Enhanced nonconvex low-rank representation for tensor completion

Haijin Zeng\textsuperscript{a}, Xiaozhen Xie\textsuperscript{a,}\textsuperscript{*} and Jifeng Ning\textsuperscript{b}

\textsuperscript{a}College of Science, Northwest A\&F University, Yangling 712100, China
\textsuperscript{b}College of Information Engineering, Northwest A\&F University, Yangling 712100, China

\begin{abstract}
Higher-order low-rank tensor arises in many data processing applications and has attracted great interests. In this paper, we propose a new low rank model for higher-order tensor completion task based on the double nonconvex $L_\gamma$ norm, which can effectively approximate the rank minimization of tensor mode-matrix. An block successive upper-bound minimization method-based algorithm is designed to efficiently solve the proposed model, and it can be demonstrated that our numerical scheme converge to the coordinatewise minimizers. Numerical results on three types of public multi-dimensional datasets have tested and shown that our algorithms can recover a variety of low-rank tensors with significantly fewer samples than the compared methods.
\end{abstract}

1. Introduction

As a generalization of vector and matrix, tensor arises in many data processing applications and has attracted great interests. For instance, video inpainting \cite{16}, magnetic resonance imaging (MRI) data recovery \cite{26}, 3D image reconstruction \cite{19}, high-order web link analysis \cite{16}, hyperspectral or multispectral data recovery \cite{27}, personalized web search \cite{21}, and seismic data reconstruction \cite{9}.

Tensor completion tries to recover a low-rank tensor from its partially observed entries. A large number of tensor completion methods have been proposed and successfully used in many applications. In all the completing methods, the tensor rank minimization based methods are considered state-of-the-art methods with promising performance, and their robustness to noisy and missing data has also been proven. Therefore, they have been universally utilized in tensor completion problems. Usually, they can be solved by replacing the rank function with its convex or non-convex relaxations in the minimization problem. This type of method can significantly reduce the deviation of rank estimation. A few notable examples are the CANDECOMP/PARAFAC rank minimization method \cite{4} and the Tucker rank minimization method \cite{22, 1}. However, both of them are NP-hard. To tackle this difficulty, Liu et al. \cite{11} extended the matrix nuclear norm to the tensor case, Based on this tensor nuclear norm, the tensor singular value decomposition (t-SVD) \cite{8} based TNN is proposed, as the tightest convex surrogate of the tensor rank, it has been widely used for low-rank tensor completion \cite{30}. Furthermore, to alleviate bias phenomena of the TNN minimization in tensor completion tasks, Jiang et al. \cite{7} propose a non-convex surrogate of the tensor rank, i.e., a partial sum of the tensor nuclear norm (PSTNN). However, these methods involved the singular value decomposition (SVD), which is time-consuming. To cope with this issue, Xu et al. \cite{25} proposed a parallel matrix factorization low-rank tensor completion model (Tmac), which obtained better results with less running time than \cite{11}. Further, combined with the Total Variation (TV) regularizer, researchers proposed TV regularized low-rank matrix factorization method (MF-TV) and TV regularized TNN low-rank tensor completion method.

Although the above-mentioned low-rank tensor completion methods show great success in dealing with various issues, three major open questions have yet to be addressed. Firstly, the low-rank priors of underlying tensors are only explored by basic low-rank decomposition, while the low-rank priors of factors obtained by the decomposition are not explored further. Secondly, TNN based methods \cite{30, 7} need to compute lots of SVDs, which become very slow or even not applicable for large-scale problems \cite{25}. Thirdly, these methods adopt single nuclear norm or partial sum minimization of singular values norm, which would cause suboptimal solution of the low-rank based problem. That is because the traditional nuclear norm based low-rank subproblem would tend to over-relaxations of rank components from the representation matrix \cite{28}. In recent works, researchers usually adopt nonconvex penalties instead of the traditional nuclear norm for low rank based problems \cite{12, 15, 20, 17}. Nonconvex penalty has decomposable approach and it could construct a more accurate low rank matrix than the traditional nuclear norm \cite{14}.

In this paper, motivated and convinced by the much better performance of models that utilize the low-ranknesses in all mode in tensors \cite{11, 25}, instead of using the single nuclear norm to represent the low-rank prior of underlying tensor directly, we first apply parallel matrix factorization to all modes of underlying tensor. Further, the novel double $L_\gamma$ norm, a kind of nonconvex penalty, is designed to represent the underlying joint-manifold drawn from the mode factorization factors. An block successive upper-bound minimization method-based algorithm is designed to efficiently solve the proposed model, and it can be demonstrated that our numerical scheme converge to the coordinatewise minimizers. The proposed models have been evaluated on three types of public datasets, which show that our algorithms can recover a variety of low-rank tensors with significantly fewer sam-
Enhanced nonconvex LR representation for tensor completion

2. Preliminary

2.1. Notations

In the rest of this paper, following [25], vectors, matrices and tensors are denoted as bold lower-case letters \( \mathbf{x} \), bold upper-case letters \( \mathbf{X} \) and caligraphic letters \( \mathcal{X} \), respectively. Let \( x_{i_1 \ldots i_N} \) represents the \((i_1, \ldots, i_N)\) -th component of an \( N \)-way tensor \( \mathcal{X} \), then, for \( \mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_N} \), their inner product is defined as

\[
\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} x_{i_1 \cdots i_N} y_{i_1 \cdots i_N}.
\]

Based on (1), the Frobenius norm of a tensor \( \mathcal{X} \) is defined as \( \| \mathcal{X} \|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle} \). Fibers of tensor \( \mathcal{X} \) are defined as a vector obtained by fixing all indices of \( \mathcal{X} \) except one, and slices of \( \mathcal{X} \) are defined as a matrix by fixing all indices of \( \mathcal{X} \) except two. The mode-\( n \) matricization/unfolding of \( \mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N} \) denoted as a matrix \( \mathbf{X}^{(n)} \in \mathbb{R}^{I_n \times \prod_{n'=1 \atop n' \neq n}^{N} I_{n'}} \) with columns being the mode-\( n \) fibers of \( \mathcal{X} \) in the lexicographical order.

