SCIENCE CHINA
Technological Sciences

• Article •

June 2022 Vol. 65 No. 6: 1300–1317
https://doi.org/10.1007/s11431-021-1976-2

Robust tensor decomposition via orientation invariant tubal nuclear norms

WANG AnDong[1,3], ZHAO QiBin[1,3], JIN Zhong[2,4], LI Chao[3] & ZHOU GuoXu[1,5*]

[1] *School of Automation, Guangdong University of Technology, Guangzhou 510006, China;*
[2] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;*
[3] *Tensor Learning Team, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan;*
[4] *Key Laboratory of Intelligent Perception and System for High-Dimensional Information, Ministry of Education, Nanjing 210094, China;*
[5] *Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, Ministry of Education, Guangzhou 510006, China*

Received August 4, 2021; accepted December 6, 2021; published online April 11, 2022

Aiming at recovering an unknown tensor (i.e., multi-way array) corrupted by both sparse outliers and dense noises, robust tensor decomposition (RTD) serves as a powerful pre-processing tool for subsequent tasks like classification and target detection in many computer vision and machine learning applications. Recently, tubal nuclear norm (TNN) based optimization is proposed with superior performance as compared with other tensorial nuclear norms for tensor recovery. However, one major limitation is its orientation sensitivity due to low-rankness strictly defined along tubal orientation and it cannot simultaneously model spectral low-rankness in multiple orientations. To this end, we introduce two new tensor norms called OITNN-O and OITNN-L to exploit multi-orientational spectral low-rankness for an arbitrary K-way ($K \geq 3$) tensors. We further formulate two RTD models via the proposed norms and develop two algorithms as the solutions. Theoretically, we establish non-asymptotic error bounds which can predict the scaling behavior of the estimation error. Experiments on real-world datasets demonstrate the superiority and effectiveness of the proposed norms.

tensor recovery, t-SVD, estimation error, tensor completion

Citation: Wang A D, Zhao Q B, Jin Z, et al. Robust tensor decomposition via orientation invariant tubal nuclear norms. Sci China Tech Sci, 2022, 65: 1300–1317, https://doi.org/10.1007/s11431-021-1976-2

1 Introduction

Tensor decomposition has become a paradigm in modern multi-way data analysis [1–4]. Due to various reasons like sensor failures, occlusion in videos, or abnormalities, the multi-way data are often corrupted by noises and gross corruptions [5–8]. For example, the embedded noises in hyper-spectral image are probably a mixture of small dense noises and sparse gross corruptions [9]. To tackle both small noises and gross corruptions, the robust tensor decomposition (RTD) [5] is studied to robustify traditional tensor decompositions like CANDECOMP/PARAFAC (CP) decomposition [10] and Tucker decomposition [11] which are sensitive to gross corruptions.

In many real-world applications, most variation of the multi-way data can be linearly dominated by a relatively small number of latent factors due to intrinsic correlations and redundancy. Thus, such data can be well approximated by a "low rank" tensor. Thanks to the multiple definitions of tensor rank function, such as CP rank [10], Tucker rank [11], tensor train rank [12], and tubal rank [13], multi-way data can be modeled with different types of low-rank structures.

To recover a low-rank tensor, one natural way is to solve the rank minimization problem (RMP) [14]. Unfortunately, RMP is NP-hard in general for matrices (2-way tensors) [15] and even harder for higher-way tensors [16]. In low-rank ma-

*Corresponding author (email: gx.zhou@gdut.edu.cn)

© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2022

trix estimation, matrix nuclear norm is proposed as the convex envelop of rank function [17] for tractable algorithms. Motivated by the great success of matrix nuclear norm, its tensor extensions have been extensively studied, like tensor trace norm [18], overlapped Schatten-1 norm (SNN) [14, 19], latent Schatten norm (LatentNN) [20], squared nuclear norm (SqNN) [21], and tubal nuclear norm (TNN) [22]. Among existing tensorial nuclear norms[1)], TNN is induced by the tensor singular value decomposition (t-SVD) [13] and has shown superior performance in various applications, such as image/video inpainting/de-noising [23–25].

In real multi-way data like images and videos, there is an ubiquitous "spatial-shifting" correlation making such data spatial-temporally smooth [1]. From a signal processing standpoint, smoothness in original domain often reflects the existence of some simple patterns in spectral domain [26]. TNN is quite suitable to capture such simple patterns since it exploits spectral low-rankness for 3-way tensors. However, by computing nuclear norms of frontal slices after 1D-DFT on the mode-3 fibers, it is strictly orientation sensitive and fails to capture the complex intra-mode and inter-mode correlations in multiple orientations for higher-way tensors. To improve the limited representation ability and flexibility of TNN in modeling multi-orientational correlations, we propose two orientation invariant tensor norms for K-way ($K \geq 3$) tensors and apply them to RTD. Main contributions of this paper are three-fold[2)].

(1) We propose two tensor norms via a novel 3d unfolding operation on K-way tensors, which are orientation invariant, and thus can be exploited for the multi-orientational spectral low-rankness.

(2) The new norms are employed to formulate RTD as two convex models, together with corresponding algorithms.

(3) Error bounds of the proposed models are analyzed and provided, which enables us to predict approximately the scaling behavior of the estimation error.

The remainder of this paper proceeds as follows. We first introduce some notation and preliminaries in Sect. 2. Then, the proposed norms are defined in Sect. 3. Two new models for the RTD based on the proposed norms are formulated in Sect. 4. The statistical performance of the proposed estimators is analyzed in Sect. 5, and we compute the estimators by using ADMM-based algorithms described in Sect. 6. Experimental results on both synthetic and real datasets are reported in Sect. 7. We summarize this paper and discuss future directions briefly in Sect. 8. The proofs of the theoretical results are given in the Appendix. The demo code in Matlab for this work can be found in https://qibinzhao.github.io.

2 Notations and preliminaries

Notations We use lowercase boldface, uppercase boldface, and calligraphy letters to denote vectors (e.g., $\mathbf{v}$), matrices (e.g., $\mathbf{M}$), and tensors (e.g., $\mathcal{T}$), respectively. Let $[n] := \{1, 2, \cdots, n\}$, $\forall n \in \mathbb{N}_+$. We use c, c', c_1 etc. to denote constants whose values can vary from line to line. Given a 3-way tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, let $\mathbf{T}^{(i)} := \mathcal{T}(:, :, i)$ denotes its i-th frontal slice. Without specification, a K-way tensor refers to a tensor of 3 or higher ways, i.e., $K \geq 3$. If the size of a tensor is not given explicitly, then it is in $\mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$. For notational simplicity, let $d_{K+1} = d_1$, $D = \prod_{k \in [K]} d_k$, $d_{\backslash k} = D/(d_k d_{k+1})$, $\tilde{d}_k = \sqrt{d_{k+1}}(\sqrt{d_k} + \sqrt{d_{\backslash k}})$, $\forall k \in [K]$.

For a matrix $\mathbf{M}$ with singular values σ_i's, define its nuclear norm and spectral norm respectively as follows:

$$\|\mathbf{M}\|_* := \sum\nolimits_i \sigma_i, \quad \|\mathbf{M}\| := \max\nolimits_i \sigma_i.$$

Given a tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, define its l_0-norm, l_1-norm, F-norm, and l_∞-norm respectively as follows:

$$\|\mathcal{T}\|_{l_0} := \|\mathrm{vec}(\mathcal{T})\|_0, \ \|\mathcal{T}\|_{l_0} := \|\mathrm{vec}(\mathcal{T})\|_0,$$
$$\|\mathcal{T}\|_{\mathrm{F}} := \|\mathrm{vec}(\mathcal{T})\|_2, \ \|\mathcal{T}\|_{l_\infty} := \|\mathrm{vec}(\mathcal{T})\|_\infty,$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization [28].

Other notations are introduced when they first appear.

2.1 Nuclear norms based on low Tucker rank structure

Originated from 1966 [11], the low Tucker rank structure has been widely studied [19, 29] and applied [14, 30, 31] in computer vision and machine learning. For a K-way tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, its Tucker rank is a K-dimensional vector whose k-th entry is the (matrix) rank of its mode-k unfolding $\mathbf{T}_{(k)}$:

$$\vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T}) := (\mathrm{rank}(\mathbf{T}_{(1)}), \cdots, \mathrm{rank}(\mathbf{T}_{(K)}))^\top \in \mathbb{R}^K, \tag{1}$$

where $\mathbf{T}_{(k)} \in \mathbb{R}^{d_k \times (D/d_k)}$ is the mode-k unfolding of $\mathcal{T}$ [28] obtained by concatenating all the mode-k fibers of $\mathcal{T}$ as column vectors. The Tucker rank is based on the matrix rank and thus can be computed efficiently using the (matrix) SVD.

By relaxing the matrix rank in eq. (1) to its convex envelop (within the unit spectral norm ball), i.e., the matrix nuclear norm, we obtain a convex relaxation of the Tucker rank, named sum of nuclear norms (SNN) [14], which is defined as follows:

$$\|\mathcal{T}\|_{\mathrm{snn}} := \sum_{k=1}^{K} \alpha_k \|\mathbf{T}_{(k)}\|_*, \tag{2}$$

1) In this paper, "tensorial nuclear norms" refer to tensor extensions of matrix nuclear norm instead of the nuclear norm induced by CP decomposition [18].

2) A short conference version of this work [27] has been presented in the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), New York, USA.

where α_k's are positive constants.

