Fast Projection Singular Value Thresholding for Low Rank Optimization and Application

Shan Wang [a], Jianhui Hu [b], Chuyi Xu [c]

School of Information Engineering East China Jiaotong University Nanchang, China.

[a]patrick_shan@163.com, [b]Jianhuihu96@gmail.com, [c]1347807680qq.com

Keywords: Matrix Completion, fast projection singular value thresholding (FPSVT), adaptive thresholding, projection operator.

Abstract: Recovering an unknown matrix from corrupted observations is known as the matrix completion problem, it is fundamental to a number of tasks. However, existing most algorithms such as Singular Value Thresholding (SVT) heavily depend on the initialization, which will bring large computational complexity. In this paper, we propose a fast projection singular value thresholding (FPSVT) method, with which we can accelerate the iteration. The key idea is using a projection operator to get an improved initialization which is closer to the unknown optimal solution and using an adaptive thresholding in each iteration of our algorithm. We demonstrate the utility of the proposed method in numerical simulations. The experimental result gives empirical evidence on efficient improvements of the proposed algorithm.

1. Introduction

In recent years, matrix completion techniques have attracted much attention from researchers in many areas; such as computer vision [1], collaborative filtering [2], image inpainting [3]. It is impossible to recover a corrupted matrix without any assumptions about the matrix, Candes proposed in [4] if the given matrix is low rank, the missing entries of the corrupted matrix can be recovered through minimizing the matrix rank. The low rank assumption is reliable and useful in many areas such as collaborative filtering and face recognition [5]. The most famous collaborative filtering problem is Netflix problem [6], Netflix is a company which want to provide recommendations to its customers based on their preferences. But customers usually rate only very few movies that they have watched so there are few entries of the movies data matrix. In this situation, the movies data matrix can be regard as low rank because it is commonly that only a few data contribute to customers preferences. So, we can use matrix completion algorithm to construct the incomplete movies data matrix. in the face recognition problem,[7] the pixel of a face image can be regard as low rank because the different columns and rows may have the same pixel. The low rank face image is that we want to gain in face recognition problem.

Unfortunately, matrix completion problem is NP-hard because the rank is nonconvex in real. To solve the problem, the authors in [8] propose convex nuclear norm to solving rank minimization, and its theoretical guarantees have been provided in [9]. Semi-definite programming (SDP) is usually used to rank minimization problem, but directly realizing the SDP will get high computational cost. So, algorithms which are more computationally efficient than the SDP-based methods have been suggested, such as singular value thresholding (SVT) [10], accelerated proximal gradient (APG) [11], Singular Value Projection [12]. The method starts with an initialization, and therefore, the performance heavily depends on that initialization. But traditional algorithms ordinarily fill the missing entries with zeros and assume that corrupted matrix equals to the real matrix. In order to improve the performance, the authors in [13] propose an improved initialization based on rank-1 updates to solve the original and give theoretical guarantees which the initialization is closer to the optimal solution than the traditional all zeros initialization.

In order to solve the problem of matrix completion in dynamic scenarios, the authors in [14] introduce an adaptive singular value thresholding algorithm which based on the online linearized Bregman iteration algorithm. In [15-16], it is showed that the proposed algorithm is fast when using an iterative method with adaptive thresholding. The adaptive thresholding can accelerate the process of iteration which can efficient decrease the great computational cost of singular value decomposition (SVD).

In this paper our main contributions are stated as followings. First, a study of the proposed existing matrix completion algorithms and propose a new framework of matrix completion method, which can recover the corrupted data. Second, we develop a fast projection singular value thresholding algorithm and claim that it outperforms than the conventional matrix completion algorithm. Third a briefly analysis of the adaptive threshold in the proposed FPSVT algorithms and how to select the appropriate threshold.

The remainder of this paper is organized as follows. We review the SVT algorithm for recovering low rank matrices, then by combining the projection operator the FPSVT algorithm is proposed and analyzed in Section II. Numerical simulations are presented and discussed in Section III and finally Section IV concludes the paper.

Notation: The notations used in this paper are presented in Table 1.

