Inverse Problems and Imaging

doi:10.3934/ipi.2024018

A NOVEL ℓ_0 MINIMIZATION FRAMEWORK OF TENSOR TUBAL RANK AND ITS MULTI-DIMENSIONAL IMAGE COMPLETION APPLICATION

Jin-Liang Xiao[1], Ting-Zhu Huang[*1], Liang-Jian Deng[*1] and Hong-Xia Dou[2]

[1]School of Mathematical Sciences,
University of Electronic Science and Technology of China, China

[2]School of Science, Xihua University, Chengdu, China

(Communicated by Gabriele Steidl)

Abstract. Recently, minimizing the tensor tubal rank based on the tensor singular value decomposition (t-SVD) has attracted significant attention in the tensor completion task. The widely-used solutions of tensor-tubal-rank minimization rely upon various convex and nonconvex surrogates of the tensor rank. However, these tensor rank surrogates usually lead to inaccurate descriptions of the tensor rank. To mitigate the limitation, we propose an innovative ℓ_0 minimization framework with guaranteed convergence to provide a novel paradigm for minimization of the tensor rank. To demonstrate the effectiveness of our framework, we develop a new tensor completion model employing a tensor adaptive sparsity-deduced rank (TASR). Subsequently, we formulate an algorithm rooted in the proposed ℓ_0 minimization framework to address this model effectively. Experimental results on multi-dimensional image data demonstrate that our method is superior to several state-of-the-art approaches. The code is accessible at https://github.com/Jin-liangXiao/L0-TC.

1. Introduction.

Multi-dimensional data, such as multispectral image (MSI), hyperspectral image (HSI), and video, can be effectively represented as tensors that are fundamental units in many fields, e.g., image processing [27, 40, 43, 44, 47], pattern recognition [30, 56, 65], and machine learning [17, 24, 41, 55]. The tensor rank minimization problem focuses on recovering the underlying tensor from the incomplete observation [53], which is generally expressed as follows,

$$\min_{\mathcal{X}} \operatorname{rank}(\mathcal{X}), \quad s.t. \quad \mathcal{P}(\mathcal{X}) = \mathcal{B}, \tag{1}$$

where $\mathcal{B}$ is the incomplete tensor, $\mathcal{X}$ is the underlying tensor, and $\mathcal{P}(\cdot)$ is a linear projector. The choice of $\mathcal{P}$ relies on the specific application [39, 64, 67].

In contrast to matrix rank, the tensor rank (i.e., $\operatorname{rank}(\mathcal{X})$ in (1)) is not uniquely defined [15]. Generally, the tensor rank is related to the corresponding tensor decomposition [42, 63]. For example, the CANDECOM/PARAFAC (CP) rank [68]

2020 *Mathematics Subject Classification.* Primary: 15A69, 68U10; Secondary: 90C26.

Key words and phrases. Rank minimization, rank surrogate, tensor completion, mathematical program with equilibrium constraints (MPEC), proximal alternating direction method of multipliers (PADMM).

*Corresponding author: Ting-Zhu Huang, Liang-Jian Deng.

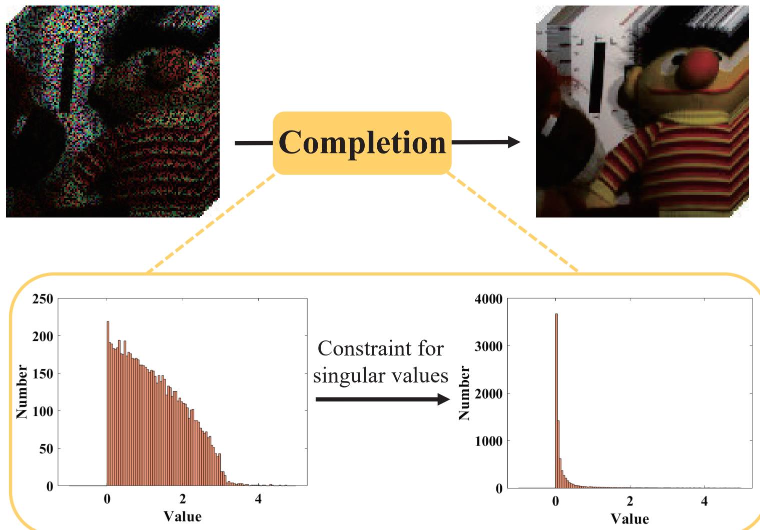


FIGURE 1. The diagram of low rank tensor completion from the observed tensor to the underlying tensor. The histograms below are the singular value distributions of the incomplete tensor and underlying tensor, respectively (the horizontal axis and vertical axis are the number and value of singular values). It is clear that the target underlying tensor with low-tensor-rank property tends to have the sparse distribution of tensor singular values. (data: MSI *Toy*, sampling rate: 20%)

is defined by the minimal number of rank-one tensors to approximate the target tensor since CP decomposition represents the tensor as the sum of rank-one tensors. The Tucker rank [58], is a vector wherein each element corresponds to the rank of the unfolding matrix of the tensor along each mode. This definition is in strong alignment with the format of Tucker decomposition. However, the calculation of the CP rank is NP-hard, and the Tucker rank inevitably destroys the internal structure of the tensor [49, 50].

Currently, tensor tubal rank [22] that is based on the tensor singular value decomposition (t-SVD) [13, 14] is developed to explore the low-rank structure of tensors in the frequency domain, which has drawn attention in many practical applications [34, 54]. Frequency transformations, e.g., fast Fourier transformation (FFT) and discrete cosine transformation (DCT), are involved in t-SVD, and the operations can separate the low-frequency and high-frequency information while maintaining the low-rank structure [28]. Also, the t-SVD can be extended to other generalized linear transformations to obtain the low-rank structure. However, the minimization of tensor tubal rank is difficult, and directly solving the tensor rank minimization problem as (1) is usually NP-hard [16]. Lu *et al.* [21] demonstrate that the low tubal rank property of tensor can be well constrained by low average rank and minimize it with a new tensor nuclear norm. Besides, various convex and nonconvex tensor rank surrogates are proposed to approximate the tensor average rank. As described in Figure 1 and the theoretical analysis of Lemma 3.1 and Remark 3.2, the target underlying tensor with low-tensor-rank property tends to have the sparse distribution of tensor singular values. Due to the strong correlation between t-SVD-based tensor ranks (e.g. tensor tubal rank and average rank) and tensor singular values (please refer to Lemma 3.1 and Remark 3.2), these tensor rank surrogates shrink singular values to constrain the low-tensor-rank property, which has applied in the tensor completion as shown in Figure 2.

In real applications, the minimization of tensor rank is usually converted into its proximal problem [8,26]. Specifically, tensor nuclear norm (TNN) [22,66] was proposed to minimize the tensor average rank in the proximal problem as follows,

$$\min_{\mathcal{X}} \ \|\mathcal{X}\|_* + \frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_F^2, \tag{2}$$

where $\mathcal{Y}$ is known, and $\|\mathcal{X}\|_*$ means the TNN of $\mathcal{X}$, which is defined by the sum of tensor singular values of $\mathcal{X}$. Nevertheless, TNN equally constrains each singular value, serving as ℓ_1-norm of singular values, which usually leads to biased solution [62]. To overcome the drawback, several nonconvex rank surrogates [1,9,33] were subsequently given to alleviate this dilemma [12], which can be denoted as follows,

$$\min_{\mathcal{X}} \ \Psi(\mathcal{X}) + \frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_F^2, \tag{3}$$

where $\Psi(\mathcal{X})$ is the tensor rank surrogate of $\mathcal{X}$ to approximate the tensor average rank. For example, Jiang *et al.* [9] proposed the partial sum of singular values to shrink the smaller singular values. Wang *et al.* [33] proposed a generalized nonconvex method to approach the tensor tubal rank and average rank. These nonconvex tensor rank surrogates can alleviate the biased solution of TNN [2].

Nonetheless, rank surrogates are actually designed by reducing the shrinkage of the larger singular values, which often suffer from limited performance in applications. As shown in Figure 2, these tensor rank surrogates cannot describe the low-rank constraint property of the rank function well, which usually leads to the over-penalization of singular values [33]. Thus, it is critical to search for a more accurate constraint of tensor singular values. Virtually, the tensor average rank is equal to the ℓ_0-norm of tensor singular values (see Lemma 3.1), which means there exists a strong connection between the low-rank property of tensor and the sparsity of tensor singular values. Motivated by the above analysis, we tend to develop a novel approach to minimize the tensor tubal rank by describing the sparsity of tensor singular values. Besides, it is also critical to design an effective algorithm to solve it [4,19,57,59].

In this paper, we propose a novel ℓ_0 minimization framework to minimize the tensor tubal rank. This framework reformulates this problem to a biconvex Mathematical Program with Equilibrium Constraints (MPEC), which provides a powerful constraint for the sparsity of tensor singular values compared with existing tensor rank surrogates. Recently, several transformed low-rank presentations based on tensor tubal rank, e.g., [10,31,45], have gained notable attention. These presentations can also be well described by corresponding revised algorithms in this framework. In addition, for the specific tensor completion task, we design a new completion model according to the sparse constraint characteristics of the ℓ_0 minimization framework. Multi-dimensional image data usually have powerful similarities in the distribution of singular values, which can be utilized to enhance the correlation between the low-rank property of tensor and the sparsity of singular values. The new model introduces an adaptive orthogonal transformation that makes the singular values of the low-rank tensor more sparse, and it can also reduce computational complexity (please refer to Figure 4 and Sect. 4.4). Then, we develop an algorithm based on the

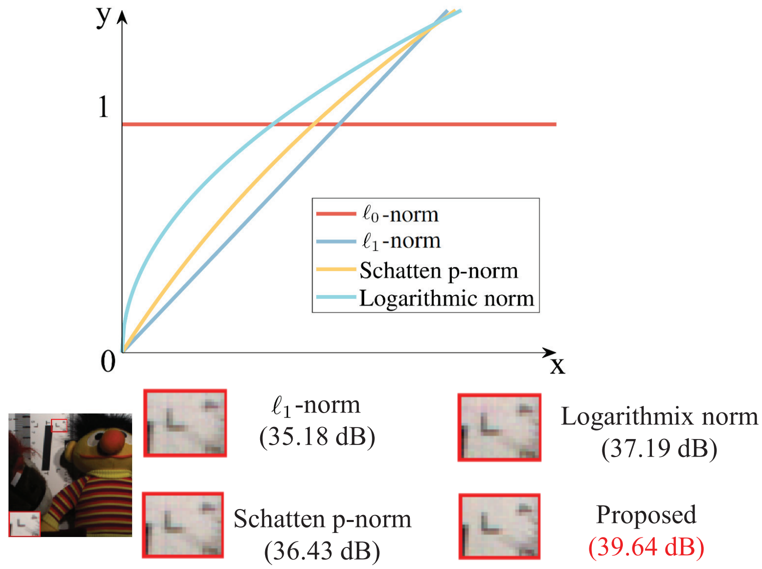


FIGURE 2. Constraint comparison of different approaches, i.e., TNN (ℓ_1-norm) [22], Schatten p-norm [25], Logarithmic norm [1], and the proposed approach, for singular values and their experimental performance. We plot the constraint curves of different approaches for singular values and apply them to the tensor completion task. Their results are assessed by the index peak signal-to-noise ratio (PSNR) [37] (data: MSI *Toy*, sampling rate: 20%)

ℓ_0 minimization framework to solve it. Numerical experiments on MSI, HSI, and video data demonstrate the effectiveness of the completion model, which also verifies the practical potential of the ℓ_0 minimization framework.

To sum them up, our contributions are listed as follows,

- We give a novel ℓ_0 minimization framework of tensor tubal rank by constraining the sparsity of tensor singular values, which can be also applicable to the minimization of other tensor ranks related to sparsity.
- For multi-dimensional image completion application, we design a new model based on the ℓ_0 minimization framework and develop an effective algorithm to solve it. This model fully explores the sparsity of tensor singular values and the low-rank property of tensor and utilizes the powerful sparse constraint ability of the proposed framework.
- Experiments display that the new completion model achieves state-of-the-art performance on multi-dimensional image data, which also verified the effectiveness of the ℓ_0 minimization framework.

