Low Rank Tensor Completion via Schatten-2/3 Norm and Schatten-1/2 Norm

Xianfeng Ding, Tingting Wei, Yiyuan Wang, and Jiaxin Li

***Abstract*—In tensor learning, recent studies have shown that the Tensor Train (TT) decomposition can capture the hidden information from tensor better than Tucker decomposition. The TT decomposition is determined by a well-balanced matricization scheme. Moreover, as a convex relaxation of the rank, the nuclear norm may cause the data recovery to differ significantly from the original value. Therefore, we propose a completion model based on TT decomposition and nonconvex Schatten-p norm. The objective function is nonconvex. It requires a large-scale singular value decomposition or eigenvalue decomposition. We introduce Schatten-2/3 and Schatten-1/2 quasi-norms to transform the problem into a convex programming with respect to norm minimization. The completion algorithms STT-2/3 and STT-1/2 are proposed by the proximal alternating linearized minimization (PALM). Experiments show that the proposed method is effective for data recovery.**

***Index Terms*—Data recovery, Proximal alternating linearized minimization, Schatten-p norm, Tensor optimization**

I. Introduction

With the explosive growth of information, data is no longer limited to a 2-order structure. Data loss is common due to errors and noises. The problem of data completion is gaining more attention. At present, 2-order data completion has been mature [1]-[3]. For higher-order problems, the usual approach is simply to transform the data into matrices or vectors. Its disadvantage is low decomposition efficiency and easy to cause spatial redundancy [4].

Essentially, a tensor is a higher-order generalization of a matrix. The tensor has the advantage that it can store structures between neighboring data. In recent years, more and more theories of tensors have been proposed by scholars [5]-[7]. The completion problem is extended to higher-dimensional spaces by using tensors. In simple terms, tensor completion is the inference of missing terms in a tensor from partial observations.

Manuscript received May 30, 2022; revised December 10, 2022.

Xianfeng Ding is an associate professor in the School of Science, Southwest Petroleum University, Chengdu 610500, PR China. (e-mail: fxxd@163.com).

Tingting Wei is a postgraduate student in the School of Science, Southwest Petroleum University, Chengdu 610500, PR China. (e-mail: WDuo155886@l63.com).

Yiyuan Wang is a postgraduate student in the School of Science, Southwest Petroleum University, Chengdu 610500, PR China. (e-mail: 1743821680@qq.com).

Jiaxin Li is a postgraduate student in the School of Science, Southwest Petroleum University, Chengdu 610500, PR China. (e-mail: 1195374462@qq.com).

In 1927, Hitchcock [8], [9] proposed the concept of n-mode rank. He demonstrated that the sum of tensors of finite rank can be expressed as a tensor. In the years since, tensor decompositions have been extensively studied.Tucker [10] introduced the well-known Tucker decomposition in 1966. The CANDECOMP (Canonical Decomposition) model was proposed by Carroll and Chang [11] in 1970. Moreover, Harshman [12] presented the PARAFAC (Parallel Factor) model. Then, Kiers HAL [13] proved that the CANDECOMP model and PARAFAC model are actually equivalent. These two models are now generally referred to as the CP decomposition. Both CP decomposition and Tucker decomposition work well for the low-order tensor. However, as the order rises, it becomes difficult to obtain the CP factor and the Tucker decomposition will grow exponentially. In light of this, after 2009 W. Hackbusch *et al.* [14]-[16] presented tree-type decomposition. It relies on spatially indexed splits and requires recursive algorithms. This may complicate the calculation of the problem. In 2011, I.V. Oseledets [17] improved the tree-type decomposition and defined the tensor-train (TT) decomposition. In contrast to other decomposition models, TT decomposition is not impacted by the "curse of dimension" and is always flexible and stable. Consequently, the TT decomposition has attracted the interest of some researchers. For instance, Bengua [18] suggested a low TT rank method in 2017. Yuan [19] proposed a gradient descent-based algorithm based on the TT decomposition completion in the same years.

At present, there are three primary methods to solve the low-rank tensor completion (LRTC) problem. The first is to extend the rank minimization framework of the matrix completion to LRTC [20]. The second is to apply the concept of tensor algebra to solve LRTC [21]. The last one is to apply the concept of manifold to solve LRTC [22]. Obviously, the rank minimization problem is NP-hard [5]. In general, the prevalent method is to substitute the nuclear norm for the rank function of the matrix. Nonetheless, Huet *et al.* [24] discovered in 2013 that the nuclear norm may excessively penalize partial singular values. In reality, several researchers have started to study non-convex surrogate functions of matrix rank. In 2012, Nie [3], [25] presented and demonstrated that the Schatten-p norm is preferable to the nuclear norm for approximating the rank function.

Even though the Schatten-p norm is gradually being used in tensor problems [26], [27], these completion models often have two drawbacks: 1) Most of them are based on the Tucker decomposition, which can cause results that deviate from the ideal value. 2) Solving LRTC typically necessitates singular value decomposition, which can result in large-scale calculations. Motivated by these flaws, the main work of this paper are as follows: 1) We define the Schatten-p norm based on the TT decomposition. 2) A novel tensor completion model is proposed. 3) By introducing the quasi-norm, the algorithms STT-2/3 and STT-1/2 are presented by the PALM method. 4) Experiments show that the proposed algorithm is significantly superior to the existing completion methods.

II. Notions

A tensor is denoted by an Euler script letter, e.g., $\mathcal{X}$. The capital letters denote matrices. Scalars are denoted by lowercase letters, e.g., λ. The element (i,j) of a matrix A is denoted by A_{ij}, element $(i_1,i_2,\cdots i_N)$ of a tensor $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$ by $x_{i_1,i_2,\cdots i_N}$.

A. Tensor Train Decomposition

Definition 1 [28] For the tensor $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$, its k-unfolding is defined as the matrix $X_{[k]}\in R^{I_1\cdots I_k\times I_{k+1}\cdots I_N}$ with entries $X_{[k]}(\overline{i_1 i_2 .. i_k},\overline{i_{k+1}...i_N})=x_{i_1 i_2 .. i_N}$.

Definition 2 [28] For the tensor $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$, its mode-k unfolding is defined as the matrix $X_{(k)}\in R^{I_k\times I_1\cdots I_{k-1}I_{k+1}\cdots I_N}$ with entries $X_{(k)}(i_k,\overline{i_1..i_{k-1}i_{k+1}..i_N})=x_{i_1 i_2..i_N}$.

Besides, $unfold_n(\mathcal{X})=X_{[n]}$ represents the matricization process of a tensor, $fold_n(unfold_n(\mathcal{X}))=\mathcal{X}$ denotes its inverse process.

Definition 3 [17] For a tensor $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$, its tensor train (TT) decomposition can be expressed as

$$\begin{aligned}&\mathcal{X}(i_1,\cdots i_N)\\&=\sum_{\alpha_0,\cdots\alpha_N}G_1(\alpha_0,i_1,\alpha_1)G_2(\alpha_1,i_2,\alpha_2)\cdots G_N(\alpha_{N-1},i_N,\alpha_N)\end{aligned}\tag{1}$$

where $G_k(\alpha_{k-1},i_k,\alpha_k)\in R^{r_{k-1}\times I_k\times r_k}$ is a core tensor, r_k is the TT rank, $r_0=r_N=1$ is the boundary condition of the rank. Fig. 1 shows the TT decomposition of the 4-order tensor.

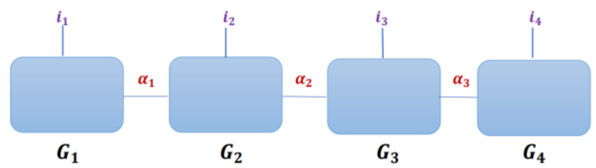


Fig. 1. Tensor train decomposition.

In fact, the rank of the matrix $X_{[k]}$ is r_k. Additionally, $rank(\mathcal{X})=(rank(X_{[1]}),...,rank(X_{[N-1]}))$ can be used to represent the multilinear rank of $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$.