Table 1. Summary of Nation

Nation	Description
$\mathbf{X} \in \boldsymbol{R}^{m\times n}$	Matrix with size $m \times n$
$\boldsymbol{P}_{\boldsymbol{\Omega}}$	the sampling operator
$\boldsymbol{X}_{\boldsymbol{ij}}$	i-th row and j-th column of X
$\boldsymbol{S}_{\boldsymbol{\tau}}$	the soft thresholding operator
$\boldsymbol{P}_{\boldsymbol{l}}$	the orthogonal projection operator
$\boldsymbol{\Omega}$	the number of known entries
$\|\boldsymbol{X}\|_*$	Nuclear norm of X
$\|\boldsymbol{X}\|_{\boldsymbol{F}}$	Forbenius norm of X

2. Problem Formulation

In this section, we first introduce the matrix completion problems, then briefly review and summarize the SVT and FPSVT algorithm.

2.1 The Form of Matrix Completion

Matrix completion is a technique to recover incomplete matrix from a subset of entries selected uniformly at random from a low rank matrix or approximately low rank matrix [2], [17-18]. The incomplete matrix M can be recovered by solving the following rank minimization problem in [2]:

$$\begin{aligned} &\text{minimize } rank(X) \\ &\text{subject to } P_\Omega(\mathrm{X}) = P_\Omega(\mathrm{M}), \end{aligned} \tag{1}$$

where rank(X) denotes the rank of a matrix X, the sampling operator $P_\Omega: R^{m\times n} \to R^{m\times n}$ is defined by following

$$P_\Omega(X) = \begin{cases} X_{ij} & (i,j) \in \Omega \\ 0 & (i,j) \in \Omega' \end{cases} \tag{2}$$

We use $|\Omega|$ represent the cardinality of Ω which is the number of known entries. For example, we suppose the matrix X is:

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \tag{3}$$

If we have three elements are known as $\Omega = \{(1,2),(2,2),(2,3)\}$, we can have:

$$X_\Omega = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 5 & 6 \end{bmatrix} \quad (4)$$

However, the problem (1) is NP-hard and impossible in practice. Candes proposed nuclear norm minimization model to solve the following rank minimization model

$$\begin{aligned} &\text{minimize } \|X\|_* \\ &\text{subject to } P_\Omega(\mathrm{X}) = P_\Omega(\mathrm{M}), \end{aligned} \quad (5)$$

where the nuclear norm $\|X\|_*$ is the summation of the singular values of X. Candes in [10] proved that if Ω is sampled uniformly at random among all subset of cardinality m, we can solve the problem (5) with large probability where the number of samples should obey $\mathrm{m} \geq 1.2\mathrm{Cn}rlogn$. In order to recover the incomplete matrix exactly, there is a restriction on the range of rank r.

2.2 The Algorithm of Singular Value Thresholding

In order to solve the matrix completion problem, Candes in [3] derived the singular value thresholding (SVT) algorithm. The singular value thresholding operator is proposed to solve the problem (5), which the soft-thresholding operator S_τ ($\tau \geq 0$) is defined as follows:

$$\begin{aligned} &S_\tau(\mathrm{X}) \coloneqq \mathrm{U}\, S_\tau(\Sigma)V^*, \\ &S_\tau(\Sigma) = diag(\{(\sigma_i - \tau)_+\}) \end{aligned} \quad (6)$$

where $(\sigma_i - \tau)_+$ is the positive part of $(\sigma_i - \tau)$, equal to $(\sigma_i - \tau)_+ = \max(0, (\sigma_i - \tau))$. The soft thresholding operator can simply apply to the singular value of X, shrinking these towards zero effectively. The singular value thresholding operator is the proximity operator conncened with the nuclear norm, which can express as follows: Starting with $Y^0 = 0 \in R^{m\times n}$, optimization variable X$\in R^{m\times n}$, k=1,2,3, …,

$$\begin{cases} X^k = S_\tau(Y^{k-1}) \\ Y^k = Y^{k-1} + \delta_k P_\Omega(M - X^k) \end{cases} \quad (7)$$

Where δ_k is the positive step sizes and the stopping criterion is determined by the step sizes. Unfortunately, SVT requires full singular value decomposition (SVD) of a $\mathrm{m} \times \mathrm{n}$ matrix at each iteration, lead to the high compleximal to deal with a large-scale matrix. In order to avoid the high computational complexity of the SVT, a fast SVT algorithm has always been wanted in various large-scale problems.