The SNN minimization encourages a low-Tucker-rank structure which requires the underlying tensor to be low-rank simultaneously in all K orientations, and has found many applications in tensor recovery [5, 14, 19, 32]. However, Tomioka et al. [20] pointed out that the low-Tucker-rank assumption may be too restricted for some tensor data that are not simultaneously low-rank along all orientations, and proposed a tensor norm (which is called latent nuclear norm (LatentNN) in this paper) to relax it as follows:

$$\|\mathcal{T}\|_{\text{latent}} := \inf_{\mathcal{T}=\sum_k \mathcal{X}^{(k)}} \sum_{k=1}^{K} \|\mathbf{X}_{(k)}^{(k)}\|_*. \tag{3}$$

The motivation of LatentNN is to model the underlying K-way tensor $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ as the sum of K component tensors $\mathcal{X}^{(1)}, \cdots, \mathcal{X}^{(K)}$, where the k-th component $\mathcal{X}^{(k)}$ is low-rank only along the k-th orientation (by minimizing the nuclear norm of the mode-k unfolding $\mathbf{X}_{(k)}^{(k)}$). For a lower generalization error in multi-task learning, a scaled version of LatentNN was proposed in ref. [33] as follows:

$$\|\mathcal{T}\|_{\text{s-latent}} := \inf_{\mathcal{T}=\sum_k \mathcal{X}^{(k)}} \sum_{k=1}^{K} \frac{1}{\sqrt{d_k}} \|\mathbf{X}_{(k)}^{(k)}\|_*. \tag{4}$$

As reported in refs. [20, 32], LatentNN is more suitable than SNN for tensor data that is not simultaneously low-rank along all orientations.

2.2 Tensor TNN

The tensor TNN is an extension of the matrix nuclear norm within the framework of t-SVD whose basic notions are given as follows.

Definition 1 (t-product [13]). Given $\mathcal{T}_1 \in \mathbb{R}^{d_1\times d_2\times d_3}$ and $\mathcal{T}_2 \in \mathbb{R}^{d_2\times d_4\times d_3}$, their t-product $\mathcal{T} = \mathcal{T}_1 * \mathcal{T}_2 \in \mathbb{R}^{d_1\times d_4\times d_3}$ is a tensor whose (i, j)-th tube is computed by

$$\mathcal{T}(i, j, :) = \sum_{k=1}^{d_2} \mathcal{T}_1(i, k, :) \bullet \mathcal{T}_2(k, j, :),$$

where $\bullet$ is the circular convolution between tubes (i.e., vectors).

Definition 2 (Tensor transpose [13]). Let $\mathcal{T}$ be a tensor of size $d_1 \times d_2 \times d_3$, then $\mathcal{T}^\top$ is the $d_2 \times d_1 \times d_3$ tensor obtained by transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through d_3.

Definition 3 (Identity tensor [13]). The identity tensor $\mathcal{I} \in \mathbb{R}^{d\times d\times d_3}$ is a tensor whose first frontal slice is the $d\times d$ identity matrix and all other frontal slices are zero.

Definition 4 (f-diagonal tensor [13]). A tensor is called f-diagonal if each frontal slice of the tensor is a diagonal matrix.

Definition 5 (Orthogonal tensor [13]). A tensor $\mathcal{Q} \in \mathbb{R}^{d\times d\times d_3}$ is orthogonal if $\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I}$.

The block diagonal matrix of 3-way tensors is further defined for the convenience of analysis.

Definition 6 (The block-diagonal matrix [13]). Let $\overline{\mathbf{T}}$ (or $\overline{\widetilde{\mathcal{T}}}$) denote the block-diagonal matrix of the tensor $\widetilde{\mathcal{T}}$ which is the Fourier version[3)] of $\mathcal{T}$, i.e.,

$$\overline{\mathbf{T}} := \begin{bmatrix} \widetilde{\mathbf{T}}^{(1)} & & \\ & \ddots & \\ & & \widetilde{\mathbf{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 d_3\times d_2 d_3}. \tag{5}$$

Then, t-SVD can be defined as follows (see Figure 1).

Definition 7 (t-SVD [13]). Any tensor $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$ has a tensor singular value decomposition (t-SVD) as

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \tag{6}$$

where $\mathcal{U} \in \mathbb{R}^{d_1\times d_1\times d_3}, \mathcal{V} \in \mathbb{R}^{d_2\times d_2\times d_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{d_1\times d_2\times d_3}$ is an f-diagonal tensor. The tubal rank of $\mathcal{T}$ is defined as the number of non-zero tubes of $\mathcal{S}$:

$$r_{\text{tb}}(\mathcal{T}) := \#\{i \,|\, \mathcal{S}(i, i, :) \neq \mathbf{0}\}. \tag{7}$$

Definition 8 (Tensor average rank, TNN [24]). Given $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times d_3}$, let $\widetilde{\mathcal{T}}$ be its Fourier version in $\mathbb{C}^{d_1\times d_2\times d_3}$. The tensor average rank $\text{rank}_{\text{avg}}(\cdot)$, TNN $\|\cdot\|_\star$ of $\mathcal{T}$ are defined as the averaged rank and nuclear norm of frontal slices of $\widetilde{\mathcal{T}}$:

$$\text{rank}_{\text{avg}}(\mathcal{T}) := \frac{1}{d_3}\sum_{i=1}^{d_3} \text{rank}(\widetilde{\mathbf{T}}^{(i)}), \quad \|\mathcal{T}\|_\star := \frac{1}{d_3}\sum_{i=1}^{d_3} \|\widetilde{\mathbf{T}}^{(i)}\|_*,$$

where tensor spectral norm $\|\cdot\|$ is the largest spectral norm:

$$\|\mathcal{T}\| := \max_{i\in[d_3]} \|\widetilde{\mathbf{T}}^{(i)}\|.$$

According to the definition of tensor average rank, it is equal to the rank of Fourier block diagonal matrix $\overline{\mathbf{T}}$ (averaged by the d_3 diagonal blocks), and thus it indeed measures

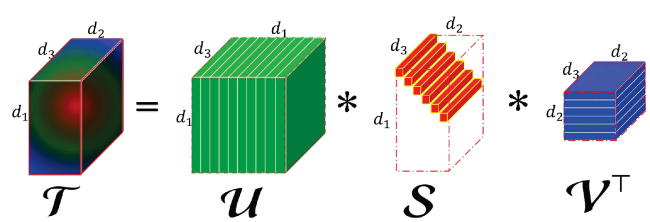

Figure 1 (Color online) An illustration of t-SVD.

3) The Fourier version $\widetilde{\mathcal{T}}$ is obtained by performing 1D-DFT on all tubes of $\mathcal{T}$, i.e., $\widetilde{\mathcal{T}} = \texttt{fft}(\mathcal{T}, [], 3) \in \mathbb{C}^{d_1\times d_2\times d_3}$ in Matlab.

low-rankness in the spectral domain defined by DFT on the mode-3 fibers. As proven in ref. [24], TNN is the convex envelop of tensor average rank in unit tensor spectral norm ball. Thus, TNN encourages a low average rank which means low-rankness in spectral domain according to Definition 8. It is strictly orientation sensitive in the sense that just mode-3 fibers are chosen to perform DFT, thus only spectral low-rankness along orientation of mode-3 can be exploited. Since TNN is orientation sensitive and defined for 3-way tensors, it has limited representation ability for higher-way tensors.

3 Orientation invariant TNNs

To overcome the orientation sensitivity of TNN, we propose two orientation invariant extensions of TNN to exploit the multi-orientational spectral low-rankness for general K-way tensors in this section. Specifically, our strategy consists of three steps.

(1) We define a new 3d-unfolding operation that can conveniently transform a K-way tensor to 3-way, such that the average rank (and TNN) can be directly utilized to model spectral low-rankness of the resulted 3-way tensor.

(2) Based on the 3d-unfolding, we propose the orientation invariant average rank (OIAR) as a suitable measure of multi-orientation spectral low-rankness for general K-way tensors.

(3) For tractable optimization, we relax the OIAR via convex proxy to define the OITNN-O, which is furthermore released to OITNN-L.

3.1 A new tensor unfolding operation

Before defining the new norms, we first propose a new tensor 3d-unfolding operation as follows.

Definition 9 (Mode-(k,t) 3d-unfolding)**.** For different integers $k,t \in [K]$, the mode-(k,t) 3d-unfolding of $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ is a 3-way tensor $\mathcal{T}_{[k,t]}$ of size $d_k\times(D/(d_kd_t))\times d_t$, obtained by the following two steps (see Figure 2).

- First, permute $\mathcal{T}$ to $\mathcal{Z} \in \mathbb{R}^{d'_1\times d'_2\times\cdots\times d'_K}$ whose 1st and Kth modes are respectively the kth and tth modes of $\mathcal{T}$, with the rest modes permuted circularly.
- Second, reshape $\mathcal{Z}$ to $\mathcal{T}_{[k,t]} \in \mathbb{R}^{d_k\times(Dd_k^{-1}d_t^{-1})\times d_t}$ obeying the equation as follows:

$$(\mathcal{T}_{[k,t]})_{i_1 j i_K} = \mathcal{Z}_{i_1 i_2\cdots i_K},$$

where $j = 1 + \sum_{l=2}^{K-1}(i_l - 1)J_l$ with $J_l = \sum_{m=2}^{l-1} d'_m$.

Intuitively, by viewing a K-way tensor $\mathcal{T}$ as a "(K-1)-way array $\mathscr{T}$" of size $d_1 \times d_2 \times \cdots \times d_{t-1} \times d_{t+1} \times \cdots d_K$ whose entries are mode-t tubes, the mode-(k,t) 3d-unfolding $\mathcal{T}_{[k,t]}$ can also be analogously viewed as a "mode-k unfolding" of $\mathscr{T}$ with size $d_k \times (Dd_k^{-1}d_t^{-1})$ whose entries are mode-t tubes.