Furthermore, to clearly represent the matricization process, we define \( \text{unfold}_n(\mathcal{X}) = \mathbf{X}^{(n)} \), \( \text{fold}_n \) is the inverse of \( \text{unfold}_n \), i.e., \( \text{fold}_n(\text{unfold}_n(\mathcal{X})) = \mathcal{X} \). The \( n \)-rank of an \( N \)-way tensor \( \mathcal{X} \), denoted as \( \text{rank}_n(\mathcal{X}) \), is the rank of \( \mathbf{X}^{(n)} \), and the rank of \( \mathcal{X} \) is defined as an array:

\[
\text{rank}(\mathcal{X}) = (\text{rank}(\mathbf{X}^{(1)}), \ldots, \text{rank}(\mathbf{X}^{(N)})).
\]

2.2. Operators

The Projecional Operator of is defined as follows:

\[
\text{prox}_f(v) := \arg \min_u f(u) + \frac{\rho}{2} \| u - v \|^2
\]

where \( f(u) \) is convex, \( \rho \) is the proximal parameter. Then, the minimization of \( f(u) \) is equivalent to

\[
\arg \min_u \left\{ f(u) + \frac{\rho}{2} \| u - u^k \|^2 \right\}.
\]

We define the Projection Operator as follows:

\[
(P_{\Omega}(\mathcal{Y}))_{i_1, \ldots, i_N} \begin{cases} y_{i_1, \ldots, i_N} & (i_1, \ldots, i_N) \in \Omega \\ 0 & \text{otherwise} \end{cases}
\]

where \( \Omega \) is the index set of observed entries. The function of \( P_{\Omega} \) is to keep the entries in \( \Omega \) and zeros out others.

3. Related Works

We first introduce related tensor completing methods based on the tensor rank minimization. Given a sample tensor \( \mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times N_3} \), tensor completion tries to recover a complete tensor from \( \mathcal{Y} \), according to the prior of underlying tensor. In the past decade, the t-SVD [8] based TNN has been widely used for low-rank tensor completion [30]. The TNN based method aims to recover a low-rank tensor by penalizing the nuclear norm of the front slice under the Fourier domain,

\[
\arg \min_{\mathcal{Y}} \frac{1}{n_3} \sum_{i=1}^{n_3} \| \mathcal{Y}^{(i)} \|_F, \quad \text{s.t. } P_{\Omega}(\mathcal{Y}) = \mathcal{F},
\]

where \( \mathcal{Y}^{(i)} \) denotes the \( i \)-th frontal slice of \( \mathcal{Y} \), \( \mathcal{F} = \text{fft}(\mathcal{Y}, [\cdot, 3]) \) denotes the fast Fourier transform of \( \mathcal{Y} \) along the third dimension, \( \mathcal{F} \) is the observed data.

Inspired by the good performance of matrix nuclear norm in representing matrix rank, Liu et al. [17] unfold the tensor into multiple modal matrices along the direction of each mode, and then use the rank of these modal matrices to describe the low-rank structure of the underlying tensor. With that definition, the completion model is formulated as follows:

\[
\min \sum_{n=1}^{N} a_n \| Y^{(n)} \|_F, \quad \text{s.t. } P_{\Omega}(\mathcal{Y}) = \mathcal{F}
\]

Furthermore, to alleviate bias phenomena of the TNN minimization in tensor completion tasks, Jiang et al. [7] represent the low-rank prior of underlying tensor by using a partial sum of the tensor nuclear norm (PSTNN). The PSTNN regularized tensor completion model is formulated as follows:

\[
\arg \min_{\mathcal{Y}} \frac{1}{n_3} \sum_{i=1}^{n_3} \| \mathcal{Y}^{(i)} \|_F, \quad \text{s.t. } P_{\Omega}(\mathcal{Y}) = \mathcal{F},
\]

where

\[
\| \mathcal{Y}^{(i)} \|_F = \min_{(n_1, n_2)} \sum_{j=M+1}^{\min(n_1, n_2)} \sigma_j(\mathcal{Y}^{(i)}),
\]

and \( \sigma_j(\mathcal{Y}^{(i)})(j = 1, \ldots, \min(n_1, n_2)) \) denotes the \( j \)-th largest singular value of \( \mathcal{Y}^{(i)} \in \mathbb{C}^{n_1 \times n_2} \).

However, both the TNN and PSTNN based methods need to calculate the SVDs, which will bring large computational complexity. To cope with this issue, Xu et al. [25] proposed a fast low-rank tensor completion model (Tmac) by using parallel matrix factorization, which obtained promising results with less computational complexity than [30, 7],

\[
\min \sum_{i=1}^{N} \frac{a_i}{2} \| Y^{(n)} - A_n X_n \|_F^2, \quad \text{s.t. } P_{\Omega}(\mathcal{Y}) = \mathcal{F}.
\]
Although the above-mentioned low-rank tensor completion methods reported success on in dealing with a large variety of tasks, there are several open issues have yet to be addressed. Firstly, the above approaches either only explored the low-rank prior lying in one mode of the underlying tensor or only utilize preliminary low-rank decomposition to explore the low-rank prior of the model, and do not further explore the prior of the factors (e.g., $A_n, X_n$ in (8)) obtained by low-rank decomposition. Secondly, TNN based methods [30, 7] need to compute lots of SVDs, which is time-consuming or even not applicable for large-scale problems [25]. Thirdly, all these methods adopt single nuclear norm or partial sum minimization of singular values norm as the approximation to rank function, which would cause suboptimal solution of the tensor completion problem.

