Hybrid Singular Value Thresholding for Tensor Completion

Xiaoqin Zhang[1*], **Zhengyuan Zhou**[2*], **Di Wang**[1], **Yi Ma**[3]
[1]Institute of Intelligent System and Decision, Wenzhou University, Zhejiang, China
[2]Department of Electrical Engineering, Stanford University, CA, USA
[3]Department of Electrical and Computer Engineering, ShanghaiTech University, Shanghai, China

Abstract

In this paper, we study the low-rank tensor completion problem, where a high-order tensor with missing entries is given and the goal is to complete the tensor. We propose to minimize a new convex objective function, based on log sum of exponentials of nuclear norms, that promotes the low-rankness of unfolding matrices of the completed tensor. We show for the first time that the proximal operator to this objective function is readily computable through a hybrid singular value thresholding scheme. This leads to a new solution to high-order (low-rank) tensor completion via convex relaxation. We show that this convex relaxation and the resulting solution are much more effective than existing tensor completion methods (including those also based on minimizing ranks of unfolding matrices). The hybrid singular value thresholding scheme can be applied to any problem where the goal is to minimize the maximum rank of a set of low-rank matrices.

Introduction

In the past few years, an explosion of massive amounts of high-dimensional data has permeated a large number of scientific and engineering fields. The curse of dimensionality has presented a challenging task to process such massive data. However, the blessing of dimensionality has provided a great opportunity: real data tends to lie in or near some low-dimensional subspace or manifold even though the ambient space is very high-dimensional. Consequently, it has been an important theme to seek and recover low-dimensional structures in high-dimensional data.

An important body of work in this area is the low-rank matrix recovery, which has been extensively studied in the recent literature and for which we provide a (highly incomplete) survey. (Candès and Recht 2009) studied the problem of recovering a low-rank matrix from incomplete observations on a given data matrix and gave theoretical conditions on when exact recovery of such a low-rank matrix is achievable by using nuclear norm minimization as the convex surrogate. A singular value thresholding algorithm was introduced in (Cai, Candès, and Shen 2010) to efficiently solve the convex program. Later, a series of articles (Keshavan, Montanari, and Oh 2010; Candès and Tao 2010; Recht 2011; Hu et al. 2013) have provided tighter bounds for the matrix completion problem. (Keshavan, Montanari, and Oh 2010) and (Candès and Plan 2010) have dealt with the case where the observed entries in the data matrix suffer from noisy corruptions. (Recht, Fazel, and Parrilo 2010) investigated the issue of matrix completion under linear sensing matrices model. (Gross 2011) gave a tight bound on matrix recovery under the model of randomly sampling expansion coefficients with respect to any given matrix basis. Another important and related line of work is Robust Principle Component Analysis (RPCA) developed in (Candès et al. 2011; Chandrasekaran et al. 2011). In this context, one seeks to decompose a data matrix into a sum of a low-rank matrix and a sparse error matrix. A series of follow-up papers including (Zhou et al. 2010; Ganesh et al. 2012; Wright et al. 2013) have made this model stable and have considered extensions where linear measurements of the data matrix are incomplete. While such theoretical endeavors have witnessed a success, algorithmic efforts to efficiently solve the proposed convex optimization programs have also surged. Fast algorithms proposed in (Lin, Chen, and Ma 2010; Lin et al. 2009) have proven effective in solving large low-rank matrix recovery problems, thus enabling the benefits guaranteed by the theoretical results to be enjoyed by real-world applications. In particular, first-order methods such as alternating proximal gradient method (Ma 2012) and alternating direction method of multipliers (Luo 2012) have regained attention in low-rank matrix recoveries due to its efficiency, especially in large-scale problems.

Although the low-rank matrix recovery problem has been well studied, there is not much work on tensor recovery. In fact, most of the existing high-dimensional data can be naturally formed as tensors (e.g. color images, videos, hyperspectral images, high-dynamical range images etc.), and directly graying and vectorizing images will lose much useful information. For tensorial data, one major challenge lies in an appropriate definition of the tensor rank. Traditionally, there are two definitions of tensor rank, which are based on CP (CANDECOMP/PARAFAC) decomposition (Kruskal 1977) and Tucker decomposition (Kolda and Bader 2009) respectively. Similar to the rank definition of a matrix, the CP rank of a tensor based on CP decompo-

*The first two authors contributed equally to this work.