3. Proposed Methods

3.1 Fast Projection Singular Value Thresholding

Considering the mathematical model (5), we apply the orthogonal projection operator $P_l(\cdot)$ proposed in [12] to deal with the original matrix, which can be derived as follows:

$$\begin{aligned} &\text{minimize } \|X'\|_* \\ &\text{subject to } P_\Omega(\mathrm{X}) = P_\Omega(\mathrm{M}), X' = P_l(X) \end{aligned} \quad (8)$$

where $X' = P_l(X) \in R^{m\times n}$ is a projected low rank matrix, which the orthogonal projection operator $P_l(\cdot)$ is defined as follows:

$$X^{t+1} \leftarrow P_l(X^t - \eta_t A^T(A(X^t) - b))$$

s.t

$$\mathrm{A}\ (X)\text{=b} \quad \mathrm{X} \in R^{m\times n}, b \in R^d \quad (9)$$

where A is an affine transformation from $R^{m\times n}$ to R^d, η_t is the step-size.

Algorithm 1. Fast projection singular value thresholding (FPSVT) algorithm.

Require: matrix sampled entries $\boldsymbol{P_\Omega}(\mathbf{M})$ and corrupted entries $\boldsymbol{P_{\Omega'}}(\mathbf{M})$, index set $\boldsymbol{\Omega}$, maximum iterations K, stopping tolerance $\boldsymbol{\epsilon}$ and step-size $\boldsymbol{\delta}$.
1: Initialization: $\boldsymbol{P_\Omega}(\mathbf{X}) = \boldsymbol{P_\Omega}(\mathbf{M})$ $\boldsymbol{Y^0} = \boldsymbol{P_l}(\boldsymbol{X})$
2: **for** k=1, …, K **do**
3: compute the top l singular vectors of $\boldsymbol{Y^k}$ to obtain **Z:** $\boldsymbol{Z^k} \leftarrow \boldsymbol{U_l \Sigma_l V_l}$
4: Update the corrupted matrix $\boldsymbol{X^{k+1}} = \boldsymbol{S_\tau}(\boldsymbol{Z^k})$
5: **if** $\left\|\boldsymbol{P_\Omega}(\boldsymbol{X^k} - \boldsymbol{M})\right\|_{\boldsymbol{F}} / \|\boldsymbol{P_\Omega}(\boldsymbol{M})\|_{\boldsymbol{F}} \leq \boldsymbol{\epsilon}$, **break;**
Set $\boldsymbol{Z^{k+1}} = \boldsymbol{Z^k} + \boldsymbol{\delta P_\Omega}(\boldsymbol{M} - \boldsymbol{X^k})$
6: **end for**
7: output: Z= $\boldsymbol{P_{\Omega'}}(\mathbf{Z}) + \boldsymbol{P_\Omega}(\mathbf{M})$

In general, SVT algorithm requires SVD computations which occupies the largest computation cost, $i, e, \mathrm{O}(\mathrm{mn}, \min(\mathrm{m}, \mathrm{n}))$ at each iterate. In algorithm 1 we apply the orthogonal projection operator $P_l(\cdot)$ to accelerate general nuclear norm minimization (NNM) problem, it can avoid the unexpected expensive computation by applying the orthogonal projection operator $P_l(\cdot)$ to deal with in completed original matrix.

The FPSVT algorithm pseudo-code is found in algorithm 1, which computation iterates the following two steps: 1) applying orthogonal projection operator $P_l(\cdot)$ to original matrix, 2) using adaptively threshold for SVD at each iteration in SVT. For SVT, the computational bottleneck in each iteration is the SVD computation in step 1. By exploiting the orthogonal projection operator $P_l(\cdot)$, our method efficiently reduces iteration count before convergence. In addition, by using adaptively threshold for SVD, we can further reduce the computation of SVT as described in Sec B.

3.2 Adaptive Thresholding

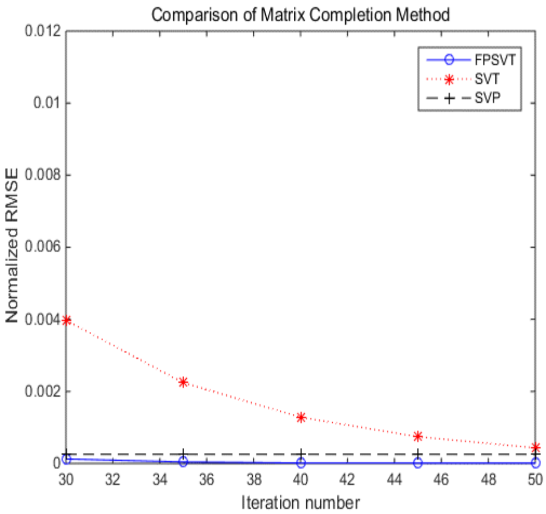
In order to accelerated the process of iteration in FPSVT, we use an iterative method with adaptive thresholding to recovery the incomplete matrix. The method of adaptive thresholding is adopted from [15] which use a thresholding operator to reduce the computation complexity. To obtain a fast and accurate approximation method for SVT, the thresholding operator can be expressed an exponential thresholding scheme as follows:

$$\tau = \mathrm{a}\sigma_i e^{-bk} \tag{10}$$

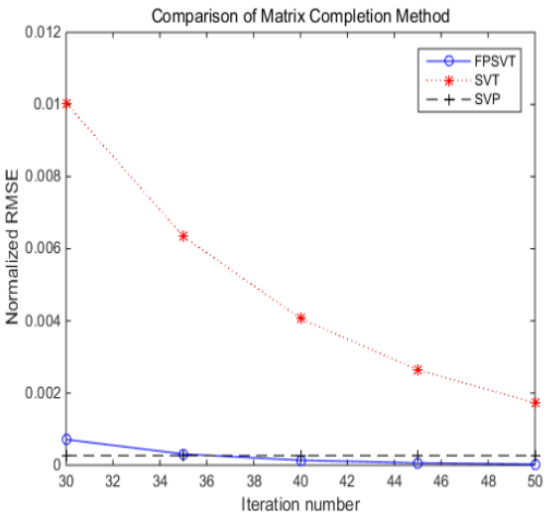
where a and b are two constants, σ_i is the largest singular value of original matrix and k is the maximum iterations. Note the b is supposed to set a value less than 1 to be sure that there has conpoent will be use at the first iteration. When we choose the constant a and b much greater, the iteration of our algorithm will reduce sooner but the performance of the algorithm will become bad corresponding. And when choose the constant smaller, the iteration of our algorithm will terminate slower and the performance will become bad too.

4. Experimental Results

In this section, we will evaluate the efficiency of the FPSVT algorithm with other algorithm using synthetic data. All the simulations were implemented in MATLAB2014b and performed on a computer with a 2.5GHz CPU and 4GB memory. The same shared parameters proposed in original paper were used in different algorithms.



(a)



(b)

Fig 1. Normalized RMSE versus iteration number. (a) Synthetic matrix (1000× 1000, k=10). (b) Synthetic matrix (1000× 1000, k=20).

4.1 Synthetic Random Data

In our experimental, we evaluate our method in comparison to other method with synthetic random data which is obtained from the synthetic matrices: $\mathrm{G} \in R^{m\times n}$ with rank k by using two random matrices $\mathrm{G}_1 \in R^{m\times k}$ and $G_2 \in R^{k\times n}$ whose entries satisfy the standard Gaussion distribution. The synthetic matrices G can be set in following: $\{1000 \times 1000(\mathrm{k} = 10)$, $1000\times 1000(\mathrm{k} = 20)$. According to [9] the minimum sample ratio for guaranteeing the exact recovery result is (O (Nklog (N))/(m× n)), so the minmun sample ratio of synthetic matrices G are: {O (6.91%), O (13.82%). In this simulated experimental, we set the maximum iterations N from 30 to 50 for all the methods to test the three algorithms and the sampling ratio is set to 30% in all simulated data. We use stopping criteria $\|X^{N+1} - X^N\|_F / \|X^N\|_F$ and terminate the algorithm if the stopping criteria or iterationd is met. We find that tol=1e-04 is small enough to obtain a reasonable result for the three algorithms, so the stopping torerance tol=1e-04 is set for all algorithm. Other parameters of the compared algorithm are set as the original paper.

To quantitatively evaluate the four reconstruction methods, the normalized root meat square error (NRMSE) is used to assess their reconstruction accuracy. The normalized root meat square error (NRMSE) is defined as following:

$$\mathrm{NRMSE(M)} = \sqrt{E\left\{\frac{\|M-G\|_F^2}{\|G\|_F^2}\right\}} \tag{11}$$

where M is the result constructed by matrix completion and G is the original matrix, the $\|\cdot\|_F$ is the Frobenius norm of matrix.