4. Double nonconvex $L_{γ}$ norm based Low Rank model on tensor completion

In the following, a novel double non-convex $L_{γ}$ norm based low-rank representation model on low-rank tensor completion. The proposed M-DNLR is formulated as

$$\min \sum_{n=1}^{N} (\tau_n \|X_n\|_γ + \lambda_n \|A_n\|_γ), \tag{9}$$

subject to $Y_n = A_n X_n$, $n = 1, 2, \ldots, N$, $P_Ω(Y) = F$,

where $\|X\|_γ = \sum_{t=1}^{\min(m,n)} (1 - e^{-\gamma t^\sigma(X)})$ is a nonconvex approximation of rank$(X)$, $\sigma_t(X)$ is the $t$-th singular value of $X \in \mathbb{R}^{m \times n}$.

4.2. Optimization Procedure of M-DNLR

In this section, the proposed model is solved by using the block successive upper-bound minimization (BSUM) [18] method.

The objective function of the proposed (9) can be formulated as follows:

$$f(X, A, Y) = \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|_F^2 + \tau_n \|X_n\|_γ\right). \tag{10}$$

According to the proximal operator, the update can be written as:

$$\rho(S, S^k) = f(X, A^k, Y^k) + \frac{\rho}{2} \|X - X^k\|_F^2 \tag{11}$$

where $\rho > 0$ is a positive constant, $S = (X, A, Y)$ and $S^k = (X^k, A^k, Y^k)$. Let

$$\begin{align*}
   g_1(X, S^k) &= f(X, A^k, Y^k) + \frac{\rho}{2} \|X - X^k\|_F^2 \\
   g_2(A, S^k) &= f(X^k, A, Y^k) + \frac{\rho}{2} \|A - A^k\|_F^2 \\
   g_3(Y, S^k) &= f(X^k, A^k, Y) + \frac{\rho}{2} \|Y - Y^k\|_F^2
\end{align*} \tag{12}$$

where

$$\begin{align*}
   S^k &= (X^k, A^k, Y^k), \\
   S_1 &= (X^k, A^k, Y^k), \\
   S_3 &= (X^k, A^k, Y^k)
\end{align*} \tag{13}$$

Then, problem (11) can be rewritten as follows

$$\begin{align*}
   X^{k+1} &= \arg\min_{X} g_1(X, S^k) \\
   A^{k+1} &= \arg\min_{A} g_2(A, S^k) \\
   Y^{k+1} &= \arg\min_{Y} g_3(Y, S^k)
\end{align*} \tag{14}$$

4.2.1. Update $X_n$ with fixing others

the $X_n$-sub-problem can be written as follows:

$$X_n^{k+1} = \arg\min_{X_n} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|_F^2 + \tau_n \|X_n\|_γ\right) + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2 \tag{15}$$

To efficiently solve the proposed model denoising model, we first introduce an auxiliary variable $X_n = Z_n$. Then, we can get the augmented Lagrangian function of (15),

$$L(X, Z, \Gamma_n) = \min \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|_F^2 + \tau_n \|Z_n\|_γ\right) + \frac{\rho_n}{2} \|X_n - X_n^k\|_F^2 + \langle \Gamma_n, X_n - Z_n \rangle$$

subject to $P_Ω(Y) = F$. \tag{16}$$

With others fixed, the minimization subproblem for $Z_n$ can be deduced from (16) as follows:

$$Z_n^{k+1} = \arg\min_{Z_n} \|Z_n\|_γ + \frac{\rho_n}{2} \|Z_n - P_{\mathcal{N}}\|_F^2 \tag{17}$$

where $\rho_n' = \rho_n / \tau_n$, $P_{\mathcal{N}} = X_n^k + \Gamma_n / \rho_n$. Let $\sigma_n' \geq \sigma_n' \geq \cdots \geq \sigma_n^n$ represent the singular values of $Z_n^k$ with $\tau_n = \min \{\tau_n, s_n\}$ and $\nabla \phi(\sigma_n^n)$ denote the gradient of $\phi$ at $\sigma_n^n$, $\phi(x) = 1 - e^{-\gamma x}$. Let

$$f(Z_n) = \frac{1}{2} \|Z_n - P_{\mathcal{N}}\|_F^2 \tag{18}$$

It is easy to prove that the gradient of $f(Z_n)$ is Lipschitz continuous by setting the Lipschitz constant being 1. As stated
in [2], considering the nonascending order of singular values and according to the antimonotone property of gradient of our nonconvex function, we have

\[
0 \leq \nabla \phi(\sigma^k_1) \leq \nabla \phi(\sigma^k_\ell) \leq \cdots \leq \nabla \phi(\sigma^k_{k_n}) \tag{18}
\]

\[
\phi(\sigma_n(Z_n)) \leq \phi(\sigma^k_1) + \nabla \phi(\sigma^k_1) (\sigma_n(Z_n) - \sigma^k_1)
\]

Following (18), the subproblem of \(Z_n\) can be written as the following relaxation problem:

\[
Z_n^{k+1} = \arg\min_{Z_n} \frac{1}{\rho_n} \sum_{k=1}^{k_n} \phi(\sigma^k_n) + \nabla \phi(\sigma^k_n) (\sigma_n(Z_n) - \sigma^k_n) + f(Z_n) = \arg\min_{Z_n} \frac{1}{\rho_n} \sum_{k=1}^{k_n} \nabla \phi(\sigma^k_n) \sigma_n(Z_n) + \frac{1}{2} \|Z_n - P_n\|^2_F
\tag{19}
\]

Then, following [13, 2], the optimum solution of \(Z_n\)-subproblem can be efficiently obtained by generalized weight singular value thresholding (WSVT) [3], as shown in Lemma 1.