2. Notations amd preliminaries.

2.1. **Notations.** In this paper, tensors, matrices, and vectors are represented by calligraphic letters, uppercase bold letters, and lowercase bold letters, e.g., $\mathcal{A}$, $\mathbf{A}$, and $\mathbf{a}$, respectively. Particularly, $\mathbf{I}$ means the identity matrix. The Matlab notation $\mathcal{A}(:,:,i)$ or $\mathbf{A}^{(i)}$ is the i-th frontal slice of $\mathcal{A}$. $\mathcal{A}(i,j,:)$ denotes the (i,j)-th tube of $\mathcal{A}$. $\mathbf{A}_{(3)} = \texttt{unfold}_3(\mathcal{A})$ means the unfolding matrix of tensor $\mathcal{A}$ along the third

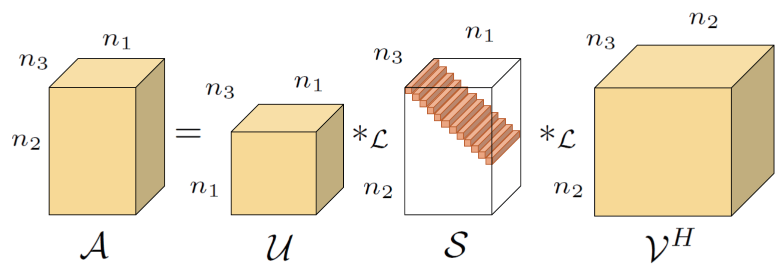


FIGURE 3. The graphical representation of t-SVD for tensor $\mathcal{A}$.

dimension, and $\texttt{fold}_3$ is the inverse operator of $\texttt{unfold}_3$. $(\cdot)^T$ denotes the matrix transpose operation. $\mathcal{C} = \mathcal{A} \triangle \mathcal{B}$ means $\mathbf{C}^{(i)} = \mathbf{A}^{(i)}\mathbf{B}^{(i)}$ [23]. Let $\mathcal{L}$ represent an invertible linear transformation, and $\mathbf{L}$ means its corresponding linear transform matrix. For the invertible linear transform matrix $\mathbf{L}$, $\overline{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{L} = \texttt{fold}_3(\mathbf{L}\mathbf{A}_{(3)})$. means $\mathcal{A}$ on the transform $\mathbf{L}$ domain. The invertible transform matrix $\mathbf{L}$ usually has the following assumption:

$$\mathbf{L}\mathbf{L}^T = \mathbf{L}^T\mathbf{L} = \gamma \mathbf{I}, \tag{4}$$

where γ is a positive fixed constant. Besides, the block circulant matrix bcirc($\mathcal{A}$) of $\mathcal{A}$ is defined as

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} & \mathbf{A}^{(n_3)} & \cdots & \mathbf{A}^{(2)} \\ \mathbf{A}^{(2)} & \mathbf{A}^{(1)} & \cdots & \mathbf{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{(n_3)} & \mathbf{A}^{(n_3-1)} & \cdots & \mathbf{A}^{(1)} \end{bmatrix}. \tag{5}$$

2.2. **Preliminaries.** Before the main results, we briefly introduce some definitions about transformation-based t-product [14] and t-SVD [13].

Definition 2.1 (T-product [13,14])**.** Denote $\mathcal{L}$ as an invertible linear transformation that satisfies (4). The transformation $\mathcal{L}$ based t-product of two tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times l \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{l \times n_2 \times n_3}$ is defined as

$$\mathcal{C} = \mathcal{A} *_{\mathcal{L}} \mathcal{B} = (\overline{\mathcal{A}} \triangle \overline{\mathcal{B}}) \times_3 \mathbf{L}^{-1}, \tag{6}$$

where $\overline{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{L}$, $\overline{\mathcal{B}} = \mathcal{B} \times_3 \mathbf{L}$, $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathbf{L} \in \mathbb{R}^{n_3 \times n_3}$ is the transform matrix, and $\mathbf{L}^{-1}$ is the inverse transform matrix of $\mathbf{L}$.

Definition 2.2 (Tensor transpose [14])**.** Denote $\mathcal{L}$ as an invertible linear transformation that satisfies (4). The transpose of tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is represented as $\mathcal{A}^H$ that satisfies $\overline{\mathbf{A}^H}^{(i)} = (\overline{\mathbf{A}}^{(i)})^T, i = 1, \cdots, n_3$.

Definition 2.3 (Identity tensor [14])**.** Denote $\mathcal{L}$ as an invertible linear transformation that satisfies (4). The tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is an identity tensor if satisfies $\overline{\mathbf{A}}^{(i)}$ is an identity matrix, $i = 1, \cdots, n_3$.

Definition 2.4 (Orthogonal Tensor [14])**.** Denote $\mathcal{L}$ as an invertible linear transformation that satisfies (4). Tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal under $\mathcal{L}$ if it satisfies $\mathcal{A} *_{\mathcal{L}} \mathcal{A}^H = \mathcal{A}^H *_{\mathcal{L}} \mathcal{A} = \mathcal{I}$.

Definition 2.5 (T-SVD [13,22])**.** Denote $\mathcal{L}$ as an invertible linear transformation that satisfies (4). Any tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be represented as

$$\mathcal{A} = \mathcal{U} *_{\mathcal{L}} \mathcal{S} *_{\mathcal{L}} \mathcal{V}^H, \tag{7}$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V}^H \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ consists of tensor singular values, and each frontal slice of $\mathcal{S}$ is a diagonal matrix.

The t-SVD of $\mathcal{A}$ can be graphically denoted as Figure 3.

Definition 2.6 (Tensor tubal rank [22])**.** Denote $\mathcal{U} *_{\mathcal{L}} \mathcal{S} *_{\mathcal{L}} \mathcal{V}^H$ as the t-SVD of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ under the invertible linear transformation $\mathcal{L}$ that satisfies (4). The tubal rank of $\mathcal{A}$ is defined as the number of nonzero singular tubes of $\mathcal{S}$, which can be represented as follows,

$$\operatorname{rank}_t(\mathcal{A}) := \sharp\{i|\mathcal{S}(i,i,:) \neq 0\}. \tag{8}$$

Definition 2.7 (Tensor average rank [21])**.** For tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor average rank of $\mathcal{A}$ is represented as follows,

$$\operatorname{rank}_a(\mathcal{A}) := \frac{1}{n_3}\operatorname{rank}(\mathtt{bcirc}(\mathcal{A})). \tag{9}$$

3. Main results.

There exists the close relationship between tensor tubal rank and tensor average rank as follows [21],

$$\operatorname{rank}_a(\mathcal{A}) \leq \operatorname{rank}_t(\mathcal{A}). \tag{10}$$

Thus, the tensor with low tubal rank always has low average rank. Besides, the minimization of tensor tubal rank can be approximated by minimizing the tensor average rank in applications [38]. Hence, we formulate an equivalent form of minimizing tensor average rank to approximate the tensor tubal rank.

3.1. Proposed ℓ_0 minimization framework.

In many practical applications, e.g., denoising [3,7], recovery [36,70], and tensor completion [20], The low-tensor-rank optimization model (1) are solved by its proximal form. Thus, we only consider to solve the following proximal version of the tensor average rank minimization problem:

$$\min_{\mathcal{X}} \lambda \operatorname{rank}_a(\mathcal{X}) + \frac{1}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2, \tag{11}$$

where the tensor $\mathcal{Y}$ is known. However, directly solving the model (11) is NP-hard [35]. Unlike most methods, we do not solve it by rank surrogates. Instead, we explore the relationship between the tensor and its singular values. Based on this model, we find that there exists a strong correlation between the low tensor-average-rank property of tensor and the sparsity of singular values. To better depict the correlation, we introduce Lemma 3.1 as follows,

Lemma 3.1. *Denote $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. $\mathcal{L}$ is an invertible linear transformation that satisfies (4), and $\mathcal{U} *_{\mathcal{L}} \mathcal{S} *_{\mathcal{L}} \mathcal{V}^H$ represents the t-SVD of $\mathcal{Y}$. Then the optimum to the following problem:*

$$\min_{\mathcal{X}} \lambda \; rank_a(\mathcal{X}) + \frac{1}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2, \tag{12}$$

can be expressed as $\mathcal{X}^* = \mathcal{R}^* \times_3 \mathbf{L}^{-1}$, *where* i-*th frontal slice of* $\mathcal{R}$ *satisfies* $\mathbf{R}^{*(i)} = \mathbf{U}^{(i)}\mathbf{W}^{*(i)}\mathbf{V}^{(i)T}$, *and the diagonal matrix* $\mathbf{W}^{*(i)}$ *is the solution of the following problem:*

$$\min_{\mathbf{W}^{(i)}} \frac{\lambda}{n_3} \left\|\mathbf{W}^{(i)}\right\|_0 + \frac{1}{2\gamma} \left\|\mathbf{W}^{(i)} - \mathbf{S}^{(i)}\right\|_F^2, \qquad i = 1, \cdots, n_3, \tag{13}$$

where $\|\cdot\|_0$ *means* ℓ_0 *norm and* γ *is a fixed constant (please refer to Section 2.1).*

Proof. Following [21], since the transform matrix $\mathbf{L}$ satisfies (4), we can get

$$(\mathbf{L} \otimes \mathbf{I}_{n_1}) \cdot (\mathtt{bcirc}(\mathcal{X})) \cdot (\mathbf{L}^{-1} \otimes \mathbf{I}_{n_2}) = \begin{bmatrix} \overline{\mathbf{X}}^{(1)} & & & \\ & \overline{\mathbf{X}}^{(2)} & & \\ & & \ddots & \\ & & & \overline{\mathbf{X}}^{(n_3)} \end{bmatrix}. \tag{14}$$

Denote $\mathcal{R} = \mathcal{X} \times_3 \mathbf{L}$, hence $\text{rank}(\mathtt{bcirc}(\mathcal{X})) = \sum_{i=1}^{n_3} \text{rank}(\mathbf{R}^{(i)})$. According to the definition of the tensor average rank and the property of $\mathbf{L}$, we have

$$\begin{aligned} &\lambda \ \text{rank}_a(\mathcal{X}) + \frac{1}{2} \|\mathcal{X} - \mathcal{Y}\|_F^2 \\ =&\frac{\lambda}{n_3} \sum_{i=1}^{n_3} \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma} \left\|\overline{\mathcal{X}} - \overline{\mathcal{Y}}\right\|_F^2 \\ =&\frac{\lambda}{n_3} \sum_{i=1}^{n_3} \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma} \left\|\mathcal{R} - \overline{\mathcal{Y}}\right\|_F^2 \\ =&\sum_{i=1}^{n_3} \frac{\lambda}{n_3} \ \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma} \left\|\mathbf{R}^{(i)}\right\|_F^2 + \frac{1}{2\gamma} \left\|\overline{\mathbf{Y}}^{(i)}\right\|_F^2 - \frac{1}{\gamma} Tr(\mathbf{R}^{(i)T}\mathbf{Y}^{(i)}). \end{aligned} \tag{15}$$

Denote $\dot{\mathbf{U}}^{(i)}\mathbf{W}^{(i)}\dot{\mathbf{V}}^{(i)T}$ as the SVD of the matrix $\mathbf{R}^{(i)}, i = 1, 2, \cdots, n_3$. We assume $n_1 \leq n_2$, and other conditions can be handled similarly. By von Neumanns trace inequality [29], $Tr(\mathbf{R}^{(i)T}\overline{\mathbf{Y}}^{(i)})$ achieves its upper bound $\sum_{j=1}^{n_1} w_j^i s_j^i$ if and only if $\dot{\mathbf{U}}^{(i)} = \mathbf{U}^{(i)}$ and $\dot{\mathbf{V}}^{(i)} = \mathbf{V}^{(i)}$, where w_j^i and s_j^i are the j-th diagonal element of matrix $\mathbf{W}^{(i)}$ and $\mathbf{S}^{(i)}$, respectively. Thus, we can obtain that the problem (12) is equal to the following problem:

$$\min_{w_j^i} \ \sum_{j=1}^{n_1} \frac{\lambda}{n_3} |w_j^i|_0 + \frac{1}{2\gamma}(w_j^i - s_j^i)^2, \ i = 1, \cdots, n_3,$$

which is also equivalent to the problem (13). Denote $\mathbf{W}^{*(i)}$ is the optimun to (13), $i = 1, \cdots, n_3$. We can get $\mathbf{R}^{*(i)} = \mathbf{U}^{(i)}\mathbf{W}^{*(i)}\mathbf{V}^{(i)T}$. Hence, $\mathcal{X}^* = \mathcal{R}^* \times_3 \mathbf{L}^{-1}$ is the solution of problem (12). □

Remark 3.2. For the fixed transformation $\mathcal{L}$, γ and n_3 are constant. The factors γ and n_3 can be integrated with λ in (13). Thus, the optimization models (13) is converted into following equivalent formats:

$$\min_{\mathbf{W}^{(i)}} \lambda \left\|\mathbf{W}^{(i)}\right\|_0 + \frac{1}{2} \left\|\mathbf{W}^{(i)} - \mathbf{S}^{(i)}\right\|_F^2, \qquad i = 1, \cdots, n_3, \tag{16}$$

Accordingly, the tensor-average-rank minimization can be equivalently converted into the ℓ_0-norm sparse constraint for its singular values. Based on the observation, we propose a new biconvex Mathematical Program with Equilibrium Constraints

(MPEC) to solve the tensor-average-rank minimization problem, which is clarified by following Lemma 3.3 and Theorem 3.4.