Figure 1 plots the NRMSE versus iteration number (30-50) with SVT, SVP, FPSVT methods for two synthetic matrices:1000$\times$ 1000(k=10) and 1000$\times$ 1000(k=20). It is clear that SVP and our method converge much faster than original SVT algorithm and only about thirty iterations are needed to obtain an accepted solution. Although, SVP have the same iteration as our method to obtain an accurate solution, the FPSVT is slightly inferior to the SVP because of having a better performance to the simulated data.

5. Conclusion

In this paper we have proposed a fast projection singular value thresholding (FPSVT) algorithm termed SVD based on random projection to recover the corrupted temperature field data (simulated by uniformly and random sampling). A projection operator is used for improving the initialization which is closer to the optimal solution, and an adaptive thresholding is used to speed up the process of iteration in FPSVT. The experimental results show that the proposed FPSVT method outperforms the SVT and SVP algorithm in solving the corresponding convex program in terms of synthetic random data.

Acknowledgments

This work was supported in part by Major Program of National Natural Science Foundation of China (No.91437220); Jiangxi Province Science Foundation for Youths (No.20171ACB21038); Jiangxi Municipal Science and Technology Project (No.20171ACG70017).

References

[1] Gu S, Zhang L, Zuo W, et al. Weighted Nuclear Norm Minimization with Application to Image Denoising[C] IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:2862-2869.

[2] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. 2009 (2009) 4:2–4:2, doi: 10.1155/2009/421425.

[3] Peng Y, Suo J, Dai Q, et al. Reweighted low-rank matrix recovery and its application in image restoration. [J]. IEEE Transactions on Cybernetics, 2014, 44(12):2418-2430.

[4] Candès E J, Recht B. Exact Matrix Completion via Convex Optimization[J]. Foundations of Computational Mathematics, 2008, 9(6):717.

[5] R. Basri and D. W. Jacobs, Lambertian reflectance and linear subsp-Aces[J]. IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 2, pp. 218–233, Feb. 2003.

[6] Proceedings of KDD Cup and Workshop. ACM SIGKDD and Netflix, 2007, proceedings available online at http://www.cs.uic.edu/∼liub/KDD-cup2007/proceedings.html.

[7] Ong F, Lustig M. Beyond Low Rank + Sparse: Multiscale Low Rank Matrix Decomposition[J]. IEEE Journal of Selected Topics in Signal Processing, 2015, 10(4):672-687.

[8] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” SIAM Rev., vol. 52, no. 3, pp. 471–501, 2010.

[9] E. J. Candès and T. Tao, “The power of convex relaxation: Near optimal matrix completion,” IEEE Trans. Inf. Theory, vol. 56, no. 5, pp. 2053–2080, May 2010.

[10] Cai J F, Candès E J, Shen Z. A Singular Value Thresholding Algorithm for Matrix Completion[J]. Siam Journal on Optimization, 2008, 20(4):1956-1982.

[11] Chen M, Ganesh A, Lin Z, et al. Fast Convex Optimization Algorithms for Exact Recovery of a Corrupted Low-Rank Matrix[J]. Journal of the Marine Biological Association of the Uk, 2009, 56(3):707-722.

[12] Meka R, Jain P, Dhillon I S. Guaranteed Rank Minimization via Singular Value Projection[J]. Nips, 2012:937-945.

[13] Douik A, Hassibi B. An Improved Initialization for Low-Rank Matrix Completion Based on Rank-L Updates[C]// 2018:3959-3963.

[14] Tripathi R, Mohan B, Rajawat K. Adaptive Low-Rank Matrix Completion[J]. IEEE Transactions on Signal Processing, 2017, 65(14):3603-3616.

[15] Zarmehi N, Marvasti F. Sparse and low-rank recovery using adaptive thresholding[J]. Digital Signal Processing, 2018, 73:145-152.

[16] Tripathi R, Mohan B, Rajawat K. Adaptive Low-Rank Matrix Completion[J]. IEEE Transactions on Signal Processing, 2017, 65(14):3603-3616.

[17] E.J. Cand, X. Li, Y. Ma, J. Wright, Robust principal component analysis? [J]. ACM 58 (3) (2011) 1–37.

[18] Keshavan R H, Montanari A, Oh S. Matrix Completion From a Few Entries[J]. IEEE Transactions on Information Theory, 2010, 56(6):2980-2998.