**Lemma 1:** For any \((1/\rho_n) > 0\), the given data \(P_n = X_n + \Gamma_n/\rho_n\), and \(0 \leq \nabla \phi(\sigma^k_1) \leq \nabla \phi(\sigma^k_\ell) \leq \cdots \leq \nabla \phi(\sigma^k_{k_n})\), a globally optimal solution \(Z_n^*\) to problem (19) is given by the WSVT [3]

\[
Z_n^* = US_{\Sigma} V^T
\tag{20}
\]

where \(P_n = US_{\Sigma} V^T\) is the SVD of \(P_n\), and

\[
S_{\Sigma} = \text{Diag}\left(\max \left(\Sigma_{nn} - \frac{\nabla \phi(\sigma^k_n)}{\rho_n}, 0\right)\right)
\]

With other variables fixed, the minimization subproblem for \(X_n\) can be deduced from (16) as follows:

\[
X_n^{k+1} = \arg\min_{X_n} \frac{\alpha_n}{2} \|Y_n - X_n\|^2_F + \frac{\rho_n}{2} \left\|X_n - \frac{Z_{n}^{k+1} - \Gamma_n/\mu_n + X_{k_n}}{2}\right\|^2_F
\tag{21}
\]

They are convex and have the following closed-form solutions

\[
X_n^{k+1} = (\alpha_n A_n^T A_n + 2\rho_n I_n)^{-1}[\alpha_n A_n^T Y_n + Z_{n}^{k+1} - \Gamma_n/\mu_n + X_{k_n}]
\tag{22}
\]

The Lagrangian multiplier \(\Gamma_n^X\) can be updated by the following equation

\[
\Gamma_n^X = \Gamma_n^X + X_n - Z_n
\tag{23}
\]

**4.2.2. Update \(A_n\) with fixing others**

The \(A_n\)-subproblem in (14) can be written as follows:

\[
A_n^{k+1} = \arg\min_{A_n} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|^2_F + \lambda_n \|A_n\|_F^2 \right) + \frac{\rho_n}{2} \left\|A_n - A_{k_n}\right\|^2_F
\tag{24}
\]

By introducing an auxiliary variable, (24) can be rewritten as

\[
\arg\min_{A_n} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|^2_F + \lambda_n \|J_n\|_F^2 \right) + \frac{\rho_n}{2} \left\|A_n - A_{k_n}\right\|^2_F
\tag{25}
\]

By the ALM method, the problem (25) can also be reformulated as

\[
\arg\min_{A_n, J_n} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|^2_F + \lambda_n \|J_n\|_F^2 \right) + \frac{\rho_n}{2} \left\|A_n - A_{k_n}\right\|^2_F + \left(\Gamma_n, A_n - J_n\right)
\tag{26}
\]

where \(\Gamma_n\) is the Lagrangian multiplier.

Firstly, with other variables fixed, the minimization subproblem for \(J_n\) can be deduced from (26) as follows:

\[
J_n^{k+1} = \arg\min_{J_n} \lambda_n \|J_n\|_F^2 + \frac{\rho_n}{2} \left\|J_n - Q_n^k\right\|^2_F
\tag{27}
\]

where \(Q_n^k = A_n + \Gamma_n/\rho_n\). Its solution can also be obtained by **Lemma 1**

\[
J_n^* = US_{\Sigma} V^T
\tag{28}
\]

where \(Q_n^k = US_{\Sigma} V^T\) is the SVD of \(Q_n^k\).

Secondly, with other variables fixed, the minimization subproblem for \(A_n\) can be deduced from (26) as follows:

\[
A_n^{k+1} = \arg\min_{A_n} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \|Y_n - A_n X_n\|^2_F \right) + \frac{\rho_n}{2} \left\|A_n - A_{k_n}\right\|^2_F
\tag{29}
\]

It is also convex and has the following closed-form solution

\[
A_n^{k+1} = \left(\frac{X_n X_n^{k+1} + 2\rho_n (X_n X_n^{k+1} + \Gamma_n/\mu_n + X_{k_n})}{2}\right)^T
\tag{30}
\]

Finally, the Lagrangian multiplier \(\Gamma_n^A\) can be updated by the following equation

\[
\Gamma_n^A = \Gamma_n^A + A_n - J_n
\tag{31}
\]
Enhanced nonconvex LR representation for tensor completion

Algorithm 1: Algorithm for the proposed M-DNLR.

**Require:** The observed tensor $F$; The set of index of observed entries; stopping criterion $\epsilon$.

**Ensure:** The completed tensor.

1. Initialize: $X_n = Z_n = 0$, $A_n = J_n = 0$, $\Gamma_n = 0$, $\Gamma^A = 0$, $n = 1, 2, 3, 4, \ldots, N$; $\mu_{\text{max}} = 10^6$, $\rho = 1.5$, and $k = p = 0$.
2. Repeat until convergence:
   3. Update $X, Z, A, J, \bar{Y}, \bar{\Gamma}, \bar{\Gamma}^A$ via
      1. 3th step: Update $Z_n$ via (20)
      2. 2th step: Update $X_n$ via (22)
      3. 3th step: Update $A_n$ via (30)
      4. 4th step: Update $J_n$ via (28)
      5. 5th step: Update $\bar{Y}$ via (33)
      6. 6th step: Update the parameter via (23), (31)
3. Check the convergence condition.

4.2.3. Update $Y$ with fixing others

With other variables fixed, the minimization subproblem for $Y_{(n)}$ in (14) can be written as

$$Y_{k+1} = \arg \min_{(n)} \sum_{n=1}^{N} \left(\frac{\alpha_n}{2} \left\| Y_{(n)} - A_n X_n \right\|_F^2 + \frac{\rho}{2} \left\| Y - Y^k \right\|_F^2 \right),$$

\hspace{2em} s.t. $P_\Omega(Y) = F$. \hspace{2em} (32)

Then, the update of $Y_{k+1}$ can be written explicitly as

$$Y_{k+1} = P_\Omega \left( \sum_{n=1}^{N} \alpha_n \text{fold}_n \left( \frac{A_n X_n + \rho_n Y_{(n)}}{1 + \rho_n} \right) \right) + F,$$ \hspace{2em} (33)

where $F$ is the observed data; $P_\Omega$ is an operator defined in subsection 2.2.

