\mathbf{\Gamma}_{i(i)}\|_{p=r_i} + \langle \mathbf{\Pi}_i, \mathbf{\Lambda} - \mathbf{\Gamma}_i \rangle + \frac{\theta}{2} \|\mathbf{\Lambda} - \mathbf{\Gamma}_i\|_F^2, \end{aligned} \tag{7}$$

where θ is a positive scalar for controlling the speed of the algorithm and $\mathbf{\Pi}_i$ is the Lagrangian multiplier tensor. It is obviously difficult to solve the function (7) directly. We can fist minimize a variable for fixed the other variables as follows:

$$\begin{aligned} \mathbf{\Gamma}_i^{k+1} &= \arg\min_{\mathbf{\Gamma}_i} F_{\theta^k}\left(\mathbf{\Lambda}^k, \mathbf{\Gamma}_i, \mathbf{\Pi}_i^k\right) \\ &= \arg\min_{\mathbf{\Gamma}_i} \frac{\alpha_i}{\theta^k} \|\mathbf{\Gamma}_i\|_{p=r_i} + \frac{1}{2} \left\| \mathbf{\Gamma}_i - \left(\mathbf{\Lambda}^k + \frac{1}{\theta^k} \mathbf{\Pi}_i^k \right) \right\|_F^2, \end{aligned} \tag{8}$$

$$\mathbf{\Lambda}^{k+1} = \arg\min_{\mathbf{\Lambda}} F_{\theta^k}\left(\mathbf{\Lambda}, \mathbf{\Gamma}_i^k, \mathbf{\Pi}_i^k\right) = \frac{\theta^k}{2} \left\| \mathbf{\Gamma}_i^k - \frac{1}{\theta^k} \mathbf{\Pi}_i^k - \mathbf{\Lambda} \right\|_F^2, \tag{9}$$

$$\mathbf{\Pi}_i^{k+1} = \mathbf{\Pi}_i^k + \theta^k \left(\mathbf{\Lambda}^{k+1} - \mathbf{\Gamma}_i^{k+1} \right).$$

Computing $\mathbf{\Gamma}_i^{k+1}$:

As Tae-Hyun Oh *et al.* [13] pointed out, the optimal solution of the sub-problem (8) in the matrix case can be described as:

$$\begin{aligned} \mathbf{\Pi}_{r_i, \tau}[\boldsymbol{Y}] &= \arg\min_{\boldsymbol{X}} \tau \|\boldsymbol{X}\|_{P=r_i} + \frac{1}{2} \|\boldsymbol{X} - \boldsymbol{Y}\|_F^2, \\ &= \boldsymbol{Y}_1 + \boldsymbol{U}_{\boldsymbol{Y}_2} \mathbf{\Sigma}_\tau[\boldsymbol{D}_{\boldsymbol{Y}_2}] \boldsymbol{V}_{\boldsymbol{Y}_2}^T, \end{aligned}$$

where $\boldsymbol{Y} = \boldsymbol{Y}_1 + \boldsymbol{Y}_2 = \boldsymbol{U}_{\boldsymbol{Y}_1} \boldsymbol{D}_{\boldsymbol{Y}_1} \boldsymbol{V}_{\boldsymbol{Y}_1}^T + \boldsymbol{U}_{\boldsymbol{Y}_2} \boldsymbol{D}_{\boldsymbol{Y}_2} \boldsymbol{V}_{\boldsymbol{Y}_2}^T$, $\boldsymbol{U}_{\boldsymbol{Y}_1}$, $\boldsymbol{V}_{\boldsymbol{Y}_1}$ are the singular vector matrices corresponding to the r_i largest singular values by SVD, and $\boldsymbol{U}_{\boldsymbol{Y}_2}$, $\boldsymbol{V}_{\boldsymbol{Y}_2}$ from the (r_i+1)th to the last singular values; $\boldsymbol{D}_{\boldsymbol{Y}_1} = diag(\sigma_1, \cdots, \sigma_{r_i}, 0, \cdots, 0)$, $\boldsymbol{D}_{\boldsymbol{Y}_2} = diag(0, \cdots, 0, \sigma_{r_i+1}, \cdots, \sigma_{\min(m,n)})$, and $\mathbf{\Sigma}_\tau[\cdot] = sign(\cdot)\max(|\cdot| - \tau, 0)$. Hence,

$$\mathbf{\Gamma}_i^{k+1} = \text{fold}_i\left[\mathbf{\Pi}_{r_i, \frac{\alpha_i}{\theta^k}} \left(\boldsymbol{L}_{(i)}^k + \frac{1}{\theta^k} \mathbf{\Pi}_{i(i)}^k \right) \right].$$

