Tensor Compressive Sensing Fused Low-Rankness and Local-Smoothness

Xinling Liu[1,2], Jingyao Hou[2], Jiangjun Peng[3], Hailin Wang[3], Deyu Meng[3,4*], Jianjun Wang[1*]

[1] School of Mathematics and Statistics, Southwest University, Chongqing 400715, China
[2] School of Mathematics and Information, China West Normal University, Nanchong 637002, China
[3] School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China
[4] Macao Institute of Systems Engineering, Macau University of Science and Technology, Taipa, Macao
{fsliiuxl,hjy17623226280}@163.com, andrew.pengjj@gmail.com, wanghailin97@163.com, dymeng@mail.xjtu.edu.cn, wjjmath@163.com

Abstract

A plethora of previous studies indicates that making full use of multifarious intrinsic properties of primordial data is a valid pathway to recover original images from their degraded observations. Typically, both low-rankness and local-smoothness broadly exist in real-world tensor data such as hyperspectral images and videos. Modeling based on both properties has received a great deal of attention, whereas most studies concentrate on experimental performance, and theoretical investigations are still lacking. In this paper, we study the tensor compressive sensing problem based on the tensor correlated total variation, which is a new regularizer used to simultaneously capture both properties existing in the same dataset. The new regularizer has the outstanding advantage of not using a trade-off parameter to balance the two properties. The obtained theories provide a robust recovery guarantee, where the error bound shows that our model certainly benefits from both properties in ground-truth data adaptively. Moreover, based on the ADMM update procedure, we design an algorithm with a global convergence guarantee to solve this model. At last, we carry out experiments to apply our model to hyperspectral image and video restoration problems. The experimental results show that our method is prominently better than many other competing ones. Our code and Supplementary Material are available at https://github.com/fsliuxl/cs-tctv.

Introduction

Ever since the pioneering work of Compressed Sensing (CS) (Candès, Romberg, and Tao 2006a; Donoho 2006), Tensor Compressive Sensing (TCS) has been continuing to attract attention (Friedland, Li, and Schonfeld 2014; Ding, Chen, and Wassell 2017). The reason behind is that TCS can make more use of the essential structure of the original tensor data, whereas both CS and its matrix version (Recht, Fazel, and Parrilo 2010; Wang, Zhang, and Wang 2021) inevitably destroy their intrinsic structure. TCS is not only a simple mathematical extension of CS to high-order data, but also had been deployed to a lot of data reconstruction applications including Background Subtraction from Compressive Measurements (BSCM) (Cao et al. 2016; Li et al. 2022), image and video compression (Wang et al. 2017; Baraniuk et al. 2017), computed tomography (Semerci et al. 2014) and so on. TCS aims to recover a clean tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ from its few linear observations coded via the degraded procedure $\boldsymbol{y} = \mathcal{A}(\boldsymbol{\mathcal{X}})+\boldsymbol{e}$, where $\mathcal{A} : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^m$ ($m \ll n_1n_2n_3$) denotes a linear degradation operator and $\boldsymbol{e} \in \mathbb{R}^m$ denotes noise. It is apparently an ill-posed problem because of the underdetermined measure system. To deal with this issue, a reasonable approach is utilizing some prior properties of the underlying needed-to-be recovered tensor data to constrain the solution space. Particularly, two representative ones are low-rankness and local-smoothness, which had been proven widely existing in real-world datasets such as Hyperspectral Image (HSI) and video (Peng et al. 2020; Chengbo Li 2011). In this paper, we are interested in the TCS problem simultaneously fused low-rank and local-smooth properties.

We first briefly introduce the TCS problem modeling on purely low-rank or local-smooth property, respectively. When the original tensor is low-rank, one often tends to the rank minimization model

$$\min_{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}} \operatorname{rank}(\boldsymbol{\mathcal{X}}),\ s.t.\ \|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2 \le \varepsilon, \quad (1)$$

where $\varepsilon \ge 0$ denotes the noise level. Tensor ranks are determined by the tensor decomposition which is not unique in general unlike the matrix rank. Typical examples include CANDECOMP/PARAFAC (CP) rank (Kiers 2000) and Tucker rank (Tucker 1966). However, the CP rank is NP-hard to compute (Hillar and Lim 2013) and the convex surrogate of Tucker rank, the Sum of Nuclear Norms (SNN), is not the tightest convex relaxation (Bernardino Romera-Paredes 2013). To circumvent these issues, the tensor tubal rank was proposed under the framework of transform-based Tensor-Singular Value Decomposition (T-SVD) (Canyi Lu 2019). Regarding model (1) under CP and Tucker decomposition, several theoretical recovery guarantees have been developed by utilizing the decomposition-based Tensor Restricted Isometry Properties (TRIP) not long ago (Grotheer et al. 2021; Rauhut, Schneider, and Stojan 2017). Under the T-SVD framework, for ease of implementation, the most often considered model refers to

$$\min_{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}} \|\boldsymbol{\mathcal{X}}\|_{\star_L},\ s.t.\ \|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2 \le \varepsilon, \quad (2)$$

*Corresponding authors.

where $\|\boldsymbol{\mathcal{X}}\|_{\star_L}$ denotes the relevant tensor nuclear norm of $\boldsymbol{\mathcal{X}}$ (see Def. 8). In theory, on the condition of subgaussian measurement ensembles, model (2) had been meticulously studied under the TRIP condition defined by Discrete Fourier Transform (DFT) based T-SVD (Zhang et al. 2020; Canyi Lu 2018). Their theoretical findings indicate that model (2) enjoys a robust recovery guarantee, which indeed reveals the positive effect of low-rank property to solve the inverse problem, just similar to results of traditional C-S (Candès, Romberg, and Tao 2006b). Moreover, modeling under the T-SVD framework had been successfully applied in many areas (Canyi Lu 2019; Zhang et al. 2021; Hou et al. 2022).

Except for the low-rank based TCS, when the original tensor is local-smooth, the most often considered model is the Total Variation (TV) (see Def. 1) minimization

$$\min_{\boldsymbol{\mathcal{X}}} \ \|\boldsymbol{\mathcal{X}}\|_{\text{TV}}, \ s.t. \ \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\mathcal{X}})\|_2 \leq \varepsilon. \tag{3}$$

Recur to the close relationship between Haar transform and TV, model (3) was also proven to enjoy a robust recovery guarantee under the conventional Sparse RIP (SRIP) condition (Needell and Ward 2013). A wide range of applications indicates that TV is vitally important to improve the recovery performance as well (Xutao Li 2017; Wang et al. 2022).

In wake of the continuously proceeding research process, it is expected to seek available mechanisms which can simultaneously take both of the aforementioned properties into consideration in modeling. A natural way is combining the regularizers from model (2) and mode (3) with a trade-off parameter. However, as far as we know, the theoretical guarantee of this approach is still unclear, and it will certainly increase more hyper-parameters to be tuned in the solution procedure. Inspired by a recent study about the Robust Principle Component Analysis (RPCA) problem which models on the gradient maps of images (Peng et al. 2022), in this paper, we use T-SVD on gradient maps of the original tensor and introduce the relevant Tensor Correlated Total Variation (TCTV). Mathematically speaking, the TCTV of $\boldsymbol{\mathcal{X}}$ is based on the T-SVD on its gradient maps

$$\|\boldsymbol{\mathcal{X}}\|_{\text{TCTV}} = \sum_{i=1}^{3} \|\nabla_i \boldsymbol{\mathcal{X}}\|_{\star_L}, \tag{4}$$

where ∇_i denotes the differential operator along the i-th dimension (see Eq. (6)). Compared with the approach of combining regularizers, the TCTV is a parameter-free way, i.e. without a trade-off coefficient. Under this setup, we are concerned with the following TCS model

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \ \|\boldsymbol{\mathcal{X}}\|_{\text{TCTV}}, \ s.t. \ \|\boldsymbol{y} - \mathcal{A}(\boldsymbol{\mathcal{X}})\|_2 \leq \varepsilon. \tag{5}$$

About model (5), what we are interested in is answering the following three questions:

- Does model (5) have a robust recovery guarantee just like model (2) and model (3) do?
- Whether model (5) benefits from the low-rank and local-smooth properties simultaneously?
- Does model (5) behave eminently in real-world applications?