B. Matrix Norm

Definition 4 [29] The Schatten-p norm $(0<p<\infty)$ of a matrix $X\in R^{m\times n}(m\geq n)$ is defined as follows

$$\|X\|_{S_p}=\left(\sum_{i=1}^{n}\sigma_i^p(X)\right)^{1/p}\tag{2}$$

where $\sigma_i(X)$ is the ith singular value of the matrix X.

Definition 5 [30] For the matrix $X\in R^{m\times n}$ with $rank(X)=r\leq d$, if $U\in R^{m\times d}$ and $V\in R^{n\times d}$ are matrices such that $X=UV^T$, the Frobenius/Nuclear Hybrid norm is defined as

$$\begin{aligned}\|X\|_{F/N}&=\min_{X=UV^T}\|U\|_*\|V\|_F\\&=\min_{X=UV^T}\left(\frac{2\|U\|_*+\|V\|_F^2}{3}\right)^{3/2}\end{aligned}\tag{3}$$

Moreover, the following equality holds

$$\|X\|_{F/N}=\|X\|_{S_{2/3}}\ (p=2/3)\tag{4}$$

Definition 6 [30] For the matrix $X\in R^{m\times n}$ with $rank(X)=r\leq d$, if $U\in R^{m\times d}$ and $V\in R^{n\times d}$ are matrices such that $X=UV^T$, the Bi-nuclear norm is defined as

$$\begin{aligned}\|X\|_{BiN}&=\min_{X=UV^T}\|U\|_*\|V\|_*\\&=\min_{X=UV^T}\left(\frac{\|U\|_*+\|V\|_*}{2}\right)^2\end{aligned}\tag{5}$$

The Bi-nuclear norm is related to the Schatten-p norm as follows

$$\|X\|_{BiN}=\|X\|_{S_{1/2}}\ (p=1/2)\tag{6}$$

III. Tensor Completion model

A. Formulation of a Model

The mode-k unfolding matrix $X_{(k)}$ is obtained by matrixing the tensor along a single mode. It captures only the correlation between the mode k and the remaining modes. Structural information for randomly combining multiple modes is ignored. In contrast, the k-unfolding matrix $X_{[k]}$ obtained via TT decomposition is much more balanced than $X_{(k)}$. From Definition 1, the rank r_k of $X_{[k]}$ globally captures the hidden correlations between the modes. Therefore, we select a completion optimization model based on TT rank.

Give a tensor $\mathcal{X}\in R^{I_1\times\cdots I_N}$, its completion based on minimizing TT rank can be formulated as

$$\begin{aligned}&\min_{\mathcal{X}} rank_{TT}(\mathcal{X})\\&s.t.\mathcal{X}_\Omega=\mathcal{T}_\Omega\end{aligned}\tag{7}$$

where $\mathcal{T}$ is the observation tensor, Ω is the index of observed entries, $\mathcal{X}_\Omega=\mathcal{T}_\Omega$ is represented as $\mathcal{X}_{i_1\cdots i_N}=\mathcal{T}_{i_1\cdots i_N}$, with $\{i_1,i_2,\cdots i_N\}\in\Omega$.

$rank_{TT}(\mathcal{X})=\left(rank(X_{[1]}),rank(X_{[2]}),\cdots rank(X_{[N-1]})\right)$ [17].

Further, the above model is written as

$$\begin{aligned}&\min_{X_{[i]}}\sum_{i=1}^{N-1}a_i rank(X_{[i]})\\&s.t.\mathcal{X}_\Omega=\mathcal{T}_\Omega\end{aligned}\tag{8}$$

where $a_i, i=1,\cdots N-1$ are the weights with $a_i\geq 0,\sum_{i=1}^{N-1}a_i=1$.

According to [3], the Schatten-p $(0<p<1)$ norm is nonconvex. But as a surrogate for rank, it is better suited than the nuclear norm. First, we define the Schatten-p norm based on the TT decomposition:

$$\|\mathcal{X}\|_{S_p}=\left(\sum_{i=1}^{N-1}a_i\|X_{[i]}\|_{S_p}^p\right)^{\frac{1}{p}}\tag{9}$$

Then, a new tensor completion model can be formulated as

$$\min_{\mathcal{X}} \sum_{i=1}^{N-1} a_i \left\| X_{[i]} \right\|_{S_p}^p \quad s.t. \mathcal{X}_\Omega = \mathcal{T}_\Omega \tag{10}$$

B. Solving the Proposed Model

The completion model (10) is a nonlinear programming that is hard to directly solve. So, we introduce the independent matrices $M_1, M_2, \cdots M_{N-1}$, the problem can be transformed into the following model:

$$\min_{M,\mathcal{X}} \sum_{i=1}^{N-1} a_i \left\| M_i \right\|_{Sp}^p + \frac{b_i}{2} \left\| M_i - X_{[i]} \right\|_F^2 \quad s.t. \mathcal{X}_\Omega = \mathcal{T}_\Omega \tag{11}$$

The model is nonconvex. The variables are separated into $\mathcal{X}$ and $(M_1, M_2, \cdots M_{N-1})$. The original problem is then split into the following two subproblems.

The first subproblem:

$$M^{k+1} \in \operatorname*{argmin}_{M} \sum_{i=1}^{N-1} a_i \left\| M_i \right\|_{Sp}^p + \frac{b_i}{2} \left\| M_i - X_{[i]}^k \right\|_F^2 + \frac{1}{2\mu_i} \left\| M_i - M_i^k \right\|_F^2 \tag{12}$$

where $0 < \mu_i < 1$.

The second subproblem:

$$\mathcal{X}^{k+1} \in \operatorname*{argmin}_{\mathcal{X}} \left(\sum_{i=1}^{N-1} \frac{b_i}{2} \left\| M_i^{k+1} - X_{[i]} \right\|_F^2 \right) + \frac{1}{2v} \left\| \mathcal{X} - \mathcal{X}^k \right\|_F^2 \tag{13}$$

where $0 < v < 1$.

1) Solving the First Subproblem

For the variable $M_i (i = 1, \cdots N-1)$, we need to solve the following problem:

$$\min_{M_i} \frac{1}{2\eta_i} \left\| M_i - M_i^k \right\|_F^2 + \frac{1}{2} \left\| M_i - X_{[i]}^k \right\|_F^2 + \gamma_i \left\| M_i \right\|_{Sp}^p \tag{14}$$

where $\gamma_i = a_i / b_i$, $\eta_i = \mu_i b_i$. (14) is further simplified:

$$\min_{M_i} \frac{1}{2} \left\| M_i - \frac{M_i^k + \sqrt{\eta_i} X_{[i]}^k}{\sqrt{\eta_i} + 1} \right\|_F^2 + \frac{\eta_i \gamma_i}{\left(\sqrt{\eta_i} + 1\right)^2} \left\| M_i \right\|_{Sp}^p \tag{15}$$

The problem is reduced to the following form for calculational ease:

$$\min_{A} \frac{1}{2} \left\| A - B \right\|_F^2 + \lambda \left\| A \right\|_{S_p}^p \tag{16}$$

For the problem (16), the key to its solution is to consider the Schatten-p norm. When $p = 2/3$, the Schatten-p norm is equivalent to the Frobenius/Nuclear Hybrid norm [30]. When $p = 1/2$, the Schatten-p norm is equivalent to the Bi-nuclear norm [30]. By substituting the Schatten-2/3 and Schatten-1/2 norms into (16), this model can be transformed into a convex optimization problem.

i) $\mathrm{p} = 2/3$, according to Definition 5

$$\left\| X \right\|_{S_p}^p = \left\| X \right\|_{S_{2/3}}^{2/3} = \frac{1}{3} \left(2 \left\| U \right\|_* + \left\| V \right\|_F^2 \right) \tag{17}$$

Let $A \in R^{m \times n}$, $U \in R^{m \times r_A}$, $V \in R^{n \times r_A}$, where r_A is the rank of matrix A. (16) is converted to

$$\min_{U,V:A=UV^T} \frac{\lambda \left(2\left\| U \right\|_* + \left\| V \right\|_F^2 \right)}{3} + \frac{1}{2} \left\| UV^T - B \right\|_F^2 \tag{18}$$

To solve for M_i is to solve for U_i and V_i. Clearly, (18) is a convex optimization problem. We use and extend the proximal alternating linearized minimization (PALM) [31].