4.3. Complexity and Convergence Analysis

The proposed algorithm for the proposed model is summarized in Algorithm 1. Further, we discuss the complexity and convergence of the proposed algorithm.

4.3.1. Complexity Analysis

The cost of computing $X_n$ is $O\left(I_n^2 + I_n^2 + \rho_n^2 s_n + r_n^2 s_n \right)$, calculating $Z_n$ has a complexity of $O\left(\sum_{j \neq n} I_j \times r_n^2 \right)$, the complexity of updating $J_n$ is $O\left(I_n^2 + r_n^2 \right)$, calculating $A_n$ has a complexity of $O\left(I_n^2 + I_n^2 + r_n^2 s_n \right)$, calculating $Y$ has a complexity of $O\left(r_n^2 + \cdots + r_n^2 \right)$. Therefore, the total complexity of the proposed algorithm can be obtained by counting the complexity of the above variables, i.e.,

$$O\left(3 I_n^2 + \sum_{j \neq n} I_j \times r_n^2 + 2 I_n^2 + 3 I_n^2 + 2 r_n^2 + 2 r_n^2 s_n \right) \hspace{2em} (34)$$

4.3.2. Convergence Analysis

In this section, we theoretically analyze the convergence of the proposed algorithm by using the block successive upper-bound minimization (BSUM)[18].

Lemma 1 [18, 6]. Given the problem min $f(x)$ s.t. $x \in \mathcal{X}$, where $\mathcal{X}$ is the feasible set. Assume $h(x, x^{k-1})$ is an approximation of $f(x)$ at the $(k-1)$-th iteration, which satisfied the following conditions:

1. $h_i(y_i, y)$ = $f(y)$, $\forall y \in \mathcal{X}$, $\forall i$;
2. $h_i(x_i, y)$ $\geq$ $f_i(y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_n)$, $\forall x_i \in \mathcal{X}_i$, $\forall y \in \mathcal{X}$, $\forall i$;
3. $h_i'(x_i, y; d_i) \bigg|_{x_i = y_i}$ = $f_i(y; d)$, $u_i = (0, \ldots, d_i \ldots 0)$ s.t. $y_i + d_i \in \mathcal{X}_i$; $\forall i$;
4. $h_i(x_i, y)$ is continuous in $(x_i, y)$. $\forall i$;

(35)

where $h_i(x_i, y)$ is the sub-problem with respect to the $i$-th block and $f_i'(y; d)$ is the direction derivative of fat the point $y$ in direction $d$. Suppose $h_i(x_i, y)$ is quasi-convex in $x_i$ for $i = 1, \ldots, n$. Furthermore, assume that each sub-problem argmin $h_i(x_i, x^{k-1})$, s.t. $x \in \mathcal{X}_i$ has a unique solution for any point $x^{k-1} \in \mathcal{X}$. Then, the iterates generated by the BSUM algorithm converge to the set of coordinatewise minimum of $f$.

Theorem 1. The iterates generated by (11) converge to the set of coordinatewise minimizers.

Proof. It is easy to verify that $g(S, s^k)$ is an approximation and a global upper bound of $f(S)$ at the $k$-th iteration, which satisfies the following conditions:

1. $g_i(S_i, S) = f(S), \forall S_i$, $i = 1, 2, 3$;
2. $g_i(S_i, S)$ $\geq$ $f(S_i, \ldots, S_i, \ldots, S_3)$, $\forall S_i$, $\forall S$, $i = 1, 2, 3$
3. $g_i'(S_i, S; M_i) \bigg|_{S_i = S_i}$ = $f_i'(S; M')$, $\forall M_i = (0, \ldots, M_i, \ldots, 0)$;
4. $g_i(S_i, S)$ is continuous in $(S_i, S)$, $i = 1, 2, 3$.

(36)

where $S = (X, A, Y)$, and $S_i$ equal $X, A, Y$ for $i = 1, 2, 3$, respectively. In addition, the sub-problem $g_i(i = 1, 2, 3)$ is strictly convex with respect to $X, A$ and $Y$ respectively and each sub-problem of $g_i$ has a unique solution. Therefore, all assumptions in Lemma 1 are satisfied. According to the conclusion of Lemma 1, the Theorem 1 is valid, and the proposed algorithm is theoretically convergent.

5. Numerical experiments

In this section, the proposed method is evaluated on three types of public datasets, i.e., video data, MRI data and hyperspectral image data, which have been frequently used to interpret the tensor completion performance of different models. To demonstrate its effectiveness, we compared the proposed method with TMac [25], TV based MF-TV method [6], single nuclear norm based TNM method [30] and partial sum of tubal nuclear norm based PSTNN method [7].

To accurately evaluate the performance of the test models, two types of standards are employed to quantitatively evaluate the quality of the recovered tensor. The first one is
the visual evaluation of the restored data, which is a qualitative evaluation standard. The second one is the five quantitative picture quality indices (PQIs), including the peak signal-to-noise ratio (PSNR) [5], structural similarity index (SSIM) [24], feature similarity (FSIM) [29], erreur relative globale adimensionnelle de synthèse (ERGAS) [23], the mean the spectral angle mapper (SAM) [10]. Larger PSNR, SSIM, FSIM and smaller ERGAS, SAM are, the better the restoration performance of the corresponding model is. Since the experimental datasets are all third-order tensors, the PQIs for each frontal slice in the restored tensor are first calculated, and then the mean of these PQIs are finally used to evaluate the performance of the models. All experiments were performed on MATLAB 2018b, the CPU of the computer is Inter core i7@2.2GHz and the memory is 64GB.