Lemma 3.3 ([60]). *For any vector* $\mathbf{w}$*, we have*

$$\|\mathbf{w}\|_0 = \min_{0 \leq \mathbf{z} \leq 1} \langle \mathbf{1}, \mathbf{1} - \mathbf{z} \rangle \,, s.t. \ \mathbf{z} \odot |\mathbf{w}| = \mathbf{0}, \tag{17}$$

where $\odot$ *means element-wise product, and* $\mathbf{z}$ *is a vector with size of* $\mathbf{w}$*.* $\mathbf{z}_* = \mathbf{1} - sign(|\mathbf{w}|)$ *is the unique solution of* (17) *and the signum function* $sign(\cdot)$ *is componentwise.*

Theorem 3.4. *Let* $\mathbf{w}_i$ *and* $\mathbf{s}_i$ *be the vector stretched by the diagonal elements of the diagonal matrices* $\mathbf{W}^{(i)}$ *and* $\mathbf{S}^{(i)}$ *in* (16)*, respectively,* $i = 1, \cdots, n_3$*. The minimization problem* (16) *is equivalent to the following problem:*

$$\begin{aligned} \min_{0 \leq \mathbf{z}_i \leq 1, \mathbf{w}_i} & \langle \mathbf{1}, \mathbf{1} - \mathbf{z}_i \rangle + \frac{1}{2\lambda} \|\mathbf{w}_i - \mathbf{s}_i\|^2 \,, \\ s.t. \ \ & \mathbf{z}_i \odot |\mathbf{w}_i| = \boldsymbol{0}, \qquad i = 1, \cdots, n_3, \end{aligned} \tag{18}$$

where $\langle \cdot \rangle$*,* $\odot$*, and* $|\cdot|$ *respectively denote the inner product, element-wise product, and absolute operator.* λ *is the parameter in* (16)*.*

Proof. Since $\mathbf{W}^{(i)}$ and $\mathbf{S}^{(i)}$ are diagonal matrices, we only need to consider the diagonal elements of $\mathbf{W}^{(i)}$ and $\mathbf{S}^{(i)}$. The problem (16) can be converted as follows,

$$\min_{\mathbf{w}_i} |\mathbf{w}_i|_0 + \frac{1}{2\lambda} \|\mathbf{w}_i - \mathbf{s}_i\|^2 \,, \qquad i = 1, \cdots, n_3,$$

where $\mathbf{w}_i$ and $\mathbf{s}_i$ are represented as the vector stretched by the diagonal elements of the matrices $diag(w_1^i, w_2^i, \cdots, w_{n_1}^i)$ and $diag(s_1^i, s_2^i, \cdots, s_{n_1}^i)$.Based on Lemma 3.3, the solution to the above problem can be obtained by the problem as follows,

$$\begin{aligned} \min_{0 \leq \mathbf{z}_i \leq 1, \mathbf{w}_i} & \langle \mathbf{1}, \mathbf{1} - \mathbf{z}_i \rangle + \frac{1}{2\lambda} \|\mathbf{w}_i - \mathbf{s}_i\|^2 \,, \\ s.t. \ \ & \mathbf{z}_i \odot |\mathbf{w}_i| = \mathbf{0}, \qquad i = 1, \cdots, n_3. \end{aligned} \tag{19}$$

The proof is completed. □

Subsequently, we design a solving algorithm for the optimation problem (18). Since $\mathbf{w}_i$, $\mathbf{z}_i$, and $\mathbf{s}_i$ in (18) are obtained from different frontal slices independently, we respectively denote them as $\mathbf{w}$, $\mathbf{z}$, and $\mathbf{s}$ for convenience. Based on the proximal alternating direction method of multipliers (PADMM) scheme [60], the augmented Lagrangian function is formulated as follows,

$$\mathtt{L}(\mathbf{w}, \mathbf{z}, \mathbf{p}) = \langle \mathbf{1}, \mathbf{1} - \mathbf{z} \rangle + \frac{1}{2\lambda} \|\mathbf{w} - \mathbf{s}\|^2 + \langle \mathbf{z} \odot |\mathbf{w}|, \mathbf{p} \rangle + \frac{\alpha}{2} \|\mathbf{z} \odot |\mathbf{w}|\|^2 \,, \tag{20}$$

where $0 \leq \mathbf{z} \leq 1$, $\mathbf{p}$ is the Lagrangian multiplier, and α is a positive penalty parameter. Hence, we can solve the problem (18) by two steps as follows:

z-update: The PADMM scheme introduces the proximal term $\left\|\mathbf{z} - \mathbf{z}^k\right\|^2$ for the variable $\mathbf{z}$. The k-th iteration of $\mathbf{z}$ is updated by

$$\mathbf{z}^{k+1} = \arg \min_{0 \leq \mathbf{z} \leq 1} \left\langle \mathbf{z}, \mathbf{c}^k \right\rangle + \frac{\alpha}{2} \left\|\mathbf{z} \odot |\mathbf{w}^k|\right\|^2 + \frac{\beta}{2} \left\|\mathbf{z} - \mathbf{z}^k\right\|^2 \,, \tag{21}$$

where β is the penalty parameter, and $\mathbf{c}^k = \mathbf{p}^k \odot |\mathbf{w}^k| - \mathbf{1}$. Thus, $\mathbf{z}^{k+1}$ is updated by

$$\mathbf{z}^{k+1} = \min(\mathbf{1}, \max(\mathbf{0}, \frac{-\mathbf{c}^k + \beta \mathbf{z}^k}{\alpha |\mathbf{w}^k| \odot |\mathbf{w}^k| + \beta})). \tag{22}$$

Algorithm 1 The PADMM-based solver for tensor average rank minimization problem (18)

Input: vector $\mathbf{s}$
Parameter: $k_{\text{mit}} = 100$, λ, α, β
Output: $\mathbf{w}$

1: Initialization $k = 0, \mathbf{z}^0 = \mathbf{p}^0 = \mathbf{0}$, and $\mathbf{w}^0 = \mathbf{s}$
2: **while** $k < k_{\text{mit}}$ **do**
3: Update $\mathbf{z}^{k+1}$ via (22)
4: Update $\mathbf{w}^{k+1}$ via (24)
5: Update $\mathbf{p}^{k+1}$ via (25)
6: $k = k + 1$
7: **end while**

w-update: Similarly, we can obtain

$$\mathbf{w}^{k+1} = \arg\min_{\mathbf{w}} \frac{\alpha}{2} \left\| \mathbf{z}^{k+1} \odot |\mathbf{w}| + \frac{\mathbf{p}^k}{\alpha} \right\|^2 + \frac{1}{2\lambda} \|\mathbf{w} - \mathbf{s}\|^2. \tag{23}$$

Therefore, $\mathbf{w}^{k+1}$ is solved by

$$\mathbf{w}^{k+1} = sign(\mathbf{s}) \odot \max(\mathbf{0}, \frac{\frac{1}{\lambda}|\mathbf{s}| - \mathbf{p}^k \odot \mathbf{z}^{k+1}}{\frac{1}{\lambda} + \alpha \mathbf{z}^{k+1} \odot \mathbf{z}^{k+1}}), \tag{24}$$

where $sign(\cdot)$ is the signum function. Subsequently, the k-th iteration of Lagrangian multiplier $\mathbf{p}$ is computed by

$$\mathbf{p}^{k+1} = \mathbf{p}^k + \alpha \mathbf{z}^{k+1} \odot |\mathbf{w}^{k+1}|. \tag{25}$$

Especially, the solving algorithm of problem (18) can be summarized in Algorithm 1. Besides, we give the convergence analysis as follows:

Theorem 3.5 (Convergence of Algorithm 1)**.** $\{\mathbf{z}^k, \mathbf{w}^k, \mathbf{p}^k\}$ *is the sequence produced by Algorithm 1. Assume* $\mathbf{p}^k$ *satisfies* $\sum_{k=0}^{\infty} \left\| \mathbf{p}^{k+1} - \mathbf{p}^k \right\|_F^2 < \infty$. *Then we have any accumulation point of the sequence that satisfies the KKT condition of* (18).

Proof. We first give the first-order KKT conditions for $\{\mathbf{z}^*, \mathbf{w}^*, \mathbf{p}^*\}$ as follows,

$$\begin{cases} 0 \in \mathbf{p}^* \odot |\mathbf{w}^*| - 1 + \partial I(\mathbf{z}^*), \\ 0 \in \frac{1}{\lambda}(\mathbf{w}^* - \mathbf{s}) + \mathbf{p}^* \odot \mathbf{z}^* \odot \partial \|\mathbf{w}^*\|_1, \\ 0 = \mathbf{z}^* \odot \mathbf{w}^*, \end{cases} \tag{26}$$

where $I(\mathbf{z})$ is the dicator function on the set $\{\mathbf{z} | \mathbf{0} \le \mathbf{z} \le \mathbf{1}\}$. The augmented Lagrangian function can be rewritten as

$$\mathtt{L}(\mathbf{w}, \mathbf{z}, \mathbf{p}) = \langle \mathbf{1}, \mathbf{1} - \mathbf{z} \rangle + \frac{1}{2\lambda} \|\mathbf{w} - \mathbf{s}\|^2 + \frac{\alpha}{2} \left\| \mathbf{z} \odot |\mathbf{w}| + \frac{\mathbf{p}}{\alpha} \right\|^2 - \frac{1}{\alpha} \|\mathbf{p}\|^2, \tag{27}$$

We denote $\mathtt{J}(\mathbf{w}, \mathbf{z}, \mathbf{p}) = \mathtt{L}(\mathbf{w}, \mathbf{z}, \mathbf{p}) + \|\mathbf{z} - \mathbf{z}'\|^2$, where $\mathbf{z}'$ is the variable $\mathbf{z}$ at the previous iteration. Thus, we have

$$\mathtt{J}(\mathbf{w}^k, \mathbf{z}^k, \mathbf{p}^k) - \mathtt{J}(\mathbf{w}^k, \mathbf{z}^{k+1}, \mathbf{p}^k) \ge \frac{\beta}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2. \tag{28}$$

Similarly, we can get

$$\mathtt{J}(\mathbf{w}^k, \mathbf{z}^{k+1}, \mathbf{p}^k) - \mathtt{J}(\mathbf{w}^{k+1}, \mathbf{z}^{k+1}, \mathbf{p}^k) \ge \frac{\beta'}{2} \left\| \mathbf{w}^k - \mathbf{w}^{k+1} \right\|^2, \tag{29}$$

where $\beta' = \min\{\alpha, \frac{1}{\lambda\gamma}\}$. According to the update of the Lagrangian multiplier, we can get

$$\mathtt{J}(\mathbf{w}^{k+1}, \mathbf{z}^{k+1}, \mathbf{p}^{k+1}) - \mathtt{J}(\mathbf{w}^{k+1}, \mathbf{z}^{k+1}, \mathbf{p}^{k}) = \frac{\alpha}{2} \left\| \mathbf{p}^k - \mathbf{p}^{k+1} \right\|^2. \tag{30}$$

Combining (28), (29), and (30), we can obtain that

$$\begin{aligned} &\mathtt{J}(\mathbf{w}^{k}, \mathbf{z}^{k}, \mathbf{p}^{k}) - \mathtt{J}(\mathbf{w}^{k+1}, \mathbf{z}^{k+1}, \mathbf{p}^{k+1}) \\ &\qquad \geq \frac{\beta}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2 + \frac{\beta'}{2} \left\| \mathbf{w}^k - \mathbf{w}^{k+1} \right\|^2 - \frac{\alpha}{2} \left\| \mathbf{p}^k - \mathbf{p}^{k+1} \right\|^2. \end{aligned} \tag{31}$$

Since $\mathtt{J}(\mathbf{w}, \mathbf{z}, \mathbf{p})$ is bounded for all $(\mathbf{w}, \mathbf{z}, \mathbf{p})$, we have

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{\beta}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2 + \frac{\beta'}{2} \left\| \mathbf{w}^k - \mathbf{w}^{k+1} \right\|^2 - \frac{\alpha}{2} \left\| \mathbf{p}^k - \mathbf{p}^{k+1} \right\|^2 \\ \leq \mathtt{J}(\mathbf{w}^0, \mathbf{z}^0, \mathbf{p}^0) - \mathtt{J}(\mathbf{w}^\infty, \mathbf{z}^\infty, \mathbf{p}^\infty) \leq \infty. \end{aligned} \tag{32}$$

Sequently, we have $\sum_{k=0}^{\infty} \lim\limits_{k\to\infty} \frac{\beta}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2 + \frac{\beta'}{2} \left\| \mathbf{w}^k - \mathbf{w}^{k+1} \right\|^2 = 0$ and $\mathbf{p}^k - \mathbf{p}^{k+1} \to 0$. According to the update of $\mathbf{z}^k$ and $\mathbf{w}^k$, we have

$$\begin{aligned} &0 = \mathbf{p}^k \odot |\mathbf{w}^k| - 1 + \partial I(\mathbf{z}^k) + \beta(\mathbf{z}^{k+1} - \mathbf{z}^k), \\ &0 \in \frac{1}{\lambda}(\mathbf{w}^* - \mathbf{s}) + \mathbf{p}^* \odot \mathbf{z}^* \odot \partial \left\| \mathbf{w}^* \right\|_1. \end{aligned} \tag{33}$$

Thus, we can get the KKT condition

$$\begin{cases} 0 \in \mathbf{p}^* \odot |\mathbf{w}^*| - 1 + \partial I(\mathbf{z}^*), \\ 0 \in \frac{1}{\lambda}(\mathbf{w}^* - \mathbf{s}) + \mathbf{p}^* \odot \mathbf{z}^* \odot \partial \left\| \mathbf{w}^* \right\|_1, \\ 0 = \mathbf{z}^* \odot \mathbf{w}^*. \end{cases} \tag{34}$$

□

Moreover, the proposed Algorithm 1 substantively provides a solving framework for minimizing tensor ranks that exist in connection with the sparsity of tensor singular values or transformed tensor singular values. Specifically, the proposed framework can effectively minimize the tensor rank, whose minimization problem can be converted into the l_0 minimization form like (13). For example, various tensor low-rank models, e.g., [10,31,45], depict the sparsity under different transformations, which can be converted into the form of (13). These models all can be solved under the proposed ℓ_0 minimization framework by minimizing the (13) form.