For U^{k+1}, let $g_k(\mathrm{U}) = \frac{1}{2} \left\| UV_k^T - B \right\|_F^2$. Then its gradient is Lipschitz continuous:

$$\begin{aligned} \left\| \nabla g_k(\mathrm{U}_1) - \nabla g_k(\mathrm{U}_2) \right\|_F &= \left\| \left(U_1 V_k^T - U_2 V_k^T \right) V_k \right\|_F \\ &\le \left\| V_k \right\|_2^2 \left\| U_1 - U_2 \right\|_F = l_{k+1}^g \left\| U_1 - U_2 \right\|_F \end{aligned} \tag{19}$$

where l_{k+1}^g is the Lipschitz constant.

Further, add a proximal term:

$$\hat{g}_k(\mathrm{U}, \mathrm{U}_k) = \mathrm{g}_k(\mathrm{U}_k) + \left\langle \nabla g_k(\mathrm{U}_k), \mathrm{U} - \mathrm{U}_k \right\rangle + \frac{l_{k+1}^g}{2} \left\| U - U_k \right\|_F^2 \tag{20}$$

According to the PALM method [31]:

$$\begin{aligned} U_{k+1} &= \arg\min_{U} \frac{2\lambda}{3} \left\| U \right\|_* + \hat{g}_k(\mathrm{U}, \mathrm{U}_k) \\ &= \arg\min_{U} \frac{2\lambda}{3} \left\| U \right\|_* + \frac{l_{k+1}^g}{2} \left\| U - U_k + \frac{\nabla g_k\left(U_k\right)}{l_{k+1}^g} \right\|_F^2 \end{aligned} \tag{21}$$

Therefore, $\nabla g_k(U_k) = (U_k V_k^T - B) V_k$, $l_{k+1}^g = \left\| V_k \right\|_2^2$.

Let $Y_1 = U_k - \frac{\nabla g_k(U_k)}{l_{k+1}^g}$,its singular value decomposition is $Y_1 = U_{Y_1} \Sigma_{Y_1} V_{Y_1}{}^T$. Based on the singular value thresholding algorithm [1], the solution is directly calculated:

$$U_{k+1} = U_{Y_1} \max(\Sigma_{Y_1} - \frac{2\lambda}{3 l_{k+1}^g} \mathrm{I}, 0) V_{Y_1}^T \tag{22}$$

where $Y_1 = U_k - \frac{\left(U_k V_k^T - B \right) V_k}{\left\| V_k \right\|_2^2}$, $B = \frac{M_i^k + \sqrt{\eta_i} X_{[i]}^k}{\sqrt{\eta_i} + 1}$, and $\lambda = \frac{\mu a}{(\sqrt{\mu b} + 1)^2}$.

Next, solve V^{k+1}. Let $h_k(\mathrm{V}) = \frac{1}{2} \left\| U_{k+1} V - B \right\|_F^2$, its gradient is also Lipschitz continuous:

$$\begin{aligned} \left\| \nabla h_k(\mathrm{V}_1) - \nabla h_k(\mathrm{V}_2) \right\|_F &= \left\| U_{k+1}^T \left(U_{k+1} (\mathrm{V}_1^T - \mathrm{V}_2^T) \right) \right\|_F \\ &\le \left\| U_{k+1} \right\|_2^2 \left\| V_1 - V_2 \right\|_F = l_{k+1}^h \left\| V_1 - V_2 \right\|_F \end{aligned} \tag{23}$$

where l_{k+1}^h is the Lipschitz constant.

Similarly:

$$V_{k+1} = \arg\min_{V} \frac{\lambda}{3} \left\| V \right\|_F^2 + \frac{l_{k+1}^h}{2} \left\| V - V_k + \frac{\nabla h_k\left(V_k\right)}{l_{k+1}^h} \right\|_F^2 \tag{24}$$

where $\nabla h_k(\mathrm{V}_k) = \left(U_{k+1} V_k{}^T - B \right) U_{k+1}$, $l_{k+1}^h = \left\| U_{k+1} \right\|_2^2$.

Let $Y_2 = V_k - \frac{\nabla h_k(V_k)}{l_{k+1}^h}$, the solution is

$$V_{k+1}=\frac{l_{k+1}^{h}Y_2}{\frac{2\lambda}{3}+l_{k+1}^{h}}=\frac{3l_{k+1}^{h}Y_2}{2\lambda+3l_{k+1}^{h}} \tag{25}$$

where $Y_2=V_k-\frac{\left(U_{k+1}V_k^T-B\right)U_{k+1}}{\|U_{k+1}\|_2^2}$, $B=\frac{M_i^k+\sqrt{\eta_i}X_{[i]}^k}{\sqrt{\eta_i}+1}$,

and $\lambda=\frac{\mu a}{(\sqrt{\mu b}+1)^2}$.

ii) $p=1/2$, according to Definition 6:

$$\|X\|_{S_p}^{p}=\|X\|_{S_{1/2}}^{1/2}=\frac{1}{2}\left(\|U\|_*+\|V\|_*\right) \tag{26}$$

(16) is converted to

$$\min_{U,V:A=UV}\frac{\lambda\left(\|U\|_*+\|V\|_*\right)}{2}+\frac{1}{2}\|UV-B\|_F^2 \tag{27}$$

Similarly, we use the PALM method [31] to convert the above problem into

$$\begin{aligned}U_{k+1}&=\arg\min_U\frac{\lambda}{2}\|U\|_*+\hat{g}_k(\mathrm{U},\mathrm{U}_k)\\&=\arg\min_U\frac{\lambda}{2}\|U\|_*+\frac{l_{k+1}^{g}}{2}\left\|U-U_k+\frac{\nabla g_k\left(U_k\right)}{l_{k+1}^{g}}\right\|_F^2\end{aligned} \tag{28}$$

Obtained from [1]:

$$U_{k+1}=U_{Y_1}\max(\Sigma_{Y_1}-\frac{\lambda}{2l_{k+1}^{g}}I,0)V_{Y_1}^T \tag{29}$$

where $Y_1=U_k-\frac{\left(U_kV_k^T-B\right)V_k}{\|V_k\|_2^2}$, $B=\frac{M_i^k+\sqrt{\eta_i}X_{[i]}^k}{\sqrt{\eta_i}+1}$, and $\lambda=\frac{\mu a}{(\sqrt{\mu b}+1)^2}$.

Similarly:

$$V_{k+1}=\arg\min_V\frac{\lambda}{2}\|V\|_*+\frac{l_{k+1}^{h}}{2}\left\|V-V_k+\frac{\nabla h_k\left(V_k\right)}{l_{k+1}^{h}}\right\|_F^2 \tag{30}$$

The optimal solution of (30) is

$$V_{k+1}=U_{Y_2}\max(\Sigma_{Y_2}-\frac{\lambda}{2l_{k+1}^{h}}I,0)V_{Y_2}^T \tag{31}$$

where $Y_2=V_k-\frac{\left(U_{k+1}V_k^T-B\right)U_{k+1}}{\|U_{k+1}\|_2^2}$, $B=\frac{M_i^k+\sqrt{\eta_i}X_{[i]}^k}{\sqrt{\eta_i}+1}$, and $\lambda=\frac{\mu a}{(\sqrt{\mu b}+1)^2}$.

According to i) and ii), the optimal solution to the subproblem (12) is $M_i^{k+1}=U_i^{k+1}(V_i^{k+1})^T$.

2) Solving the Second Subproblem

After updating all M_k, the tensor $\mathcal{X}^{k+1}$ is computed. Subproblem (13) is a convex optimization. The optimal solution according to the KKT condition is

$$\mathcal{X}^{k+1}=\begin{cases}t_{i_1\cdots i_N},(\mathrm{i}_1\cdots i_N)\in\Omega\\ \left(\frac{\sum_{i=1}^{N-1}b_i\,fold_i(\mathrm{M}_i)-\mathcal{X}^k/v}{\sum_{i=1}^{N-1}b_i-1/v}\right),(\mathrm{i}_1\cdots i_N)\notin\Omega\end{cases} \tag{32}$$

where $\mathcal{T}=(t_{i_1\cdots i_N})$ is the observation data.