The main contributions of this paper concentrate on the above three key questions. In theory, we show that model (5) indeed enjoys a robust recovery guarantee. Moreover, we theoretically prove that the error bound derived by the TCTV can be upper bounded by the smaller one of the error bounds derived by tensor nuclear norm and TV seminorm, which indicates that model (5) can benefit from both properties adaptively. Based on the Alternating Direction Method of Multipliers (ADMM) (Boyd et al. 2011), we implement an optimization algorithm with a global convergence guarantee. We also apply model (5) to HSI and video restoration problems, and our model is superior to many competing methods both numerically and visually all along. In summary, both theoretical and experimental results illustrate that the two properties may reinforce each other under the framework of model (5).

Related work. There are some other practices that also simultaneously take both properties into account in many image restoration problems. The most representative two methods are the weighted summation of both regularizers and factorization-based methods. For example, (Madathil and N.George 2018) used the weighted summation of both regularizers for image and video completion problems and (Chen, Wang, and Zhou 2018) applied it to image denoising problem in a similar way; (Wang et al. 2017) took the Tucker factorization into modeling for TCS problem and (Li et al. 2022) used an analogical strategy to BSCM problem. These studies experimentally reveal the prominent advantage of both approaches in efficaciously improving recovery performance. However, the disadvantages are also obvious: (1) both are not theoretically guaranteed so far; (2) the weighted summation one introduces more hard-to-select hyper-parameters and the factorization one needs to preset the tensor rank, which lacks a general approach. Compared with these studies, our approach is not only experimentally verified to behave well but does not have the shortcomings mentioned above.

Notations and Preliminaries

Notations

We denote scalars by lowercase letters, e.g., x, vectors by boldface lowercase letters, e.g., $\boldsymbol{x}$, matrices by boldface capital letters, e.g., $\boldsymbol{X}$, and tensors by boldface Euler script letters, e.g., $\boldsymbol{\mathcal{X}}$. For a three-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote its (i,j,k)-th entry as x_{ijk} or $\boldsymbol{\mathcal{X}}_{ijk}$ and use $\boldsymbol{\mathcal{X}}(i,:,:)$, $\boldsymbol{\mathcal{X}}(:,i,:)$ and $\boldsymbol{\mathcal{X}}(:,:,i)$ to represent the i-th horizontal, lateral, and frontal slice, respectively. Generally, we also denote $\boldsymbol{X}^{(i)}$ or $\boldsymbol{\mathcal{X}}^{(i)}$ as the i-th frontal slice $\boldsymbol{\mathcal{X}}(:,:,i)$ and $\boldsymbol{\mathcal{X}}(i,j,:)$ as the (i,j)-th tube.

The inner product between two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ of the same size is defined as $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \text{tr}(\boldsymbol{X}^T \boldsymbol{Y})$. The inner product between two tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ of the same size is defined as $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle = \sum_{i,j,k} x_{ijk} y_{ijk}$. For a matrix $\boldsymbol{X}$, its nuclear norm $\|\boldsymbol{X}\|_*$ is defined as sum of singular values of $\boldsymbol{X}$ and its Frobenius norm is defined as $\|\boldsymbol{X}\|_F = \sqrt{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}$. For a three-order tensor $\boldsymbol{\mathcal{X}}$, we denote its Frobenius norm as $\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}} \rangle}$ and the ℓ_1 norm as $\|\boldsymbol{\mathcal{X}}\|_1 = \sum_{i,j,k} |x_{ijk}|$. For $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, we

denote $\boldsymbol{\mathcal{X}} \triangle \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ by $(\boldsymbol{\mathcal{X}} \triangle \boldsymbol{\mathcal{Y}})^{(i)} = \boldsymbol{\mathcal{X}}^{(i)} \boldsymbol{\mathcal{Y}}^{(i)}$ for all $i \in [n_3]$, where $[n_3] = \{1, 2, \cdots, n_3\}$.

Total Variation and Haar Wavelet Transform

It is well known that total variation can be used to express the local-smooth property of images. At the same time, the total variation of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is closely related to its discrete gradients. Mathematically, the gradients along the i-th dimension denoted as the gradient tensor $\nabla_i(\boldsymbol{\mathcal{X}})$ is defined as

$$\nabla_i(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{X}} \times_i \boldsymbol{D}_i, \tag{6}$$

where $\times_i$ denotes the mode-i product (Kolda and Bader 2009) and $\boldsymbol{D}_i = \operatorname{circ}([-1, 1, \cdots, 0]) \in \mathbb{R}^{(n_i - 1) \times n_i}$ is a differential matrix.

Definition 1. *The total variation seminorm $\|\boldsymbol{\mathcal{X}}\|_{\text{TV}}$ [1] is defined as*

$$\|\boldsymbol{\mathcal{X}}\|_{\text{TV}} \triangleq \sum_{i=1}^{3} \|\nabla_i(\boldsymbol{\mathcal{X}})\|_1. \tag{7}$$

In order to transform a tensor to its gradient domain, we further introduce two augmented tensors $\boldsymbol{\mathcal{X}}^{0_1}, \boldsymbol{\mathcal{X}}_{0_1} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ for $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{(n_1 - 1) \times n_2 \times n_3}$ defined by

$$x^{0_1}_{ijk} = \begin{cases} 0, & i = 1, \\ x_{i-1,jk}, & 2 \le i \le n_1, \end{cases} (x_{0_1})_{ijk} = \begin{cases} 0, & i = n_1, \\ x_{ijk}, & i \in [n_1 - 1], \end{cases}.$$

Meanwhile, we can define $\boldsymbol{\mathcal{X}}^{0_2}, \boldsymbol{\mathcal{X}}_{0_2}$ for $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times (n_2 - 1) \times n_3}$ and $\boldsymbol{\mathcal{X}}^{0_3}, \boldsymbol{\mathcal{X}}_{0_3}$ for $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times (n_3 - 1)}$ in a similar way. It has been shown in (Needell and Ward 2013) that the inner product between the gradient tensor $\nabla_1(\boldsymbol{\mathcal{Y}})$ and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{(n_1 - 1) \times n_2 \times n_3}$ obeys the following rule

$$\langle \boldsymbol{\mathcal{X}}, \nabla_1(\boldsymbol{\mathcal{Y}}) \rangle = \langle \boldsymbol{\mathcal{X}}^{0_1}, \boldsymbol{\mathcal{Y}} \rangle - \langle \boldsymbol{\mathcal{X}}_{0_1}, \boldsymbol{\mathcal{Y}} \rangle. \tag{8}$$

Similar equations of (8) hold true for $\nabla_2(\boldsymbol{\mathcal{Y}})$ and $\nabla_3(\boldsymbol{\mathcal{Y}})$. Moreover, this rule can be extended to a linear operator $\mathcal{A} : \mathbb{R}^{(n_1 - 1) \times n_2 \times n_3} \to \mathbb{R}^m$ with $[\mathcal{A}(\boldsymbol{\mathcal{X}})]_j = \langle \boldsymbol{\mathcal{A}}_j, \boldsymbol{\mathcal{X}} \rangle$. To this end, we denote $\mathcal{A}^{0_1} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$ with components $[\mathcal{A}^{0_1}(\boldsymbol{\mathcal{X}})]_j = \langle \boldsymbol{\mathcal{A}}^{0_1}_j, \boldsymbol{\mathcal{X}} \rangle$, and the definition of $\mathcal{A}_{0_1}$ is similarly. Thus, we conclude from (8) that $\mathcal{A}(\nabla_1 \boldsymbol{\mathcal{X}}) = \mathcal{A}^{0_1}(\boldsymbol{\mathcal{X}}) - \mathcal{A}_{0_1}(\boldsymbol{\mathcal{X}})$. In the same way, we can derive homologous results for $\mathcal{A}(\nabla_i \boldsymbol{\mathcal{X}})$ $(i = 2, 3)$. Lastly, for two operators $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{m_1}$ and $\mathcal{B} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{m_2}$, we denote their concatenate operator as $[\mathcal{A}, \mathcal{B}] : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{m_1 + m_2}$.