4) Using circular order of modes, let $d_{K+1} = d_{(K+1)\mathrm{mod}K} = d_1$.

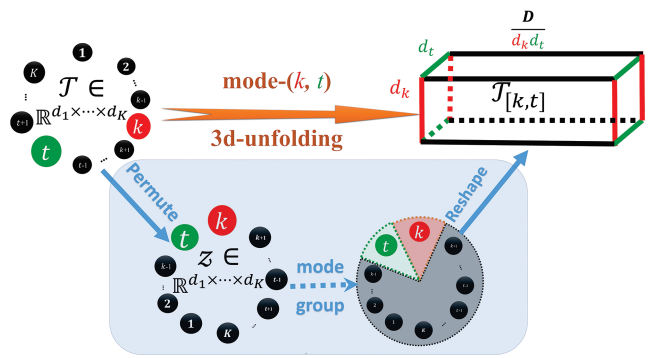

Figure 2 (Color online) Illustration of 3d-unfolding.

Generally, the mode t of this 3d-unfolding can be any mode except k. In the sequel, we simply set $t = k + 1$, such that mode t traverses all the K orientations when k slides from 1 to K, by which some orientation invariant measures can be defined. For the ease of presentation, we simply let

$$\mathcal{T}_{[k]} := \mathcal{T}_{[k,k+1]},$$

and name it the mode-k 3d-unfolding[4)] of $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$.

3.2 Multi-orientational spectral low-rankness

3.2.1 *The low OIAR structure*

Based on the proposed 3d-unfolding operation, two orientation invariant tensor ranks are defined.

Definition 10 (OITR and OIAR)**.** For any tensor $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$, its orientation invariant tubal rank (OITR) $\vec{\mathrm{r}}_{\mathrm{t}}$ and orientation invariant average rank (OIAR) $\vec{\mathrm{r}}_{\mathrm{a}}$ are defined as the K-dimensional vectors whose k-th entries are respectively the tubal rank and average rank of the mode-k 3d-unfolding $\mathcal{T}_{[k]}$ as follows:

$$\begin{aligned}\vec{\mathrm{r}}_{\mathrm{t}}(\mathcal{T}) &:=(r_{\mathrm{tb}}(\mathcal{T}_{[1]}),\cdots,r_{\mathrm{tb}}(\mathcal{T}_{[K]}))^{\top} \in \mathbb{R}^K,\\ \vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) &:=(\mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[1]}),\cdots,\mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[K]}))^{\top} \in \mathbb{R}^K.\end{aligned} \tag{8}$$

As shown in eq. (8), OITR serves as a complexity measure in the original domain in all orientations, whereas the OIAR measures multi-orientational low-rankness in the spectral domain. They have the following relationship with the classical Tucker rank $\vec{\mathrm{r}}_{\mathrm{Tucker}}$.

Lemma 1. It holds for any tensor $\mathcal{T} \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ that

$$\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) \le \min\{\vec{\mathrm{r}}_{\mathrm{t}}(\mathcal{T}), \vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T})\}, \tag{9}$$

where the partial order "$\le$" is defined entry-wisely.

Proof. Given any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, let $\mathcal{K} = \mathcal{T}_{[k]} \in \mathbb{R}^{d_k \times (D d_k^{-1} d_{k+1}^{-1}) \times d_{k+1}}$ denote its mode-k 3d-unfolding ($k \in [K]$).

We first show that $\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) \leq \vec{\mathrm{r}}_{\mathrm{t}}(\mathcal{T})$. Indeed, it holds that

$$(\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}))_k = \mathrm{rank}_{\mathrm{avg}}(\mathcal{K}) \overset{(i)}{\leq} r_{\mathrm{tb}}(\mathcal{K}) = (\vec{\mathrm{r}}_{\mathrm{t}}(\mathcal{T}))_k, \tag{10}$$

where inequality (i) holds due to the property of DFT which indicates that the tubal rank of $\mathcal{T}$ defined in eq. (7) is lower bounded by the average rank:

$$\begin{aligned} r_{\mathrm{tb}}(\mathcal{T}) &:= \#\{i \,|\, \mathcal{S}(i,i,:) \neq \mathbf{0}\} \\ &= \#\{i \,|\, \widetilde{\mathcal{S}}(i,i,:) \neq \mathbf{0}\} \\ &= \max_{l \in [d_3]} \mathrm{rank}(\mathbf{S}^{(l)}) \\ &\geq \frac{1}{d_3} \sum_{l=1}^{d_3} \mathrm{rank}(\mathbf{S}^{(l)}) \\ &\geq \mathrm{rank}_{\mathrm{avg}}(\mathcal{S}). \end{aligned} \tag{11}$$

Then, we show $\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) \leq \vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T})$. On the one hand, according to ref. [24], we have

$$(\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}))_k = \mathrm{rank}_{\mathrm{avg}}(\mathcal{K}) \leq \mathrm{rank}(\mathbf{K}_{(1)}), \tag{12}$$

where $\mathbf{K}_{(1)}$ is the mode-1 unfolding of $\mathcal{K}$. On the other hand, since $\mathcal{K}$ is the mode-k 3d-unfolding of $\mathcal{T}$, it holds that

$$\mathrm{rank}(\mathbf{K}_{(1)}) = \mathrm{rank}(\mathbf{T}_{(k)}) = (\vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T}))_k, \tag{13}$$

where matrix $\mathbf{T}_{(k)}$ is the mode-k unfolding of $\mathcal{T}$. Thus, we have $\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) \leq \vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T})$. Putting things together, we obtain $\vec{\mathrm{r}}_{\mathrm{a}}(\mathcal{T}) \leq \min\{\vec{\mathrm{r}}_{\mathrm{t}}(\mathcal{T}), \vec{\mathrm{r}}_{\mathrm{Tucker}}(\mathcal{T})\}$. ■

3.2.2 *Ubiquity of the low OIAR structure*

Lemma 1 indicates that low OITR or Tucker rank results in low OIAR. Thus, the low OIAR assumption is weaker than the popular low Tucker rank assumption. As the low-Tucker-rankness is an intrinsic low-dimensional structure of many visual data [14], it is also interesting to ask: is the low-OIAR structure ubiquitous for real tensor data?

Next, we will show that the low-OIAR structure is a ubiquitous property of many typical visual data like 3-way color images (exampled by Figure 3(a)) and 4-way color videos (exampled by the YUV video akiyo[5)] in Figure 3(b)).

As discussed in Sect. 2, the tensor average rank indeed measures low-rankness of the rank of Fourier block diagonal matrix $\overline{\mathbf{M}}$ of a 3-way tensor $\mathcal{M} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. Thus, according to Definition 10, OIAR indeed measures low-rankness of the rank of Fourier block diagonal matrix $\overline{\mathbf{T}_{[k]}}$ of the mode-k 3d-unfolding $\mathcal{T}_{[k]}$ of a given tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ for all $k \in [K]$. Therefore, if all the Fourier block diagonal matrices $\overline{\mathbf{T}_{[k]}}$ are (approximately) low-rank for all $k \in [K]$, then we can say $\mathcal{T}$ has a low OIAR structure.

For the 3-way color image $\mathcal{T}$ in Figure 3(a), we plot the singular values of $\overline{\mathbf{T}_{[k]}}$ ($k = 1, 2, 3$) in Figure 4. It can be found that for all $k = 1, 2, 3$, the singular values of $\overline{\mathbf{T}_{[k]}}$ decay dramatically, indicating nice low-rankness of $\overline{\mathbf{T}_{[k]}}$. Thus, the 3-way color image $\mathcal{T}$ in Figure 3(a) has low OIAR structure.

For the 4-way akiyo video tensor $\mathcal{T}$ illustrated in Figure 3(b), we also plot the singular values of $\overline{\mathbf{T}_{[k]}}$ ($k = 1, 2, 3, 4$) in Figure 5. We can see that for all $k = 1, 2, 3, 4$, the singular values of $\overline{\mathbf{T}_{[k]}}$ also decay dramatically, indicating the low-rankness of $\overline{\mathbf{T}_{[k]}}$ and thus the low-OIAR-ness of the 4-way akiyo video.

3.3 The proposed OITNNs

By relaxing average rank to its convex envelop in each orientation, we naturally define the following norm as a convex relaxation of OIAR.

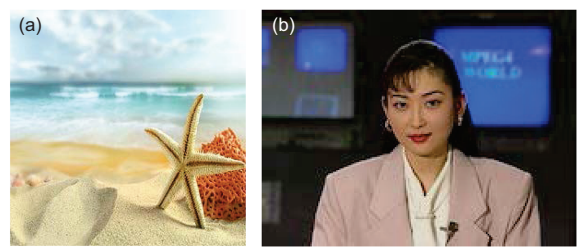

Figure 3 (Color online) Color image and videos as example tensors with low-OIAR-structure. (a) A 3-way color images of size $256 \times 256 \times 3$; (b) one frame of the 4-way akiyo video of size $142 \times 176 \times 3 \times 300$.