For the proposed tensor completion model (10), the algorithms of $p=2/3$ and $p=1/2$ are referred to as STT-2/3 and STT-1/2, respectively. The pseudocode for STT-2/3 is as follows. STT-1/2 simply needs to adjust the formulas for U^{k+1} and V^{k+1} on the basis.

Algorithm：STT-2/3

Input：$\mathcal{T}$, Ω, a, b, μ, v, λ, η, ε
Output: $\mathcal{X}$
1：while not converged do：
2： for $i=1,2,\ldots \mathrm{N}-1$ do
3： U_i^{k+1} by（22）；V_i^{k+1} by（25）；
4： $M_i^{k+1}=U_i^{k+1}(V_i^{k+1})^T$；
5： end for
6： update $\mathcal{X}^{k+1}$ by （32）；
7： if $\frac{\|\mathcal{X}^{k+1}-\mathcal{X}^k\|_F}{\|\mathcal{X}^k\|_F}<\varepsilon$ then
8：   break;
9： end if
10：end while

IV. Complexity and Convergence

For a tensor $\mathcal{X}\in R^{I_1\times I_2\times\cdots I_N}$, we assume that $I_1=I_2=\cdots I_N=I$ and $r_1=r_2=\cdots r_{N-1}=r$. The original model (10) is divided into sub-problems (12) and (13). In this paper, we analyze the complexity of each subproblem. Since the computational complexity of (15) is $o\left(I^Nr\right)$, it follows that the complexity of (12) is $o\left((N-1)I^Nr\right)$. For (13), the computational complexity is $o\left((N-1)I^N\right)$. Suppose the proposed algorithm requires K iterates in total, its computation complexity is $o\left(K(N-1)I^Nr\right)$.

Let $\{\psi^k\}_{k=0,1,..}$ ($\{\xi^k\}_{k=0,1,..}$) be the sequence generated by STT-2/3 (STT-1/2), then each limit point of $\{\psi^k\}$ ($\{\xi^k\}$) is a critical point of the objective function. Further, we assume that $\{\psi^k\}$ ($\{\xi^k\}$) is bounded and then $\{\psi^k\}$ ($\{\xi^k\}$) is converges to a critical point of the function. In other words, the algorithm is globally convergent.

The convergence proof may refer to the convergence analysis in [26], which is omitted here.

V. Numerical Experiment

In this section, STT-2/3 and STT-1/2 are used to reconstruct the lost synthetic data and color image. We compare the results with SpBCD[26], SiLRTC-TT[18], SiLRTC[32], FaLRTC[32], GeomCG[33], and TMac[34].

The algorithms are compared under different data missing ratios (mr). The missing ratio is defined as follows:

$$mr=\frac{s}{\prod_{k=1}^{N}I_k} \tag{33}$$

where s is the number of missing entries and its selection follows a uniform distribution.

In addition, the performance of the algorithm is evaluated by the relative square error (RSE). The RSE is defined as

$$RSE = \frac{\|\mathcal{X} - \mathcal{X}_{true}\|_F}{\|\mathcal{X}_{true}\|_F} \quad (34)$$

where $\mathcal{X}_{true}$ is the original tensor, $\mathcal{X}$ is the recovered tensor.

In the experiments, the weight a_i is defined as $a_i = \delta / \sum_{i=1}^{N-1} \delta_i$, where $\delta_i = \min(\prod_{l=1}^{k} I_l, \prod_{l=k+1}^{N} I_l)$. The parameter b_i is set to be $b_i = k \cdot a_i$, where k selects a value from $[0.01, 0.05, 0.1, 0.5, 1]$ that makes the best performance. μ_i is $\mu_i = 100 / b_i$. v is set to be $v = (10(N-1)) / \sum_{i=1}^{N-1} b_i$. λ_i is set to be $\lambda_i = \frac{\mu_i a_i}{(\sqrt{\mu_i b_i} + 1)^2}$, and η_i is set to be $\eta_i = \mu_i b_i$. The iteration termination condition of the algorithm is set as $\varepsilon = 10^{-4}$. The parameters of SpBCD, SiLRTC-TT, SiLRTC, FaLRTC, GeomCG and TMac are all set to their default values.

A. *Color Image Inpainting*

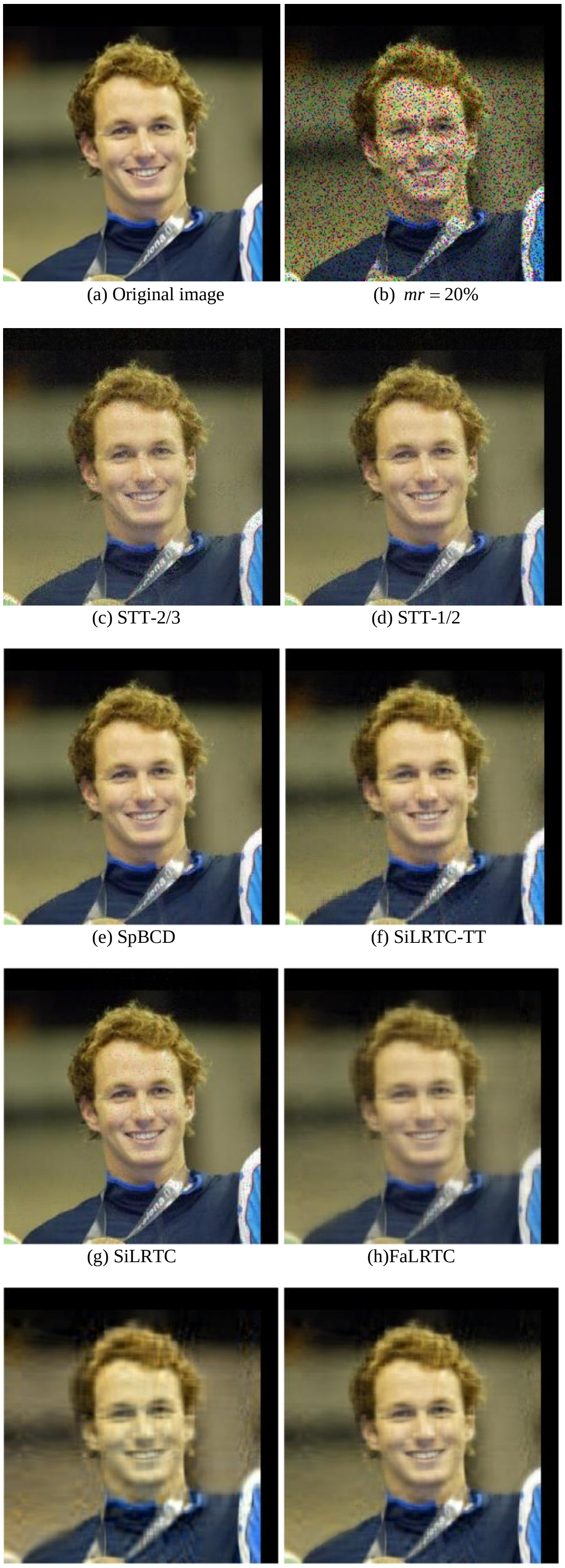
Two color images are used to test the methods. One is from the LFW face dataset, which contains 13233 faces in total. The other is from CBSD68, a collection of 68 color photographs. The face image and airplane image sizes are $250 \times 250 \times 4$ and $160 \times 240 \times 4$, respectively. The corresponding 3-order tensor $a \times b \times c$ of an image represents its $Height \times width \times RGBA$ values. We use STT-2/3, STT-1/2, SpBCD, SilRTC-TT, SiLRTC, FaLRTC, GeomCG, and TMac to complete face and airplane. Figs. 2, 3, 4, and 5 show the recovery of the two images under $mr = 20\%$ and $mr = 40\%$, respectively. Figs. 6 and 7 illustrate RSE, iteration, and runtime (s) for different mr. Tables I and II show the RSE about these methods. The best results are highlighted in bold black text.

TABLE I
THE RSE ON THE FACE IMAGE.