sition is defined as the minimum number of rank-one decompositions of a given tensor. However, this rank definition is a nonconvex function, and direct minimization of this function is an NP-hard problem. As a result, Tucker decomposition of a tensor has been proposed, resulting in a new rank definition of a tensor, the Tucker rank, which is simply a vector of ranks of the unfolding matrices of a given tensor. Due to the recent breakthroughs in low-rank recovery of matrix (Cai, Candès, and Shen 2010), the latter definition has received more attention. (Gandy, Recht, and Yamada 2011) adopted the sum of the ranks of the different unfolding matrices as the rank of the tensor data, in which the rank of the unfolding matrix is approximated by the nuclear norm. (Signoretto, De Lathauwer, and Suykens 2010) proposed a general Shatten-$\{p, q\}$ norm for tensors and discussed the relationship between the proposed metric and the nuclear norm. A weighted sum of the ranks of the unfolding matrices was introduced by (Liu et al. 2013) and three optimization algorithms were proposed to estimate missing values in tensors of visual data. Three different strategies were developed to extend the trace-norm regularization to tensors (Tomioka, Hayashi, and Kashima 2010): (1) tensors treated as matrices; (2) constrained optimization of low-rank tensors; (3) mixture of low-rank tensors. The above-mentioned methods all adopt a weighted sum of the nuclear norms of the unfolding matrices to represent the tensor rank. However, there is no universal agreement on why such a metric is the best objective function to minimize, nor is there any consensus on what the best objective function is for tensor completion problems (Mu et al. 2013; Oymak et al. 2012). Moreover, the weights play an important role in tensor completion, but it is not clear how the weights should be *a priori* chosen since the nuclear norm of the unfolding matrix of each mode is not known given the incomplete data tensor.

Motivated by the above discussions, we propose to minimize the maximum rank of all unfolding matrices. To this end, we propose an alternative metric that serves as a convex surrogate for the maximum rank. This metric is based on the log sum of exponential of the nuclear norms of all unfolding matrices. It aims to minimize the maximum of the nuclear norms (and hence minimize the maximum rank). Somewhat surprisingly, we show that this convex objective function allows an efficiently computable proximal operator, which is given by an interesting hybrid singular value thresholding scheme, unlike any singular value thresholding scheme that we have seen before. In addition, experiments show that this new objective function can better promote the low-rankness of the recovered tensors than existing methods based on tensor nuclear norm, tractional matrix norm, and the newly proposed square matrix norm. We believe the hybrid thresholding solution is an interesting result of its own right: It suggests that there might be other low-rank promoting convex objective functions for high-order tensors that could also lead to efficient proximal operators of a similar kind. This result can also be independently applied to any problem where the goal is to minimize the maximum rank of a set of matrices.

Problem Formulation

We adopt the following notation conventions throughout the paper. Scalars $(a, b, c \cdots)$ are denoted as lowercase letters and vectors $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \cdots)$ are denoted as bold lowercase letters. We use capital letters for matrices $(A, B, C \cdots)$ and use calligraphic letters for tensors $(\mathcal{A}, \mathcal{B}, \mathcal{C} \cdots)$. In the following subsections, we provide the motivation of our work, followed by a presentation of the problem statement.

Motivation

In tensor completion literature, a commonly adopted formulation is as follows:

$$\min_{\mathcal{X}} \sum_{i=1}^{m} \alpha_i \|X_{(i)}\|_*, \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{1}$$

where $X_{(i)}$ is the ith unfolding matrix of tensor $\mathcal{X}$, and $\|X_{(i)}\|_*$ is the nuclear norm (the sum of the singular values) of the matrix $X_{(i)}$. Ω is the set of observed elements and $\bar{\Omega}$ represents the missing entries.

Even though certain empirical success has been achieved under the formulation in Eq. (1), there is no universal agreement on why $\sum_{i=1}^{m} \alpha_i \|X_{(i)}\|_*$ is the best objective function to minimize, nor is there any consensus on what the best objective function is for tensor completion (Mu et al. 2013; Oymak et al. 2012). Moreover, it is not clear how the weights should be *a priori* chosen since the nuclear norm of the unfolding matrix of each mode is not known given the incomplete data tensor. Motivated by the above-mentioned concerns, we propose in this paper a worst-case metric: $|||\mathcal{X}||| = \max_i \|X_{(i)}\|_*$, where it can be easily verified that $|||\cdot|||$ thus defined is a norm for tensors. Hence, we can interpret this norm as a characterization of the low-rankness of a tensor. The corresponding optimization problem can hence be written as follows.

$$\min_{\mathcal{X}} \max_i \|X_{(i)}\|_*, \quad \text{s.t.} \quad \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{2}$$

In addition to having a clear interpretation of a worst-case metric, which can be interesting for study in its own right, our proposed formulation can in fact be complementary with the traditional weighted sums of nuclear norms formulation. If one decides to use the formulation in Eq. (1), then one can first solve the optimization problem in Eq. (2) and obtain an estimate, in the worst-case mindset, the values of the nuclear norms for all the unfolding matrices. Those values can provide a basis on which one can select their choice of the weights. Finally, $\max_i \operatorname{rank}(X_{(i)})$ lower bounds the CP rank of the tensor $\mathcal{X}$, where the CP rank of $\mathcal{X}$ is the minimum number of rank-one tensors that sum up to $\mathcal{X}$ (a rank-one k-way tensor in $\mathbf{R}^{n_1 \times n_2 \times \cdots \times n_k}$ is the outer product of k vectors, where the ith vector has dimension n_i). In special cases where the ranks of all unfolding matrices are equal, they all equal the CP rank of the tensor. Consequently, when $|||\mathcal{X}|||$ is viewed as a convex surrogate for $\max_i \operatorname{rank}(X_{(i)})$, solving the optimization problem in Eq. (2) amounts to approximating a feasible solution with low CP rank.