5.1. Video
This subsection compares the proposed model, MF-TV, Tmac, TNN and PSTNN on two video datasets: "suzie" and "uni0102/uni0132/uni0102/uni0130“, both of which are colored using YUV format. Their sizes both are 144×176×150. We tested all three methods on a series of sampling rates (SR): 5%, 10% and 20%, and all the test models are evaluated in terms of quantitative comparison and visual evaluation.

For quantitative comparison, Table 1 and Table 2 report the PQIs of the results recovered by different methods. The best result for each PQI are marked in bold. From Table 1 and Table 2, it can be found that in all SR cases the proposed method obtains the highest indices, compared to other compared methods. The Tmac obtains the second best PQIs, when the SR is set to 5% or 10%, while MF-TV obtains the second best PQIs when SR is set to 20%. The margins between the results by our method and the second best results are more than 5dB considering the PSNR.

5.2. MRI
To further verify the versatility of the proposed model for different datasets, this subsection compares the proposed model, MF-TV, Tmac, TNN and PSTNN on MRI dataset, i.e., the cubical MRI data. The size of the dataset is 181×217×150. We tested all three methods on a series of sampling rates: 5%, 10%, 20% and 30%.

For quantitative comparison, Table 3 reports the PQIs of the results recovered by different methods. The best result for each PQI are marked in bold. From Table 3, it can be found that in all SR cases the proposed method obtains the highest indices, compared to other compared methods.

5.3. Hyperspectral image
This section compares the proposed model, MF-TV, Tmac, TNN and PSTNN on two HSIs datasets: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Cuprite data. The size of AVIRIS Cuprite data is 150×150×188. We tested all three methods on a series of sampling rates: 2.5%, 5% and 10%.

For quantitative comparison, Table 4 reports the average results of each tested method with three different sampling rates. From the results, we see again that the proposed method not only obtain the highest PQIs, but also recover the more structural information of the image, and restore more

---

Table 1
The averaged PSNR, SSIM, FSIM, ERGA and SAM of the recovered results on video "suzie" by Tmac, MF-TV, TNN, PSTNN and the proposed model with different sampling rates. The best value is highlighted in bolder fonts.

| SR   | method  | nosiy | our model | MF-TV | Tmac | PSTNN | TNN |
|------|---------|-------|-----------|-------|------|-------|-----|
| 0.05 | PSNR    | 7.259 | 29.464    | 13.801| 23.385| 17.447| 22.005|
|      | SSIM    | 0.009 | 0.807     | 0.094 | 0.622 | 0.192 | 0.563 |
|      | FSIM    | 0.454 | 0.885     | 0.42  | 0.792 | 0.59  | 0.776 |
|      | ERGA    | 1057.282| 83.571    | 501.117| 167.927| 327.678| 194.844|
|      | MSAM    | 77.324| 3.622     | 24.095| 6.927 | 7.797 |
| 0.1  | PSNR    | 7.493 | 32.056    | 22.356| 26.189| 26.647| 26.032|
|      | SSIM    | 0.014 | 0.878     | 0.605 | 0.74  | 0.68  | 0.692 |
|      | FSIM    | 0.426 | 0.924     | 0.758 | 0.838 | 0.843 | 0.846 |
|      | ERGA    | 1029.096| 62.314    | 196.059| 124.369| 117.104| 124.923|
|      | MSAM    | 71.725| 2.764     | 6.99  | 5.423 | 5.171 | 5.405 |
| 0.2  | PSNR    | 8.005 | 34.378    | 32.064| 27.274| 30.566| 30.561|
|      | SSIM    | 0.02  | 0.918     | 0.872 | 0.782 | 0.829 | 0.831 |
|      | FSIM    | 0.391 | 0.948     | 0.916 | 0.853 | 0.91  | 0.911 |
|      | ERGA    | 970.285| 47.877    | 66.692| 109.627| 75.472| 75.598|
|      | MSAM    | 63.522| 2.183     | 2.81  | 4.812 | 3.399 | 3.395 |

---

1http://trace.eas.asu.edu/yuv/
2http://brainweb.bic.mni.mcgill.ca/brainweb/selection_normal.html
3http://aviris.jpl.nasa.gov/html/aviris.freedata.html
Table 2
The averaged PSNR, SSIM, FSIM, ERGA and SAM of the recovered results on video "hall" by Tmac, MF-TV, TNN, PSTNN and the proposed model with different sampling rates. The best value is highlighted in bolder fonts.

| method | nosiy | our model | SR = 0.05 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   |       |           | MF-TV     |   | 13.539 | 22.101 | 16.075 | 20.78 |
| SSIM   | 0.007 | 0.894     |           |   | 0.412 | 0.675 | 0.36 | 0.636 |
| FSIM   | 0.387 | 0.920     |           |   | 0.612 | 0.789 | 0.672 | 0.792 |
| ERGA   | 1225.779 | 83.146 | 452.351 | 168.866 | 335.52 | 195.315 |
| MSAM   | 77.299 | 2.360     |           |   | 12.865 | 3.818 | 8.64 | 4.299 |

| method | nosiy | our model | SR = 0.1 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   | 5.055 | 31.804    |           |   | 24.855 | 26.936 | 29.014 | 28.433 |
| SSIM   | 0.013 | 0.935     |           |   | 0.829 | 0.854 | 0.892 | 0.905 |
| FSIM   | 0.393 | 0.950     |           |   | 0.873 | 0.888 | 0.934 | 0.936 |
| ERGA   | 1193.075 | 56.998 | 131.422 | 97.185 | 77.395 | 82.259 |
| MSAM   | 71.7 | 1.904     |           |   | 3.669 | 2.404 | 2.417 | 2.46 |

| method | nosiy | our model | SR = 0.2 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   | 5.567 | 33.941    |           |   | 33.006 | 27.648 | 33.629 | 33.691 |
| SSIM   | 0.025 | 0.952     |           |   | 0.94 | 0.869 | 0.961 | 0.962 |
| FSIM   | 0.403 | 0.964     |           |   | 0.954 | 0.897 | 0.973 | 0.974 |
| ERGA   | 1124.737 | 44.581 | 50.971 | 89.271 | 46.123 | 45.851 |
| MSAM   | 63.507 | 1.574     |           |   | 1.779 | 2.226 | 1.584 | 1.565 |