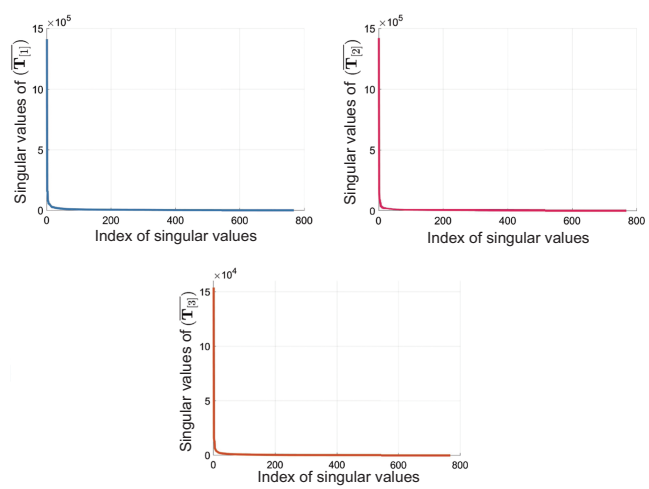

Figure 4 (Color online) Plots of singular values of the Fourier block diagonal matrices $\overline{\mathbf{T}_{[k]}}$ of the mode-k 3d-unfolding $\mathcal{T}_{[k]}$ ($k = 1, 2, 3$) of the 3-way color image tensor $\mathcal{T}$ in Figure 3(a).

5) The video is available at https://media.xiph.org/video/derf/y4m/akiyo_qcif.y4m.

Definition 11 (OITNN-O)**.** The overlapped orientation invariant tubal nuclear norm (OITNN-O) of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is defined as

$$\|\mathcal{T}\|_{\star o} := \sum_{k=1}^{K} w_k \|\mathcal{T}_{[k]}\|_{\star}, \tag{14}$$

where w_k's are positive weights satisfying $\sum_k w_k = 1$.

OITNN-O encourages a low OIAR structure, which means low-rankness in spectral domain of all orientations. Thus in the original domain, it models a data tensor as simultaneously low tubal rank in all orientations (see Figure 6(a)).

Although the assumption of low OIAR is weaker than low Tucker rank, it may still be strict for some real data tensors. In ref. [20], it is pointed out that LatentNN induced by a mixture model is more suitable than SNN for tensors only low rank in certain modes. Motivated by this, we define the latent OITNN to relax the low OIAR assumption.

Definition 12 (OITNN-L)**.** The latent orientation invariant tubal nuclear norm (OITNN-L) of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is defined as

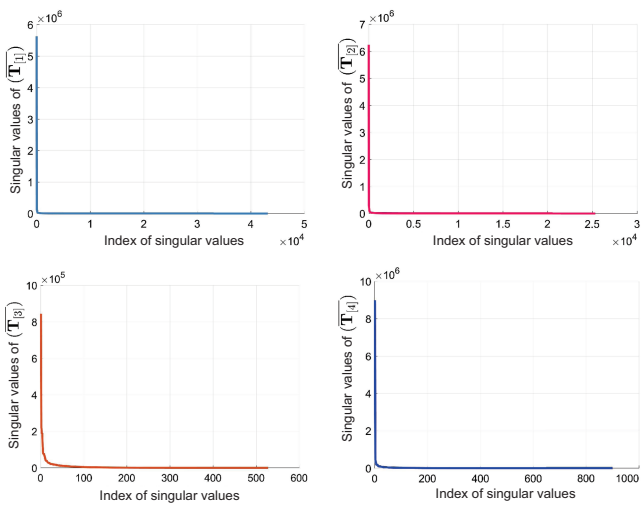

Figure 5 (Color online) Plots of singular values of the Fourier block diagonal matrices $\overline{\mathbf{T}_{[k]}}$ of the mode-k 3d-unfolding $\mathcal{T}_{[k]}$ $(k = 1, 2, 3, 4)$ of the 4-way akiyo video tensor $\mathcal{T}$ in Figure 3(b).

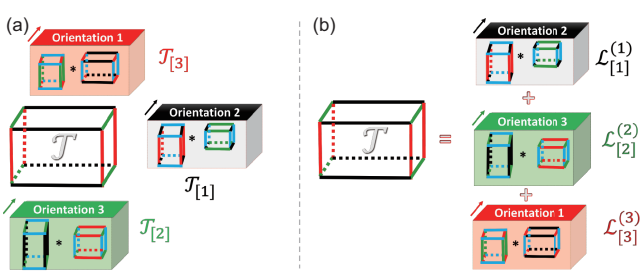

Figure 6 (Color online) Illustration of two OITNNs for 3D tensors $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. (a) OITNN-O encourages simultaneously low tubal rank structure in all orientations; (b) OITNN-L models $\mathcal{T}$ as a mixture of three low tubal rank tensors $\{\mathcal{L}^{(k)}\}$.

$$\|\mathcal{T}\|_{\star l} := \inf_{\sum_k \mathcal{L}^{(k)} = \mathcal{T}} \sum_{k=1}^{K} v_k \|\mathcal{L}^{(k)}_{[k]}\|_{\star}, \tag{15}$$

where v_k's are non-negative weights satisfying $\sum_k v_k = 1$, and $\mathcal{L}^{(k)}_{[k]}$ is the mode-k 3d-unfolding of latent component $\mathcal{L}^{(k)}$, $\forall k \in [K]$.

OITNN-L seeks K latent components $\{\mathcal{L}^{(k)}\}$ to minimize a weighted sum of their TNNs in each orientation. Thus, it models $\mathcal{T}$ as a mixture of K low tubal rank tensors in original domain (see Figure 6(b)). According to Definitions 11 and 12, both OITNN-O and OITNN-L can exploit spectral low-rankness in all orientations and are invariant to circular permutations. Since TNN has been shown to be more powerful than the matrix nuclear norm [22, 24], we expect that OITNN-O and OITNN-L outperform SNN and LatentNN in some applications, respectively. This expectation will be verified by experiments on real datasets in the experiment section.

We now give the dual norms of OITNN-O and OITNN-L.

Lemma 2. The dual norms of $\|\cdot\|_{\star o}$ and $\|\cdot\|_{\star l}$, denoted by $\|\cdot\|^*_{\star o}$ and $\|\cdot\|^*_{\star l}$ respectively, can be computed as follows:

$$\begin{aligned} \|\mathcal{T}\|^*_{\star o} &= \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \max_k \left\{ w_k^{-1} \|\mathcal{X}^{(k)}_{[k]}\| \right\}, \\ \|\mathcal{T}\|^*_{\star l} &= \max_{k \in [K]} \left\{ v_k^{-1} \|\mathcal{T}_{[k]}\| \right\}. \end{aligned} \tag{16}$$

Proof. According to the definition of dual norm [34], the dual norms of $\|\cdot\|_{\star o}$ and $\|\cdot\|_{\star l}$ can be formulated as the following maximization problems:

$$\|\mathcal{T}\|^*_{\star o} := \sup_{\mathcal{M}} \langle \mathcal{M}, \mathcal{T} \rangle, \text{ s.t. } \|\mathcal{M}\|_{\star o} \le 1, \tag{17}$$

and

$$\|\mathcal{T}\|^*_{\star l} := \sup_{\mathcal{M}} \langle \mathcal{M}, \mathcal{T} \rangle, \text{ s.t. } \|\mathcal{M}\|_{\star l} \le 1. \tag{18}$$

They are constrained maximization problems. We prove the first part. Since Problem (17) satisfies Slatter's condition, the strong duality holds: the supremum of Problem (17) is equal to the infimum of its dual problem. Thus, by comparing the definition in eq. (17) and the target formulation in the first line of eq. (16), we only need to show that the dual problem of Problem (17) agrees with

$$\inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \max_k \{ w_k^{-1} \|\mathcal{X}^{(k)}_{[k]}\| \}. \tag{19}$$

The Lagrangian of Problem (17) is

$$L(\mathcal{M}, \lambda) = \langle \mathcal{T}, \mathcal{M} \rangle - \lambda(\|\mathcal{M}\|_{\star o} - 1),$$

where $\lambda \geq 0$ is the Lagrangian multiplier. As strong duality holds, we solve the dual problem:

$$\begin{aligned}
&\inf_{\lambda} \sup_{\mathcal{M}} L(\mathcal{M}, \lambda) \\
&= \inf_{\lambda} \Big(\sup_{\mathcal{M}} \langle \mathcal{T}, \mathcal{M} \rangle - \lambda \|\mathcal{M}\|_{\star\mathrm{o}} \Big) + \lambda \\
&= \inf_{\lambda} \Big(\sup_{\mathcal{M}} \langle \mathcal{T}, \mathcal{M} \rangle - \lambda \sum_k w_k \|\mathcal{M}_{[k]}^{(k)}\|_{\star} \Big) + \lambda \\
&\overset{(i)}{=} \inf_{\lambda} \lambda + \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \sum\nolimits_k \delta(\lambda^{-1} w_k^{-1} \|\mathcal{X}_{[k]}^{(k)}\| \leq 1) \\
&= \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \inf_{\lambda} \Big(\lambda + \delta\Big(\lambda \geq \max_k w_k^{-1} \|\mathcal{X}_{[k]}^{(k)}\|\Big)\Big) \\
&= \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \max_k w_k^{-1} \|\mathcal{X}_{[k]}^{(k)}\|,
\end{aligned} \tag{20}$$

where $\delta(C)$ is the indicator of condition C (0 if C is true and $+\infty$ otherwise). Equality (i) in eq. (20) holds due to the proposition "the conjugate of the sum equals the infimal convolution of the conjugates" (see, e.g., Theorem 16.4 in ref. [34]), the fact that "the conjugate of a norm equals the indicator of its unit dual norm ball", and the proposition "the dual norm of TNN is the tensor spectral norm" (see, e.g., ref. [24]). In this way, the first part is proved. The second part can be proved similarly, and is thus omitted. ■

The dual norms play key roles in the statistical analysis of OITNN-based RTD models.