Computing $\mathbf{\Lambda}^{k+1}$:

Solving sub-problem (9) directly, we have

$$\mathbf{\Lambda}_{\Omega}^{k+1} = \frac{1}{N}\left(\sum_{i}^{N}\mathbf{\Gamma}_{i}^{k} - \frac{1}{\theta^{k}}\mathbf{\Pi}_{i}^{k}\right)_{\overline{\Omega}}.$$

The above results can be summarized in the **Algorithm 1**.

Algorithm 1 PSSV-LRTC: Low Rank Tensor Completion Based On **PSSV**

1: **Input:** $\mathbf{O}_{\Omega}$, the target rank r_i, and the maximum iterations K.

2: **Initialization:** $\mathbf{\Lambda}_{\Omega}^{0} = \mathbf{\Gamma}_{\Omega}^{0} = \mathbf{O}_{\Omega}$, $\mathbf{\Lambda}_{\overline{\Omega}}^{0} = \mathbf{\Gamma}_{\overline{\Omega}}^{0} = mean(\mathbf{O}_{\Omega})$, $\mathbf{\Pi}^{0} = 0$, $\theta^{0} > 0$, $\rho > 1$, and $k = 0$.

3: **for** $k = 0$ to K **do**

4: **for** $i = 0$ to N **do**

5: $\mathbf{\Gamma}_{i}^{k+1} = \text{fold}_{i}\left[\mathbf{\Pi}_{r_i, \frac{\alpha_i}{\theta^k}}\left(\boldsymbol{L}_{(i)}^{k} + \frac{1}{\theta^{k}}\mathbf{\Pi}_{i(i)}^{k}\right)\right]$

6: **end for**

7: $\mathbf{\Lambda}_{\Omega}^{k+1} = \frac{1}{N}\left(\sum_{i}^{N}\mathbf{\Gamma}_{i}^{k} - \frac{1}{\theta^{k}}\mathbf{\Pi}_{i}^{k}\right)_{\overline{\Omega}}$

8: $\mathbf{\Pi}_{i}^{k+1} = \mathbf{\Pi}_{i}^{k} + \theta^{k}\left(\mathbf{\Lambda}^{k+1} - \mathbf{\Gamma}_{i}^{k+1}\right)$

9: $\theta^{k+1} = \rho\theta^{k}$

10: **end for**

11: **Output:** $\mathbf{\Lambda}^{k}$

Obviously, Computing $\mathbf{\Gamma}_{i}^{k+1}$ is a non-convex problem. when **ADMM** solves both the convex and non-convex problems at the same time, its convergence property has not been explained. But our each sub-problem has a closed form solution and our algorithm has a great performance in the experiment. More convergence analysis about **PSSV** can be obtained in [13].

III. Experimental Results

In this section, we compare the performance of the proposed **PSSV-LRTC** algorithm against synthetic data and real word data with two related algorithms: simple low rank tensor **completion** (**SiLRTC**) and high accuracy low rank tensor completion (**HaLRTC**) [18]. **SiLRTC** solves a simple convex structure via block coordinate descent. **HaLRTC** solves the convex problem of **SiLRTC** based on **ADMM**. All experiments are performed in Matlab on a 2.4 GHz Intel Core i3 machine with 3 GB RAM.

A. Synthetic Data

For synthetic data, we first sysnthesize a ground-truth low rank 3-order tensor $\mathbf{O} \in \mathbf{P}^{125\times125\times125}$ whose rank $(r_1, r_2, r_3) = (1,1,1)$ based on the Tucker decomposition [17]. In all tests, we set $\theta = 1e^{-}3$, $\rho = 1.05$, $K = 500$, tolerance error $tol = 1e^{-}5$ and other parameters are same as [18]. The reconstruction performance is defined by the relative standard error: $RSE = \frac{\|\mathbf{\Lambda} - \mathbf{O}\|_F}{\|\mathbf{O}\|_F}$. We perform 10 time against each test and report the average result.