MR	10%	20%	30%	40%
Uncomplement	0.316	0.446	0.548	0.631
STT-2/3	0.033	0.050	0.065	0.083
STT-1/2	**0.029**	**0.043**	**0.057**	0.071
SpBCD	0.052	0.051	0.064	**0.056**
SiLRTC-TT	0.046	0.067	0.077	0.085
SiLRTC	0.032	0.053	0.114	0.177
FaLRTC	0.064	0.091	0.078	0.083
GeomCG	0.093	0.092	0.093	0.093
TMac	0.058	0.063	0.067	0.073

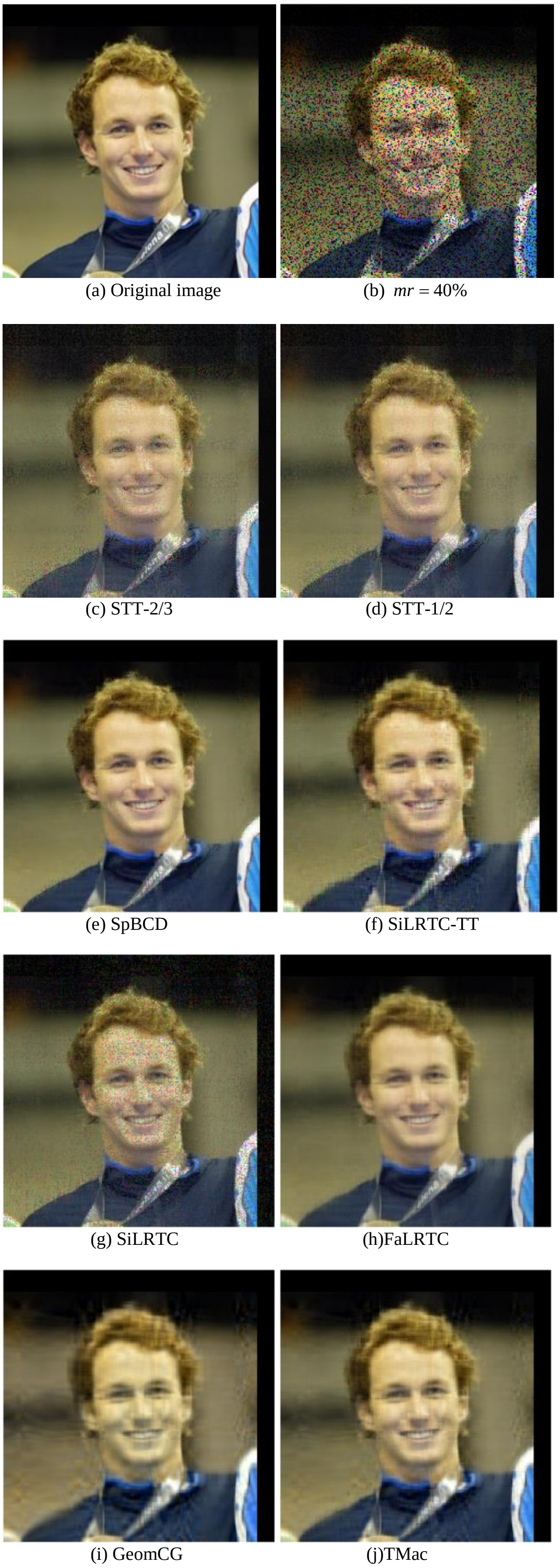
TABLE II
THE RSE ON THE AIRPLANE IMAGE.

MR	10%	20%	30%	40%
Uncomplement	0.322	0.417	0.548	0.611
STT-2/3	0.023	0.042	0.055	0.067
STT-1/2	**0.019**	**0.037**	**0.046**	0.051
SpBCD	0.026	0.039	0.048	**0.045**
SiLRTC-TT	0.036	0.056	0.066	0.079
SiLRTC	0.028	0.042	0.048	0.058
FaLRTC	0.050	0.069	0.105	0.069
GeomCG	0.044	0.044	0.048	0.049
TMac	0.030	0.038	0.047	0.047



(a) Original image (b) $mr = 20\%$
(c) STT-2/3 (d) STT-1/2
(e) SpBCD (f) SiLRTC-TT
(g) SiLRTC (h)FaLRTC
(i) GeomCG (j) TMac

Fig. 2. Results of face image recovery with 20% missing ratio using different algorithms. (a) is the undamaged original image. (b) is the image with missing ratio is 20% . (c) - (j) are the recovery results.



(a) Original image (b) $mr = 40\%$

(c) STT-2/3 (d) STT-1/2

(e) SpBCD (f) SiLRTC-TT

(g) SiLRTC (h)FaLRTC

(i) GeomCG (j)TMac

Fig. 3. Results of face image recovery with 40% missing ratio using different algorithms. (a) is the undamaged original image. (b) is the image with missing ratio is 40% . (c) - (j) are the recovery results.

(a) Original image (b) $mr = 20\%$

(c) STT-2/3 (d)STT-1/2

(e) SpBCD (f) SiLRTC-TT

(g) SiLRTC (h)FaLRTC

(i) GeomCG (j) TMac

Fig. 4. Results of airplane image recovery with 20% missing ratio using different algorithms. (a) is the undamaged original image. (b) is the image with missing ratio is 20% . (c) - (j) are the recovery results.

(a) Original image (b) $mr = 40\%$

(c) STT-2/3 (d) STT-1/2

(e) SpBCD (f) SiLRTC-TT

(g) SiLRTC (h) FaLRTC

(i) GeomCG (j) TMac

Fig. 5. Results of airplane image recovery with 40% missing ratio using different algorithms. (a) is the undamaged original image. (b) is the image with missing ratio is 40% . (c) - (j) are the recovery results.

The numerical results show that STT-2/3 and STT-1/2 are optimal in most instances. From the RSE in Tables I and II, we conclude that the proposed method is superior in recovering missing data. In addition, SpBCD performs the best among the other six methods. It is proved that the Schatten-p norm can be used to make the solution closer to the actual value.

In Figs. 6 and 7, the relative squared error of most methods increases as the mr increases. Therefore, the performance of the algorithm is affected by the severity of missing data. The algorithm with the fewest iterations is SiLRTC-TT, which directly calculates M_k^{l+1} through $X_{[k]}^l$ to reduce the number of calculation steps. In terms of running time, TMac is the most efficient. Since the final result is dependent on the initial parameters chosen, this experiment does not mean that comparison methods are flawed. Overall, the data recovery results are satisfactory.

Moreover, in order to study the convergence of the algorithm, we define the convergence condition as follows:

$$Error = \left\| \mathcal{X}^{k+1} - \mathcal{X}^k \right\|_F / \left\| \mathcal{X}^k \right\|_F \tag{35}$$

Fig. 8 shows the convergence curves of STT-1/2 and STT-2/3 under the face and airplane.

It can be observed from the steepness of the curve that STT-2/3 and STT-1/2 converge rapidly. After 5 iterations, the curve tends to flatten out and approach 0 . After 9 iterations, $Error < 10^{-4}$. This shows that STT-2/3 and STT-1/2 converge well.