The total variation seminorm is closely related to the Haar wavelet transform. In general, the discrete three-dimensional Haar wavelet transform $\mathcal{H} : \mathbb{R}^{n \times n \times n} \to \mathbb{R}^{n \times n \times n}$ $(n = 2^p, p \in \mathbb{N}_+)$ is introduced by

$$\mathcal{H}(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{X}} \times_1 \boldsymbol{H} \times_2 \boldsymbol{H} \times_3 \boldsymbol{H}, \tag{9}$$

where $\boldsymbol{H} \in \mathbb{R}^{n \times n}$ is the Haar matrix satisfying $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{H}^T\boldsymbol{H} = \boldsymbol{I}$ (Morton and Petersen 1997). It is easy to check that $\|\mathcal{H}(\boldsymbol{\mathcal{X}})\|_F = \|\mathcal{H}^{-1}(\boldsymbol{\mathcal{X}})\|_F = \|\boldsymbol{\mathcal{X}}\|_F$.

[1] Here we mean the anisotropic total variation, and the definition of the isotropic total variation is omitted. The dimension of gradient maps considered here is smaller than the original tensor, we can pad zeros along the i-th dimension such that both $\nabla_i \boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{X}}$ are of the same size. We use the same notations without confusion.

Transform Based T-product Induced Tensor Nuclear Norm

Let $\boldsymbol{L} \in \mathbb{C}^{n_3 \times n_3}$ be an arbitrary invertible matrix. We use the notation $\bar{\boldsymbol{\mathcal{X}}}$ to denote the tensor in the transformed domain specified by $\boldsymbol{L}$

$$\bar{\boldsymbol{\mathcal{X}}} = L(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mathcal{X}} \times_3 \boldsymbol{L}. \tag{10}$$

In this paper, we are interested in the following case

$$\boldsymbol{L}^T\boldsymbol{L} = \boldsymbol{L}\boldsymbol{L}^T = \ell \boldsymbol{I}_{n_3}. \tag{11}$$

Transform-based T-product and related concepts come from (Kernfeld, Kilmer, and Aeron 2015).

Definition 2 (T-product)**.** *Let L be any invertible linear transform, and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$. Then the transform L based T-product denoted as $\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{X}} \star_L \boldsymbol{\mathcal{Y}}$, is defined such that $L(\boldsymbol{\mathcal{Z}}) = L(\boldsymbol{\mathcal{X}}) \triangle L(\boldsymbol{\mathcal{Y}})$.*

Definition 3 (Tensor transpose)**.** *Let L be any invertible linear transform, and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then the tensor transpose $\boldsymbol{\mathcal{X}}^T$ satisfies $L(\boldsymbol{\mathcal{X}}^T)^{(i)} = (L(\boldsymbol{\mathcal{X}}^{(i)}))^T$ $(i \in [n_3])$.*

Definition 4 (Identity tensor)**.** *Let L be any invertible linear transform, $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n \times n \times n_3}$ be such that each frontal slice of $L(\boldsymbol{\mathcal{I}}) = \bar{\boldsymbol{\mathcal{I}}}$ is a $n \times n$ sized identity matrix. Then the identity tensor is defined as $\boldsymbol{\mathcal{I}} = L^{-1}(\bar{\boldsymbol{\mathcal{I}}})$.*

Definition 5 (Orthogonal tensor)**.** *Let L be an invertible linear transform. We say that a tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is orthogonal if it satisfies $\boldsymbol{\mathcal{Q}}^T \star_L \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} \star_L \boldsymbol{\mathcal{Q}}^T = \boldsymbol{\mathcal{I}}$.*

Definition 6 (F-diagonal tensor)**.** *We call a tensor F-diagonal if each of its frontal slices is a diagonal matrix.*

With the above definitions, we introduce the transform-based T-SVD in the following lemma.

Lemma 1 (T-SVD)**.** *Let L be an invertible linear transform, and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factorized as*

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star_L \boldsymbol{\mathcal{S}} \star_L \boldsymbol{\mathcal{V}}^T, \tag{12}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an F-diagonal tensor.

Based on the T-SVD, (Canyi Lu 2019) gave the related definitions of tensor tubal rank and tensor nuclear norm.

Definition 7 (Tensor tubal rank and tensor average rank)**.** *Let L be any invertible linear transform. For a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with* T-SVD *$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star_L \boldsymbol{\mathcal{S}} \star_L \boldsymbol{\mathcal{V}}^T$, the tensor tubal rank, denoted as $\operatorname{rank}_t(\boldsymbol{\mathcal{X}})$, is defined as the number of nonzero singular tubes of $\boldsymbol{\mathcal{S}}$. That is to say, $\operatorname{rank}_t(\boldsymbol{\mathcal{X}}) = \#\{i, \boldsymbol{\mathcal{S}}(i, i, :) \neq \boldsymbol{0}\} = \max_{i \in [n_3]}(\operatorname{rank}(\bar{\boldsymbol{\mathcal{X}}})^{(i)})$. The tensor average rank is defined as $\operatorname{rank}_a(\boldsymbol{\mathcal{X}}) = \frac{1}{\ell} \operatorname{rank}(\bar{\boldsymbol{X}})$.*

Definition 8 (Tensor Nuclear Norm (TNN))**.** *Let L be an invertible linear transform satisfying* (11)*. The tensor nuclear norm of $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\|\boldsymbol{\mathcal{X}}\|_{\star_L} = \frac{1}{\ell}\|\bar{\boldsymbol{X}}\|_* = \frac{1}{\ell}\sum_{i=1}^{n_3} \|(\bar{\boldsymbol{\mathcal{X}}})^{(i)}\|_*$, where $\bar{\boldsymbol{X}} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$ is denoted as a block diagonal matrix whose i-th block on the diagonal is the i-th frontal slice $\bar{\boldsymbol{X}}^{(i)}$ of $\bar{\boldsymbol{\mathcal{X}}}$, i.e.*

$$\bar{\boldsymbol{X}} = \operatorname{bdiag}(\bar{\boldsymbol{\mathcal{X}}}) = \begin{bmatrix} \bar{\boldsymbol{X}}^{(1)} & & & \\ & \bar{\boldsymbol{X}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\boldsymbol{X}}^{(n_3)} \end{bmatrix}. \tag{13}$$

Main Theoretical Guarantees

In order to recover $\boldsymbol{\mathcal{X}}$ from its linear observations $\boldsymbol{y}$, it is necessary to ensure that these measurements contain enough information from $\boldsymbol{\mathcal{X}}$. Nowadays, the widely recognized property of $\mathcal{A}$ for this requirement is the TRIP. In this section, we first derive the T-SVD based TRIP [2] condition and further obtain a robust recovery guarantee for model (5). Proofs of these theoretical results are delayed to the Supplementary Material.

