

# Limitations of Matrix Completion via Trace Norm Minimization

Xiaoxiao Shi  
Computer Science Department  
University of Illinois at Chicago  
xiaoxiao@cs.uic.edu

Philip S. Yu  
Computer Science Department  
University of Illinois at Chicago  
psyu@cs.uic.edu

## ABSTRACT

In recent years, compressive sensing attracts intensive attentions in the field of statistics, automatic control, data mining and machine learning. It assumes the sparsity of the dataset and proposes that the whole dataset can be reconstructed by just observing a small set of samples. One of the important approaches of compressive sensing is trace norm minimization, which can minimize the rank of the data matrix under some conditions. For example, in collaborative filtering, we are given a small set of observed item ratings of some users and we want to predict the missing values in the rating matrix. It is assumed that the users' ratings are affected by only a few factors and the resulting rating matrix should be of low rank. In this paper, we analyze the issues related to trace norm minimization and find an unexpected result that trace norm minimization often does not work as well as expected.

## 1. INTRODUCTION

Compressive sensing has been proposed to solve sparse matrix problem in many fields including machine learning, automatic control, and image compression. It applies the sparsity of the datasets and assumes that the whole datasets can be recovered by just observing a small set of the data. One of the important approaches is minimizing the rank of a matrix variable subject to certain constraints. For example, in collaborative filtering (CF), the objective is to predict the missing values of a rating matrix. It is assumed that the ratings are affected by only a few of factors and the rating matrix should be of low rank. Given that directly minimizing the rank of the matrix is NP hard, a commonly-used convex relaxation of the rank function is the trace norm, defined as the sum of the singular values of the matrix.

Recently, a number of work has shown that the low rank solution can be recovered exactly by minimizing the trace norm [12; 6]. Furthermore, trace norm minimization starts to be widely used in many scenarios, such as multi-task learning [1; 3], multivariate linear regression [20], matrix classification formulation [16; 4], multi-class classification [2], etc. Among all the applications, an important one is matrix completion [15; 13; 17; 11; 5; 9], which aims at recovering a low-rank matrix based on a small portion of observed entries. Recently in [10], the two-way matrix completion problem is further generalized into a tensor completion problem and trace norm minimization is adapted to solve the problem.

However, in this paper, we present an unexpected result that despite its popularity, the trace norm minimization does not work well in general. We show the limitations of trace norm minimization in several situations and analyze the reasons behind the failures. We mainly focus on two issues related to trace norm minimization. First, only with a small portion of observed samples, it is difficult to determine whether or not the real matrix is of low rank. The results will deteriorate significantly for a high rank matrix. Second, we analyze that rank minimization or trace norm minimization can lead to multiple solutions, which may produce unstable results.

The paper is organized as follows. A formal introduction of trace norm minimization is first given, followed by the argument of its limitations with examples. Three sets of experiments are then performed to analyze the empirical results in applications including collaborative filtering and image recovery. A conclusion is made with possible future works of trace norm minimization.

## 2. TRACE NORM MINIMIZATION

We use capital-bold letters such as  $\mathbf{X}$ ,  $\mathbf{M}$  as matrices, and  $\mathbf{X}_{ij}$  denotes the  $(i, j)$ -th element of  $\mathbf{X}$ . In the problem definition, we further denote the observed matrix as  $\mathbf{M}$  and the observed entries as  $\mathbf{M}_{ij}$  where  $(i, j) \in \Omega$ . In other words, if  $(m, n) \notin \Omega$ , the value of  $\mathbf{M}_{mn}$  is missing.

In compressive sensing [7], matrix completion can be usually stated as

$$\begin{aligned} & \text{minimize} && \text{rank}(\mathbf{X}) \\ & \text{subject to} && \mathbf{X}_{ij} = \mathbf{M}_{ij} \quad (i, j) \in \Omega \end{aligned} \quad (1)$$

where  $\mathbf{M}$  is the observed matrix with some missing values, and  $\mathbf{X}$  is the resulting matrix with all missing values filled. The objective is to find a matrix  $\mathbf{X}$  which has low rank after selectively filling the missing values. However, the optimization function in Eq. 1 is not convex and may require intensive computation to obtain the result. As an alternative, one usually solves the following optimization problem which is the tightest convex relaxation of Eq. 1.

$$\begin{aligned} & \text{minimize} && \|\mathbf{X}\|_* \\ & \text{subject to} && \mathbf{X}_{ij} = \mathbf{M}_{ij} \quad (i, j) \in \Omega \end{aligned} \quad (2)$$

where  $\|\mathbf{X}\|_*$  is the nuclear norm, or trace norm of the matrix  $\mathbf{X}$ , which is the sum of its singular values. A more general trace norm minimization problem can be stated as

$$\text{minimize} \quad f(\mathbf{X}) + \lambda \|\mathbf{X}\|_* \quad (3)$$

where  $\lambda$  is a parameter to control the confidence that the

Table 1: Summary of Jester Recommendation datasets.