Problem Relaxation

Although the objective function $\max(\cdot)$ in Problem (2) is convex, it is highly non-smooth. Common gradient descent methods, when applied to solving this problem, can be quite inefficient, with accuracy not guaranteed. Hence we seek a relaxation which can be solved very efficiently with high accuracy. We replace the objective function with the following function to obtain again a convex problem.

$$\min_{\mathcal{X}} \ \log\left(\mathrm{e}^{\|X_{(1)}\|_*} + \mathrm{e}^{\|X_{(2)}\|_*} + \cdots + \mathrm{e}^{\|X_{(m)}\|_*}\right), \quad \text{s.t.} \ \ \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{3}$$

We note that this is a rather tight approximation of the original function $|||\mathcal{X}|||$, since the following inequality holds: $\max_i \|X_{(i)}\|_* \leq \log\left(\mathrm{e}^{\|X_{(1)}\|_*} + \mathrm{e}^{\|X_{(2)}\|_*} + \cdots + \mathrm{e}^{\|X_{(m)}\|_*}\right) \leq \max_i \|X_{(i)}\|_* + \log m$. The second inequality is tight when the nuclear norms of all the unfolding matrices are equal. In addition, since $\log(\cdot)$ is a monotonically strictly increasing function, it suffices to solve the following optimization program:

$$\min_{\mathcal{X}} \ \mathrm{e}^{\|X_{(1)}\|_*} + \mathrm{e}^{\|X_{(2)}\|_*} + \cdots + \mathrm{e}^{\|X_{(m)}\|_*}, \quad \text{s.t.} \ \ \mathcal{X}_\Omega = \mathcal{T}_\Omega. \tag{4}$$

It is easily seen that the optimal solutions to Problem (3) and Problem (4) are the same.

A Hybrid Thresholding Solution

In this section, we provide a key building block that will prove essential in solving the optimization problem in (4). Given a matrix $Y \in \mathbf{R}^{n_1 \times n_2}$ and a $\tau \geq 0$, consider the function

$$h(X) = \tau \mathrm{e}^{\|X\|_*} + \frac{1}{2}\|X - Y\|_F^2, \ X \in \mathbf{R}^{n_1 \times n_2}.$$

The ultimate goal of this section is to find the matrix $\hat{X}$ that achieves the minimum of the function $h(X)$, i.e. $\hat{X} = \arg\min_X h(X)$. This is also known as the *proximal operator* for the convex function $\mathrm{e}^{\|X\|_*}$. We note that when $\tau = 0$, this problem is trivial, and hence we will assume $\tau > 0$ for the rest of the paper. Cai et al. in (Cai, Candès, and Shen 2010) have considered the problem of finding $\arg\min_X \tau\|X\|_* + \frac{1}{2}\|X - Y\|_F^2$. There, they have shown that the singular value shrinkage gives the proximal operator. We will briefly describe it here since our method makes use of it. However, as we will see the proximal operator for $\mathrm{e}^{\|X\|_*}$ is a nontrivial extension of the soft thresholding shrinkage operator.

Consider a matrix $Y \in \mathbf{R}^{n_1 \times n_2}$ of rank r, whose condensed singular value decomposition (SVD) is:

$$Y = U\Sigma V^*, \Sigma = \mathrm{diag}(\{\sigma_Y^i\}_{1 \leq i \leq r}),$$

where V^* is the transpose of the matrix V, U and V are $n_1 \times r$ and $n_2 \times r$ matrices respectively with orthonormal columns. The positive singular values of Y are σ_Y^i's. Unless specified otherwise, all SVDs in this paper will be assumed to be of this condensed form. In addition, we assume that the singular values σ_Y^i are always ordered in decreasing values.

For each $\tau > 0$, the soft thresholding shrinkage operator $\mathcal{D}_\tau$ introduced in (Cai, Candès, and Shen 2010) is defined to be:

$$\mathcal{D}_\tau(Y) = U\mathcal{D}_\tau(\Sigma)V^*, \ \mathcal{D}_\tau(\Sigma) = \mathrm{diag}((\sigma_Y^i - \tau)_+),$$

where $t_+ = \max(0, t)$.