Guaranteed RIP Conditions

Our interested tensors are equipped with low-tubal-rank property. We first derive the TRIP condition for tensors appertain to the set $S_r = \{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3} : \operatorname{rank}_t(\boldsymbol{\mathcal{X}}) \leq r, \|\boldsymbol{\mathcal{X}}\|_F = 1\}$.

Definition 9 (TRIP)**.** *We call the linear map $\mathcal{A} : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^m$ satisfies the TRIP, if*

$$(1-\delta_r)\|\boldsymbol{\mathcal{X}}\|_F^2 \leq \|\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2 \leq (1+\delta_r)\|\boldsymbol{\mathcal{X}}\|_F^2, \quad (14)$$

holds for any $\boldsymbol{\mathcal{X}} \in S_r$, where the Restricted Isometry Constant (RIC) δ_r is the smallest constant such that (14) *holds.*

Our first main theorem contributes to the question of which kind of linear operators satisfies TRIP.

Theorem 1. *Let $\mathcal{A} : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^m$ be a linear map, $0 < t < 1$ be fixed. If for any $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the following concentration inequality holds*

$$P\left(\left|\|\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2 - \|\boldsymbol{\mathcal{X}}\|_F^2\right| > t\|\boldsymbol{\mathcal{X}}\|_F^2\right) \leq c_1\exp(-c_2 m), \quad (15)$$

further, if

$$m > \frac{1}{c_2} r(n_1+n_2+1)n_3 \ln\frac{36\sqrt{2}}{\delta_r},$$

then $\mathcal{A}$ satisfies the TRIP with probability greater than $1 - c_1\exp(-c_3 m)$, where c_3 is a constant which may depend on δ_r.

We do not give the TRIP guarantee with special sensing tensors, but instead present it under the concentration inequality (15). Actually, it has the advantage of being broadly suitable to many types of linear operators such as the sensing tensors are composed of random tensors or structurally random tensors (Krahmer and Ward 2011; Rudelson and Vershynin 2007; Do et al. 2012). Moreover, another widely used SRIP defined for tensors from the set $S_s = \{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3} : \|\boldsymbol{\mathcal{X}}\|_0 \leq s, \|\boldsymbol{\mathcal{X}}\|_F = 1\}$ (Candès 2008) is also satisfied under the concentration inequality (15) if $m \gtrsim s\log\frac{n_1 n_2 n_3}{s}$ [3] (Baraniuk et al. 2008). What's worth mentioning, if (15) holds, direct computation shows

$$P\left(\left|\|\mathcal{A}\mathcal{H}^{-1}(\boldsymbol{\mathcal{X}})\|_2^2 - \|\boldsymbol{\mathcal{X}}\|_F^2\right| > t\|\boldsymbol{\mathcal{X}}\|_F^2\right) \leq c_1\exp(-c_2 m).$$

This indicates that if $\mathcal{A}$ satisfies some RIP on the condition that (15) holds, then the same RIP holds for $\mathcal{A}\mathcal{H}^{-1}$.

[2] In the rest of this paper, we use the notation TRIP to stand for the T-SVD based TRIP without confusion.

[3] $a \gtrsim b$ means that there exists an absolute constant c such that $a \geq cb$; $a \lesssim b$ means that there exists an absolute constant c such that $a \leq cb$.

Recovery Guarantee of TCTV Minimization

The tensors that we care about have simultaneously low-tubal-rank and local-smooth properties. To deduce the robust recovery guarantee of model (5), we shall adopt both TRIP and SRIP to construct the recoverability conditions. Based on the TRIP and SRIP above, we demonstrate the theoretical recovery guarantee of model (5) below.

Theorem 2. *Suppose $n = 2^p$, $r \in [n]$, $s \in [n^3]$, and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n\times n\times n}$ is a simultaneously low-rank and gradient sparse tensor. Let $\mathcal{A}_1 : \mathbb{R}^{(n-1)\times n\times n} \to \mathbb{R}^{m_1}$, $\mathcal{A}_2 : \mathbb{R}^{n\times(n-1)\times n} \to \mathbb{R}^{m_1}$ and $\mathcal{A}_3 : \mathbb{R}^{n\times n\times(n-1)} \to \mathbb{R}^{m_1}$ be satisfied that $\mathcal{A}_i$ ($i \in [3]$) have the TRIP with RIC δ_{4r}. Let $\mathcal{H} : \mathbb{R}^{n\times n\times n} \to \mathbb{R}^{n\times n\times n}$ be the three-dimensional Haar transform and $\mathcal{B} : \mathbb{R}^{n\times n\times n} \to \mathbb{R}^{m_2}$ be satisfying that $\mathcal{B}\mathcal{H}^{-1}$ has the SRIP with RIC δ_{2s}. Let $m = 6m_1 + m_2$, and consider the concatenate linear operator given by $\mathcal{M}(\boldsymbol{\mathcal{X}}) = [\mathcal{A}^{0_1}, \mathcal{A}_{0_1}, \mathcal{A}^{0_2}, \mathcal{A}_{0_2}, \mathcal{A}^{0_3}, \mathcal{A}_{0_3}, \mathcal{B}](\boldsymbol{\mathcal{X}})$. If $\delta_{4r} < \frac{1}{1+\sqrt{2n}}$, $\delta_{2s} < 1$ and the noisy measurements $\boldsymbol{y} = \mathcal{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{e}$ are observed with noise level $\|\boldsymbol{e}\|_2 \leq \varepsilon$, then the solution $\hat{\boldsymbol{\mathcal{X}}}$ of model* (5) *satisfies*

$$\frac{\|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}\|_F}{n^{3/2}} \leq C\Big(\frac{\|\boldsymbol{\mathcal{X}}^c_{r,s}\|_{\text{TCTV}}}{\sqrt{rs}} + \frac{\varepsilon}{\sqrt{s}}\Big)\log n, \quad (16)$$

where C is a constant dependent only on the RIC, $\boldsymbol{\mathcal{X}}^c_{r,s} = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}_{r,s}$ and $\boldsymbol{\mathcal{X}}_{r,s}$ is a tensor satisfying simultaneously tubal rank at most r and the number of nonzero gradients at most s.

If $\boldsymbol{\mathcal{X}}$ has tubal-rank at most r and nonzero gradients at most s, Theorem 2 presents a robust recovery guarantee of model (5), and it is exact if $\varepsilon = 0$. Furthermore, even though our theoretical findings are given in terms of three-order tensor, it can be easily extended to the high-order case by using a recent work from (Qin et al. 2022). We omit it here. Besides, the additive concatenate sampling operator is just used to fulfill the need of proof process, which will not change the sampling order. At last, we require the tensor side-length to be a power of 2, i.e. $n = 2^p$. This should not be a restriction, as a tensor $\boldsymbol{\mathcal{X}}$ of arbitrary side-length $n_1 \times n_2 \times n_3$ can be reflected across each dimension by producing an at most $n \times n \times n$ tensor without increasing the TCTV.

Corollary 1. *Suppose the same conditions of Theorem 2 hold, then*

$$\frac{\|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}\|_F}{n^{3/2}} \lesssim C\Big(\frac{\min\{\|\boldsymbol{\mathcal{X}}^c_{r,s}\|_{\text{TV}}, \|\boldsymbol{\mathcal{X}}^c_{r,s}\|_{\star_L}\}}{\sqrt{rs}} + \frac{\varepsilon}{\sqrt{s}}\Big)\log n. \quad (17)$$

The theoretical findings from Corollary 1 are intriguing. From the 'min' formulation of the first term of the error bound, we can make two conclusions: if $\boldsymbol{\mathcal{X}}_{r,s}$ is the best tubal rank r approximation of $\boldsymbol{\mathcal{X}}$ (Kilmer et al. 2021), the error bound (17) degenerates into the theoretical guarantee for model (2) (Canyi Lu 2018); if $\boldsymbol{\mathcal{X}}_{r,s}$ is the best gradient s approximation of $\boldsymbol{\mathcal{X}}$, the error bound (17) degenerates into the theoretical guarantee for model (3) (Needell and Ward 2013). In other words, we theoretically obtain that model (5) will be better than model (2) and model (3). Our results profoundly indicate that model (5) may make both low-tubal-rank and local-smooth properties mutually reinforcing, and

our model indeed benefits from both properties of $\boldsymbol{\mathcal{X}}$ adaptively.