3.2. **Tensor completion application.** To demonstrate the effectiveness of the proposed ℓ_0 minimization framework, we apply it to the tensor completion task. Tensor completion refers to recovering the underlying tensor from the missing observation [53]. A common solving scheme resorts to the low-tensor-rank model as follows,

$$\min_{\mathcal{X}} \operatorname{rank}(\mathcal{X}), \quad s.t. \quad \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}), \tag{35}$$

where $\mathcal{M}$ is the missing tensor, $\mathcal{X}$ is the underlying tensor, $\mathcal{P}_\Omega(\cdot)$ is a projector, and $\mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M})$ means the values of $\mathcal{X}$ and $\mathcal{M}$ in the area Ω are equal. Different from directly utilizing it to minimize the tensor tubal rank, we develop a novel tensor completion model according to the characteristics of the proposed ℓ_0 minimization. First, we define a new tensor adaptive sparsity-deduced rank (TASR) to retain the advantage of the frequency domain and enhance the connection between the sparsity of tensor singular values and the low-rank property as follows,

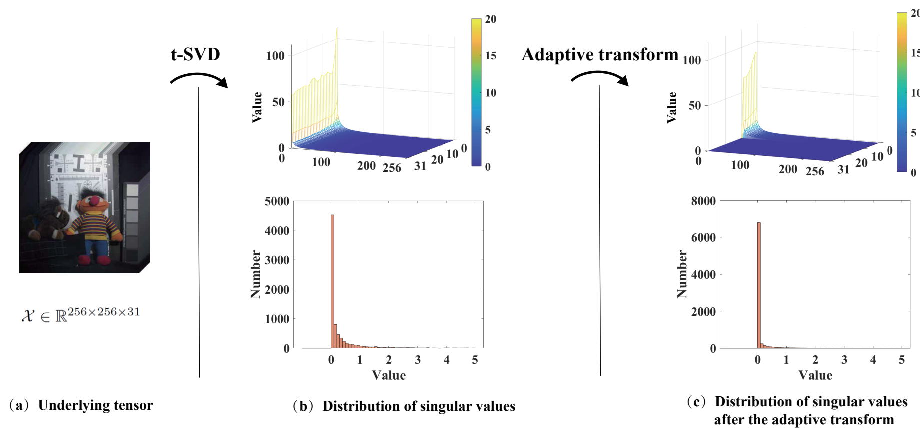


FIGURE 4. The effect of transformation $\mathbf{E}$. The MSI data *Toy* (size: $256 \times 256 \times 31$), denoted as $\mathcal{X}$, is decomposed with t-SVD. We provide the distribution curve and histogram of singular values before and after the transform. It is clear that the sparsity of singular values of $\mathcal{X}$ is effectively enhanced by the adaptive transformation.

Definition 3.6 (TASR)**.** Denote $\mathcal{F}$ as a frequency transformation, e.g. discrete cosine transformation (DCT) and discrete Fourier transformation (DFT), that satisfies $\mathbf{F}\mathbf{F}^T = \mathbf{F}^T\mathbf{F} = \gamma\mathbf{I}$, γ is a fixed constant. Then the TASR of $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is defined as follows,

$$\mathrm{rank}_{\mathrm{TASR}}(\mathcal{A}) := \sum_{i=1}^{r} \mathrm{rank}(\mathbf{R}^{(i)}), \tag{36}$$

where $\mathcal{R} = \overline{\mathcal{A}} \times_3 \mathbf{E} = \mathcal{A} \times_3 \mathbf{F} \times_3 \mathbf{E}$, $\mathbf{F}$ is the transform matrix of $\mathcal{F}$ and $\mathbf{E} \in \mathbb{R}^{r\times n_3}$ is an adaptive transform matrix that satisfies $\mathbf{E}^T\mathbf{E} = \mathbf{I}$. ($r$ is a hyperparameter.)

Remark 3.7. TASR is similar to the tensor average rank under the coupled transformation ($\mathbf{FE}$). Compared with the original transformation $\mathbf{L}$ in (4), the transformation $\mathbf{E}$ only satisfies $\mathbf{E}^T\mathbf{E} = \mathbf{I}$, and the frequency transformation $\mathbf{F}$ can retain the advantage of frequency domain. Besides, the TASR is more flexible since the transformation $\mathbf{E}$ is adaptively updated in applications. The TASR can enhance the sparse distribution of singular values by the transformation $\mathbf{E}$, which can fully utilize the structural similarity of singular values in different frontal slices of tensor (please refer to Figure 4). The value of r usually satisfies $r < n_3$, which means that the transformation $\mathbf{E}$ can effectively reduce running time.

Thus, we can build the tensor completion model as follows,

$$\min_{\mathcal{X}} \ \mathrm{rank}_{\mathrm{TASR}}(\mathcal{X}) \quad s.t.\ \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{M}). \tag{37}$$

Based on the PADMM scheme, we can effectively solve the above model. By introducing the auxiliary variable $\mathcal{D} = \mathcal{X}$, the model (37) can be represented as follows,

$$\min_{\mathcal{X},\mathcal{D},\mathbf{E}} \ \mathrm{rank}_{\mathrm{TASR}}(\mathcal{D}) + \mathtt{I}_{\mathbf{\Phi}}(\mathcal{X}) \quad s.t.\ \mathcal{D} = \mathcal{X}, \tag{38}$$

where $\mathtt{I}_{\mathbf{\Phi}}(\mathcal{X})$ means

$$\mathtt{I}_{\mathbf{\Phi}} = \begin{cases} 0 & \mathcal{X} \in \mathbf{\Phi}, \\ \infty & \text{otherwise}, \end{cases} \tag{39}$$

and $\mathbf{\Phi} := \{\mathcal{X} \mid \mathcal{P}_{\Omega}(\mathcal{X}) = \mathcal{P}_{\Omega}(\mathcal{M})\}$. Subsequently, the augmented Lagrangian function of (38) can be deduced as follows,

$$\mathtt{L}(\mathcal{X}, \mathcal{D}, \mathcal{O}, \mathbf{E}, \mu_1) = \mathrm{rank}_{\mathrm{TASR}}(\mathcal{D}) + \mathtt{I}_{\mathbf{\Phi}}(\mathcal{X}) + \frac{\mu_1}{2} \left\| \mathcal{X} - \mathcal{D} + \frac{\mathcal{O}}{\mu_1} \right\|_F^2, \tag{40}$$

where $\mathcal{O}$ is the Lagrangian multiplier, and μ_1 represents the penalty parameter. Under the PADMM framework, we can solve the model (37) by the following three sub-problems.

$\mathcal{X}$ **sub-problem**:

We address this sub-problem by fixing other variables except the variable $\mathcal{X}$. Thus, the $\mathcal{X}$ sub-problem can be represented as follows,

$$\min_{\mathcal{X}} \mathtt{I}_{\mathbf{\Phi}}(\mathcal{X}) + \frac{\mu_1}{2} \left\| \mathcal{X} - \mathcal{D} + \frac{\mathcal{O}}{\mu_1} \right\|_F^2. \tag{41}$$

It is clear that $\mathcal{X}$ at k-th iteration can be updated by

$$\mathcal{X}^{k+1} = \mathcal{P}_{\Omega}(\mathcal{M}) + \mathcal{P}_{\Omega^C}(\mathcal{D}^k - \frac{\mathcal{O}^k}{\mu_1^k}), \tag{42}$$

where Ω^C means the complement of Ω.

$\mathcal{D}$ **sub-problem:**

According to the augmented Lagrange function (40), the $\mathcal{D}$ sub-problem can be denoted as follows,

$$\min_{\mathcal{D}} \ \mathrm{rank}_{\mathrm{TASR}}(\mathcal{D}) + \frac{\mu_1}{2} \left\| \mathcal{X} - \mathcal{D} + \frac{\mathcal{O}}{\mu_1} \right\|_F^2. \tag{43}$$

Similar to Lemma 3.1, we introduce Theorem 3.8 that is based on the proposed ℓ_0 minimization framework to solve the sub-problem.

Theorem 3.8. *Denote $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{E} \in \mathbb{R}^{r \times n_3}$. The adaptive orthogonal transformation $\mathbf{E}$ stisfies $\mathbf{E}^T\mathbf{E} = \mathbf{I}$. $\mathbf{F}$ is the frequency transformation that meets $\mathbf{F}\mathbf{F}^T = \mathbf{F}^T\mathbf{F} = \gamma\mathbf{I}$. $\mathbf{U}^{(i)}\mathbf{S}^{(i)}{\mathbf{V}^{(i)}}^T$ is the SVD of $\mathbf{G}^{(i)}, i = 1, \cdots, r$, where $\mathcal{G} = \overline{\mathcal{Y}} \times_3 \mathbf{E}$, $\overline{\mathcal{Y}}$ is $\mathcal{Y}$ on frequency domain. Then the optimum to the following problem:*

$$\min_{\mathcal{X}} \lambda \ rank_{TASR}(\mathcal{X}) + \frac{1}{2} \left\| \mathcal{X} - \mathcal{Y} \right\|_F^2, \tag{44}$$

can be expressed as $\mathcal{X}^* = \mathcal{R}^* \times_3 \mathbf{E}^T \times_3 \mathbf{F}^{-1}$, where i-th frontal slice of $\mathcal{R}^*$ satisfies $\mathbf{R}^{*(i)} = \mathbf{U}^{(i)}\mathbf{W}^{*(i)}{\mathbf{V}^{(i)}}^T$, and the diagonal matrix $\mathbf{W}^{*(i)}$ is the solution of the following problem:

$$\min_{\mathbf{W}^{(i)}} \lambda \left\| \mathbf{W}^{(i)} \right\|_0 + \frac{1}{2\gamma} \left\| \mathbf{W}^{(i)} - \mathbf{S}^{(i)} \right\|_F^2, \qquad i = 1, \cdots, r. \tag{45}$$

Algorithm 2 The PADMM-based solver for the proposed tensor completion model (37)

Input: Observed image $\mathcal{M}$
Parameter: μ_1, α, β, μ_2, ρ, r , $\mu_{\max} = 10^{10}$, $\varepsilon = 10^{-5}$, and $k_{\text{mit}} = 100$
Output: $\mathcal{X}$

1: Initialization $k = 0, \mu_1^0 = \mu_1, \mathcal{X}^0 = \mathcal{M}$, $\mathbf{E}^0 = \mathbf{I}$, and $\mathcal{D}^0 = \mathcal{O}^0 = \mathbf{0}$
2: **while** $k < k_{\text{mit}}$ and RelCha$> \varepsilon$ **do**
3: Update $\mathcal{X}^{k+1}$ via (42)
4: Update $\mathcal{D}^{k+1}$ via (47)
5: Update $\mathbf{E}^{k+1}$ via (49)
6: Update Lagrange multiplier $\mathcal{O}^{k+1}$ via (50)
7: Update μ_1^{k+1} via $\mu_1^{k+1} = \min\{\rho\mu_1^k, \mu_{\max}\}$
8: $k = k + 1$
9: **end while**