Name	#Users (#Rows)	#Jokes (#Columns)	Descriptions
Jester I	24,983	101	Data from 24,983 users who have rated 36 or more jokes
Jester II	23,500	101	Data from 23,500 users who have rated 36 or more jokes
Jester III	24,938	101	Data from 24,938 users who have rated between 15 and 35 jokes

matrix  $\mathbf{X}$  is of low rank. And  $f(\mathbf{X})$  can be any evaluation function on the matrix  $\mathbf{X}$ , and in matrix completion,  $f(\mathbf{X})$  can be defined as the difference between  $\mathbf{X}_{ij}$  and  $\mathbf{M}_{ij}$  for  $(i, j) \in \Omega$ .

There are various approaches proposed to solve Eq. 2 or Eq. 3. For example, they can be formulated as a semidefinite program [8; 15]; or some approximation can be employed to deal with the non-smooth trace norm term [13; 17; 18; 11]. Recently, [9] proposes a subgradient method to approach the optimal solution; and [5] proposes a fast and simple algorithm which performs a thresholding in the singular values to obtain the optimal result. In the experiment section, we will use these two algorithms as examples to study trace norm minimization.

## 2.1 Limitations of Trace Norm Minimization

Trace norm minimization has strong mathematical foundations and has been used in various applications such as matrix completion. However, there are still issues related to this category of approaches.

First, it is not clear what type of matrices can be well recovered by trace norm minimization. In practice, we usually have to assume that the resulting matrix has low rank and we expect that trace norm minimization can lead to a good recovery. However, just observing a small set of samples (e.g., 10% of the matrix entries can be observed), it is difficult to guarantee that it is of low rank. In [19], it is further assumed that those rating matrices<sup>1</sup> in collaborative filtering are of low rank because it is assumed to be affected by just a few factors. Later experiment in Section 3 shows that this assumption may not be valid in all cases. A more precise criterion is needed to calibrate which type of matrices are suitable to be recovered by trace norm minimization.

Second, trace norm minimization or rank minimization can lead to multiple alternative solutions, which make the result unstable. We illustrate an example in Eq. 4. It is an incomplete matrix with one missing value (denoted as “?”). Note that most of the traditional matrix completion approach will assign the missing value as 2 because it has more 2s than 4s in the second column. However, to complete the matrix with rank minimization, we can either fill in the missing value with 2 or 4. Both solutions lead to a rank-2 matrix. In other words, there may be multiple solutions that can induce the recovered matrix to be of the same rank. Intuitively, with more missing values, it has higher probability that there are multiple solutions that can lead to the same matrix rank. Hence, the result can be very unstable because it can end up in any of the solutions. We further study this

problem in Section 3.1 and Section 3.3.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 4 & 3 \\ 1 & ? & 3 \end{bmatrix} \quad (4)$$

## 3. EXPERIMENTS

In this section, three sets of experiments are devised to analyze the performance of trace norm minimization.

### 3.1 Synthetic Dataset

We first design a simple matrix as in Eq. 5 and drop part of the entries with “?” as in Eq. 6. We then apply one of the state-of-the-art trace norm minimization approaches, singular value thresholding (SVT [5]), to recover the matrix as in Eq. 7. It is clear that none of the entry has been recovered correctly. However, the resulting matrix is of the same rank as the original matrix. Hence, in the view of low rank approximation or trace norm minimization, the objective has been achieved. But in the view of ground truth, the result is far from satisfactory. This phenomenon is also analyzed in Section 2.1: trace norm minimization can lead to multiple unstable solutions. Similar phenomenon is also studied in tensor completion case in Section 3.3.

$$\begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ ? & ? & ? & 1 \\ 1 & ? & ? & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix} \quad (6)$$

$$\begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix} \quad (7)$$

### 3.2 Collaborative Filtering via Matrix Completion

In this section, we apply two state-of-the-art trace norm minimization approaches to perform matrix completion in collaborative filtering (CF). The two approaches are singular value thresholding (SVT [5]) and multivariate linear regression described in [9] (MultiRegression). Both approaches solve matrix completion with the constraint to minimize the rank of the final matrix. SVT works well in a simulated

<sup>1</sup>those matrices record users’ ratings on a set of items. Each row describes all the item ratings of one user; each column represents the ratings of one item among all users.