Table 3
The averaged PSNR, SSIM, FSIM, ERGA and SAM of the recovered results on MRI by Tmac, MF-TV, TNN, PSTNN and the proposed model with different sampling rates. The best value is highlighted in bolder fonts.

| method | nosiy | our model | SR = 0.05 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   | 10.258 | 26.414    |           |   | 12.332 | 20.51 | 15.859 | 18.218 |
| SSIM   | 0.228 | 0.722     |           |   | 0.099 | 0.45 | 0.224 | 0.27 |
| FSIM   | 0.473 | 0.834     |           |   | 0.52 | 0.711 | 0.642 | 0.646 |
| ERGA   | 1030.203 | 184.279 | 814.747 | 339.385 | 545.77 | 434.774 |
| MSAM   | 76.54 | 20.411    |           |   | 55.603 | 31.367 | 36.355 | 31.11 |

| method | nosiy | our model | SR = 0.1 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   | 10.492 | 32.652    |           |   | 15.406 | 21.411 | 22.061 | 22.535 |
| SSIM   | 0.241 | 0.912     |           |   | 0.25 | 0.531 | 0.482 | 0.536 |
| FSIM   | 0.511 | 0.926     |           |   | 0.587 | 0.732 | 0.764 | 0.78 |
| ERGA   | 1002.8 | 89.116    | 584.827 | 308.655 | 275.473 | 266.753 |
| MSAM   | 70.986 | 14.637    |           |   | 41.826 | 29.345 | 24.585 | 24.6 |

| method | nosiy | our model | SR = 0.2 | | Tmac | PSTNN | TNN |
|--------|-------|-----------|-----------|---|-----|-----|-----|
| PSNR   | 11.003 | 36.529    |           |   | 27.062 | 22.33 | 29.152 | 28.571 |
| SSIM   | 0.271 | 0.962     |           |   | 0.737 | 0.586 | 0.804 | 0.802 |
| FSIM   | 0.564 | 0.963     |           |   | 0.84 | 0.754 | 0.895 | 0.891 |
| ERGA   | 945.583 | 57.037 | 173.636 | 276.269 | 127.133 | 136.182 |
| MSAM   | 62.887 | 11.559    |           |   | 21.792 | 27.267 | 17.513 | 17.855 |

spatial details than comparison methods, especially at low sampling rates than compared methods. Therefore, one can see that the recovered data obtained by our models get the best visual evaluation and PQIs.

6. Conclusions
In this paper, we propose a new low-rank model based on multi-mode matrix decomposition for tensor completion. Instead of using the traditional single nuclear norm to represent the low-rank prior of underlying tensor directly, we first
Enhanced nonconvex LR representation for tensor completion

Table 4
The averaged PSNR, SSIM, FSIM, ERGA and SAM of the recovered results on hyperspectral image "Cuprite" by Tmac, MF-TV, TNN, PSTNN and the proposed model with different sampling rates. The best value is highlighted in bolder fonts.

| method | nosiy | our model | MF-TV | Tmac | PSTNN | TNN |
|--------|-------|-----------|-------|------|-------|-----|
| PSNR   | 7.666 | 34.595    | 26.115| 21.25| 13.387| 22.783|
| SSIM   | 0.007 | 0.861     | 0.539 | 0.412| 0.124 | 0.554|
| FSIM   | 0.48  | 0.916     | 0.765 | 0.755| 0.613 | 0.775|
| ERGA   | 1043.633 | 50.383 | 237.074 | 235.594 | 539.574 | 245.333 |
| MSAM   | 81.221| 1.662     | 12.913| 7.842| 17.98 | 9.156|

| method | nosiy | our model | MF-TV | Tmac | PSTNN | TNN |
|--------|-------|-----------|-------|------|-------|-----|
| PSNR   | 7.779 | 38.202    | 34.684| 28.945| 20.621| 26.579|
| SSIM   | 0.01  | 0.928     | 0.845 | 0.712| 0.31  | 0.663|
| FSIM   | 0.471 | 0.960     | 0.915 | 0.846| 0.735 | 0.836|
| ERGA   | 1030.139 | 41.898 | 89.372| 93.325| 234.445| 154.292|
| MSAM   | 77.268| 1.559     | 4.386 | 3.278| 7.886 | 5.413|

| method | nosiy | our model | MF-TV | Tmac | PSTNN | TNN |
|--------|-------|-----------|-------|------|-------|-----|
| PSNR   | 8.013 | 39.056    | 40.888| 35.627| 35.51 | 35.015|
| SSIM   | 0.014 | 0.939     | 0.957 | 0.885| 0.907 | 0.897|
| FSIM   | 0.451 | 0.966     | 0.978 | 0.931| 0.951 | 0.943|
| ERGA   | 1002.75 | 34.544 | 34.263| 44.518| 54.421| 57.537|
| MSAM   | 71.695| 1.299     | 1.46  | 1.445| 2.072 | 2.192|

apply parallel matrix factorization to all modes of underlying tensor, then, a novel double non-convex $L_p$ norm is designed to represent the low-rank structure in all modes of underlying tensor. An BSUM-based algorithm is designed to efficiently solve the proposed model, and it can be demonstrated that our numerical scheme converge to the coordinatewise minimizers. The proposed models have been evaluated on three types of public datasets, which show that our algorithms can recover a variety of low-rank tensors with significantly fewer samples than the compared methods.