However, since the function of interest now is $\tau \mathrm{e}^{\|X\|_*} + \frac{1}{2}\|X - Y\|_F^2$, the soft thresholding shrinkage operator cannot be directly applied to obtain the minimizer $\hat{X}$. In fact, it is not even *a priori* clear why $\hat{X}$ should have the same left and right singular vectors as Y. As we shall see later, a properly generalized method, the hybrid thresholding scheme, readily finds the minimizer $\hat{X}$. We first state a lemma that will be used later in establishing an algorithm to find $\hat{X}$. For convenience, we augment the list of singular values $\{\sigma_Y^i\}_{1 \leq i \leq r}$ of Y with $\sigma_Y^{r+1} = -\infty$.

Lemma 1. *Suppose the rank of Y is r and $\tau < \sigma_Y^1$. There exists a unique integer j, with $1 \leq j \leq r$, such that the solution t_j to the following equation*

$$\ln(t_j) + jt_j = \ln(\tau) + \sum_{i=1}^{j} \sigma_Y^i \tag{5}$$

satisfies the constraint

$$\sigma_Y^{j+1} \leq t_j < \sigma_Y^j. \tag{6}$$

Proof. We first show that if at least one such j exists, then such a j (and hence t_j) is unique. Consider the set $J = \{j \mid t_j$ satisfies (5) and (6)$\}$. Assume J is not empty, let j^* be the smallest element in J. Now we argue that no j^*+k, $1 \leq k \leq r - j^*$, can be in J. Consider any k with $1 \leq k \leq r - j^*$. Suppose for the sake of contradiction, $j^* + k \in J$. That is

$$\ln(t_{j^*+k}) + (j^* + k)t_{j^*+k} = \ln(\tau) + \sum_{i=1}^{j^*+k} \sigma_Y^i, \tag{7}$$

$$\sigma_Y^{j^*+k+1} \leq t_{j^*+k} < \sigma_Y^{j^*+k}. \tag{8}$$

Expanding on the right side of (7), we have

$$\ln(\tau) + \sum_{i=1}^{j^*+k} \sigma_Y^i \geq \ln(\tau) + \sum_{i=1}^{j^*} \sigma_Y^i + k\sigma_Y^{j^*+k}$$

$$= \ln(t_{j^*}) + j^* t_{j^*} + k\sigma_Y^{j^*+k} \geq \ln(\sigma_Y^{j^*+k}) + (j^* + k)\sigma_Y^{j^*+k}$$

where the first inequality follows from the decreasing values of the singular values, the equality follows from that t_{j^*} satisfies (5) and the last inequality follows from that $t_{j^*} \geq \sigma_Y^{j^*+1} \geq \sigma_Y^{j^*+k}$ due to (6). Hence we have

$$\ln(t_{j^*+k}) + (j^*+k)t_{j^*+k} \geq \ln(\sigma_Y^{j^*+k}) + (j^*+k)\sigma_Y^{j^*+k} \tag{9}$$

by (7).

Since the function $f(t) = \ln(t) + mt$ is a strictly increasing function for positive m, (9) implies that $t_{j^*+k} \geq \sigma_Y^{j^*+k}$. However, this contradicts with the second assumption in (8), establishing that such a k does not exist.

Next, we prove that J is indeed not empty.

First, we note that by monotonicity of $f(t) = \ln(t) + mt$, a unique solution exists for $\ln(t) + jt = \ln(\tau) + \sum_{i=1}^{j} \sigma_Y^i$,

for each j satisfying $1 \leq j \leq r$. We denote by t_j the unique solution corresponding to each j. Hence, it suffices to show at least one t_j satisfies $\sigma_Y^{j+1} \leq t_j < \sigma_Y^j$.

Again by monotonicity of $f(t) = \ln(t) + mt$, it is easily seen that $\tau < \sigma_Y^1$ if and only if $t_1 < \sigma_Y^1$. Hence, by assumption of the lemma, we have $t_1 < \sigma_Y^1$. Now suppose it also holds that $\sigma_Y^2 \leq t_1$, then we are done. Otherwise, we have $t_1 < \sigma_Y^2$. Under this assumption, we claim that $t_1 < t_2$ and $t_2 < \sigma_Y^2$. This is readily verified by the following inequalities.

$$\ln(t_2) + 2t_2 = \ln(\tau) + \sigma_Y^1 + \sigma_Y^2 \quad (10)$$

$$= \ln(t_1) + t_1 + \sigma_Y^2 > \ln(t_1) + 2t_1, \quad (11)$$

which implies that $t_2 > t_1$. On the other hand, we also have

$$\ln(t_2) + 2t_2 = \ln(t_1) + t_1 + \sigma_Y^2 < \ln(\sigma_Y^2) + 2\sigma_Y^2, \quad (12)$$

which implies that $t_2 < \sigma_Y^2$. If $t_2 \geq \sigma_Y^3$, then the claim is established. If not, we can repeat this process inductively. More formally, suppose we have just finished the j-th iteration (note that the induction basis $j = 1$ is already verified) and we have $t_j < \sigma_Y^j$. If it also holds that $t_j \geq \sigma_Y^{j+1}$, then the claim follows. If not, then

$$\ln(t_{j+1}) + (j+1)t_{j+1} = \ln(\tau) + \sum_{i=1}^{j+1} \sigma_Y^i$$

$$= \ln(t_j) + jt_j + \sigma_Y^{j+1} > \ln(t_j) + (j+1)t_j,$$

which implies that $t_{j+1} > t_j$. On the other hand, we also have

$$\begin{aligned} \ln(t_{j+1}) + (j+1)t_{j+1} &= \ln(t_j) + jt_j + \sigma_Y^{j+1} \\ &< \ln(\sigma_Y^{j+1}) + (j+1)\sigma_Y^{j+1}, \end{aligned}$$

which implies that $t_{j+1} < \sigma_Y^{j+1}$.