3.4 Connections and differences of OITNNs with existing works

In this subsection, we briefly discuss the connections and differences of the proposed OITNNs with tightly related tensorial nuclear norms including TNN [23], SNN [14], LatentNN [20, 33], Triple TNN [35], and Tensor N-TNN [36].

- TNN: As the motivation is to alleviate the orientation sensitivity of TNN, the proposed OITNNs can be seen as orientation invariant extensions of the vanilla TNN.
- SNN: Both the proposed OITNN-O and SNN are designed to model the tensor that is simultaneously structured: OITNN-O imposes simultaneous spectral low-rankness in all orientations via mode-k 3d unfoldings, whereas SNN encourages low-rankness in the original domain in all orientations via mode-k unfoldings.
- LatentNN: Both the proposed OITNN-L and LatentNN are proposed to relax the simultaneously structured restriction respectively imposed by OITNN-O and SNN. The difference lies in that OITNN-L models low-rankness in the spectral domain, whereas LatentNN works in the original domain.
- Triple TNN: When $K = 3$, one of the proposed norms, namely OITNN-O, is exactly the triple TNN [35] obtained by first letting TNN work on mode-1, mode-2, and mode-3 fibers, and then averaging their TNNs with weights, which effectively abates the orientation sensitivity for 3-way tensors. Since OITNN-O degenerates to Triple TNN for 3-way tensors, Triple TNN can be seen as a special case of the proposed OITNN-O. When compared with Triple TNN, the main innovations of the proposed OITNNs lie in that (i) thanks to the newly defined mode-k 3d-unfolding, OITNNs can be applied to the much broader K-way tensors rather than 3-way tensors on which Triple TNN is defined, and (ii) the proposed OITNN-O is further released to OITNN-L for relaxation the low OIAR assumptions which may be too strict for some real tensor data.
- Tensor N-TNN: Recently, the N-TNN is defined via a new reshaping operator named tensor-k_1k_2 unfolding to generalize TNN to higher-way tensors [36]. The motivation of it is similar to the proposed OITNN-O, and both of them degenerate to Triple TNN for 3-way tensors. The main differences are two-fold. First, the permutation of resulted 3-way tensors are different: for examples, the mode-12 unfolding in ref. [36] reshapes a $d_1 \times d_2 \times d_3 \times d_4$ tensor to the one of size $d_1 \times d_2 \times (d_3 d_4)$, whereas our proposed mode-1 3d-unfolding reshapes it to $d_1 \times (d_3 d_4) \times d_2$. Second, the low-rank patterns modeled by N-TNN and the proposed OITNN-O are different: for general K-way tensors, N-TNN considers $K(K-1)/2$ TNNs by enumerating all mode-k_1k_2 unfoldings, whereas the proposed OITNN-O uses K TNNs by considering the K orientations.

4 Robust tensor decomposition via OITNNs

4.1 Observation model

Suppose we observe a K-way tensor $\mathcal{Y} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ generated by the following model:

$$\mathcal{Y} = \mathcal{L}^* + \mathcal{S}^* + \mathcal{E}, \tag{21}$$

where $\mathcal{L}^*$ is the low-rank signal tensor, $\mathcal{S}^*$ stores entry-wisely sparse outliers, and tensor $\mathcal{E}$ represents dense small noises. See Figure 7 for an illustration for 3-way tensors.

The goal of RTD is to reconstruct the low-rank $\mathcal{L}^*$ and sparse $\mathcal{S}^*$ from the noise observation $\mathcal{Y}$. If the noises $\mathcal{E} = \mathbf{0}$, RTD degenerates to the tensor robust PCA [24]; If the outlier tensor $\mathcal{S}^* = \mathbf{0}$, then RTD becomes the noisy tensor decomposition [20].

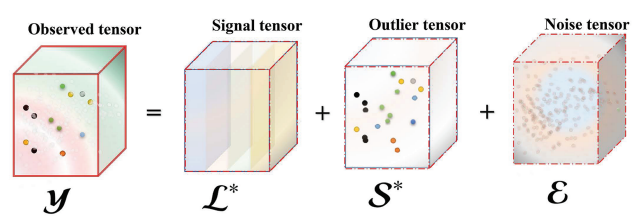

Figure 7 (Color online) The observation model of RTD.

4.2 Incoherence conditions

First, to guarantee separability of low-rank $\mathcal{L}^*$ and sparse $\mathcal{S}^*$, we suppose $\mathcal{L}^*$ satisfies the non-spiky condition with parameter α [37]:

$$\|\mathcal{L}^*\|_{l_\infty} \le \alpha. \tag{22}$$

Second, let $\{\mathcal{L}^{(k)*}\}$ be any latent components obtained while computing $\|\mathcal{L}^*\|_{\star\iota}$ in eq. (15). Then, the signal $\mathcal{L}^*$ can be written as

$$\mathcal{L}^* = \sum\nolimits_{k=1}^{K} \mathcal{L}^{(k)*}. \tag{23}$$

For separability of latent tensors $\mathcal{L}^{(k)*}$'s, an incoherence condition with parameter β is further assumed to hold:

$$\|\mathcal{L}^{(k)*}_{[l]}\| \le \beta\tilde{d}_l,\ \forall l \ne k \in [K], \tag{24}$$

where $\tilde{d}_l = \sqrt{d_{l+1}}(\sqrt{d_l} + \sqrt{d_{\backslash l}})$. The motivation of eq. (24) is to force each latent component $\mathcal{L}^{(k)*}$ to be low tubal rank only in mode-k 3d-unfolding, and behave like a Gaussian random tensor[6)] in any mode-l 3d-unfolding ($l \ne k$).

4.3 Proposed RTD models

Using the proposed tensor norms and the l_1-norm to encourage multi-orientational spectral low-rankness and sparsity respectively, we propose the following two models for RTD.

Model I: RTD-OITNN-O

$$\begin{aligned} &\min_{\mathcal{L},\mathcal{S}}\ l(\mathcal{L},\mathcal{S}) + \lambda_{\rm o}\|\mathcal{L}\|_{\star\rm o} + \mu_{\rm o}\|\mathcal{S}\|_{l_1}, \\ &\text{s.t.}\quad \|\mathcal{L}\|_{l_\infty} \le \alpha. \end{aligned} \tag{25}$$

Model II: RTD-OITNN-L

$$\begin{aligned} &\min_{\{\mathcal{L}^{(k)}\},\mathcal{S}}\ l\Big(\sum\nolimits_k \mathcal{L}^{(k)},\mathcal{S}\Big) + \lambda_\iota \sum\nolimits_k v_k\|\mathcal{L}^{(k)}_{[k]}\|_\star + \mu_\iota\|\mathcal{S}\|_{l_1}, \\ &\text{s.t.}\quad \|\sum\nolimits_k \mathcal{L}^{(k)}\|_{l_\infty} \le \alpha,\ \|\mathcal{L}^{(l)}_{[k]}\| \le \beta\tilde{d}_k,\ \forall l \ne k, \end{aligned} \tag{26}$$

where $\lambda_{\rm o}, \mu_{\rm o}, \lambda_\iota, \mu_\iota$ denote regularization parameters, square loss $l(\mathcal{L},\mathcal{S}) = \|\mathcal{Y} - \mathcal{L} - \mathcal{S}\|_{\rm F}^2/2$ is the data fitting term. Model I explicitly uses OITNN-O as the regularizer of $\mathcal{L}$, whereas Model II implicitly adopts OITNN-L with incoherent latent components $\{\mathcal{L}^{(k)}\}$.

5 Statistical performance

We analyze statistical performance of the proposed models. Let $(\hat{\mathcal{L}}_{\rm o}, \hat{\mathcal{S}}_{\rm o})$ and $(\{\hat{\mathcal{L}}^{(k)}\}, \hat{\mathcal{S}}_\iota)$ be any solution to Problems (25) and (26), respectively. We establish both deterministic and non-asymptotic bounds on the estimation errors, i.e., $\mathfrak{E}^{\rm o}, \mathfrak{E}^\iota, \mathfrak{E}^\iota_{\rm com}$ (defined in Table 1 due to space limitation), of the low-rank component $\mathcal{L}^*$ and the sparse component $\mathcal{S}^*$ in their sum[7)].

5.1 Deterministic bounds

When $\mathcal{E}$ in the observation model (21) represents any (deterministic or random) noise, we bound the estimation error in the following theorems where the dual norms in Lemma 2 are used.

Theorem 1. If $\lambda_{\rm o} \ge 2\|\mathcal{E}\|^*_{\star\rm o}$ and $\mu_{\rm o} \ge 2(\|\mathcal{E}\|_{l_\infty} + 2\alpha)$ in Problem (25), then any solution $(\hat{\mathcal{L}}_{\rm o}, \hat{\mathcal{S}}_{\rm o})$ satisfies

$$\mathfrak{E}^{\rm o} \le c_1\lambda_{\rm o}^2\Big(\sum\nolimits_k w_k\sqrt{\underline{r}^{\rm o}_k}\Big)^2 + c_2\mu_{\rm o}^2 s.$$

Theorem 1 indicates that once parameters $(\lambda_{\rm o}, \mu_{\rm o})$ exceed certain quantities of the noise $\mathcal{E}$, estimation error of Model I can be upper bounded linear by the OITR of $\mathcal{L}^*$ and the sparsity of $\mathcal{S}^*$.