Algorithm Implementations

Rather than solving model (5) directly, we design an algorithm based on the ADMM for the unconstrained model

$$\min_{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}} \frac{1}{2}\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2+\lambda\|\boldsymbol{\mathcal{X}}\|_{\text{TCTV}}, \tag{18}$$

where λ is a properly chosen regularization parameter. Let $\boldsymbol{\mathcal{Z}}_i=\nabla_i\boldsymbol{\mathcal{X}}$ $(i\in[3])$, problem (18) can be transformed to

$$\begin{aligned}\min_{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}} &\frac{1}{2}\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2+\lambda\sum_{i=1}^{3}\|\boldsymbol{\mathcal{Z}}_i\|_{\star_L}\\ &s.t.\ \nabla_i\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{Z}}_i, i\in[3].\end{aligned} \tag{19}$$

Hence, the Lagrangian function is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\mathcal{X}},\{\boldsymbol{\mathcal{Z}}_i\}_{i=1}^3,\{\boldsymbol{\mathcal{D}}_i\}_{i=1}^3,\mu)&=\frac{1}{2}\|\boldsymbol{y}-\mathcal{A}(\boldsymbol{\mathcal{X}})\|_2^2+C\\ &+\lambda\sum_{i=1}^{3}\|\boldsymbol{\mathcal{Z}}_i\|_{\star_L}+\frac{\mu}{2}\sum_{i=1}^{3}\Big\|\boldsymbol{\mathcal{Z}}_i-\nabla_i\boldsymbol{\mathcal{X}}+\frac{\boldsymbol{\mathcal{D}}_i}{\mu}\Big\|_F^2,\end{aligned} \tag{20}$$

where C is only multipliers dependent squared items. According to the update rule of ADMM, variables are updated alternately.

For term $\boldsymbol{\mathcal{X}}$, we update it by solving the following linear system obtained by taking derivative on both sides of the Lagrangian function with respect to $\boldsymbol{\mathcal{X}}$

$$(\mathcal{A}^T\mathcal{A}+\sum_{i=1}^{3}\mu^k\nabla_i^T\nabla_i)(\boldsymbol{\mathcal{X}})=\mathcal{A}^T(\boldsymbol{y})+\sum_{i=1}^{3}\nabla_i^T(\mu^k\boldsymbol{\mathcal{Z}}_i^k+\boldsymbol{\mathcal{D}}_i^k). \tag{21}$$

Equations (21) can be solved by off-the-shelf techniques such as the preconditioned conjugate gradient method (Axelsson and Lindskog 1986).

For terms $\boldsymbol{\mathcal{Z}}_i$ $(i\in[3])$, we update it by

$$\begin{aligned}\boldsymbol{\mathcal{Z}}_i^{k+1}&=\arg\min_{\boldsymbol{\mathcal{Z}}_i}\lambda\|\boldsymbol{\mathcal{Z}}_i\|_{\star_L}+\frac{\mu^k}{2}\Big\|\boldsymbol{\mathcal{Z}}_i-\nabla_i\boldsymbol{\mathcal{X}}^{k+1}+\frac{\boldsymbol{\mathcal{D}}_i^k}{\mu}\Big\|_F^2\\ &=\mathcal{D}_{\lambda/\mu^k}\big(\nabla_i\boldsymbol{\mathcal{X}}^{k+1}-\frac{\boldsymbol{\mathcal{D}}_i^k}{\mu^k}\big),\end{aligned} \tag{22}$$

where $\mathcal{D}_{\lambda/\mu}(\cdot)$ denotes the Tensor Singular Value Thresholding (T-SVT) operator which can be found in (Canyi Lu 2019).

We summarize the whole update procedures in Algorithm 1. It is easy to check that the time complexity of Algorithm 1 is $\mathcal{O}((n_1n_2n_3)^2)$. Besides, it seems that problem (19) involves four variable blocks. However, the convergence of the corresponding ADMM algorithm still remains unclear in general (Lin, Ma, and Zhang 2016; Chen et al. 2019). Fortunately, the proposed algorithm can be transformed into the two-block ADMM whose convergence is provable. We formally state the conclusion in the following theorem and give detailed proof in the Supplementary Material.

Theorem 3. *Let $\{\boldsymbol{\mathcal{X}}^k, k\geq 0\}$ be a sequence generated by Algorithm 1, then it converges to the global minimum of problem* (19).

Algorithm 1: ADMM for problem (19)

Input: Linear operator $\mathcal{A}$, measurements $\boldsymbol{y}$.

1: Initialization: $\boldsymbol{\mathcal{X}}^0$, $\boldsymbol{\mathcal{Z}}_i^0$ and $\boldsymbol{\mathcal{D}}_i^0$ $(i\in[3])$, regularization parameter λ, $\mu>0$, tolerate error $tol=10^{-8}$, $\rho=1.1$, $\mu_{max}=10^6$ and $k=0$.

while not convergent **do**

2: Update $\boldsymbol{\mathcal{X}}^{k+1}$ by (21);

3: Update $\boldsymbol{\mathcal{Z}}_i^{k+1}$ by (22);

4: $\boldsymbol{\mathcal{D}}_i^{k+1}=\boldsymbol{\mathcal{D}}_i^k+\mu^k\big(\boldsymbol{\mathcal{Z}}_i^{k+1}-\nabla_i\boldsymbol{\mathcal{X}}^{k+1}\big)$;

5: $\mu^{k+1}=\min\{\rho\mu^k,\mu_{max}\}$.

6: Check the convergence condition $\|\boldsymbol{\mathcal{X}}^{k+1}-\boldsymbol{\mathcal{X}}^k\|_F/\max\{\|\boldsymbol{\mathcal{X}}^k\|_F,1\}<tol$.

7: Update $k\leftarrow k+1$.

8: **end while**

Output: $\hat{\boldsymbol{\mathcal{X}}}=\boldsymbol{\mathcal{X}}^{k+1}$.

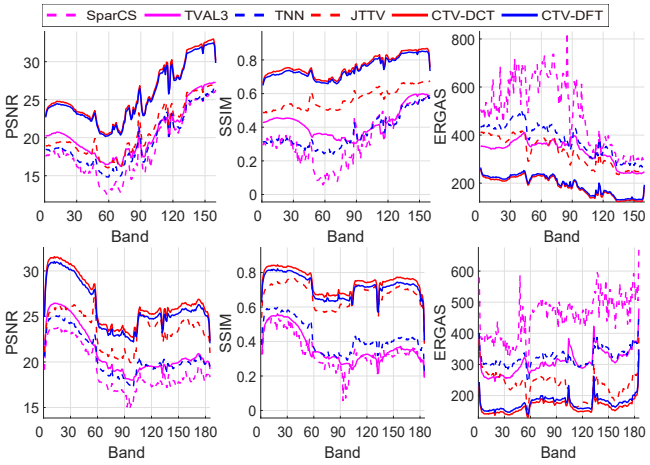

Figure 1: PSNR, SSIM, and ERGAS comparisons of the proposed approaches over four competing methods on all bands of two HSI datasets under SR=1%. Up: HYDICE Washington DC Mall, Down: HYDICE Urbanpart.