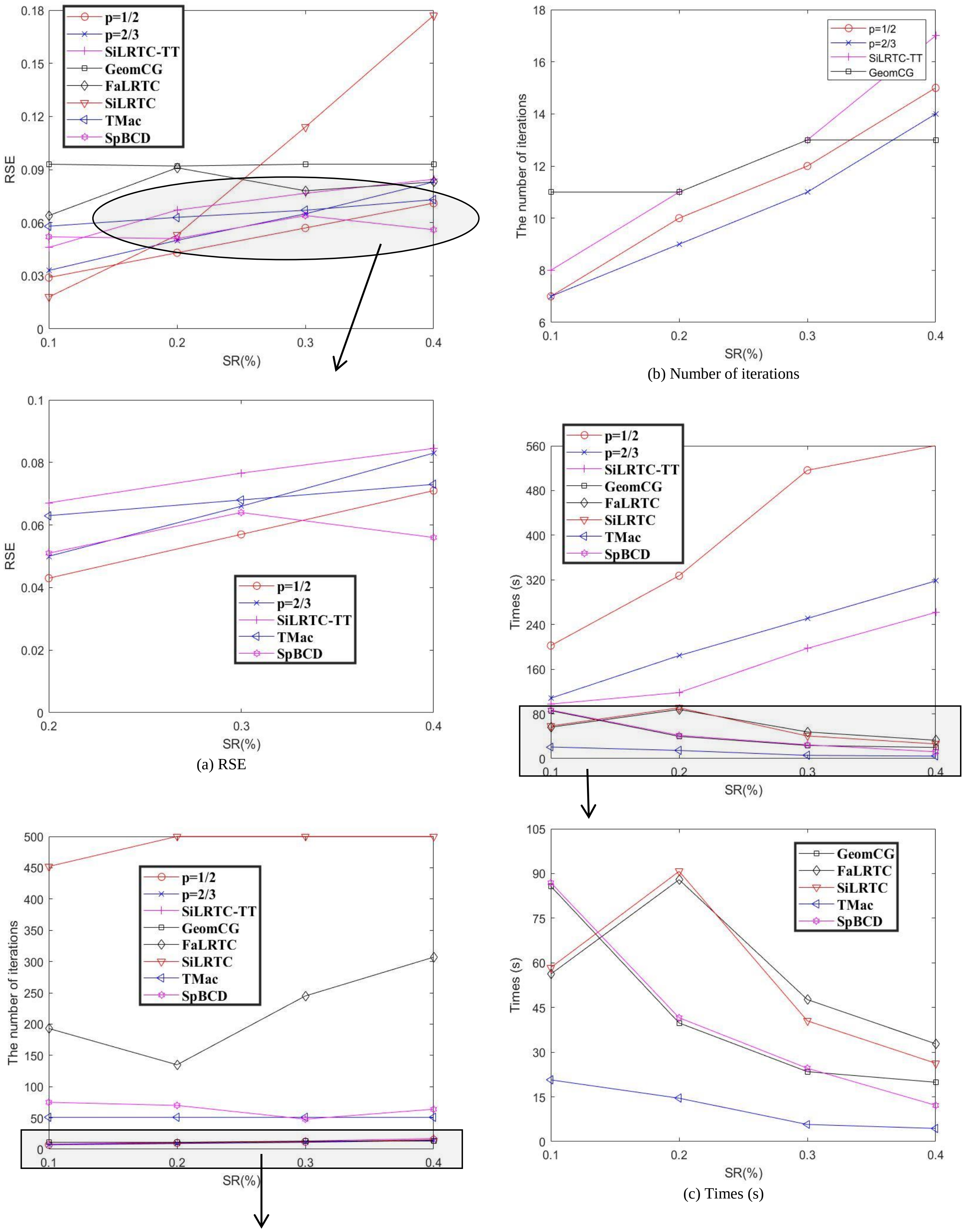
B. Synthetic Data Completion

The Schatten-p norm has significant advantages over SiLRTC, FaLRTC and SILRTC-TT methods. Overall, the three most effective techniques are STT-1/2, STT-2/3, and SpBCD. SpBCD is based on the Tucker decomposition, whereas the proposed method is based on the TT decomposition. STT-2/3, STT-1/2, and SpBCD are used to recover randomly generated multidimensional data. Next, we study the impact of different tensor decompositions on the completion problem.

We generate three tensors of different sizes, $15\times15\times15\times15(4D)$, $10\times10\times10\times10\times10(5D)$, and $5\times5\times5\times5\times5\times5(6D)$, whose elements respect the uniform distribution on the interval $(0,1)$. Assuming $mr = 20\%$, reconstruct three synthesized tensors. Fig. 9 shows the plots of RSE, iteration, and runtime (s), respectively. Table III displays the specific RSE values. Highlight the best results in bold. The convergence curves for STT-1/2 and STT-2/3 are plotted in Fig. 10.

TABLE III
THE RSE OF ALGORITHM ON THE SYNTHETIC TENSOR WITH $mr = 20\%$.

	4D	5D	6D
STT-2/3	0.072	0.098	0.157
STT-1/2	**0.065**	**0.094**	**0.156**
SpBCD	0.444	0.450	0.447



(a) RSE

(b) Number of iterations

(c) Times (s)

Fig. 6. The RSE, iteration and runtime (s) of algorithms on face data under different mr. (a) is the RSE. (b) is the number of algorithm iterations. (c) is the time consumed.

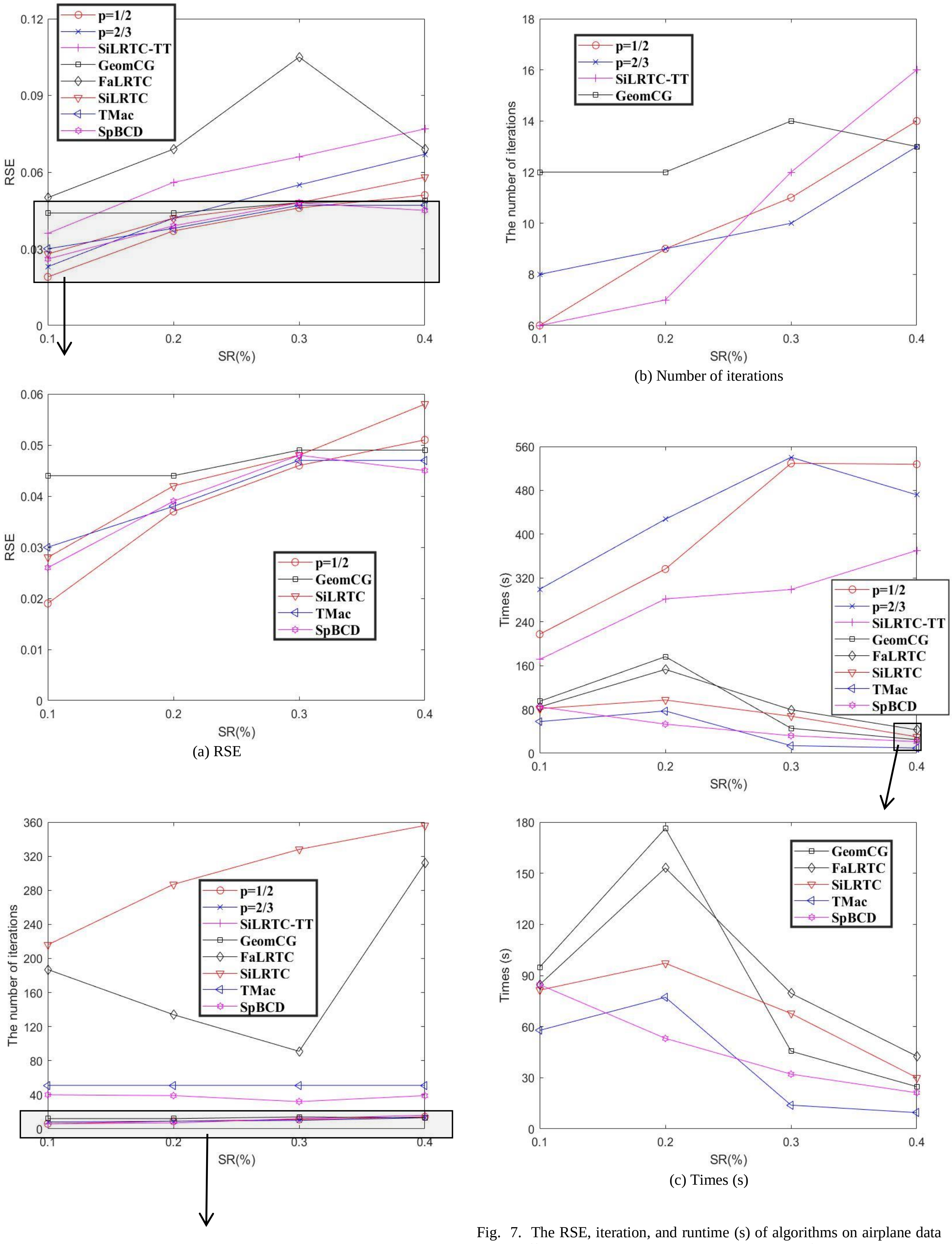
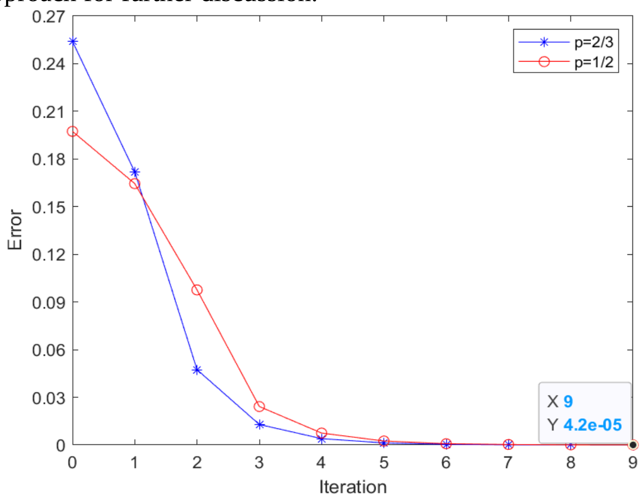