Proof. Denote $\mathcal{R} = \overline{\mathcal{X}} \times_3 \mathbf{E}$. According to the definition of the TASR and the property of $\mathbf{E}$, we have

$$
\begin{aligned}
&\lambda \ \text{rank}_{\text{TASR}}(\mathcal{X}) + \frac{1}{2}\|\mathcal{X} - \mathcal{Y}\|_F^2 \\
=&\lambda \sum_{i=1}^{r} \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma}\|\overline{\mathcal{X}} - \overline{\mathcal{Y}}\|_F^2 \\
=&\sum_{i=1}^{r} \lambda \ \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma}\left\|\mathbf{R}^{(i)} - \mathbf{G}^{(i)}\right\|_F^2 \\
=&\sum_{i=1}^{r} \lambda \ \text{rank}(\mathbf{R}^{(i)}) + \frac{1}{2\gamma}\left\|\mathbf{G}^{(i)}\right\|_F^2 + \frac{1}{2\gamma}\left\|\mathbf{R}^{(i)}\right\|_F^2 - \frac{1}{\gamma}Tr(\mathbf{G}^{(i)^T}\mathbf{R}^{(i)}).
\end{aligned}
\tag{46}
$$

Denote $\dot{\mathbf{U}}^{(i)}\mathbf{W}^{(i)}\dot{\mathbf{V}}^{(i)^T}$ as the SVD of the matrix $\mathbf{R}^{(i)}, i = 1, 2, \cdots, r$. We assume $n_1 \leq n_2$, and other conditions can be handled similarly. By von Neumanns trace inequality [29], $Tr(\mathbf{G}^{(i)^T}\mathbf{R}^{(i)})$ achieves its upper bound $\sum_{j=1}^{n_1} w_j^i s_j^i$ if and only if $\dot{\mathbf{U}}^{(i)} = \mathbf{U}^{(i)}$ and $\dot{\mathbf{V}}^{(i)} = \mathbf{V}^{(i)}$, where w_j^i and s_j^i are the j-th diagonal element of matrix $\mathbf{W}^{(i)}$ and $\mathbf{S}^{(i)}$, respectively. Thus, we can obtain that the problem (44) is equal to the following problem:

$$
\min_{w_j^i} \ \sum_{j=1}^{n_1} \lambda |w_j^i|_0 + \frac{1}{2\gamma}(w_j^i - s_j^i)^2, \ i = 1, \cdots, r,
$$

which is also equivalent to the problem (45). Denote $\mathbf{W}^{*(i)}$ is the optimun to (45). We can get $\mathbf{R}^{*(i)} = \mathbf{U}^{(i)}\mathbf{W}^{*(i)}\mathbf{V}^{(i)^T}$. Since $\mathcal{R} = \overline{\mathcal{X}} \times_3 \mathbf{E}$, we have $\mathcal{X}^* = \mathcal{R}^* \times_3 \mathbf{E}^T \times_3 \mathbf{F}^{-1}$ is the solution of problem (44). □

Remark 3.9. Theorem 3.8 builds the relationship between the TASR and the ℓ_0-norm, and Theorem 3.4 further converts the ℓ_0-norm minimization (45) into an equivalent biconvex problem (18). Thus, we can solve the TASR minimization by the proposed ℓ_0 minimization framework.

Assuming $\mathbf{U}_1^{(i)}\mathbf{S}_1^{(i)}\mathbf{V}_1^{(i)^T}$ is the SVD of $\mathbf{H}^{(i)}, i = 1, \cdots, r$, where $\mathcal{H} = (\overline{\mathcal{X}}^{k+1} + \frac{\overline{\mathcal{O}}^k}{\mu_1^k}) \times_3 \mathbf{E}^k$, we can update $\mathcal{D}$ by

$$\mathcal{D}^{k+1} = \mathcal{Q}^{k+1} \times_3 \mathbf{E}^{k^T} \times_3 \mathbf{F}^{-1}, \tag{47}$$

where $\mathbf{Q}^{(i)^{k+1}} = \mathbf{U}_1^{(i)}\mathbf{T}^{(i)^{k+1}}\mathbf{V}_1^{(i)^T}$, and $\mathbf{T}^{(i)^{k+1}}$ is obtained by Algorithm 1.

E sub-problem:

The adaptive transformation $\mathbf{E}$ is essential for the proposed TASR. To improve the data adaptation of the transformation $\mathbf{E}$, we adaptively compute the transformation. By introducing a proximal term, $\mathbf{E}$ is updated steadily, and the k-th iteration result of the transformation $\mathbf{E}$ is obtained by the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{E}} \left\| \mathcal{Q}^{k+1} - (\overline{\mathcal{X}}^{k+1} + \frac{\overline{\mathcal{O}}^k}{\mu_1^k}) \times_3 \mathbf{E} \right\|_F^2 + \frac{\mu_2}{2} \left\| \mathbf{E} - \mathbf{E}^k \right\|_F^2, \\ &s.t. \quad \mathbf{E}^T\mathbf{E} = \mathbf{I}, \end{aligned} \tag{48}$$

where μ_2 is a fixed penalty parameter, $\mathbf{E}^k$ is the $\mathbf{E}$ obtained by the (k-1)-th iteration. Because of orthogonal constraint [48], we obtain the closed-form solution of $\mathbf{E}$ as follows,

$$\mathbf{E}^{k+1} = \mathbf{U}\mathbf{V}^T \tag{49}$$

where the matrices $\mathbf{U}$ and $\mathbf{V}^T$ are from the SVD of $\mu_2\mathbf{E}^k + \mathbf{Q}_{(3)}(\mathtt{unfold}_{(3)}(\overline{\mathcal{X}}^{k+1} + \frac{\overline{\mathcal{O}}^k}{\mu_1^k}))^T$.

$\mathcal{O}$ update:

Under the PADMM framework, the multiplier $\mathcal{O}$ can be updated as

$$\mathcal{O}^{k+1} = \mathcal{O}^k + \mu_1^k(\mathcal{X}^{k+1} - \mathcal{D}^{k+1}), \tag{50}$$

and the parameter μ_1 is updated by $\mu_1^{k+1} = \rho\mu_1^k$, where ρ is a fixed parameter such that $\rho > 1$.

The relative change (RelCha) and the number of iterations k_{mit} are used as the termination condition of the algorithm [46], where the RelCha is defined as

$$\text{RelCha} = \left\| \mathcal{X}^{k+1} - \mathcal{X}^k \right\|_F / \left\| \mathcal{X}^k \right\|_F. \tag{51}$$

The whole solving algorithm is summarized in Algorithm 2, where k_{mit} is the maximum number of iterations, $\mu_{\max}$ denotes the upper bound of μ_1, and ε is a tolerance value.

4. **Numerical experiments.** In this section, to demonstrate the effectiveness of the proposed method, we conduct numerical experiments on multi-dimensional image data, i.e., MSI data, HSI data, and video data. The benchmark includes HaLRTC (12'TPAMI) [18], TNN (14'CVPR) [66], TNN-DCT (19'CVPR) [22], PSTNN (20'JCAM) [9], IRTNN (22'TNNLS) [33], DTNN (23'TNNLS) [11]. These methods are executed on the PC with 32Gb RAM, Intel(R) Core(TM) i7-8700K CPU @3.70GHz, and NVIDIA GeForce GTX 1650. All the parameters are chosen according to the recommendation of the authors. Besides, we provide some discussions. In the proposed model, discrete cosine transformation (DCT) is employed as $\mathcal{F}$ for the TASR (please refer to Definition 3.6). Peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [37] as the evaluation indices. Besides, for MSI and HSI data, we use the spectral angle mapper (SAM) [61] and the relative dimensionless global error in synthesis (ERGAS) [32] for accurate assessment.

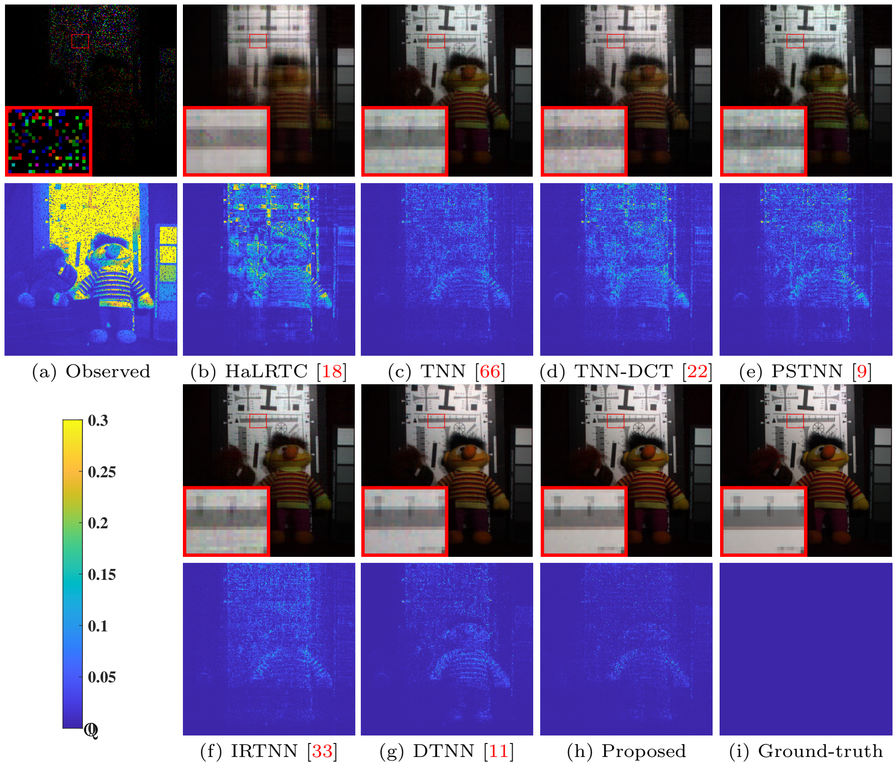


FIGURE 5. Multi-dimensional data completion results on MSI *Toy* with SR = 10%. The first row is the visual comparisons (R: 31-th, G: 20-th, B: 10-th spectral bands, respectively), and the second row is the residual images for better visualization (20-th spectral band). From left to right are the observed image, results of different methods, and the ground-truth.

4.1. **Results on MSI data completion.** In this part, we conduct the methods of the benchmark on the MSI data with different sampling rates (SRs). We choose two images, i.e., *Toy* and *Cloth*, from the CAVE dataset [1]. The spatial size of all the images is reshaped into 256×256, which is widely applied in multi-dimensional image processing e.g., [35]. As shown in Figure 5, compared with other methods, our approach can recover more details of the underlying information. Clearly, one can observe that our method achieves the best visual performance. On the different SRs, our approach also shows superior abilities, which can be found in Tab. 1. Besides, the time cost of our method is satisfactory, which is critical in practical applications. Although HaLRTC [18] consumes the least time, the performance of the method is limited. The results obtained by TNN-DCT [22] and PSTNN [9] have some spatial and spectral distortions because of their limited constraint for singular values. It is worth noting that DTNN [11] is the state-of-the-art method with data-dependent dictionary learning. Compared with DTNN [11], our approach is better in terms of running time and the evaluation index, i.e., PSNR and SSIM [37], which verifies the validity of the proposed method.