References
[1] Cao, W., Wang, Y., Sun, J., Meng, D., Yang, C., Cichocki, A., Xu, Z., 2016. Total variation regularized tensor pca for background subtraction from compressive measurements. IEEE Transactions on Image Processing 25, 4075–4090.
[2] Chen, Y., Guo, Y., Wang, Y., Wang, D., Peng, C., He, G., 2017. Denoising of hyperspectral images using nonconvex low rank matrix approximation. IEEE Transactions on Geoscience and Remote Sensing 55, 5366–5380.
[3] Gaïffas, S., Lecué, G., 2011. Weighted algorithms for compressed sensing and matrix completion. arXiv preprint arXiv:1107.1638 .
[4] Harshman, R.A., Lundy, M.E., 1994. Parafac: Parallel factor analysis. Computational Statistics & Data Analysis 18, 39–72.
[5] Huyynh-Thu, Q., Ghanbari, M., 2008. Scope of validity of psnr in image/video quality assessment. Electronics letters 44, 800–801.
[6] Ji, T.Y., Huang, T.Z., Zhao, X.L., Ma, T.H., Liu, G., 2016. Tensor completion using total variation and low-rank matrix factorization. Information Sciences 326, 243–257.
[7] Jiang, T.X., Huang, T.Z., Zhao, X.L., Deng, L.J., 2020. Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm. Journal of Computational and Applied Mathematics 372, 112680.
[8] Kilmer, M.E., Braman, K., Hao, N., Hoover, R.C., 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM Journal on Matrix Analysis and Applications 34, 148–172.
[9] Kreimer, N., Sacchi, M.D., 2012. A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation. Geophysics 77, V113–V122.
[10] Kruse, F., Lefkoff, A., Dietz, J., 1993. Expert system-based mineral mapping in northern death valley, california/nevada, using the airborne visible/infrared imaging spectrometer (aviris). Remote Sensing of Environment 44, 309–316.
[11] Liu, J., Musialski, P., Wonka, P., Ye, J., 2012. Tensor completion for estimating missing values in visual data. IEEE transactions on pattern analysis and machine intelligence 35, 208–220.
[12] Lou, Y., Yin, P., He, Q., Xin, J., 2015. Computing sparse representation in a highly coherent dictionary based on difference of 11 and 12. Journal of Scientific Computing 64, 178–196.
[13] Lu, C., Tang, J., Yan, S., Lin, Z., 2015. Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm. IEEE Transactions on Image Processing 25, 829–839.
[14] Malek-Mohammadi, M., Babaie-Zadeh, M., Skoglund, M., 2015. Performance guarantees for schatten-p quasi-norm minimization in recovery of low-rank matrices. Signal Processing 114, 225–230.
[15] Nie, F., Wang, H., Cai, X., Huang, H., Ding, C., 2012. Robust matrix completion via joint schatten p-norm and lp-norm minimization, in: 2012 IEEE 12th International Conference on Data Mining, IEEE, pp. 566–574.
[16] Patwardhan, K.A., Sapiro, G., Bertalmio, M., 2007. Video inpainting under constrained camera motion. IEEE Transactions on Image Processing 16, 545–553.
[17] Piao, X., Hu, Y., Gao, J., Sun, Y., Yin, B., 2019. Double nuclear norm based low rank representation on grassmann manifolds for clustering, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12075–12084.
[18] Razaviyayn, M., Hong, M., Luo, Z.Q., 2013. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. SIAM Journal on Optimization 23, 1126–1153.
[19] Sauve, A.C., Hero, A.O., Rogers, W.L., Wilderman, S.J., Clithorne,
Enhanced nonconvex LR representation for tensor completion

N.H., 1999. 3d image reconstruction for a compton spect camera model. IEEE Transactions on Nuclear Science 46, 2075–2084.

[20] Shang, F., Cheng, J., Liu, Y., Luo, Z.Q., Lin, Z., 2017. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. IEEE transactions on pattern analysis and machine intelligence 40, 2066–2080.

[21] Sun, J.T., Zeng, H.J., Liu, H., Lu, Y., Chen, Z., 2005. Cubesvd: a novel approach to personalized web search, in: Proceedings of the 14th international conference on World Wide Web, pp. 382–390.

[22] Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31, 279–311.

[23] Wald, L., 2002. Data fusion: definitions and architectures: fusion of images of different spatial resolutions. Presses des MINES.

[24] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13, 600–612.

[25] Xu, Y., Hao, R., Yin, W., Su, Z., 2013. Parallel matrix factorization for low-rank tensor completion. arXiv preprint arXiv:1312.1254.

[26] Yuan, J., 2019. Mri denoising via sparse tensors with reweighted regularization. Applied Mathematical Modelling 69, 552–562.

[27] Zeng, H.J., Xie, Z.M., Wen-Feng, K., Cui, S., Ning, J.F., 2020. Hyperspectral image denoising via combined non-local self-similarity and local low-rank regularization. IEEE Access 8, 50190–50208.

[28] Zhang, H., Yang, J., Shang, F., Gong, C., Zhang, Z., 2018. Lrr for subspace segmentation via tractable schatten-p norm minimization and factorization. IEEE transactions on cybernetics 49, 1722–1734.

[29] Zhang, L., Zhang, L., Mou, X., Zhang, D., 2011. FSIM: A feature similarity index for image quality assessment. IEEE transactions on Image Processing 20, 2378–2386.

[30] Zhang, Z., Aeron, S., 2016. Exact tensor completion using t-svd. IEEE Transactions on Signal Processing 65, 1511–1526.