Fig. 7. The RSE, iteration, and runtime (s) of algorithms on airplane data under different mr. (a) is the RSE. (b) is the number of algorithm iterations. (c) is the time consumed.

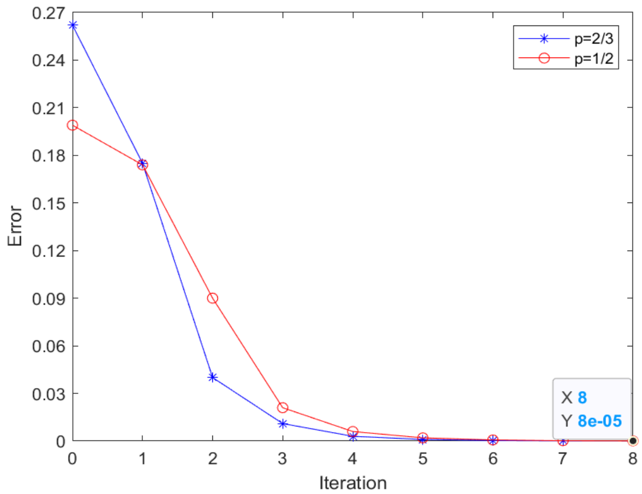
Fig. 9 and Table III show that STT-1/2 has the best recovery results. STT-1/2 has a higher global solution than STT-2/3. But it is also more difficult to compute because the model $p = 1/2$ adds a nuclear norm term. Moreover, the RSE indicates that the proposed method significantly outperforms SpBCD. This shows that TT decomposition is more suitable than Tucker decomposition to replace low-rank tensor computation in experiments.

VI. Conclusion

In this paper, we derive a new tensor completion model based on TT decomposition and Schatten-p norm. To solve the model, we first split the original problem into two subproblems. Then, we solve the norm minimization problem using the PALM method and the singular value thresholding method. Finally, the STT-2/3 and STT-1/2 algorithms are proposed. In contrast to existing methods, our algorithm does not require a large number of singular value decompositions. Experiments show that TT rank yields superior completion results. The proposed algorithm is effective. However, as the image size rises, the memory consumption of calculation will continue to increase. Therefore, the algorithm must be further optimized. Besides, different parameters directly affect the recovery results. Therefore, we will consider the adaptive approach for further discussion.

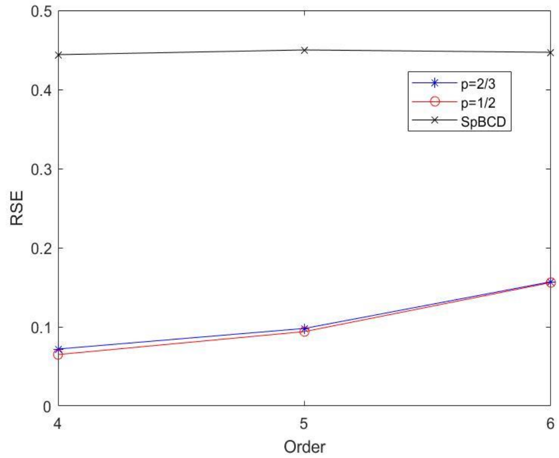


(a) face image

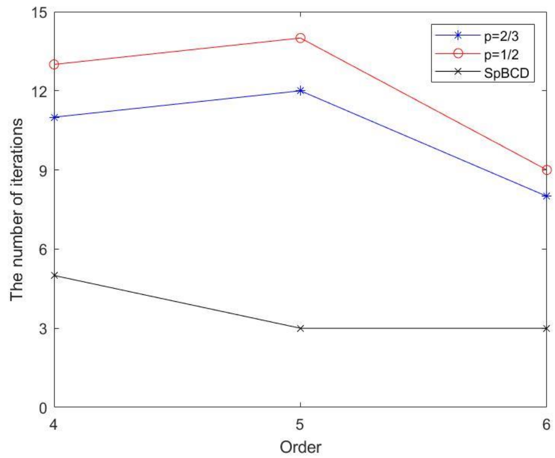


(b) airplane image

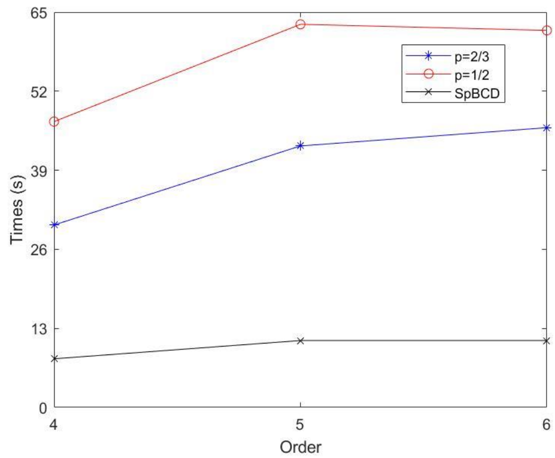
Fig. 8. Convergence plots for STT-2/3 and STT-1/2 with 20% missing ratio.



(a) RSE

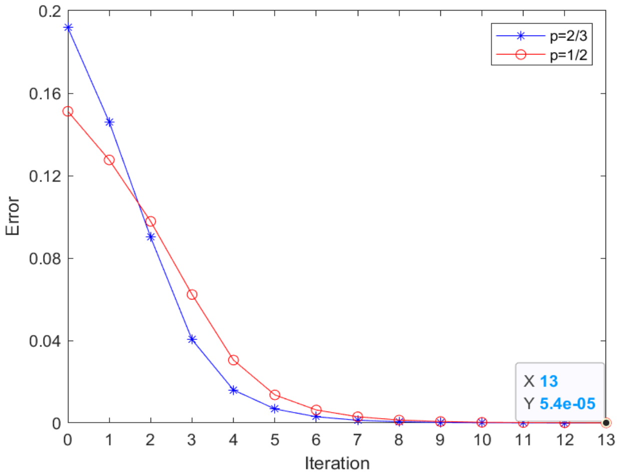


(b) Number of iterations

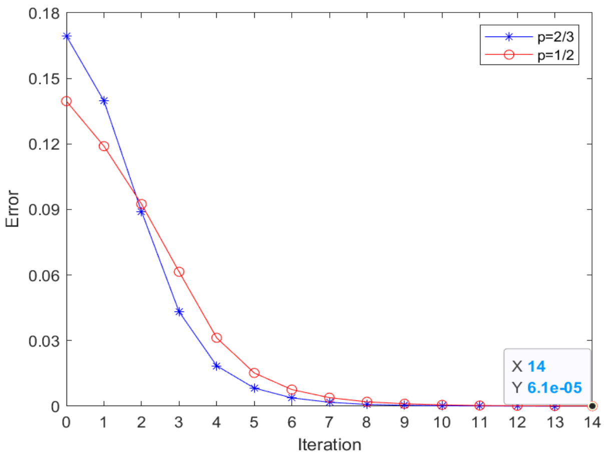


(c) Times

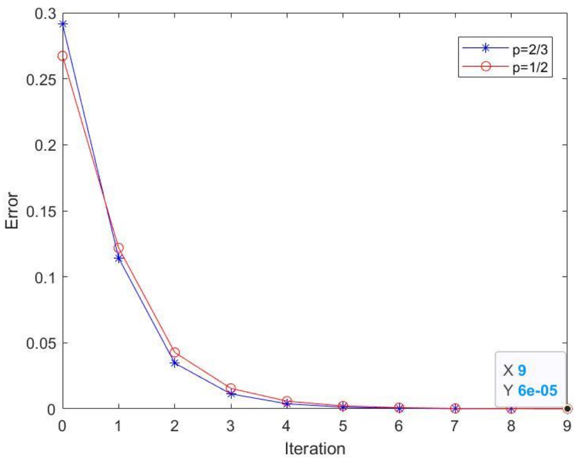
Fig. 9. The RSE, iteration and runtime (s) of algorithms on the data of different orders. (a) is the RSE. (b) is the number of algorithm iterations. (c) is the time consumed.