Thus, we have a strictly increasing sequence $\{t_j\}$ with $t_j < \sigma_Y^j$. If it holds that $\sigma_Y^{j+1} \leq t_j < \sigma_Y^j$ at some iteration j, then such a j certifies that J is not empty. If $\sigma_Y^{j+1} \leq t_j < \sigma_Y^j$ never holds for j up to $r-1$, then it must hold for $j = r$, since $-\infty = \sigma_Y^{r+1} \leq t_r < \sigma_Y^r$, also certifying that J is not empty.

□

Remark 1. *The above proof suggests a sequential search through all j's to find such a j that satisfies the constraints. However, a sequential search is inefficient especially if such systems need to be solved a large number of times for different τ's and Y's. It turns out binary search can be used to improve the efficiency of the search. The algorithm is given in Algorithm 1, followed by the proof of its correctness.*

Note that the step "Compute t_M" can be easily done very efficiently by numerically solving $\ln(x) + Mx = \ln(\tau) + \sum_{i=1}^M \sigma_Y^i$ to an arbitrary precision. Next, we give the proof of the algorithm's correctness.

Proof. From the first part of proof for Lemma 1, we know that if j^* is the unique j guaranteed by Lemma 1, then for all $k > j^*$, we have $t_k \geq \sigma_Y^k$. Thus, if $t_M < \sigma_Y^M$, then we know that j^* cannot be less than M. That is, j^* must be in the second half of the unsearched space.

Conversely, if we hypothetically do a sequential search, then it follows immediately from the second part of proof of Lemma 1 that before j reaches j^*, $t_M < \sigma_Y^M$ must hold. This establishes that if in the while loop we encounter $t_M \geq \sigma_Y^M$, then it must be the case that $j^* \leq M$. That is, j^* must lie in the first part of the unsearched space.

It then follows that j^* always lies between L and R, establishing that while loop will eventually halt, returning t_{j^*} and j^*.

□

Algorithm 1 Hybrid Threshold Computation

Input: Y, τ
Compute $\{\sigma_Y^i\}_{1 \leq i \leq r}$
if $\tau \geq \sigma_Y^1$ **then**
  RETURN σ_Y^1.
end if
Initialize $L = 1, R = r$,
while $L \leq R$ **do**
  $M = \lceil \frac{L+R}{2} \rceil$.
  Compute t_M.
  if $t_M < \sigma_Y^M$ **then**
    if $\sigma_Y^{M+1} \leq t_M$ **then**
      RETURN t_M and M.
    end if
    if $\sigma_Y^{M+1} > t_M$ **then**
      $L = M$.
      CONTINUE
    end if
  end if
  if $t_M \geq \sigma_Y^M$ **then**
    $R = M$.
    CONTINUE
  end if
end while

We now proceed to define a *hybrid thresholding operator* $\mathcal{H}_\tau$:

Definition 1. *Given a $\tau > 0$, the hybrid thresholding operator $\mathcal{H}_\tau$ is define to be:*

$$\mathcal{H}_\tau(Y) = U\mathcal{D}_{t_{j^*}}(\Sigma)V^*, Y = U\Sigma V^* \in \mathbf{R}^{n_1 \times n_2},$$

where t_{j^*} is the threshold computed by Algorithm 1.

Lemma 1 guarantees that $\mathcal{H}_\tau$ is well-defined and Algorithm 1 guarantees that $\mathcal{H}_\tau$ is efficiently computable. Having defined $\mathcal{H}_\tau$, we are ready to state our main result promised at the beginning of the section.