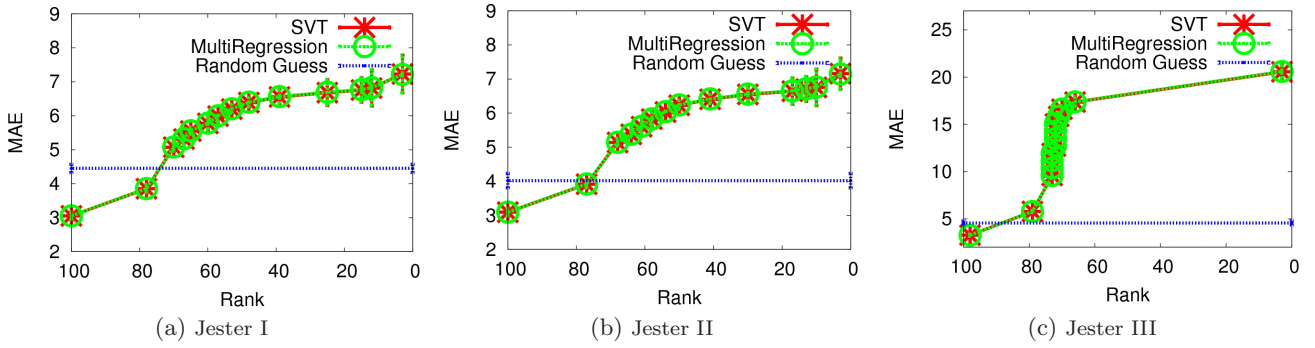


Figure 1: Parameter sensitivity

matrix completion dataset [5] and MultiRegression works well in a multi-task learning application which minimizes the trace norm of the shared parameter matrix among tasks [9]. In this experiment, we apply the two approaches to tackle the matrix completion problem in CF. The dataset used in the experiment is the Jester Recommendation dataset<sup>2</sup>, which is known as one of the benchmarks in CF. The dataset contains three separate sub-datasets which are summarized in Table 1. The Jester dataset describes millions of continuous ratings (-10.00 to +10.00) of 100 jokes from 73,421 users, which are collected between April 1999 to May 2003. As shown in Table 1, most of the users only rate a small portion of the items (sparsity); hence the final matrix should have good chance to be of low rank, and the two approaches should work well. In the experiment, 70% of the ratings are blocked, and we use trace norm minimization to recover the whole matrix by using the 30% of known ratings. The experiment results are evaluated by the Mean Absolute Error (MAE). If  $p_{ij}$  is the prediction for how user  $i$  will rate item  $j$ , the MAE for user  $i$  is defined as

$$\text{MAE} = \frac{1}{c} \sum_{j=1}^c |\tilde{r}_{ij} - p_{ij}|$$

where  $c$  is the number of items user  $i$  has rated and  $\tilde{r}_{ij}$  is the ground truth. MAE for a set of users is the average MAE over all members of that set. Hence, a small MAE value indicates a good prediction. Fig. 1 summarizes the results on 10 runs where in each run we randomly sample 70% of the matrix entries to be blocked. Since both approaches can be tuned to control the confidence that the final matrix is of low rank, we plot the result with the decreasing rank. Meanwhile, we plot the result of random guessing, which randomly assign values to the blocked entries.

There are two conclusions can be drawn from Fig. 1. First, the results given by SVT and MultiRegression almost overlap because both of them solve the same optimization function and MultiRegression embeds SVT as a sub-function<sup>3</sup>. Second, the errors evaluated by MAE grow with the decreasing ranks. The best results are given when the ranks are large. When the ranks are smaller than 80, the results are worse than random guessing. It is important to emphasize that although this benchmark dataset should have good chance to be of low rank (sparse collaborative filtering matrix as analyzed in [19]), its best result is given with full

rank. Thus, a more precise criterion is needed to determine what type of matrices is suitable to be recovered by trace norm minimization.