[1]https://www.cs.columbia.edu/CAVE/databases/multispectral/

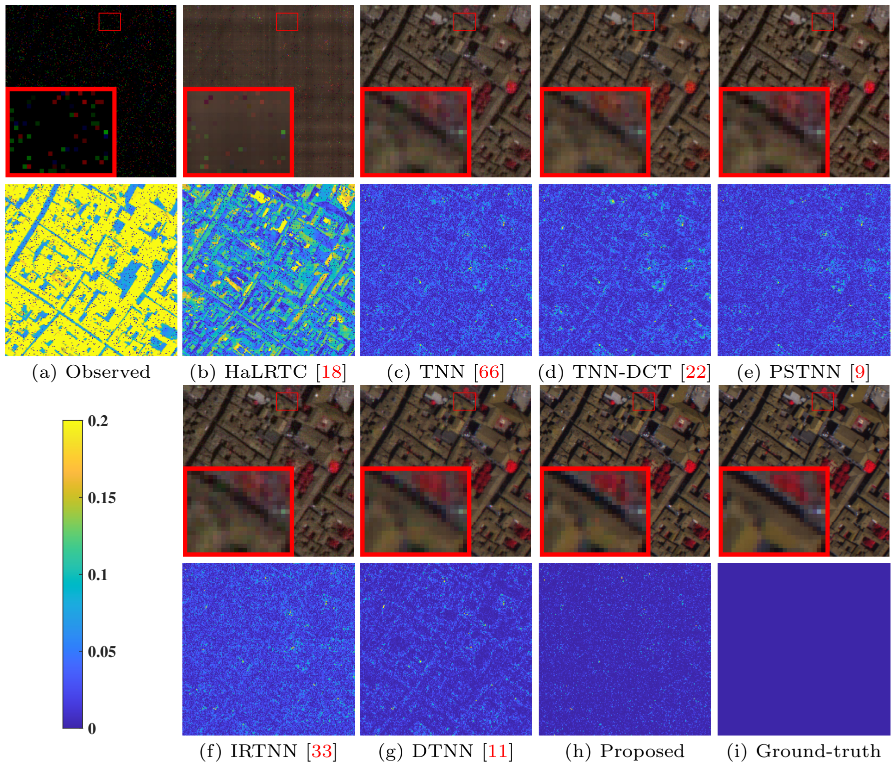
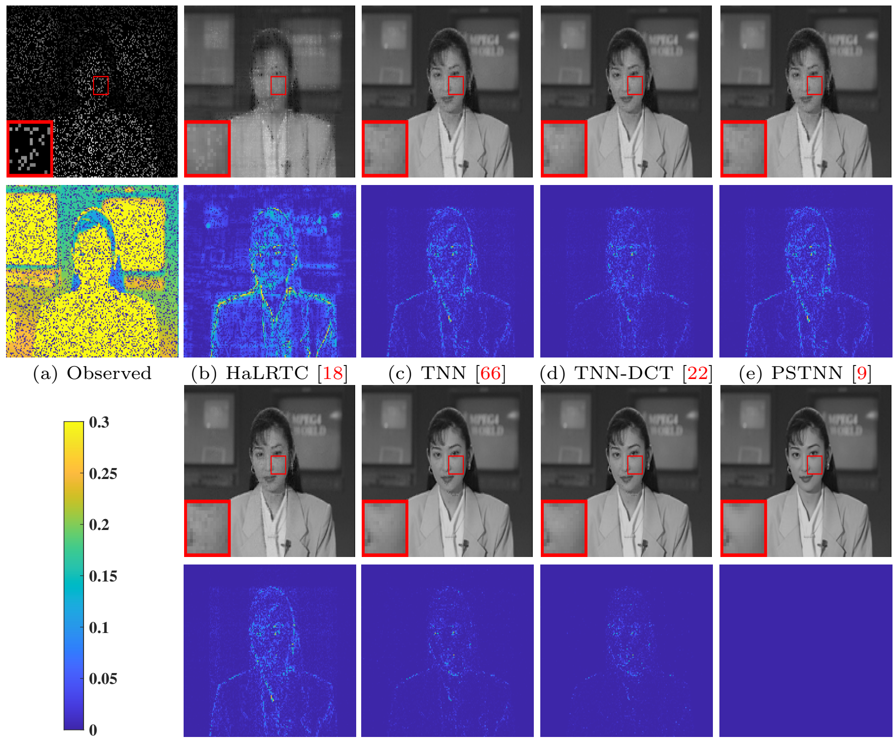


FIGURE 6. Multi-dimensional data completion results on HSI *Pavia* with SR = 5%. The first row is the visual comparisons (R: 68-th, G: 40-th, B: 10-th spectral bands, respectively), and the second row is the residual images for better visualization (32-th spectral band). From left to right are the observed image, results of different methods, and the ground-truth.

4.2. **Results on HSI data completion.** For the HSI data, we test different methods on the *Pavia* dataset and *Washington DC* dataset [2] with the size of $200 \times 200 \times 80$ and $256 \times 256 \times 191$, respectively. On the one hand, Figure 6 shows the visual comparison of different methods on the *Pavia* dataset with SR = 5%, which demonstrates the effectiveness of the proposed approach on the qualitative assessment. On the other hand, Tab. 2 shows the quantitative results of methods at different SRs. Although HaLRTC [18] spends the least time, the details cannot be preserved well. The time consumption of IRTNN [33] and DTNN [11] is huge. The proposed technique shows superiority in both completion results and time consumption compared with other methods that use different transformations and tensor rank surrogates.

4.3. **Results on video data completion.** For the video data, we cut the 70 frames from *Salesman* and *Akiyo* [3]. Although the tensor low-rank property of video data is not stronger than MSI and HSI data, our method also achieves excellent results.

[2]https://rslab.ut.ac.ir/data
[3]http://trace.eas.asu.edu/yuv/



(a) Observed (b) HaLRTC [18] (c) TNN [66] (d) TNN-DCT [22] (e) PSTNN [9]

(f) IRTNN [33] (g) DTNN [11] (h) Proposed (i) Ground-truth

FIGURE 7. Tensor completion results on video image *Akiyo* with SR = 20% (size: $144 \times 176 \times 70$). The first row is the visual comparisons for the 4-th frame, and the second row is the residual images for better visualization (the 4-th frame). From left to right are the observed image, results of different methods, and the ground-truth.

As shown in Figure 7, we choose the 4-th frame to display the visual effects, which demonstrate the significance of our method for the video data. The performance of TNN-DCT [22] and PSTNN [9] is limited since their rank approximations cannot accurately deal with different singular values. On the data with different SRs, our approach can get the best numerical results, which is revealed in Tab. 3.

4.4. **Discussions.**

Parameters analysis: In this part, we analyze the robustness of the five parameters in the proposed method, i.e., μ_1, α, β, ρ, and μ_2. We test these parameters with $r = 9$ on the MSI data *Toy*. As shown in Figure 8, the proposed model is more robust for the parameter ρ. When ρ increases to a critical point, the influence of the parameter ρ would become weak. Obviously, the choice of μ_1, α, and β is sensitive, which should be carefully adjusted for better performance.

Ablation study: In this area, we conduct the ablation study to demonstrate the effectiveness of the proposed solving algorithm of the TASR and the update of the adaptive transformation $\mathbf{E}$. Specifically, to verify the effect of the ℓ_0 minimization, we replace it with the widely used TNN technique [66]. Similarly, we execute the proposed tensor completion model without the improved transformation $\mathbf{E}$ to explore the impact of $\mathbf{E}$. As displayed in Tab. 4, the performance of the proposed model is

Data	SR / Method	3% PSNR	SSIM	SAM	ERGAS	Time (s)	5% PSNR	SSIM	SAM	ERGAS	Time (s)
	HaLRTC [18]	14.60	0.556	23.37	777.6	**6.826**	18.65	0.629	18.36	489.3	**6.685**
	TNN [66]	25.10	0.747	22.03	254.1	338.7	27.45	0.815	18.71	198.8	337.0
	TNN-DCT [22]	26.15	0.781	20.35	224.6	35.84	28.82	0.852	16.56	165.6	35.42
	PSTNN [9]	25.54	0.740	22.19	244.4	84.82	28.73	0.837	17.46	174.2	84.48
	IRTNN [33]	25.01	0.726	23.33	254.0	405.9	27.97	0.818	18.50	185.5	272.7
	DTNN [11]	<u>27.82</u>	**0.864**	**14.76**	<u>198.1</u>	173.3	<u>30.47</u>	<u>0.908</u>	<u>12.90</u>	<u>155.4</u>	159.9
	Proposed	**28.31**	<u>0.805</u>	<u>17.42</u>	**159.1**	<u>14.88</u>	**33.42**	**0.917**	**12.02**	**88.63**	<u>14.91</u>
	SR / Method	10% PSNR	SSIM	SAM	ERGAS	Time (s)	20% PSNR	SSIM	SAM	ERGAS	Time (s)
	HaLRTC [18]	25.46	0.802	11.78	224.4	**6.790**	29.20	0.895	8.421	146.1	**6.527**
	TNN [66]	31.46	0.898	14.09	128.2	341.8	36.56	0.956	9.515	74.85	344.5
	TNN-DCT [22]	27.95	0.834	16.41	182.2	35.33	36.90	0.962	8.740	71.17	32.68
	PSTNN [9]	28.95	0.850	16.48	179.5	84.79	36.19	0.948	10.12	76.21	78.61
	IRTNN [33]	32.58	0.906	13.76	110.4	158.7	38.22	0.964	8.873	62.04	97.08
	DTNN [11]	<u>35.39</u>	<u>0.957</u>	**8.252**	<u>91.73</u>	159.4	<u>43.32</u>	<u>0.989</u>	<u>4.485</u>	<u>33.33</u>	172.5
	Proposed	**38.96**	**0.968**	<u>8.390</u>	**47.34**	<u>15.05</u>	**47.29**	**0.994**	**4.447**	**19.24**	<u>15.07</u>
Data	SR / Method	3% PSNR	SSIM	SAM	ERGAS	Time (s)	5% PSNR	SSIM	SAM	ERGAS	Time (s)
	HaLRTC [18]	18.04	0.293	17.73	514.5	**7.188**	19.04	0.321	16.88	462.0	**6.519**
	TNN [66]	21.58	0.360	17.73	347.2	124.8	22.98	0.483	15.15	295.1	144.1
	TNN-DCT [22]	22.55	0.421	14.29	314.5	35.00	23.91	0.540	12.59	268.5	33.25
	PSTNN [9]	21.81	0.378	16.68	338.0	85.41	23.34	0.512	14.18	282.9	78.35
	IRTNN [33]	21.32	0.326	18.53	358.0	337.4	23.04	0.480	15.27	293.1	464.3
	DTNN [11]	<u>24.37</u>	**0.620**	<u>10.69</u>	<u>249.6</u>	168.2	<u>26.18</u>	<u>0.736</u>	<u>8.442</u>	<u>202.6</u>	161.4
	Proposed	**24.47**	<u>0.616</u>	**10.30**	**248.2**	<u>14.99</u>	**26.56**	**0.755**	**8.133**	**194.2**	<u>14.92</u>
	SR / Method	10% PSNR	SSIM	SAM	ERGAS	Time (s)	20% PSNR	SSIM	SAM	ERGAS	Time (s)
	HaLRTC [18]	22.23	0.414	13.38	334.7	**6.846**	23.87	0.564	11.32	275.6	**6.418**
	TNN [66]	25.84	0.685	11.39	211.0	155.1	30.24	0.860	7.620	126.6	194.9
	TNN-DCT [22]	27.05	0.747	9.390	184.3	29.37	31.78	0.899	6.162	106.2	28.74
	PSTNN [9]	26.76	0.732	10.36	189.3	74.44	31.87	0.886	6.844	107.5	66.95
	IRTNN [33]	26.40	0.710	10.89	197.4	144.1	31.12	0.877	7.069	114.9	90.05
	DTNN [11]	<u>29.98</u>	<u>0.884</u>	<u>5.953</u>	<u>129.9</u>	154.2	<u>36.01</u>	<u>0.963</u>	<u>3.521</u>	<u>66.75</u>	156.8
	Proposed	**31.19**	**0.897**	**5.331**	**113.1**	<u>15.19</u>	**37.81**	**0.969**	**3.075**	**55.36**	<u>15.21</u>
	Ideal value	$+\infty$	1	0	0	0	$+\infty$	1	0	0	0

TABLE 1. Quantitative results on MSI data *Toy* and *Cloth*, respectively. (Bold: best; Underline: second best; Time: running time)

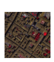
Data	SR / Method	2% PSNR	SSIM	SAM	ERGAS	5% PSNR	SSIM	SAM	ERGAS	10% PSNR	SSIM	SAM	ERGAS	AverT (s)
	HaLRTC [18]	14.36	0.152	18.76	637.5	20.12	0.318	8.657	329.0	22.28	0.457	7.875	256.4	**22.68**
	TNN [66]	24.58	0.649	9.134	199.7	28.58	0.836	7.358	130.2	32.61	0.920	5.601	86.91	519.9
	TNN-DCT [22]	25.66	0.697	7.365	174.0	29.67	0.868	5.646	109.5	38.49	0.980	2.377	40.13	104.9
	PSTNN [9]	25.33	0.701	9.968	184.3	29.43	0.857	7.578	120.6	33.99	0.932	5.589	77.88	187.3
	IRTNN [33]	24.34	0.636	9.130	205.6	28.97	0.841	7.553	124.8	32.58	0.921	5.545	87.34	374.1
	DTNN [11]	<u>26.80</u>	<u>0.748</u>	<u>6.114</u>	<u>153.7</u>	<u>32.79</u>	<u>0.936</u>	<u>3.326</u>	<u>76.57</u>	<u>41.89</u>	<u>0.991</u>	<u>1.294</u>	<u>27.91</u>	125.7
	Proposed	**28.13**	**0.827**	**4.905**	**130.9**	**36.17**	**0.963**	**2.580**	**53.38**	**43.41**	**0.993**	**1.118**	**24.74**	<u>64.21</u>
Data	SR / Method	2% PSNR	SSIM	SAM	ERGAS	5% PSNR	SSIM	SAM	ERGAS	10% PSNR	SSIM	SAM	ERGAS	AverT (s)
	HaLRTC [18]	30.38	0.541	26.31	399.2	33.42	0.663	17.49	274.4	36.45	0.751	11.97	185.0	**87.87**
	TNN [66]	31.58	0.702	12.81	494.2	34.96	0.834	8.940	336.6	38.57	0.914	6.129	219.3	2204
	TNN-DCT [22]	31.79	0.706	12.24	299.0	35.76	0.862	8.002	203.3	39.09	0.929	5.601	146.7	479.0
	PSTNN [9]	32.35	0.738	12.24	468.9	36.06	0.866	8.287	299.1	39.38	0.929	5.769	196.6	797.9
	IRTNN [33]	32.00	0.714	12.83	485.2	35.33	0.845	8.559	298.9	39.59	0.932	5.447	177.3	1410
	DTNN [11]	<u>40.81</u>	<u>0.897</u>	<u>7.424</u>	<u>111.8</u>	<u>47.67</u>	<u>0.974</u>	<u>3.767</u>	<u>50.86</u>	<u>55.69</u>	<u>0.995</u>	<u>1.803</u>	<u>22.05</u>	9716
	Proposed	**42.47**	**0.910**	**6.704**	**90.05**	**50.01**	**0.976**	**3.390**	**42.99**	**57.65**	**0.996**	**1.457**	**17.17**	<u>135.7</u>
	Ideal value	$+\infty$	1	0	0	$+\infty$	1	0	0	$+\infty$	1	0	0	0

TABLE 2. Quantitative results on HSI data *Pavia* and *Washington DC*, respectively. (Bold: best; Underline: second best; AverT: average running time)

affected by the condition without the ℓ_0 approach or the improved transformation. In addition, by comparing (a) and (c), we can find the proposed ℓ_0 solving algorithm effectively shrinks the singular values. The validity of the improved transformation is proofed by (c) and (d) of Tab. 4.