Applications

CS had been successfully applied to many data types scattered in multiple fields, typical examples include the HSI data and gray video data, which can be seen as three-order tensors in our experiments (Wang et al. 2017; Baraniuk et al. 2017). The Walsh-Hadamard sampling strategy is adopted: $\mathcal{A}(\boldsymbol{\mathcal{X}})=\boldsymbol{A}\,\text{vec}(\boldsymbol{\mathcal{X}})$, where $\boldsymbol{A}=\boldsymbol{DFR}$ and $\boldsymbol{R}$ is a random permutation matrix, $\boldsymbol{F}$ is the Walsh-Hadamard transform, $\boldsymbol{D}$ is a randomly downsampling operator. This strategy has two typical practical advantages: it is of easy implementation on hardware, fast transformation, and satisfactory recoverability (Chengbo Li 2011); it satisfies the concentration inequality (15) (Krahmer and Ward 2011; Rudelson and Vershynin 2007), and thus both the TRIP and SRIP hold according to Theorem 1. Three quantitative quality indices, i.e. PSNR, SSIM (Wang et al. 2004) and ERGAS (Lucien 2002) are employed to display the recovery effect. PSNR and SSIM are two conventional spatial-based metrics, and ERGAS is a spectral-based evaluation measure for HSI. The higher PSNR and SSIM values and the lower ERGAS value are, the better quality of the recovered images become.

SR	Quality indices	HYDICE Washington DC Mall						HYDICE Urbanpart					
		SparCS	TVAL3	JTTV	TNN	TCTV-DCT	TCTV-DFT	SparCS	TVAL3	JTTV	TNN	TCTV-DCT	TCTV-DFT
0.3%	PSNR	17.278	19.099	18.364	17.619	**22.195**	21.906	18.241	19.850	20.167	18.663	**23.005**	22.623
	SSIM	0.1887	0.2968	0.3284	0.2182	**0.5217**	0.5028	0.3126	0.2794	0.4680	0.3257	**0.5440**	0.5203
	ERGAS	591.37	404.54	439.6	478.64	**285.33**	294.99	513.59	358.85	348.37	409.92	**249.75**	261.10
	TIME(s)	2023	**313**	1181	781	2690	2500	2695	**453**	1887	1273	2545	4037
1%	PSNR	18.608	20.662	20.820	19.534	**25.800**	25.481	19.204	21.189	23.628	20.701	**26.585**	26.036
	SSIM	0.3340	0.4253	0.5695	0.3691	**0.7640**	0.7493	0.3440	0.3683	0.6762	0.4258	**0.7472**	0.7188
	ERGAS	516.01	337.13	334.43	386.8	**190.99**	197.83	460.34	308.92	236.05	323.97	**166.44**	177.11
	TIME(s)	1856	**390**	1293	985	2921	2394	2296	**506**	1935	1395	4874	2207
5%	PSNR	24.418	23.895	26.222	29.353	**35.250**	35.232	25.973	24.947	34.364	31.846	**36.449**	36.338
	SSIM	0.7195	0.6311	0.8375	0.9051	**0.9701**	0.9692	0.6933	0.6200	0.9486	0.9134	**0.9578**	0.9550
	ERGAS	266.73	235.31	200.85	124.92	**64.926**	65.25	195.27	203.59	74.07	93.208	**60.21**	61.02
	TIME(s)	1163	**428**	1261	865	2343	1310	1515	**525**	1932	1205	3890	2001
10%	PSNR	29.299	26.061	28.494	34.768	41.088	**41.629**	31.531	27.506	37.334	37.429	41.684	**41.882**
	SSIM	0.8731	0.7468	0.8880	0.9723	0.9921	**0.9928**	0.8709	0.7474	0.9683	0.9686	0.9803	**0.9798**
	ERGAS	153.41	184.19	167.23	66.48	33.07	**31.23**	104.25	153.73	57.27	54.28	41.22	**41.10**
	TIME(s)	909	**374**	1228	783	2259	1266	1215	**502**	1937	1059	1993	1785

Table 1: Comparisons about quality indices of different models under different SRs on two HSI datasets.

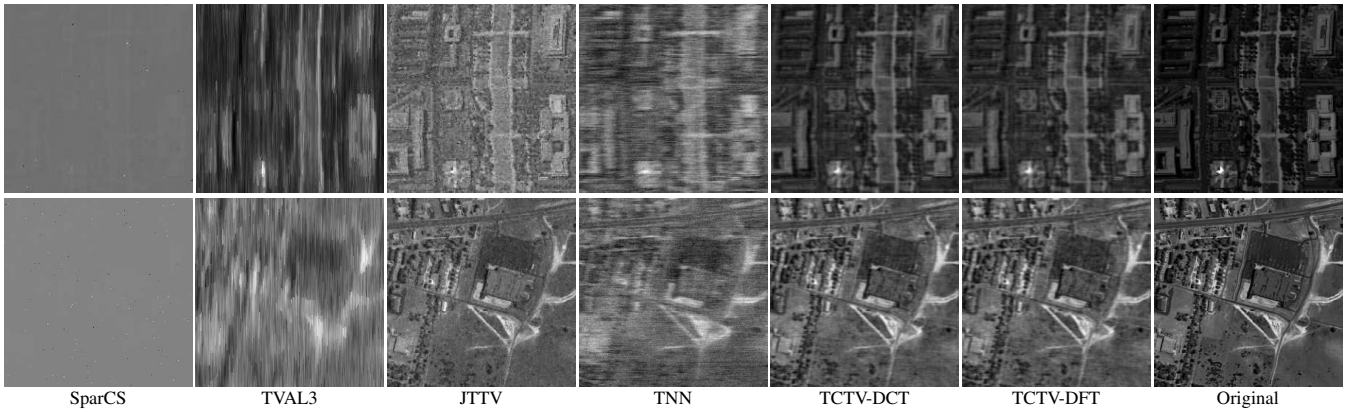

Figure 2: Visual comparisons of the proposed approaches over four other competing methods on band 130 of both HSI datasets under SR=1%. Up: HYDICE Washington DC Mall, Down: HYDICE Urbanpart.

We consider two cases of $\boldsymbol{L}$: Discrete Cosine Transform (DCT) (Kernfeld, Kilmer, and Aeron 2015), Discrete Fourier Transform (DFT) (Kernfeld, Kilmer, and Aeron 2011), and name the corresponding models as TCTV-DCT and TCTV-DFT, respectively. Before experiments, pixels of the HSIs and videos are normalized to [0,1] via the max-min formula. Regularization parameters λ for our models are chosen empirically from the set $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{1}, 10^{2}, 10^{3}\}$ by cross-validation. Due to space constraints, we deposit some more experimental results in the Supplementary Material. All experiments are run in MATLAB R2016a on a 64-bit PC with an E7-4820 2.00GHz CPU and 64GB memory.

Application to HSI Compressive Sensing

We choose two typical datasets, HYDICE Washington DC Mall and HYDICE Urbanpart in this experiment. After removing the seriously polluted bands and cropping images, the images used in this experiment are of size $200\times200\times160$ and $256\times256\times185$, respectively. We also set $\varepsilon = 0$ and vary Sampling Ratio (SR) in the set $\{0.3\%, 1\%, 5\%, 10\%\}$. Four other models are used for comparisons: SparCS (Andrew Waters 2011), TVAL3 (Li et al. 2013), JTTV (Vandergheynst 2012), TNN (Canyi Lu 2018). Here, SparCS, TVAL3, TNN, JTTV represent four most common methods to regularize low-rankness, local-smoothness, tensor low-rankness and simultaneously low-rank and local-smooth properties of a tensor.