(a) 4-order



(b) 5-order



(c) 6-order

Fig. 10. Convergence plots for STT-2/3 and STT-1/2 with 20% missing ratio.

Fig. 9 and Table III show that STT-1/2 has the best recovery results. STT-1/2 has a higher global solution than STT-2/3. But it is also more difficult to compute because the model $p = 1/2$ adds a nuclear norm term. Moreover, the RSE indicates that the proposed method significantly outperforms SpBCD. This shows that TT decomposition is more suitable than Tucker decomposition to replace low-rank tensor computation in experiments.

VII. Conclusion

In this paper, we derive a new tensor completion model based on TT decomposition and Schatten-p norm. To solve the model, we first split the original problem into two subproblems. Then, we solve the norm minimization problem using the PALM method and the singular value thresholding method. Finally, the STT-2/3 and STT-1/2 algorithms are proposed. In contrast to existing methods, our algorithm does not require a large number of singular value decompositions. Experiments show that TT rank yields superior completion results. The proposed algorithm is effective. However, as the image size rises, the memory consumption of calculation will continue to increase. Therefore, the algorithm must be further optimized. Besides, different parameters directly affect the recovery results. Therefore, we will consider the adaptive approach for further discussion.

References

[1] Cai, Jian-Feng, Candè; s, Emmanuel J, Shen, Zuowei. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, Vol.20: 1956-1982.
[2] Lin, XF; Wei, G. Accelerated reweighted nuclear norm minimization algorithm for low rank matrix recovery[J]. Signal Processing, 2015, Vol.114: 24-33.
[3] Feiping Nie; Heng Huang; Chris Ding. Low-Rank matrix recovery via efficient Schatten p-Norm minimization[A]. Twenty-Sixth AAAI Conference on Artificial Intelligence[C], 2012.
[4] Shashua, Amnon; Hazan, Tamir. Non-negative tensor factorization with applications to statistics and computer vision[A]. ICML 2005: 22nd International Conference on Machine Learning[C], 2005.
[5] T.G. Kolda, B.W. Bader. Tensor decompositions and applications[J]. Siam Review, 2009, 51(3): 455-500.
[6] Vasilescu, M.A.O.; Terzopoulos, D. Multilinear subspace analysis of image ensembles[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, Vol.2: II93-II99.
[7] Thomas Franz; Antje Schultz; Sergej Sizov; Steffen Staab. TripleRank: ranking semantic web data by tensor decomposition[A]. International Semantic Web Conference (ISWC 2009) [C], 2009.
[8] Frank L. Hitchcock. The expression of a tensor or a polyadic as a sum of products[J]. Studies in Applied Mathematics, 1927, Vol.6: 164-189.
[9] Frank L. Hitchcock. Multiple invariants and generalized rank of a p-way matrix or tensor[J]. Studies in Applied Mathematics, 1928, Vol.7: 39-79.
[10] Tucker, L R. Some mathematical notes on three-mode factor analysis[J]. Psychometrika, 1966, Vol.31(3): 279-311.
[11] J. Douglas Carroll; Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition[J]. Psychometrika, 1970, Vol.35(3): 283-319.
[12] Harshman R A, Lundy M E. PARAFAC: Parallel factor analysis[J]. Computational Statistics & Data Analysis, 1994, 18(1): 39-72.
[13] Henk A.L.Kiers. A three-step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity[J]. Journal of Chemometrics, 1998, Vol.12(3): 155-171.
[14] Oseledets, I.V; Tyrtyshnikov, E.E. Breaking the curse of dimensionality, or how to use SVD in many dimensions[J]. SIAM Journal on Scientific Computing, 2010, Vol.31(5): 3744-3759.
[15] Hackbusch, W.; Kühn, S. A new scheme for the tensor representation[J]. Journal of Fourier Analysis and Applications, 2009, Vol.15(5): 706-722.
[16] Grasedyck, Lars. Hierarchical singular value decomposition of tensors[J]. SIAM Journal on Matrix Analysis and Applications, 2010, Vol.31(4): 2029-2054.
[17] Oseledets, I.V. Tensor-Train decomposition[J]. Siam Journal on Scientific Computing, 2011, Vol.33(5): 2295-2317.
[18] Bengua, Johann A; Phien, Ho N; Tuan, Hoang Duong; Do, Minh N. Efficient tensor completion for color image and video recovery:

low-rank tensor train[J]. IEEE Transactions on Image Processing, 2017, Vol.26(5): 2466-2479.

[19] Yuan, Longhao; Zhao, Qibin; Cao, Jiangting. Completion of high order tensor data with missing entries via Tensor-Train decomposition[J]. Neural Information Processing, 2017, Vol.10634: 222-229.

[20] Long, Z; Liu, YP; Chen, LX; Zhu, C. Low rank tensor completion for multiway visual data[J]. Signal Processing, 2019, Vol.155: 301-316.

[21] Shi, Jiarong; Jiao, Licheng; Shang, Fanhua. Tensor completion algorithm and its application in face recognition[J]. Pattern Recognition and Artificial Intelligence, 2011, Vol.24(2): 255-261.

[22] Kressner, Daniel; Steinlechner, Michael; Vandereycken, Bart. Low-rank tensor completion by Riemannian optimization[J]. BIT: Numerical Mathematics, 2014, Vol.54(2): 447-468.

[23] Benjamin Recht; Maryam Fazel; Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization[J]. SIAM Review, 2010, Vol.52(3): 471-501.

[24] Hu, Y; Zhang, DB; Ye, JP; Li, XL; He, XF. Fast and accurate matrix completion via truncated nuclear norm regularization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, Vol.35(9): 2117-2130.

[25] Nie, Feiping; Wang, Hua; Cai Xiao; Huang, Heng; Ding, Chris. Robust matrix completion via joint Schatten p-Norm and lp-Norm minimization[A]. 2012 IEEE 12th International Conference on Data Mining[C], 2012.

[26] Gao, Shangqi; Fan, Qibin. Robust Schatten-p Norm based approach for tensor completion[J]. Journal of Scientific Computing, 2020, Vol.82(1): 1-30.

[27] Kong, H; Xie, XY; Lin, ZC. t-Schatten-p Norm for low-rank tensor recovery[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, Vol.12(6): 1405-1419.

[28] Andrzej Cichocki. Tensor networks for big data analytics and large-scale optimization problems[J]. Computer Science, 2014.

[29] Schatten, Robert. A theory of cross-spaces[J]. Princeton University Press, Princeton, N.J., 1950.

[30] Fanhua Shang; Yuanyuan Liu; James Cheng. Scalable algorithms for tractable Schatten Quasi-Norm minimization[J]. Learning, 2016.

[31] Jérôme Bolte; Shoham Sabach; Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems[J]. Mathematical Programming, 2013, Vol.146: 459-494.

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

[33] Kressner, Daniel; Steinlechner, Michael; Vandereycken, Bart. Low-rank tensor completion by Riemannian optimization[J]. BIT Numerical Mathematics, 2014, Vol.54(2): 447–468.

[34] Xu, YY; Hao, RR; Yin, WT; Su, ZX. Parallel matrix factorization for low-rank tensor completion[J]. Inverse Problems and Imaging, 2015, Vol.9(2): 601–624.