Data	SR / Method	10% PSNR	10% SSIM	20% PSNR	20% SSIM	30% PSNR	30% SSIM	AverT (s)
	HaLRTC [18]	31.81	0.935	37.64	<u>0.980</u>	41.57	<u>0.990</u>	**8.152**
	TNN [66]	33.54	0.945	36.76	0.972	39.49	0.984	429.2
	TNN-DCT [22]	34.26	0.954	37.76	0.979	40.72	0.989	<u>33.23</u>
	PSTNN [9]	33.68	0.944	37.10	0.974	39.99	0.986	50.23
	IRTNN [33]	33.77	0.944	37.19	0.974	40.01	0.985	117.4
	DTNN [11]	<u>36.00</u>	<u>0.972</u>	<u>40.29</u>	**0.989**	<u>43.52</u>	**0.994**	569.9
	Proposed	**36.33**	**0.973**	**40.70**	**0.989**	**43.64**	**0.994**	47.57
	HaLRTC [18]	22.84	0.571	25.91	0.736	28.02	0.822	**12.26**
	TNN [66]	30.72	0.894	33.70	0.941	35.92	0.962	583.0
	TNN-DCT [22]	31.06	0.901	34.16	0.946	36.45	0.966	39.43
	PSTNN [9]	31.24	0.901	34.10	0.944	36.21	<u>0.964</u>	111.0
	IRTNN [33]	30.82	0.893	33.94	0.942	36.22	<u>0.964</u>	156.6
	DTNN [11]	<u>32.26</u>	**0.928**	<u>35.22</u>	<u>0.959</u>	<u>37.92</u>	**0.976**	545.4
	Proposed	**32.44**	<u>0.927</u>	**35.55**	**0.960**	**38.11**	**0.976**	<u>32.20</u>
	Ideal value	$+\infty$	1	$+\infty$	1	$+\infty$	1	0

TABLE 3. Quantitative results on video data *Akiyo* and *Salesman*, respectively. (Bold: best; Underline: second best; AverT: average running time)

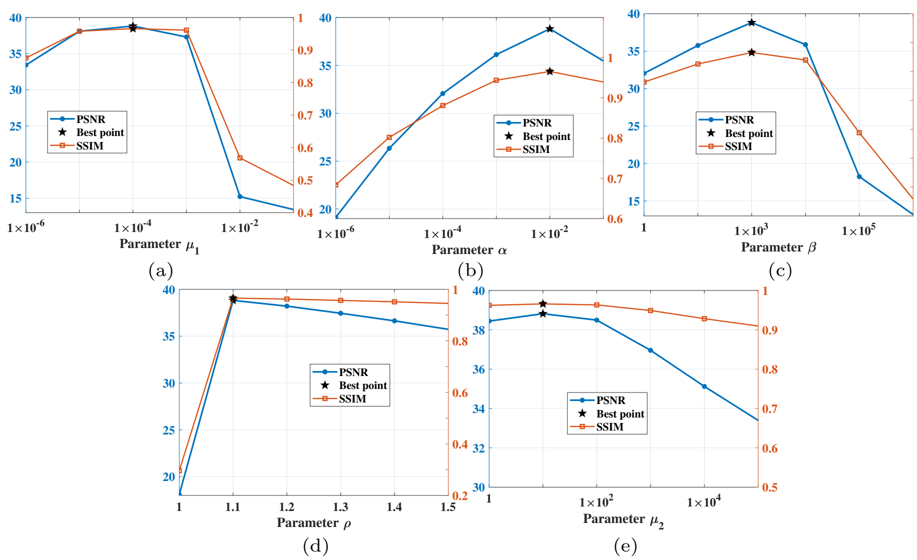


FIGURE 8. Robustness analysis for the five parameters (a) μ_1, (b) α, (c) β, (d) ρ, and (e) μ_2. (data: MSI *Toy*, SR = 10%)

Time consumption analysis: The designed algorithm involves the transformation $\mathbf{E}$. The adaptive transformation $\mathbf{E}$ is updated iteratively, and the size of the transformation is critical. In this part, we analyze the effect of the parameter r on the algorithm. Specifically, as displayed in Figure 9, we calculate the change in PSNR and running time as the parameter r varies. We can observe that PSNR becomes stable when r reaches 9, while the running time increases as r increases. Thus, the appropriate choice of r can effectively reduce running time without compromising the quality of the result.

Numerical convergence: The numerical convergence analysis of the proposed tensor completion model is provided in this part. We calculate the RelCha at each

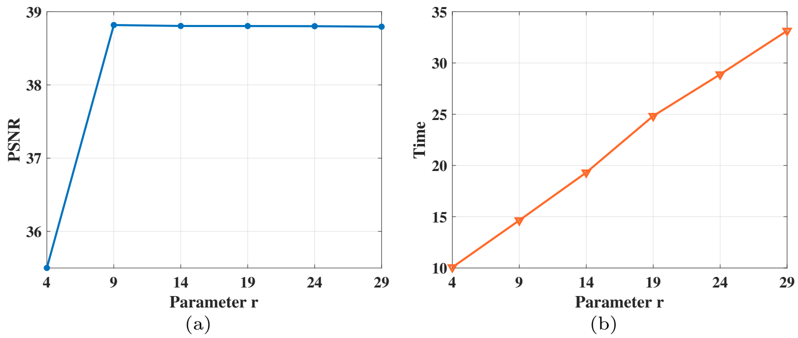


FIGURE 9. The PSNR and running time for different parametr r on the MSI *Toy* with SR = 10%.

Method	PSNR	SSIM	Time (s)
(a) TNN [66]	27.95	0.834	<u>17.08</u>
(b) w/o ℓ_0 approach	32.12	0.918	52.49
(c) w/o improved transformation	<u>34.26</u>	<u>0.920</u>	93.13
(d) Proposed	**38.96**	**0.968**	**14.77**

TABLE 4. Ablation experiment results on the MSI *Toy* with SR = 10%. (Bold: best; Underline: second best)

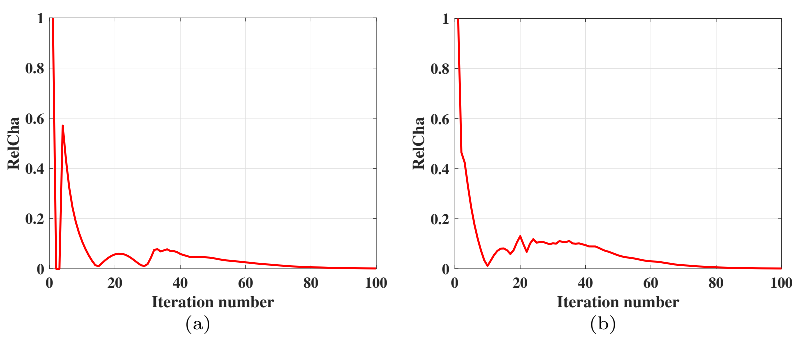
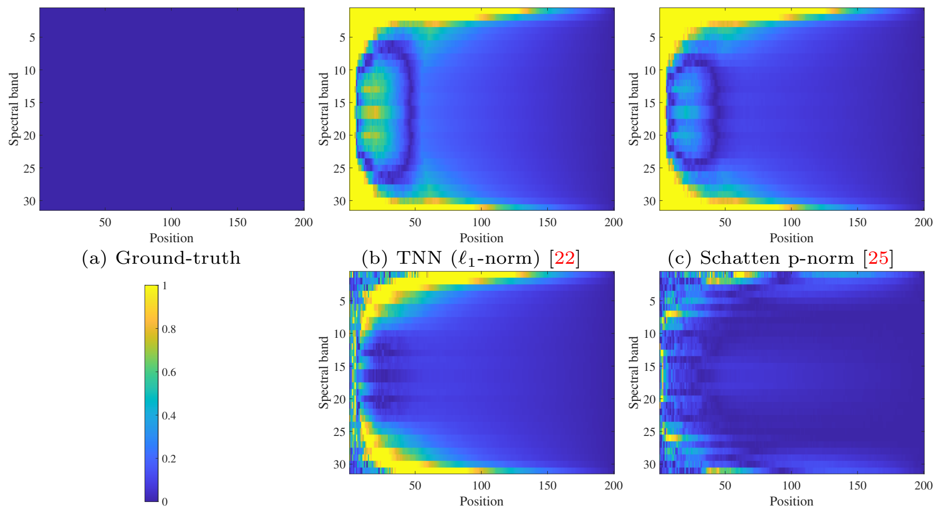


FIGURE 10. Convergence analysis for the proposed tensor completion model on the HSI *Pavia* with SR = 5%, 10%, respectively.

iteration of Algorithm 2, where RelCha is defined in (51). As shown in Figure 10, the RelCha curves show the convergence behavior of the proposed algorithm at different SR conditions. We can observe that the algorithm tends to converge when the iteration number is more than 60. This numerical experiment shows that the proposed multi-dimensional data completion model has excellent convergence ability.

Singular value constraint analysis: In this part, we discuss the singular value constraint of different rank approximations. As depicted in Figure 11, we illustrate the disparity between the tensor singular value distributions of the underlying tensor and the results reconstructed by various rank surrogates. It is evident that the larger singular values constructed by TNN [22] and Schatten p-norm [25] differ significantly from those of the underlying tensor. Although the Logarithmic norm [1] can handle larger singular values, there still exist some singular values that are



(a) Ground-truth (b) TNN (ℓ_1-norm) [22] (c) Schatten p-norm [25]

(d) Logarithmic norm [1] (e) The proposed

FIGURE 11. The difference between tensor singular value distributions of underlying tensor and the results reconstructed by different rank surrogates. The i-th row of the image means the difference in the i-th spectral band of the data. (data: MSI *Toy*, size: $200 \times 200 \times 31$, sampling rate: 20%)

difficult to constrain. Our method effectively addresses the constraint of different tensor singular values due to its robust ability to handle various singular values.

5. **Conclusions.** In this article, we propose a novel ℓ_0 minimization framework of tensor tubal rank, which can also be extended to minimize other sparsity-related tensor ranks. Different from other rank surrogates, the proposed framework formulates an equivalent form of tensor average rank minimization and displays the powerful constraint ability for the sparsity of tensor singular values. A convergent algorithm, i.e., Algorithm 1, is developed to solve it. In addition, we propose TASR by an adaptive transformation and then give a new model for the multi-dimensional image completion application, which can fully explore the sparse constraint of the ℓ_0 minimization framework. Based on the scheme of PADMM, we design an effective algorithm, i.e., Algorithm 2, to solve the completion model. Numerical experiments on multi-dimensional data, e.g., MSI, HSI, and video data, verify the excellent performance of the ℓ_0 minimization framework for the completion task. The proposed model achieves state-of-the-art results. In the future, the applications of the TASR on image denoising [6], restoration [51, 69], and super-resolution [5, 52] can be considered.

Acknowledgments. This research is supported by the National Natural Science Foundation of China (Grant No. 12171072, 12271083), Natural Science Foundation of Sichuan Province (2022NSFSC0501, 2023NSFSC1341), Key Projects of Applied Basic Research in Sichuan Province (Grant No. 2020YJ0216), and National Key Research and Development Program of China (Grant No. 2020YFA0714001).

REFERENCES

[1] L. Chen, X. Jiang, X. Liu and Z. Zhou, Logarithmic norm regularized low-rank factorization for matrix and tensor completion, *IEEE Trans. Image Process.*, **30** (2021), 3434-3449.

[2] W. Chen and N. Song, Low-rank tensor completion: A pseudo-bayesian learning approach, in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, 3305-3313.

[3] Y. Chen, W. He, X.-L. Zhao, T.-Z. Huang, J. Zeng and H. Lin, Exploring nonlocal group sparsity under transform learning for hyperspectral image denoising, *IEEE Trans. Geosci. Remote Sens.*, **60** (2022), 1-18.