Theorem 1. *Given a $\tau > 0$ and a $Y \in \mathbf{R}^{n_1 \times n_2}$, we have:*

$$\mathcal{H}_\tau(Y) = \arg\min_X \{\tau \mathrm{e}^{\|X\|_*} + \frac{1}{2}\|X - Y\|_F^2\}.$$

Remark 2. *The proof is rather lengthy and hence omitted due to space limitation. However, in what follows, we give a rough proof sketch as a guideline.*

First, note that the function $h(X) = \tau \mathrm{e}^{\|X\|_*} + \frac{1}{2}\|X - Y\|_F^2$ is strictly convex, and hence the minimizer $\hat{X}$ to $h(X)$

is unique and it suffices to show that $\mathcal{H}_\tau(Y)$ is one minimizer. One can check that if there exists such a W satisfying $U_{\hat{X}}^* W = 0, W V_{\hat{X}} = 0, \|W\|_2 \leq 1$ such that:

$$Y - \hat{X} = \tau \mathrm{e}^{\|\hat{X}\|_*}(U_{\hat{X}} V_{\hat{X}}^* + W), \tag{13}$$

(where $U_{\hat{X}}$ and $V_{\hat{X}}$ are obtained from the SVD of $\hat{X}$), then $\hat{X}$ is a minimizer (hence the unique minimizer) to $h(X)$. Finally, one can show that with $\hat{X} = \mathcal{H}_\tau(Y)$, Eq. (13) does hold, with W satisfying the given constraints.

Optimization Algorithm

Although the problem in (4) is convex, it is still difficult to solve due to the interdependent nuclear norm terms. To remove these interdependencies and optimize these terms independently, we may introduce a set of auxiliary matrices $\{M_i, i = 1, 2, \ldots, m\}$ to replace $\{X_{(i)}, i = 1, 2, \ldots, m\}$, and the optimization problem changes to

$$\min_{\mathcal{X}, M_i} \sum_{i=1}^{m} \mathrm{e}^{\|M_i\|_*}, \ \text{s.t.}\ \mathcal{X}_\Omega = \mathcal{T}_\Omega, \ X_{(i)} = M_i, \forall i. \tag{14}$$

To relax the above equality constraints, we apply the Augmented Lagrange Multiplier (ALM) method (Lin, Chen, and Ma 2010), and obtain the following augmented Lagrangian function:

$$f_\mu(M_i, X_{(i)}, Q_i) = \sum_{i=1}^{m} (\mathrm{e}^{\|M_i\|_*} + \langle Q_i, X_{(i)} - M_i \rangle + \frac{\mu_i}{2} \|X_{(i)} - M_i\|_F^2). \tag{15}$$

In $f_\mu(M_i, X_{(i)}, Q_i)$, there are several terms need to be optimized. To optimize these terms in a separated way, we adopt the alternating direction method (ADM) (Lin, Chen, and Ma 2010), which is effective to solve optimization problems with multiple terms. According to the framework of ADM, the above optimization problem can be iteratively solved as follows.

$$\begin{cases} M_i^{k+1}: & = \arg\min_{M_i} f_\mu(M_i, X_{(i)}^k, Q_i^k), \forall i; \\ X_{(i)}^{k+1}: & = \arg\min_{X_{(i)}} f_\mu(M_i^{k+1}, X_{(i)}, Q_i^k), \forall i; \\ Q_i^{k+1}: & = Q_i^k + \mu_i (X_{(i)}^{k+1} - M_i^{k+1}), \forall i. \end{cases} \tag{16}$$

In detail, the solutions of each term are obtained as follows.

- For term M_i^{k+1}:

  $$\arg\min_{M_i} \sum_{i=1}^{k} (\mathrm{e}^{\|M_i\|_*} + \langle Q_i^k, X_{(i)}^k - M_i \rangle + \frac{\mu_i}{2} \|X_{(i)}^k - M_i\|_F^2),$$

  Since $\{M_i, i = 1, 2, \ldots, k\}$ are independent, so the above problem can be simplified as follows,

  $$\arg\min_{M_i} \mathrm{e}^{\|M_i\|_*} + \langle Q_i^k, X_{(i)}^k - M_i \rangle + \frac{\mu_i}{2} \|X_{(i)}^k - M_i\|_F^2,$$
  $$= \arg\min_{M_i} \frac{1}{\mu_i} \mathrm{e}^{\|M_i\|_*} + \frac{1}{2} \|X_{(i)}^k + \frac{1}{\mu_i} Q_i^k - M_i\|_F^2$$

  According to the hybrid thresholding scheme, the optimal solution of M_i is given by

  $$M_i^{k+1} = \mathcal{H}_\tau(Y),$$

  where $Y = X_{(i)}^k + \frac{1}{\mu_i} Q_i^k$ and $\tau = \frac{1}{\mu_i}$.