Theorem 2. If $\lambda_\iota \ge 2\max_k\{\|\mathcal{E}\|^*_{\star\iota} + v_k^{-1}(K-1)\beta\tilde{d}_k\}$ and $\mu_\iota \ge 2(\|\mathcal{E}\|_{l_\infty} + 2\alpha)$ in Problem (26), then it holds that

$$\begin{aligned} &\mathfrak{E}^\iota_{\rm com} \le c_3\lambda_\iota^2\sum\nolimits_k v_k^2\overline{r}^\iota_k + c_4\mu_\iota^2 s, \\ &\mathfrak{E}^\iota \le c_3\lambda_\iota^2\min_k v_k^2\underline{r}^{\rm o}_k + c_4\mu_\iota^2 s. \end{aligned}$$

Theorem 2 shows that when $(\lambda_\iota, \mu_\iota)$ exceed some thresholds in terms of $\mathcal{E}$, estimation error $\mathfrak{E}^\iota_{\rm com}$ involving the latent components $\{\mathcal{L}^{(k)}\}$ is upper bounded by the "latent tubal ranks $\overline{r}^\iota_k$" of $\mathcal{L}^*$ and the sparsity of $\mathcal{S}^*$, whereas the error $\mathfrak{E}^\iota$ for $(\mathcal{L}^*, \mathcal{S}^*)$ is bounded by the "minimal" OITR of $\mathcal{L}^*$ and the sparsity of $\mathcal{S}^*$.

5.2 Non-asymptotic bounds

For a typical setting where the noise tensor $\mathcal{E}$ represents the tensor of *i.i.d.* $N(0,\sigma^2)$ entries, we have the following two Theorems.

Table 1 List of some notations

Error of $(\mathcal{L}^*, \mathcal{S}^*)$ by Model I	$\mathfrak{E}^{\rm o} = \|\hat{\mathcal{L}}_{\rm o} - \mathcal{L}^*\|_{\rm F}^2 + \|\hat{\mathcal{S}}_{\rm o} - \mathcal{S}^*\|_{\rm F}^2$
OITR of true tensor $\mathcal{L}^*$	$\vec{\mathfrak{r}}^{\rm o} = (\underline{r}^{\rm o}_1, \cdots, \underline{r}^{\rm o}_K),\ \underline{r}^{\rm o}_k = r_{\rm tb}(\mathcal{L}^*_{[k]})$
Error of $(\mathcal{L}^*, \mathcal{S}^*)$ by Model II	$\mathfrak{E}^\iota = \|\sum_k \hat{\mathcal{L}}^{(k)} - \mathcal{L}^*\|_{\rm F}^2 + \|\hat{\mathcal{S}} - \mathcal{S}^*\|_{\rm F}^2$
Error of $(\{\mathcal{L}^{(k)*}\}, \mathcal{S}^*)$	$\mathfrak{E}^\iota_{\rm com} = \sum_k \|\hat{\mathcal{L}}^{(k)} - \mathcal{L}^{(k)*}\|_{\rm F}^2 + \|\hat{\mathcal{S}}_\iota - \mathcal{S}^*\|_{\rm F}^2$
Tubal rank of component $\mathcal{L}^{(k)*}$	$\vec{\mathfrak{r}}^\iota = (\overline{r}^\iota_1, \cdots, \overline{r}^\iota_k),\ \overline{r}^\iota_k = r_{\rm tb}(\mathcal{L}^{(k)*}_{[k]})$
Sparsity of corruption $\mathcal{S}^*$	$s = \|\mathcal{S}^*\|_{l_0}$

6) Note that a random $d_1 \times d_2 \times d_3$ tensor with *i.i.d.* standard Gaussian entries has full tubal rank with high probability and its tensor spectral norm scales as $O\Big(\sqrt{d_3}(\sqrt{d_1} + \sqrt{d_2})\Big)$ (see Lemma A.3).

7) Separated bounds on $\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\rm F}^2$ and $\|\hat{\mathcal{S}} - \mathcal{S}^*\|_{\rm F}^2$ can also be obtained via more strict assumptions like Assumption 1 in ref. [5]. However, it requires two additional unknown separation parameters κ_1 and κ_2. Thus, only the summed error is considered in this paper.

Theorem 3. If parameters $\lambda_{\rm o} = 2\sigma K^{-2}\sum_k(\tilde{d}_k/w_k)$ and $\mu_{\rm o} = 8\sigma\sqrt{\log D} + 16\alpha$ in Problem (25), then with high probability it holds that

$$\mathfrak{E}^{\rm o} \le \frac{c_1\sigma^2}{K^4}\Big(\sum_k \frac{\tilde{d}_k}{w_k}\Big)^2\Big(\sum_k w_k\sqrt{\underline{r}_k^{\rm o}}\Big)^2 + c_2(\sigma^2\log D + \alpha^2)s.$$

Theorem 4. If parameters $\lambda_\iota = c\sigma\max_k\{\tilde{d}_k/v_k\}$ and $\mu_\iota = 8\sigma\sqrt{\log D} + 16K\alpha$ in Problem (26), then with high probability it holds that

$$\mathfrak{E}^\iota_{\rm com} \le c_5\sigma^2\Big(\max_k\Big\{\frac{\tilde{d}_k}{v_k}\Big\}\Big)^2\sum_k v_k^2\overline{r}_k^\iota + c_6(\sigma^2\log D + \alpha^2)s,$$

$$\mathfrak{E}^\iota \le c_5\sigma^2\Big(\max_k\Big\{\frac{\tilde{d}_k}{v_k}\Big\}\Big)^2\min_k v_k^2\underline{r}_k^{\rm o} + c_6(\sigma^2\log D + \alpha^2)s.$$

To understand Theorems 3 and 4 intuitively, we have the following remark whose correctness is verified in the experiment section.

Remark 1. Given a K-way cubical tensor $\mathcal{L}^* \in \mathbb{R}^{d\times d\times\cdots\times d}$, suppose its OITR is $(\underline{r}_1^{\rm o},\cdots,\underline{r}_K^{\rm o})$. Letting parameters $w_k = v_k = 1/K, \forall k\in[K]$, then we have the following bounds on the entry-wise estimation error with high probability:

$$\frac{\mathfrak{E}^{\rm o}}{D} \lesssim \sigma^2(\mathfrak{r}_{\rm o} + \mathfrak{s}\log D), \text{and } \frac{\mathfrak{E}^\iota}{D} \lesssim \sigma^2(\mathfrak{r}_\iota + \mathfrak{s}\log D), \tag{27}$$

where $\mathfrak{r}_{\rm o} = (K^{-1}\sum_k\sqrt{\underline{r}_k^{\rm o}/d})^2$ and $\mathfrak{r}_\iota = \min_k\{\overline{r}_k^\iota/d\}$ act as the "averaged" and "minimal" OITR complexities of the signal $\mathcal{L}^*$, respectively, and $\mathfrak{s} = s/D$ is the sparse ratio of the corruption $\mathcal{S}^*$.

As discussed in Remark 1, OITNN-L tends to find the orientation with lowest tubal rank, whereas OITNN-O considers the tubal rank in all orientations. The upper bounds in Theorems 1–4 are consistent with the intuition that RTD for complex $\mathcal{L}^*$ (with higher OITR) and denser $\mathcal{S}^*$ is harder. We give the following remark on the identifiability issue in our analysis.

Remark 2. In the noiseless setting (i.e., $\mathcal{E} = \mathbf{0}$), the proposed error bounds do not vanish, which means our analysis cannot guarantee exact recovery. This is because the incoherence conditions in our analysis are less strict than the ones defined in terms of singular vectors [24, 38] which intrinsically ensure separability between low-rank and sparse components.

6 Optimization algorithms

We develop two algorithms (Algorithms 1 and 2) to solve the proposed Model I and Model II, respectively. For simplicity, we also define the 3d-unfolding operator $\mathfrak{F}_k(\cdot)$ for any $\mathcal{T}\in\mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ and its inverse operator $\mathfrak{F}_k^{-1}(\cdot)$ as

$$\mathfrak{F}_k(\mathcal{T}) := \mathcal{T}_{[k]}, \quad \text{and} \quad \mathfrak{F}_k^{-1}(\mathcal{T}_{[k]}) = \mathcal{T}.$$

Before giving solutions to the sub-problems in Algorithms 1 and 2, we briefly give the proximal operator of TNN as follows.

Lemma 3 (Proximal operator of TNN [39])**.** Let tensor $\mathcal{T}\in\mathbb{R}^{d_1\times d_2\times d_3}$ with t-SVD $\mathcal{T} = \mathcal{U}*\mathcal{S}*\mathcal{V}^\top$, where $\mathcal{U}\in\mathbb{R}^{d_1\times r\times d_3}$, and $\mathcal{V}\in\mathbb{R}^{d_2\times r\times d_3}$ are orthogonal tensors, and $\mathcal{S}\in\mathbb{R}^{r\times r\times d_3}$ is the f-diagonal tensor of singular tubes. Then the proximal operator of function $\tau\|\cdot\|_\star$ at point $\mathcal{T}_0$, denoted by $\mathrm{Prox}_\tau^{\|\cdot\|_\star}(\mathcal{T}_0)$, can be computed as follows:

$$\begin{aligned}\mathrm{Prox}_\tau^{\|\cdot\|_\star}(\mathcal{T}_0) &= \underset{\mathcal{T}}{\mathrm{argmin}}\ \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\rm F}^2 + \tau\|\mathcal{T}\|_\star \\ &= \mathcal{U}*\mathrm{idft}_3(\max(\mathrm{dft}_3(\mathcal{S}) - \tau, 0))*\mathcal{V}^\top,\end{aligned} \tag{28}$$

where $\mathrm{dft}_3(\cdot)$ and $\mathrm{idft}_3(\cdot)$ denote the operators of performing DFT and inverse DFT on all mode-3 fibers of a 3-way tensor, respectively.