We collect numerical results of all methods in Table 1. It can be seen that our two models always outperform the others apparently and achieve satisfactory numerical values even though SR=0.3%, 1%. Particularly, under the extremely low SR=1%, we make further comparisons. Figure 1 shows the quality indices on each band, and our two models obviously behave better on all bands of both datasets. Figure 2 shows the recovered images on band 130 of both datasets, where the restored images by ours are also closer to the original in visual.

These experimental results further demonstrate the advantage of our proposed model. In addition, both TCTV-DCT and TCTV-DFT behave similarly except that TCTV-DCT seems slightly better in lower SRs while TCTV-DFT seems slightly better in higher SRs. It seems that our algorithm takes more time. The reason behind may be that Algorithm 1 needs three T-SVT computations in each iteration.

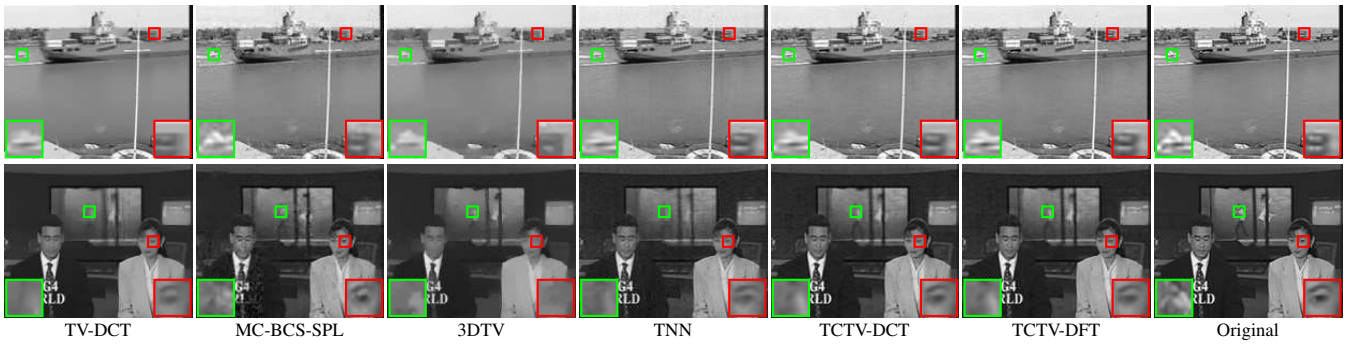

Figure 3: Visual comparisons of the proposed approaches to four other competing methods on frame 40 of both video cubes under SR=10%. Up: Container, Down: News.

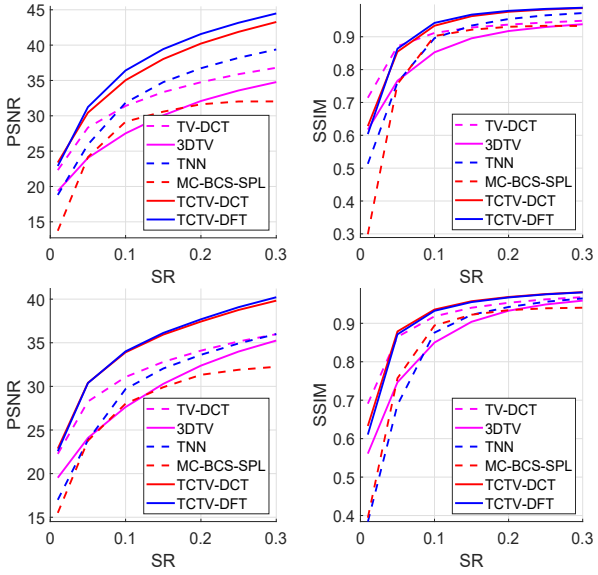

Figure 4: PSNR and SSIM comparisons of the proposed approaches over four competing methods with varying SRs. Up: Container, Down: News.

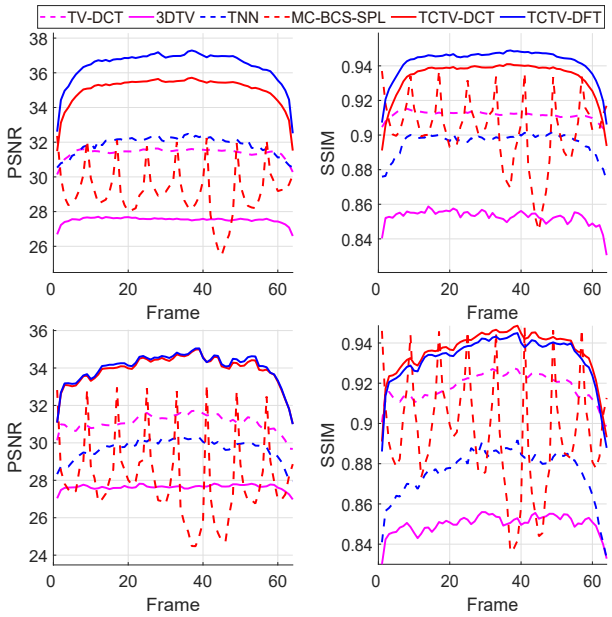

Figure 5: PSNR and SSIM comparisons of the proposed approaches over four competing methods on all frames of two video cubes under SR=10%. Up: Container, Down: News.

Application to Video Compressive Sensing

We choose two video datasets Container and News[4] in this experiment. For simplicity, we treat each video cube as a whole just for the purpose of verifying the effectiveness of our models. After cropped, both video cubes used here are of size $144\times176\times64$. We set $\varepsilon = 0$, and four other methods are used for comparisons: TV-DCT (Chengbo Li 2011), MC-BCS-SPL (Fowler 2012), 3DTV (Chengbo Li 2011), TNN (Canyi Lu 2018).

Figure 4 gathers the PSNR and SSIM values of the two video cubes on varying SRs. Clearly, our methods obtain higher quality indices in both video cubes about almost all SRs. Under SR=10%, we do further comparisons. In Figure 5, we compare the quality values on frames of both video cubes for different methods. As before, our methods present dramatically better performances in almost all frames. The MC-BCS-SPL relies on key frames with higher SR (we use its default setting 0.7), and its recovery results fluctuate up and down among frames. Even so, our methods still behave better in almost all frames than MC-BCS-SPL even in the key frames. Figure 3 further displays the 40-th recovered frames of different methods on both video cubes in visual, where the local area is enlarged by a ratio of 3. Obviously, our methods exhibit better recovery details compared with other methods. Additionally, TCTV-DFT behaves slightly better than TCTV-DCT on both video cubes.

[4]https://media.xiph.org/video/derf/

Conclusions and Future Work

In this paper, we have given positive answers to the three questions from the first section about model (5). In theory, we prove that this model indeed enjoys a robust recovery guarantee and benefits from both low-tubal-rank and local-smooth properties of the original data adaptively. In the experiment, we apply our model to HSI and video CS problems, and the experimental results further show its superiority to many other competing ones. The findings of our work may encourage more studies by modeling on gradient maps of data to capture the local-smooth property.

Acknowledgments

This research was funded in part by National Key Research and Development Program of China under Grant 2021YFB3101500, National Natural Science Foundation of China (Grants 12071380, 12101454, 12101512, 12201505), Macao Science and Technology Development Fund under Grant 061/2020/A2, Fundamental Research Funds for the Central Universities (Grant No. SWU120078), and China Postdoctoral Science Foundation (Grant No. 2021M692681).

References

Andrew Waters, R. B., Aswin Sankaranarayanan. 2011. SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In *Advances in Neural Information Processing Systems 24 (NIPS 2011)*.

Axelsson, O.; and Lindskog, G. 1986. On the rate of convergence of the preconditioned conjugate gradient method. *Numerische Mathematik*, 48(5): 499–523.

Baraniuk, R.; Davenport, M.; DeVore, R.; and Wakin, M. 2008. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3): 253–263.