### 3.3 Image Recovery via Tensor Completion

In [10], trace norm minimization is further extended to handle multidimensional matrix completion, or tensor completion. More specifically, trace norm minimization is applied to recover images which are expressed as 3-dimensional tensors. The experiment results in [10] show that the approach can effectively recover some blurred images. In this experiment, we change to another set of images from Caltech 256<sup>4</sup> benchmark dataset. Fig. 2(a) and Fig. 3(a) show the two images used in the experiment. We further erase a small part of the images as shown in Fig. 2(b) and Fig. 3(b). The approach in [10] is applied to recover the images by filling pixels in the blanks, and the results are shown in Fig. 2(c) and Fig. 3(c) after 2000 iterations. We further plot Fig. 2(d) and Fig. 3(d) to show that the convergence of the optimization function of the approach, where x-axis denotes the number of iterations and y-axis is the value of the objective function. Hence, although the algorithm has reached the optimal or near optimal solution as shown in Fig. 2(d) and Fig. 3(d), the results of the recovered images are not encouraging. It is important to mention that both of the images are quite simple especially the one in Fig. 2(a), which has clear patterns which are similar as the matrix in Eq. 5. However, traditional trace norm minimization approach fails to recover the images correctly. The reason of the poor performance of Fig. 2 is similar as the matrix in Eq. 5. Trace norm minimization can lead to multiple optimal solutions and end up in any of them; the result can be arbitrary.

## 4. CONCLUSION

We study and analyze the limitations of trace norm minimization in matrix completion. First, it is difficult to guarantee that the real matrix should be of low rank. Second, trace norm minimization can lead to arbitrarily satisfied matrix with low rank. Three sets of experiments are analyzed in the application of collaborative filtering and image recovery. For example, in collaborative filtering, the performance of trace norm minimization drops with decreasing matrix rank. In summary, the paper aims at studying the unexpected is-

<sup>2</sup><http://eigentaste.berkeley.edu/dataset/>

<sup>3</sup><http://www.public.asu.edu/~jye02/Software/SLEP/index.htm>

<sup>4</sup><http://www.vision.caltech.edu/Image-Datasets/Caltech256/>

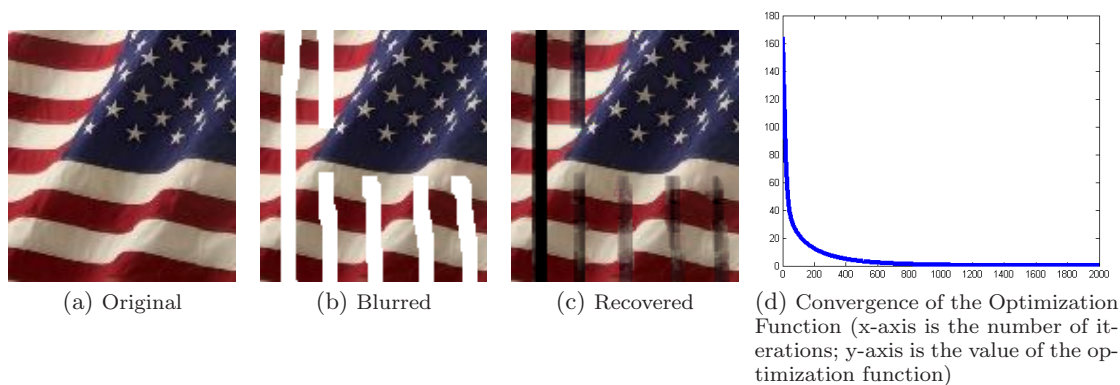


Figure 2: Flag Image Recovery

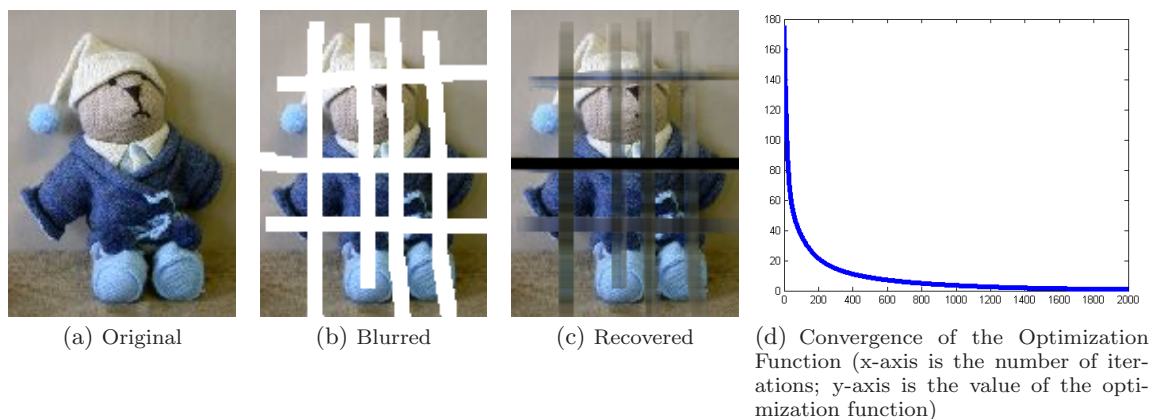


Figure 3: Teddybear Image Recovery

sues when intending to put trace norm minimization into practice:

- What was unexpected: Although trace norm minimization attracts intensive attentions in the field of statistics and machine learning, it only works under a very strict assumption; in other words, it does not work in general.
- Why it was unexpected: (1) Trace norm minimization methods make the assumption that the matrices are of low rank, which may not be true in reality; (2) trace norm minimization can lead to multiple unstable solutions.
- What was learned: Trace norm minimization does not work in general. A more in-depth criterion is needed to determine when it works before using it in practice.
- Possible advice for practitioners: If one intends to make trace norm minimization into practice, one may have to study deeply on the objective matrix and make sure (1) the matrix follows the assumption of trace norm minimization; (2) it does not have multiple low rank solutions.

As a future work, a more precise criterion is needed to determine what type of matrices is suitable to be recovered via trace norm minimization. One of the possibility is to

develop a lower bound to study the number of possible low rank solutions to a given incomplete matrix, and only use trace norm minimization when the value of the bound is small. Another extension is to bring in more constraints to lead to a better and unique low rank solution. For instance, the missing value can be filled in with the majority value, in addition of satisfying the low rank constraint. As an example, the missing value in Eq. 4 can be filled with “2” under the majority-value constraint.

## 5. REFERENCES

- [1] J. Abernethy, F. Bach, T. Evgeniou, and J. P. Vert. Low-rank matrix factorization with attributes (*Technical Report N24/06/MM*). Ecole des Mines de Paris. 2006
- [2] Y. Amit, M. Fink, N. Srebro, and S. Ullman. Uncovering shared structures in multiclass classification. *In Proceedings of the International Conference on Machine Learning*. 2007
- [3] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Machine Learning*, 73, 243~272. 2008
- [4] F. R. Bach. Consistency of trace norm minimization. *J. Mach. Learn. Res.*, 9, 1019~1048. 2008

- [5] J. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. on Optimization* 20(4), 1956-1982. 2008
- [6] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. (*Technical Report 08-76*). *UCLA Computational and Applied Math.* 2008
- [7] M. Fazel. Matrix Rank Minimization with Applications. *PhD thesis.* 2002
- [8] M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, 4734~4739. 2001
- [9] S. Ji and J. Ye. An accelerated gradient method for trace norm minimization. In *Proceedings of the 26th Annual International Conference on Machine Learning.* 2009
- [10] J. Liu, P. Musialski, P. Wonka, and J. Ye. Tensor completion for estimating missing values in visual data. *ICCV'09.* 2009
- [11] S. Ma, D. Goldfarb, and L. Chen. Fixed point and Bregman iterative methods for matrix rank minimization (*Technical Report 08-78*). *UCLA Computational and Applied Math.* 2008
- [12] B. Recht, W. Xu, and B. Hassibi. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *Proceedings of the 47th IEEE Conference on Decision and Control.* 2008
- [13] J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. In *Proceedings of the International Conference on Machine Learning.* 2005
- [14] N. Srebro. Learning with Matrix Factorizations. *PhD Thesis, MIT.* 2004
- [15] N. Srebro, J. D. M. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems.* 2005
- [16] R. Tomioka and K. Aihara. Classifying matrices with a spectral regularization. In *Proceedings of the International Conference on Machine Learning.* 2007
- [17] M. Weimer, A. Karatzoglou, Q. Le, and A. Smola. COFIRank—maximum margin matrix factorization for collaborative ranking. In *Advances in Neural Information Processing System.* 2008
- [18] M. Weimer, A. Karatzoglou, and A. Smola. Improving maximum margin matrix factorization. *Machine Learning*, 72, 263~276. 2008
- [19] K. Yu, S. Zhu, J. Lafferty, and Y. Gong. Fast Nonparametric Matrix Factorization for Large-scale Collaborative Filtering. *SIGIR'09.* 2009
- [20] M. Yuan, A. Ekici, Z. Lu, and R. Monteiro. Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B*, 69, 329~346. 2007