We randomly select 10%, 30%, 50%, 70% and 90% of $\mathbf{O}$ as known samples. The comparison results are shown in Fig.1.

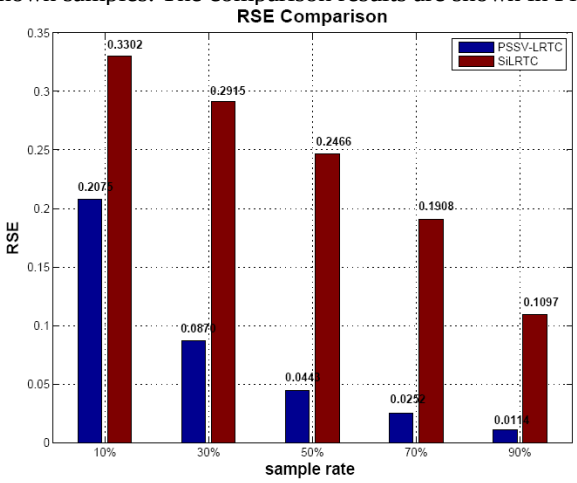

Fig. 1. RSE comparison for synthetic data between PSSV-LRTC and **SiLRTC** for sample rate-10%, 30%, 50%, 70%, 90% cases.

Fig.1 shows that our proposed **PSSV-LRTC** has a higher accuracy than SiLRTC for recovering the missing values of synthetic tensors under different sampling rates. Even in the case of 90% of the data is lost, we also can restore the original tensor successfully and *RSE* is about 0.2075.

B. Real Word Data

As Y Hu, D Zhang *ea al*. [20] remarked, the information of a picture is dominated by the top singular values, so most images can be viewed as approximately low rank matrices. Similarly, the color images also can be regarded as approximately 3-order low rank tensors and each color channel of it is defined as a mode of tensors. The followings are two recovery example of a baboon image with 50% missing entries and a facade image with block missing entries.

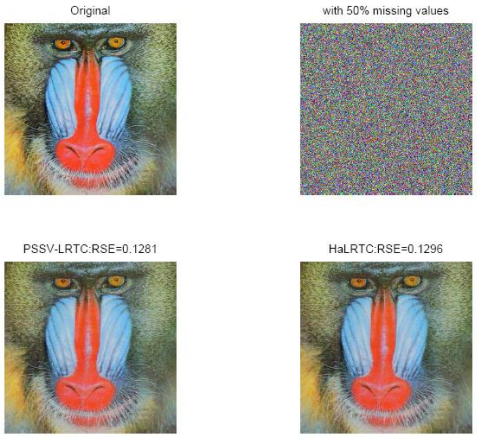

Fig. 2. Reconstruction results of the boboon image by **PSSV-LRTC** and **HaLRTC**.

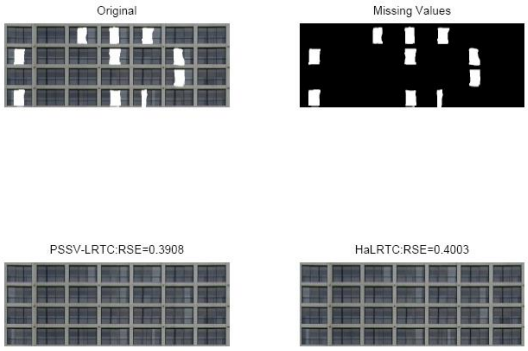

Fig. 3. Reconstruction results of the facade image by **PSSV-LRTC** and **HaLRTC**.

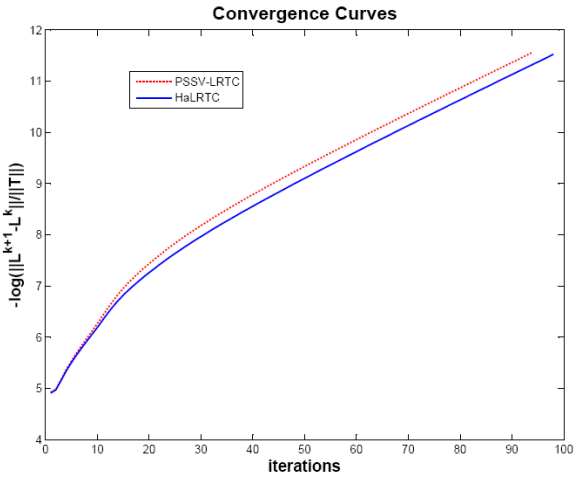

Fig. 4. Convergence curves of **PSSV-LRTC** and **HaLRTC** based on the example of Fig. 3.