[4] L.-J. Deng, R. Glowinski and X.-C. Tai, A new operator splitting method for the euler elastica model for image smoothing, *SIAM J. Imaging Sci.*, **12** (2019), 1190-1230.

[5] L.-J. Deng, W. Guo and T.-Z. Huang, Single-image super-resolution via an iterative reproducing kernel hilbert space method, *IEEE Trans. Circuits Syst. Video Technol.*, **26** (2015), 2001-2014.

[6] H.-X. Dou, T.-Z. Huang, L.-J. Deng, X.-L. Zhao and J. Huang, Directional ℓ_0 sparse modeling for image stripe noise removal, *Remote Sens.*, **10** (2018), 361.

[7] Y. Duan, Q. Zhong, X.-C. Tai and R. Glowinski, A fast operator-splitting method for beltrami color image denoising, *J. Sci. Comput.*, **92** (2022), Paper No. 89, 28 pp.

[8] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang and L. Zhang, Non-local meets global: An iterative paradigm for hyperspectral image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2020), 2089-2107.

[9] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao and L.-J. Deng, Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm, *J. Comput. Appl. Math.*, **372** (2020), 112680, 15 pp.

[10] T.-X. Jiang, M. K. Ng, X.-L. Zhao and T.-Z. Huang, Framelet representation of tensor nuclear norm for third-order tensor completion, *IEEE Trans. Image Process.*, **29** (2020), 7233-7244.

[11] T.-X. Jiang, X.-L. Zhao, H. Zhang and M. K. Ng, Dictionary learning with low-rank coding coefficients for tensor completion, *IEEE Trans. Neural Netw. Learn. Syst.*, **34** (2023), 932-946.

[12] Z.-F. Jin, Z. Wan, Y. Jiao and X. Lu, An alternating direction method with continuation for nonconvex low rank minimization, *J. Sci. Comput.*, **66** (2016), 849-869.

[13] E. Kernfeld, M. Kilmer and S. Aeron, Tensor–tensor products with invertible linear transforms, *Linear Alg. Appl.*, **485** (2015), 545-570.

[14] M. E. Kilmer, K. Braman, N. Hao and R. C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, *SIAM J. Matrix Anal. Appl.*, **34** (2013), 148-172.

[15] H. Kong, C. Lu and Z. Lin, Tensor Q-rank: New data dependent definition of tensor rank, *Mach. Learn.*, **110** (2021), 1867-1900.

[16] H. Kong, X. Xie and Z. Lin, t-Schatten-p norm for low-rank tensor recovery, *IEEE J. Sel. Top. Signal Process.*, **12** (2018), 1405-1419.

[17] C. Li, J. Zeng, Z. Tao and Q. Zhao, Permutation search of tensor network structures via local sampling, in *Proc. Int. Conf. Mach. Learn.*, PMLR, (2022), 13106-13124.

[18] J. Liu, P. Musialski, P. Wonka and J. Ye, Tensor completion for estimating missing values in visual data, *IEEE Trans. Pattern Anal. Mach. Intell.*, **35** (2012), 208-220.

[19] Z. Long, C. Zhu, J. Liu, P. Comon and Y. Liu, Trainable subspaces for low rank tensor completion: Model and analysis, *IEEE Trans. Signal Process.*, **70** (2022), 2502-2517.

[20] C. Lu, Transforms based tensor robust pca: Corrupted low-rank tensors recovery via convex optimization, in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, 1145-1152.

[21] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin and S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 925-938.

[22] C. Lu, X. Peng and Y. Wei, Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms, in *Proc. of the IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2019, 5996-6004.

[23] Y. Luo, X. Zhao, D. Meng and T. Jiang, HLRTF: Hierarchical low-rank tensor factorization for inverse problems in multi-dimensional imaging, in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, 2022, 19303-19312.

[24] M. K. Ng, C.-P. Tam and F. Wang, Multi-view foreground segmentation via fourth order tensor learning, *Inverse Probl. Imaging*, **7** (2013), 885-906.

[25] F. Nie, H. Huang and C. Ding, Low-rank matrix recovery via efficient schatten p-norm minimization, *Proc. AAAI Conf. Artif. Intell.*, **26** (2012), 655-661.

[26] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo and T. Huang, Low-rank high-order tensor completion with applications in visual data, *IEEE Trans. Image Process.*, **31** (2022), 2433-2448.

[27] D. Qiu, M. Bai, M. K. Ng and X. Zhang, Robust low transformed multi-rank tensor methods for image alignment, *J. Sci. Comput.*, **87** (2021), Paper No. 24, 40 pp.

[28] H. Qiu, Y. Wang, S. Tang, D. Meng and Q. Yao, Fast and provable nonconvex tensor RPCA, in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2022, 18211-18249.

[29] D. Rhea, The case of equality in the von neumann trace inequality, *preprint*.

[30] J. Salmi, A. Richter and V. Koivunen, Sequential unfolding SVD for tensors with applications in array signal processing, *IEEE Trans. Signal Process.*, **57** (2009), 4719-4733.

[31] G.-J. Song, M. K. Ng and X. Zhang, Tensor completion by multi-rank via unitary transformation, *Appl. Comput. Harmon. Anal.*, **65** (2023), 348-373.

[32] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*, Presses des MINES. Paris, France: Ecole des Mines, 2002.

[33] H. Wang, F. Zhang, J. Wang, T. Huang, J. Huang and X. Liu, Generalized nonconvex approach for low-tubal-rank tensor recovery, *IEEE Trans. Neural Netw. Learn. Syst.*, **33** (2022), 3305-3319.

[34] J. Wang, J. Hou and Y. C. Eldar, Tensor robust principal component analysis from multilevel quantized observations, *IEEE Trans. Inf. Theory*, **69** (2022), 383-406.

[35] J.-L. Wang, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang and M. K. Ng, Multi-dimensional visual data completion via low-rank tensor representation under coupled transform, *IEEE Trans. Image Process.*, **30** (2021), 3581-3596.

[36] W. Wang, C. Wu and X.-C. Tai, A globally convergent algorithm for a constrained non-lipschitz image restoration model, *J. Sci. Comput.*, **83** (2020), Paper No. 14, 29 pp.

[37] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. Image Process.*, **13** (2004), 600-612.

[38] Z. Wang, J. Dong, X. Liu and X. Zeng, Low-rank tensor completion by approximating the tensor average rank, in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, 4612-4620.

[39] K. Wei, J.-F. Cai, T. F. Chan and S. Leung, Guarantees of riemannian optimization for low rank matrix completion, *Inverse Probl. Imaging*, **14** (2020), 233-265.

[40] R. Wen, L.-J. Deng, Z.-C. Wu, X. Wu and G. Vivone, A novel spatial fidelity with learnable nonlinear mapping for panchromatic sharpening, *IEEE Trans. Geosci. Remote Sens.*, **61** (2023), 1-15.

[41] C. Wu, X. Guo, Y. Gao and Y. Xue, A general non-lipschitz infimal convolution regularized model: Lower bound theory and algorithm, *SIAM J. Imaging Sci.*, **15** (2022), 1499-1538.

[42] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, H.-X. Dou and D. Meng, Tensor wheel decomposition and its tensor completion application, *Proc. Adv. Neural Inf. Process. Syst.*, **35** (2022), 27008-27020.

[43] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J. Huang, J. Chanussot and G. Vivone, Lrtcfpan: Low-rank tensor completion based framework for pansharpening, *IEEE Trans. Image Process.*, **32** (2023), 1640-1655.

[44] Z.-C. Wu, T.-Z. Huang, L.-J. Deng and G. Vivone, A framelet sparse reconstruction method for pansharpening with guaranteed convergence, *Inverse Probl. Imaging*, **17** (2023), 1277-1300.

[45] S. Xia, D. Qiu and X. Zhang, Tensor factorization via transformed tensor-tensor product for image alignment, *Numer. Algorithms*, **95** (2024), 1251-1289.

[46] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu and G. Vivone, A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening, *IEEE Trans. Geosci. Remote Sens.*, **60** (2022), 1-15.

[47] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, X. Wu and G. Vivone, Variational pansharpening based on coefficient estimation with nonlocal regression, *IEEE Trans. Geosci. Remote Sens.*, **61** (2023), 1-15.

[48] Q. Xie, Q. Zhao, D. Meng and Z. Xu, Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery, *IEEE Trans. Pattern Anal. Mach. Intell.*, **40** (2017), 1888-1902.

[49] T. Xu, T.-Z. Huang, L.-J. Deng, H.-X. Dou and N. Yokoya, Tr-stf: A fast and accurate tensor ring decomposition algorithm via defined scaled tri-factorization, *Comput. Appl. Math.*, **42** (2023), Paper No. 234, 25 pp.

[50] T. Xu, T.-Z. Huang, L.-J. Deng and N. Yokoya, An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution, *IEEE Trans. Geosci. Remote Sens.*, **60** (2022), 1-16.

[51] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao and J.-F. Hu, Exemplar-based image inpainting using adaptive two-stage structure-tensor based priority function and nonlocal filtering, *J. Vis. Commun. Image Represent.*, **83** (2022), 103430.

[52] T. Xu, T.-Z. Huang, L.-J. Deng, X.-L. Zhao and J. Huang, Hyperspectral image superresolution using unidirectional total variation with tucker decomposition, *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, **13** (2020), 4381-4398.

[53] Y. Xu, R. Hao, W. Yin and Z. Su, Parallel matrix factorization for low-rank tensor completion, *Inverse Probl. Imaging*, **9** (2015), 601-624.

[54] J. Xue, Y. Zhao, W. Liao, J. C.-W. Chan and S. G. Kong, Enhanced sparsity prior model for low-rank tensor completion, *IEEE Trans. Neural Netw. Learn. Syst.*, **31** (2019), 4567-4581.

[55] M. Yan and Y. Duan, Nonlocal elastica model for sparse reconstruction, *J. Math. Imaging Vis.*, **62** (2020), 532-548.

[56] M. Yang, Q. Luo, W. Li and M. Xiao, Nonconvex 3D array image data recovery and pattern recognition under tensor framework, *Pattern Recognit.*, **122** (2022), 108311.

[57] Y. Yang, Y. Li, C. Wu and Y. Duan, A convergent iterative support shrinking algorithm for non-lipschitz multi-phase image labeling model, *J. Sci. Comput.*, **96** (2023), Paper No. 47, 29 pp.

[58] T. Yokota, B. Erem, S. Guler, S. K. Warfield and H. Hontani, Missing slice recovery for tensors using a low-rank model in embedded space, in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2018, 8251-8259.

[59] G. Yuan and B. Ghanem, $\ell_0$tv: A new method for image restoration in the presence of impulse noise, in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, 5369-5377.

[60] G. Yuan and B. Ghanem, ℓ_0TV: A sparse optimization method for impulse noise image restoration, *IEEE Trans. Pattern Anal. Mach. Intell.*, **41** (2019), 352-364.

[61] R. H. Yuhas, A. F. Goetz and J. W. Boardman, Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm, in *Proc. JPL Airborne Geosci. Workshop; AVIRIS Workshop, Pasadena, CA, USA*, 1992, 147-149.

[62] C. Zeng, Proximal linearization methods for schatten p-quasi-norm minimization, *Numer. Math.*, **153** (2023), 213-248.

[63] C. Zeng, Rank properties and computational methods for orthogonal tensor decompositions, *J. Sci. Comput.*, **94** (2023), Paper No. 6, 24 pp.

[64] C. Zeng, C. Wu and R. Jia, Non-lipschitz models for image restoration with impulse noise removal, *SIAM J. Imaging Sci.*, **12** (2019), 420-458.

[65] D. Zhang and K. Chen, 3d orientation-preserving variational models for accurate image registration, *SIAM J. Imaging Sci.*, **13** (2020), 1653-1691.

[66] Z. Zhang, G. Ely, S. Aeron, N. Hao and M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-svd, in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, 3842-3849.

[67] M. Zhao, Y.-W. Wen, M. Ng and H. Li, A nonlocal low rank model for poisson noise removal, *Inverse Probl. Imaging*, **15** (2021), 519-537.

[68] Q. Zhao, L. Zhang and A. Cichocki, Bayesian CP factorization of incomplete tensors with automatic rank determination, *IEEE Trans. Pattern Anal. Mach. Intell.*, **37** (2015), 1751-1763.

[69] Q. Zhong, R. W. Liu and Y. Duan, Spatially adapted first and second order regularization for image reconstruction: From an image surface perspective, *J. Sci. Comput.*, **92** (2022), Paper No. 33, 36 pp.

[70] Q. Zhong, K. Yin and Y. Duan, Image reconstruction by minimizing curvatures on image surface, *J. Math. Imaging Vis.*, **63** (2021), 30-55.

Received October 2023; revised March 2024; early access April 2024.