The proximal operator of l_1-norm $\|\cdot\|_{l_1}$ is given as

$$\begin{aligned}\mathrm{Prox}_\tau^{\|\cdot\|_{l_1}}(\mathcal{T}_0) &= \underset{\mathcal{T}}{\mathrm{argmin}}\ \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\rm F}^2 + \tau\|\mathcal{T}\|_{l_1} \\ &= \mathrm{sgn}(\mathcal{T}_0)\circledast\max(|\mathcal{T}_0| - \tau, 0),\end{aligned} \tag{29}$$

where $\mathrm{sgn}(\cdot)$ exacts the sign of a tensor element-wisely with $\mathrm{sgn}(0) = 0$, and $\circledast$ denotes the element-wise tensor product.

The proximal operator of indicator function of l_∞-norm ball $\delta_\alpha^{l_\infty}(\cdot)$ is a projector:

$$\begin{aligned}\mathrm{Proj}_\alpha^{\|\cdot\|_{l_\infty}}(\mathcal{T}_0) &= \underset{\mathcal{T}}{\mathrm{argmin}}\ \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_{\rm F}^2 + \delta_\alpha^{l_\infty}(\mathcal{T}_0) \\ &= \mathrm{sgn}(\mathcal{T}_0)\circledast\min(|\mathcal{T}_0|, \alpha).\end{aligned} \tag{30}$$

6.1 Solutions to sub-problems in Algorithm 1

In this subsection, we derive solutions to sub-problems in Algorithm 1.

First, adding auxiliary variables $\mathcal{T}, \mathcal{K}$ and $\{\mathcal{K}_k\}_k$ to Problem (25), we get

$$\begin{aligned}\min_{\substack{\mathcal{L},\mathcal{S},\mathcal{T},\\ \mathcal{K},\{\mathcal{K}_k\}_k}}\ & \frac{1}{2}\|\mathcal{Y} - \mathcal{L} - \mathcal{S}\|_{\rm F}^2 + \lambda_{\rm o}\sum\nolimits_k w_k\|\mathcal{K}_k\|_\star \\ & + \mu_{\rm o}\|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K}), \\ \text{s.t. }\ & \mathcal{K}_k = \mathfrak{F}_k(\mathcal{L}), \forall k; \mathcal{T} = \mathcal{S}; \mathcal{K} = \mathcal{L}.\end{aligned} \tag{31}$$

Then, the augmented Lagrangian is given as follows:

$$\begin{aligned}&L_\rho^1(\mathcal{L},\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{K}_k\}_k,\{\mathcal{Y}_k\}_k,\mathcal{Z},\mathcal{W}) \\ =&\frac{1}{2}\|\mathcal{Y} - \mathcal{L} - \mathcal{S}\|_{\rm F}^2 + \lambda_{\rm o}\sum\nolimits_k w_k\|\mathcal{K}_k\|_\star + \mu_{\rm o}\|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K}) \\ &+ \sum_k(\langle\mathcal{Y}_k, \mathcal{K}_k - \mathfrak{F}_k(\mathcal{L})\rangle + \frac{\rho}{2}\|\mathcal{K}_k - \mathfrak{F}_k(\mathcal{L})\|_{\rm F}) \\ &+ \langle\mathcal{Z}, \mathcal{T} - \mathcal{S}\rangle + \frac{\rho}{2}\|\mathcal{T} - \mathcal{S}\|_{\rm F}^2 + \langle\mathcal{W}, \mathcal{K} - \mathcal{L}\rangle + \frac{\rho}{2}\|\mathcal{K} - \mathcal{L}\|_{\rm F}^2.\end{aligned}$$

Algorithm 1 ADMM for Model I

Require: Observation $\mathcal{Y}$, parameters $\lambda_o, \mu_o, \{w_k\}_k, \rho > 0, \epsilon > 0$.

1: Initialize $\mathcal{L}^0 = \mathcal{K}^0 = \mathcal{W}^0 = \mathcal{S}^0 = \mathcal{T}^0 = \mathcal{Z}^0 = \mathbf{0}, \mathcal{K}_k^0 = \mathcal{Y}_k^0 = \mathbf{0}, \forall k$.

2: **while** not converged **do**

3: Update $(\mathcal{L}^{t+1}, \mathcal{S}^{t+1})$ simultaneously by

$$\min_{\mathcal{L},\mathcal{S}} l(\mathcal{L},\mathcal{S}) + \sum_k \frac{\rho}{2}\left\|\mathcal{L} - \mathfrak{F}_k^{-1}(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho})\right\|_F^2 + \frac{\rho}{2}\left\|\mathcal{S} - \left(\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}\right)\right\|_F^2 + \frac{\rho}{2}\left\|\mathcal{L} - \left(\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}\right)\right\|_F^2;$$

4: Update $\{\mathcal{K}_k^{t+1}\}_k, \mathcal{T}^{t+1}$ and $\mathcal{K}^{t+1}$ simultaneously by

$$\min_{\mathcal{K}_k} \lambda_o w_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\left\|\mathcal{K}_k - \mathfrak{F}_k(\mathcal{L}^{t+1}) + \frac{\mathcal{Y}_k^t}{\rho}\right\|_F^2, \quad \min_{\mathcal{T}} \mu_o \|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\left\|\mathcal{T} - \left(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}\right)\right\|_F^2, \quad \min_{\mathcal{K}} \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\left\|\mathcal{K} - \left(\mathcal{L}^{t+1} - \frac{\mathcal{W}^t}{\rho}\right)\right\|_F^2;$$

5: Dual update: $\mathcal{Z}^{k+1} = \mathcal{Z}^t + \rho(\mathcal{T}^{t+1} - \mathcal{S}^{t+1})$, $\mathcal{W}^{k+1} = \mathcal{W}^t + \rho(\mathcal{K}^{t+1} - \mathcal{L}^{t+1})$ and $\mathcal{Y}_k^{t+1} = \mathcal{Y}_k^t + \rho(\mathcal{K}_k^{t+1} - \mathfrak{F}_k(\mathcal{L}^{t+1}))$, $\forall k \in [K]$;

6: Check the convergence conditions:
$\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_{l_\infty} \le \epsilon$, $\forall \mathcal{X} \in \{\mathcal{L}, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}\}$; $\|\mathcal{T}^{t+1} - \mathcal{S}^{t+1}\|_{l_\infty} \le \epsilon$ $\|\mathcal{K}^{t+1} - \mathcal{L}^{t+1}\|_{l_\infty} \le \epsilon$; $\|\mathcal{K}_k^{t+1} - \mathfrak{F}_k(\mathcal{L}^{t+1})\|_{l_\infty} \le \epsilon, \forall k \in [K]$;

7: $t = t + 1$.

8: **end while**

Algorithm 2 ADMM for Model II

Require: Observation $\mathcal{Y}$, parameters $\lambda_t, \mu_t, \{v_k\}_k, \rho > 0, \epsilon > 0$.

1: Initialize $\mathcal{S}^0 = \mathcal{T}^0 = \mathcal{Z}^0 = \mathcal{K}^0 = \mathcal{W}^0 = \mathbf{0}, (\mathcal{L}^{(k)})^0 = \mathcal{K}_k^0 = \mathcal{Y}_k^0 = \mathbf{0}, \forall k$.

2: **while** not converged **do**

3: Update $\{(\mathcal{L}^{(k)})^{t+1}\}_k$ and $\mathcal{S}^{t+1}$ simultaneously by

$$\min_{\{\mathcal{L}^{(k)}\}_k,\mathcal{S}} l\left(\sum_k \mathcal{L}^{(k)}, \mathcal{S}\right) + \sum_k \frac{\rho}{2}\left\|\mathcal{L}^{(k)} - \mathfrak{F}_k^{-1}(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho})\right\|_F^2 + \frac{\rho}{2}\left\|\mathcal{S} - \left(\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}\right)\right\|_F^2 + \frac{\rho}{2}\left\|\sum_k \mathcal{L}^{(k)} - \left(\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}\right)\right\|_F^2;$$

4: Update $\{\mathcal{K}_k^{t+1}\}_k, \mathcal{T}^{t+1}$ and $\mathcal{K}^{t+1}$ simultaneously by

$$\min_{\mathcal{K}_k} \lambda_t v_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\left\|\mathcal{K}_k - \mathfrak{F}_k((\mathcal{L}^{(k)})^{t+1}) + \frac{\mathcal{Y}_k^t}{\rho}\right\|_F^2, \quad \min_{\mathcal{T}} \mu_t \|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\left\|\mathcal{T} - \left(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}\right)\right\|_F^2, \quad \min_{\mathcal{K}} \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\left\|\mathcal{K} - \sum_k (\mathcal{L}^{(k)})^{t+1} + \frac{\mathcal{W}^t}{\rho}\right\|_F^2;$$

5: Dual update: $\mathcal{Z}^{t+1} = \mathcal{Z}^t + \rho(\mathcal{T}^{t+1} - \mathcal{S}^{t+1})$, $\mathcal{W}^{t+1} = \mathcal{W}^t + \rho(\mathcal{K}^{t+1} - \sum_k (\mathcal{L}^{(k)})^{t+1})$ and $\mathcal{Y}_k^{t+1} = \mathcal{Y}_k^t + \rho(\mathcal{K}_k^{t+1} - \mathfrak{F}_k((\mathcal{L}^{(k)})^{t+1}))$, $\forall k \in [K]$;

6: Check the convergence conditions:
$\|\mathcal{X}^{t+1} - \mathcal{X}^t\|_{l_\infty} \le \epsilon$, $\forall \mathcal{X} \in \{\{\mathcal{L}^{(k)}\}_k, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}\}$; $\|\mathcal{T}^{t+1} - \mathcal{S}^{t+1}\|_{l_\infty} \le \epsilon$; $\|\mathcal{K}^{t+1} - \sum_k (\mathcal{L}^{(k)})^{t+1}\|_{l_\infty} \le \epsilon$; $\|\mathcal{K}_k^{t+1} - \mathfrak{F}_k((\mathcal{L}^{(k)})^{t+1})\|_{l_\infty} \le \epsilon, \forall k \in [K]$;

7: $t = t + 1$.