Baraniuk, R. G.; Goldstein, T.; Sankaranarayanan, A. C.; Studer, C.; Veeraraghavan, A.; and Wakin, M. B. 2017. Compressive video sensing: Algorithms, architectures, and applications. *IEEE Signal Processing Magazine*, 34(1): 52–66.

Bernardino Romera-Paredes, M. P. 2013. A new convex relaxation for tensor completion. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 26.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; and Eckstein, J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1): 1–122.

Candès, E.; Romberg, J.; and Tao, T. 2006a. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2): 489–509.

Candès, E. J. 2008. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9-10): 589–592.

Candès, E. J.; Romberg, J. K.; and Tao, T. 2006b. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8): 1207–1223.

Canyi Lu, Y. W., Xi Peng. 2019. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5996–6004.

Canyi Lu, Z. L. S. Y., Jiashi Feng. 2018. Exact low tubal rank tensor recovery from Gaussian measurements. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*, 2504–2510. IJCAI Organization.

Cao, W.; Wang, Y.; Sun, J.; Meng, D.; Yang, C.; Cichocki, A.; and Xu, Z. 2016. Total variation regularized tensor RPCA for background subtraction from compressive measurements. *IEEE Transactions on Image Processing*, 25(9): 4075–4090.

Chen, L.; Li, X.; Sun, D.; and Toh, K.-C. 2019. On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming. *Mathematical Programming*, 185(1): 111–161.

Chen, Y.; Wang, S.; and Zhou, Y. 2018. Tensor nuclear norm-based low-rank approximation with total variation regularization. *IEEE Journal of Selected Topics in Signal Processing*, 12(6): 1364–1377.

Chengbo Li, P. W. Y. Z., Hong Jiang. 2011. Video coding using compressive sensing for wireless communications. In *2011 IEEE Wireless Communications and Networking Conference*, 12031610.

Ding, X.; Chen, W.; and Wassell, I. J. 2017. Joint sensing matrix and sparsifying dictionary optimization for tensor compressive sensing. *IEEE Transactions on Signal Processing*, 65(14): 3632–3646.

Do, T. T.; Gan, L.; Nguyen, N. H.; and Tran, T. D. 2012. Fast and efficient compressive sensing using structurally random matrices. *IEEE Transactions on Signal Processing*, 60(1): 139–154.

Donoho, D. 2006. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4): 1289–1306.

Fowler, S. M. J. E. 2012. Residual reconstruction for block-based compressed sensing of video. In *2011 Data Compression Conference*, 11933088.

Friedland, S.; Li, Q.; and Schonfeld, D. 2014. Compressive sensing of sparse tensors. *IEEE Transactions on Image Processing*, 23(10): 4438–4447.

Grotheer, R.; Li, S.; Ma, A.; Needell, D.; and Qin, J. 2021. Iterative hard thresholding for low CP-rank tensor models. *Linear and Multilinear Algebra*, 1–17.

Hillar, C. J.; and Lim, L.-H. 2013. Most tensor problems are NP-hard. *Journal of the ACM*, 60(6): 1–39.

Hou, J.; Zhang, F.; Qiu, H.; Wang, J.; Wang, Y.; and Meng, D. 2022. Robust low-tubal-rank tensor recovery from binary measurements. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8): 4355–4373.

Kernfeld, E.; Kilmer, M.; and Aeron, S. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.

Kernfeld, E.; Kilmer, M.; and Aeron, S. 2015. Tensortensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485(15): 545–570.

Kiers, H. A. L. 2000. Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3): 105–122.

Kilmer, M. E.; Horesh, L.; Avron, H.; and Newman, E. 2021. Tensor-tensor algebra for optimal representation and compression of multiway data. *Proceedings of the National Academy of Sciences*, 118(28): e2015851118.

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

Krahmer, F.; and Ward, R. 2011. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3): 1269–1281.

Li, C.; Yin, W.; Jiang, H.; and Zhang, Y. 2013. An efficient augmented Lagrangian method with applications to total variation minimization. *Computational Optimization and Applications*, 56(3): 507–530.

Li, Z.; Wang, Y.; Zhao, Q.; Zhang, S.; and Meng, D. 2022. A tensor-based online RPCA model for compressive background subtraction. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Lin, T.; Ma, S.; and Zhang, S. 2016. Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity. *Journal of Scientific Computing*, 69(1): 52–81.

Lucien, W., ed. 2002. *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Presses des MINES.

Madathil, B.; and N.George, S. 2018. Twist tensor total variation regularized-reweighted nuclear norm based tensor completion for video missing area recovery. *Information Sciences*, 423: 376–397.

Morton, P.; and Petersen, A. 1997. Image compression using the Haar wavelet transform. *Spelman Science and Mathematics Journal*, 1(1): 22–31.

Needell, D.; and Ward, R. 2013. Near-optimal compressed sensing guarantees for total variation minimization. *IEEE Transactions on Image Processing*, 22(10): 3941–3949.

Peng, J.; Wang, Y.; Zhang, H.; Wang, J.; and Meng, D. 2022. Exact decomposition of joint low rankness and local smoothness plus sparse matrices. arXiv:2201.12592v1.

Peng, J.; Xie, Q.; Zhao, Q.; Wang, Y.; Yee, L.; and Meng, D. 2020. Enhanced 3DTV regularization and its applications on HSI denoising and compressed sensing. *IEEE Transactions on Image Processing*, 29: 7889–7903.

Qin, W.; Wang, H.; Zhang, F.; Wang, J.; Luo, X.; and Huang, T. 2022. Low-rank high-order tensor completion with applications in visual data. *IEEE Transactions on Image Processing*, 31: 2433–2448.

Rauhut, H.; Schneider, R.; and Stojan, Z. 2017. Low rank tensor recovery via iterative hard thresholding. *Linear Algebra and its Applications*, 523(15): 220–262.

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

Rudelson, M.; and Vershynin, R. 2007. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8): 1025–1045.

Semerci, O.; Hao, N.; Kilmer, M. E.; and Miller, E. L. 2014. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Transactions on Image Processing*, 23(4): 1678–1693.

Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3): 279–311.

Vandergheynst, M. G. P. 2012. Joint trace/TV norm minimization: A new efficient approach for spectral compressive imaging. In *19th IEEE International Conference on Image Processing*, 13356653.

Wang, W.; Zhang, F.; and Wang, J. 2021. Low-rank matrix recovery via regularized nuclear norm minimization. *Applied and Computational Harmonic Analysis*, 54: 1–19.

Wang, Y.; Han, Y.; Wang, K.; and Zhao, X. 2022. Total variation regularized nonlocal low-rank tensor train for spectral compressive imaging. *Signal Processing*, 195: 108464.

Wang, Y.; Lin, L.; Zhao, Q.; Yue, T.; Meng, D.; and Leung, Y. 2017. Compressive sensing of hyperspectral images via joint tensor tucker decomposition and weighted total variation regularization. *IEEE Geoscience and Remote Sensing Letters*, 14(12): 2457–2461.

Wang, Z.; Bovik, A.; Sheikh, H.; and Simoncelli, E. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.

Xutao Li, X. X., Yunming Ye. 2017. Low-rank tensor completion with total variation for visual data inpainting. In *Thirty-First AAAI Conference on Artificial Intelligence*, 12031610. AAAI Press.

Zhang, F.; Wang, J.; Wang, W.; and Xu, C. 2021. Low-tubal-rank plus sparse tensor recovery with prior subspace information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10): 3492–3507.

Zhang, F.; Wang, W.; Huang, J.; Wang, J.; and Wang, Y. 2020. RIP-based performance guarantee for low-tubal-rank tensor recovery. *Journal of Computational and Applied Mathematics*, 374(15): 112767.