Fig.2 and Fig.3 show that **PSSV-LRTC** and **HaLRTC** can successfully restore the picture with random missing values or block missing values. But our algorithm achieves much smaller *RSE* than **HaLRTC**. Fig.4 shows that although the gap between the *RSE* obtained by our algorithm and the **HaLRTC** algorithm is not large, our convergence rate outperforms the **HaLRTC** algorithm.

IV. Conclusion

In this paper, we extend the low rank matrix completion problem to the low rank tensor completion problem. Taking full advantage of the prior information about the rank of tensors, we creatively propose the partial sum minimization of singular values model of the low rank tensor completion and the **PSSV-LRTC** algorithm. By comparing our work with **SiLRTC** and **HaLRTC** in the simulation experiment and the practical application, it is demonstrated that our algorithm has a high ability of data recovery.

Acknowledgment

This work was supported by Natural Science Foundation of China under Grant number 61273020, Fundamental Research Funds for the Central Universities under Grant number XDJK2015A007.

References

[1] Candès E J, and Recht B, Exact matrix completion via convex optimization, Foundations of Computational Mathematics, 2008, 9(6):717-772.

[2] Keshavan R H, Oh S, and Montanari A, Matrix completion from a few entries, IEEE Transactions on Information Theory, 2009, 56(6):2980 - 2998.

[3] Cai J F, Candès, E J, and Shen Z, A singular value thresholding algorithm for matrix completion, Siam Journal on Optimization, 2010, 20(4):1956-1982.

[4] Bugeau A, Bertalmío M, Caselles V, and Sapiro G, A comprehensive framework for image inpainting, IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2010, 19(10):2634-45.

[5] Guillemot C, Meur O L, Image inpainting : overview and recent advances, IEEE Signal Processing Magazine, 2014, 31(1):127-144.

[6] ACM SIGKDD, Netflix, Proceedings of KDD Cup and Workshop 2007, Proceedings available online at http://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings.html.

[7] Lee C, Li Y, Monga V, Ghost-free high dynamic range imaging via rank minimization, IEEE Signal Processing Letters, 2014, 21(9):1045-1049.

[8] Zhou J, Robles-Kelly A, Image inpainting based on local optimisation, 2010:4440-4443.

[9] Zarif S, Faye I, Rohaya D, Fast and efficient multichannel image completion using local similarity, 2012:3116-3119.

[10] Shi J, Qi C, Sparse modeling based image inpainting with local similarity constraint, 2013:1371-1375.

[11] Candès E J, Tao T, The power of convex relaxation: near-optimal matrix completion, IEEE Transactions on Information Theory, 2010, 56(5):2053-2080.

[12] Wen Z, Yin W, Zhang Y, Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm, Mathematical Programming Computation, 2012, 4(4):333-361.

[13] Oh T H, Tai Y W, Bazin J C, Kim H, and Kweon I S, Partial sum minimization of singular values in robust PCA: algorithm and applications, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(4):744-758.

[14] Cai J F, Candès, E J, and Shen Z, A singular value thresholding algorithm for matrix completion, Siam Journal on Optimization, 2010, 20(4):1956-1982.

[15] Ma S, Goldfarb D, Chen L. Fixed point and Bregman iterative methods for matrix rank minimization, Mathematical Programming, 2011, 128(1-2):321-353.

[16] Toh, Kim-Chuan, Yun, Sangwoon, An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems, Pacific Journal of Optimization, 2010, 6(3):615-640.

[17] Kolda T G, Bader B W, Tensor decompositions and applications, Siam Review, 2009, 51(3):455-500.

[18] Liu J, Musialski P, Wonka P, Ye J, Tensor completion for estimating missing values in visual data, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):208-220.

[19] Lin Z, Chen M, Ma Y, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, Eprint Arxiv, 2010, 9.

[20] Hu Y, Zhang D, Ye J, Li X, and He X, Fast and accurate matrix completion via truncated nuclear norm regularization, IEEE Transactions on Software Engineering, 2013, 35(9):2117-30