- For term $X_{(i)}^{k+1}$:

  $$\arg\min_{\mathcal{X}} \sum_{i=1}^{k} \left(\langle Q_i^k, X_{(i)}^k - M_i \rangle + \frac{\mu_i}{2} \|X_{(i)}^k - M_i\|_F^2 \right),$$
  $$= \arg\min_{\mathcal{X}} \sum_{i=1}^{k} \frac{\mu_i}{2} \|X_{(i)} + \frac{1}{\mu_i} Q_i^k - M_i^{k+1}\|_F^2.$$

  Together with the equality constraint $\mathcal{X}_\Omega = \mathcal{T}_\Omega$, the optimal solution of $\mathcal{X}$ can be obtained as follows.

  $$\mathcal{X}_{\bar{\Omega}} = \frac{\sum_i \mu_i \left(\mathrm{fold}_i (M_i^{k+1} - \frac{1}{\mu_i} Q_i^k) \right)_{\bar{\Omega}}}{\sum_i \mu_i}$$

Experimental Evaluation

To validate the effectiveness of the proposed tensor completion algorithm, we conduct two comparison experiments as follows: (1) the proposed metric versus tensor nuclear norm; (2) the proposed metric versus matrix nuclear norm. All the experiments are conducted with MATLAB on a platform with Pentium IV 3.2GHz CPU and 1G memory. The source code and data will be released once the paper is accepted for publication.

Proposed Metric vs Tensor Nuclear Norm

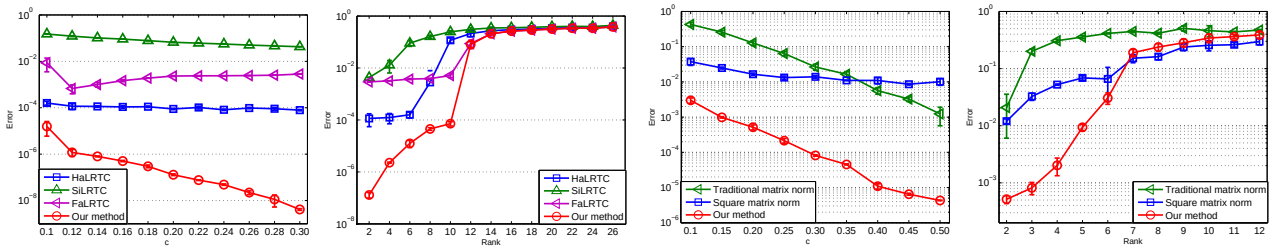
In this section, we compare the proposed metric with the tensor nuclear norm adopted in (Liu et al. 2013), and all the algorithms in their work are used for comparison. These algorithms are referred to as SiLRTC (simple low-rank tensor completion), FaLRTC (fast low-rank tensor completion), and HaLRTC (high accuracy low-rank tensor completion).

We first randomly generate a pure low-rank tensor $\mathcal{L}_o \in \mathbb{R}^{50 \times 50 \times 50}$ whose Tucker rank (defined in the introduction) is (2,2,2) (the same set is adopted in (Liu et al. 2013)), and then we sample a fraction c of elements in $\mathcal{L}_o$ as the observations while the rest elements are missing. Here, Ω is the set of observed elements and $\bar{\Omega}$ represents the set of missing entries. All the above algorithms are applied to recover the low-rank structure of $\mathcal{L}_o$ from the observed elements, which is represented as $\mathcal{L}_r$. Therefore, the reconstruction error is defined as $\text{error} = \frac{\|\mathcal{L}_o - \mathcal{L}_r\|_F}{\|\mathcal{L}_o\|_F}$. The result of a single run is a random variable, because the data are randomly generated, so the experiment is repeated 50 times to generate statistical averages.

Fig. 1(a) shows the reconstruction error of all the algorithms when the observed elements changes from 1% to 30%, from which we can see that our method achieves the highest accuracy of reconstruction among all the algorithms. For SiLRTC, FaLRTC and HaLRTC, if the weight parameters are not properly chosen[1], the performance of the other three algorithms will be worse.

To show the influence of the tensor rank on the performance, this part investigates the sensitivity of the algorithm to changes of this factor. Fig. 1(b) shows the reconstruction

[1] In our experiment, we choose the parameters with the best performance.



(a) (b) (c) (d)

Figure 1: (a),(c): reconstruction errors versus sampling rate c for different tensor completion methods; (b)(d): reconstruction errors versus tensor rank for different tensor completion methods.

error curve using different rank numbers (take $c = 20\%$ as an example). We can see that the performance of all algorithms degrades when the tensor rank increases. According to the theoretical results in (Candès and Recht 2009), the degradation of performance is inevitable because we need to observe more elements to recover the tensor structure when the tensor rank increases. However, our method always obtains better results than the other three algorithms.

Proposed Metric vs Matrix Norm

As a baseline, we compare the proposed metric to the matrix norm. Here, two kinds of matrix norm are chosen for comparison: (1) Traditional matrix nuclear norm (Tomioka, Hayashi, and Kashima 2010). We can convert the tensor completion problem into a matrix completion problem by unfolding the tensor in the kth mode, and solve the matrix completion problem with the traditional matrix nuclear norm. To make this norm more competitive, we perform matrix completion on each mode of the tensor and take the mode that has the minimal reconstruction error. (2) Square matrix nuclear norm (Mu et al. 2013). Mu et al. defined a new unfolding operator which can generate a balanced matrix, and then define the nuclear norm on that matrix. In their work, they claim that the square norm can achieve better performance than the traditional tensor nuclear norm (Liu et al. 2013).

Since the square matrix nuclear norm can only work well on the 4-order or higher tensor, we generate a low-rank tensor $\mathcal{L}_o \in \mathbb{R}^{30\times30\times30\times30}$ with Tucker rank (2,2,2,2) (the same set is adopted in (Mu et al. 2013)). Other settings are the same as that in Section 'Proposed Metric vs Tensor Nuclear Norm'. Fig. 1(c) and Fig. 1(d) show the reconstruction errors when c and the tensor rank changes, respectively. The results show that the proposed metric can recover the low-rank tensor more accurately than the other two norms.

Conclusion

We have considered the low-rank tensor completion problem in this paper. We have shown that minimizing the maximum rank of all unfolding matrices of a tensor is a feasible thing to do in the following sense: Firstly, there exists a convex surrogate for this goal based on log sum of exponentials of nuclear norms. Secondly, this convex objective function allows an efficiently computable proximal operator. The proximal operator can be computed through a hybrid singular value thresholding scheme which has some rather interesting properties. Through extensive comparison, we have demonstrated the effectiveness of this new surrogate for completing low-rank tensors over existing methods. As future work, it would be interesting both to identify new problems that can leverage the hybrid thresholding scheme and to study the other norms or metrics whose proximal operator can be better dealt with in a similar hybrid thresholding framework. Another line of future work can be theoretically understanding the exact recovery conditions for tensor completion. In addition, it would also be interesting to look for other characterizations of tensor rank that can model useful structures of tensorial data.

Acknowledgment

This work is partly supported by NSFC (Grant Nos. 61100147, 61203241 and 61305035), Zhejiang Provincial Natural Science Foundation (Grants Nos. LY12F03016, LQ12F03004 and LQ13F030009).

References

Cai, J.; Candès, E.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Candès, E., and Plan, Y. 2010. Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936.

Candès, E., and Recht, B. 2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics* 9(6):717–772.

Candès, E., and Tao, T. 2010. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory* 56(5):2053–2080.

Candès, E.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):11.

Chandrasekaran, V.; Sanghavi, S.; Parrilo, P.; and Willsky, A. 2011. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization* 21(2):572–596.

Gandy, S.; Recht, B.; and Yamada, I. 2011. Tensor completion and low-n-rank tensor recovery via convex optimization. *Inverse Problems* 27(2):025010.

Ganesh, A.; Min, K.; Wright, J.; and Ma, Y. 2012. Principal component pursuit with reduced linear measurements. In *Proceedings of International Symposium on Information Theory*, 1281–1285.

Gross, D. 2011. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory* 57(3):1548–1566.

Hu, Y.; Zhang, D.; Ye, J.; Li, X.; and He, X. 2013. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(9):2117–2130.

Keshavan, R.; Montanari, A.; and Oh, S. 2010. Matrix completion from a few entries. *IEEE Transactions on Information Theory* 56(6):2980–2998.

Kolda, T., and Bader, B. 2009. Tensor decompositions and applications. *SIAM review* 51(3):455–500.

Kruskal, J. 1977. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications* 18(2):95–138.

Lin, Z.; Ganesh, A.; Wright, J.; Wu, L.; Chen, M.; and Ma, Y. 2009. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *Computational Advances in Multi-Sensor Adaptive Processing* 61.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

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

Luo, Z. 2012. On the linear convergence of the alternating direction method of multipliers. *arXiv preprint arXiv:1208.3922*.

Ma, S. 2012. Alternating proximal gradient method for convex minimization. *Preprint of Optimization Online*.

Mu, C.; Huang, B.; Wright, J.; and Goldfarb, D. 2013. Square deal: Lower bounds and improved relaxations for tensor recovery. *arXiv preprint arXiv:1307.5870*.

Oymak, S.; Jalali, A.; Fazel, M.; Eldar, Y.; and Hassibi, B. 2012. Simultaneously structured models with application to sparse and low-rank matrices. *arXiv preprint arXiv:1212.3753*.

Recht, B.; Fazel, M.; and Parrilo, P. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* 52(3):471–501.

Recht, B. 2011. A simpler approach to matrix completion. *Journal of Machine Learning Research* 3413–3430.

Signoretto, M.; De Lathauwer, L.; and Suykens, J. 2010. Nuclear norms for tensors and their use for convex multilinear estimation. *Linear Algebra and Its Applications*.

Tomioka, R.; Hayashi, K.; and Kashima, H. 2010. Estimation of low-rank tensors via convex optimization. *arXiv preprint arXiv:1010.0789*.

Wright, J.; Ganesh, A.; Min, K.; and Ma, Y. 2013. Compressive principal component pursuit. *Information and Inference* 2(1):32–68.

Zhou, Z.; Li, X.; Wright, J.; Candès, E.; and Ma, Y. 2010. Stable principal component pursuit. In *Proceedings of International Symposium on Information Theory*, 1518–1522.