8: **end while**

Further, we update blocks $(\mathcal{L}, \mathcal{S})$ and $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$ alternatively by fixing the other variables.

Update $(\mathcal{L}, \mathcal{S})$ Fixing $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$, we update $(\mathcal{L}, \mathcal{S})$ by minimizing the augmented Lagrangian $L_r^{\mathrm{I}} ho$ with respect to $(\mathcal{L}, \mathcal{S})$, which can be simplified as follows:

$$\min_{\mathcal{L},\mathcal{S}} l(\mathcal{L},\mathcal{S}) + \sum_k \frac{\rho}{2}\left\|\mathcal{L} - \left(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho}\right)k\right\|_F^2 + \frac{\rho}{2}\left\|\mathcal{S} - \left(\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}\right)\right\|_F^2 + \frac{\rho}{2}\left\|\mathcal{L} - \left(\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}\right)\right\|_F^2. \tag{32}$$

Taking the derivatives with respect to $\mathcal{L}$ and $\mathcal{S}$ and setting the derivatives to zero, we obtain

$$(K\rho + \rho + 1)\mathcal{L} + \mathcal{S} = \rho\tilde{\mathcal{K}} + \rho\sum_k \tilde{\mathcal{K}}_k + \mathcal{Y} \tag{33}$$

and

$$\mathcal{L} + (1 + \rho)\mathcal{S} = \mathcal{Y} + \mu\tilde{\mathcal{T}}, \tag{34}$$

where

$$\tilde{\mathcal{K}} = \mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}, \tilde{\mathcal{K}}_k = \mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho}, \text{ and } \tilde{\mathcal{T}} = \mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}.$$

By solving matrix equation group, we get the closed-form solution of $\mathcal{L}^{t+1}$ and $\mathcal{S}^{t+1}$:

$$\mathcal{L}^{t+1} = \frac{(1+\rho)\tilde{\mathcal{K}} + (1+\rho)\sum_k \tilde{\mathcal{K}}_k + \mathcal{Y} - \mathcal{T}}{(K+1)(\rho+1)+1},$$

$$\mathcal{S}^{t+1} = \frac{(K+1)\mathcal{Y} + (K\rho+\rho+1)\tilde{\mathcal{T}} - \tilde{\mathcal{K}} - \sum_k \tilde{\mathcal{T}}_k}{(K+1)(\rho+1)+1}.$$

Update $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$ Fixing $(\mathcal{L}, \mathcal{S})$, we update $\{\mathcal{K}_k\}_k$, $\mathcal{T}$, and $\mathcal{K}$ by minimizing the augmented Lagrangian L_ρ^{I} with respect to $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$. The problem can be solved separately as follows:

$$\begin{aligned}\mathcal{K}_k^{t+1} &= \operatorname*{argmin}_{\mathcal{K}_k} \lambda_o w_k \|\mathcal{K}_k\|_\star + \frac{\rho}{2}\left\|\mathcal{K}_k - \mathfrak{F}_k(\mathcal{L}^{t+1}) + \frac{\mathcal{Y}_k^t}{\rho}\right\|_F^2 \\ &= \mathrm{Prox}_{\lambda_o w_k/\rho}^{\|\cdot\|_\star}(\mathfrak{F}_k(\mathcal{L}^{t+1}) - \frac{\mathcal{Y}_k^t}{\rho}),\end{aligned}$$

$$\begin{aligned}\mathcal{T}^{t+1} &= \underset{\mathcal{T}}{\operatorname{argmin}}\ \ \mu_{\mathrm{o}}\|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\left\|\mathcal{T} - \left(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}\right)\right\|_{\mathrm{F}}^2 \\ &= \operatorname{Prox}_{\mu_{\mathrm{o}}/\rho}^{\|\cdot\|_{l_1}}(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}),\end{aligned}$$

$$\begin{aligned}\mathcal{K}^{t+1}\ \underset{\mathcal{K}}{\operatorname{argmin}}\ \ \delta_{\alpha}^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\left\|\mathcal{K} - \left(\mathcal{L}^{t+1} - \frac{\mathcal{W}^t}{\rho}\right)\right\|_{\mathrm{F}}^2 \\ = \operatorname{Proj}_{\alpha}^{\|\cdot\|_{l_\infty}}(\mathcal{L}^{t+1} - \frac{\mathcal{W}^t}{\rho}).\end{aligned}$$

6.2 Solutions to sub-problems in Algorithm 2

We solve the sub-problems in Algorithm 2 as follows. First, adding auxiliary variables $\mathcal{K}, \mathcal{T}$ and $\{\mathcal{K}_k\}_k$ to Problem (26) yields

$$\begin{aligned}&\min_{\substack{\{\mathcal{L}^{(k)}\}_k, \mathcal{S},\\ \{\mathcal{K}_k\}_k, \mathcal{T}, \mathcal{K}}} l\left(\sum\nolimits_k \mathcal{L}^{(k)}, \mathcal{S}\right) + \lambda_{\mathrm{o}} \sum\nolimits_k v_k\|\mathcal{K}_k\|_\star + \mu_\iota\|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K}), \\ &\text{s.t. } \mathcal{K}_k = \mathfrak{F}_k(\mathcal{L}^{(k)}), \forall k; \mathcal{T} = \mathcal{S}; \mathcal{K} = \sum\nolimits_k \mathcal{L}^{(k)}. \end{aligned} \tag{35}$$

Then, the augmented Lagrangian is given as follows:

$$\begin{aligned}&L_\rho^{\mathrm{II}}(\{\mathcal{L}^{(k)}\}_k, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}_k, \{\mathcal{Y}_k\}_k, \mathcal{Z}, \mathcal{W}) \\ &= \frac{1}{2}\|\mathcal{Y} - \sum_k \mathcal{L}^{(k)} - \mathcal{S}\|_{\mathrm{F}}^2 + \lambda_{\mathrm{o}} \sum\nolimits_k w_k\|\mathcal{K}_k\|_\star + \mu_\iota\|\mathcal{T}\|_{l_1} \\ &+ \sum_k \left(\langle \mathcal{Y}_k, \mathcal{K}_k - \mathfrak{F}_k(\mathcal{L}^{(k)})\rangle + \frac{\rho}{2}\|\mathcal{K}_k - \mathfrak{F}_k(\mathcal{L}^{(k)})\|_{\mathrm{F}}\right) \\ &+ \delta_\alpha^{l_\infty}(\mathcal{K}) + \langle \mathcal{Z}, \mathcal{T} - \mathcal{S}\rangle + \frac{\rho}{2}\|\mathcal{T} - \mathcal{S}\|_{\mathrm{F}}^2 \\ &+ \langle \mathcal{W}, \mathcal{K} - \mathcal{L}^{(k)}\rangle + \frac{\rho}{2}\left\|\mathcal{K} - \sum_k \mathcal{L}^{(k)}\right\|_{\mathrm{F}}^2.\end{aligned}$$

Further, we update blocks $(\{(\mathcal{L}^{(k)})\}, \mathcal{S})$ and $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$ alternatively by fixing the other variables.

Update $(\{(\mathcal{L}^{(k)})\}, \mathcal{S})$ Fixing $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$, we update $(\{(\mathcal{L}^{(k)})\}, \mathcal{S})$ by minimizing the augmented Lagrangian $L_r^{\mathrm{II}} ho$ with respect to $(\mathcal{L}, \mathcal{S})$, which can be simplified to the following problem:

$$\begin{aligned}&\min_{\{\mathcal{L}^{(k)}\}_k, \mathcal{S}} l\left(\sum\nolimits_k \mathcal{L}^{(k)}, \mathcal{S}\right) + \sum\nolimits_k \frac{\rho}{2}\left\|\mathcal{L}^{(k)} - \left(\mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho}\right)k\right\|_{\mathrm{F}}^2 \\ &+ \frac{\rho}{2}\left\|\mathcal{S} - \left(\mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}\right)\right\|_{\mathrm{F}}^2 + \frac{\rho}{2}\left\|\sum_k \mathcal{L}^{(k)} - \left(\mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}\right)\right\|_{\mathrm{F}}^2.\end{aligned}$$

Taking the derivatives with respect to $\mathcal{L}^{(k)}$ and $\mathcal{S}$ and setting the derivatives to zero, we obtain

$$\sum_k \mathcal{L}^{(k)} + \mathcal{S} - \mathcal{Y} + \rho\mathcal{L}^{(k)} - \rho\tilde{\mathcal{K}}_k + \rho\sum_k \mathcal{L}^{(k)} - \rho\tilde{\mathcal{K}} = 0 \tag{36}$$

and

$$\sum_k \mathcal{L}^{(k)} + \mathcal{S} - \mathcal{Y} + \rho\mathcal{S} - \mu\tilde{\mathcal{T}} = 0, \tag{37}$$

where

$$\tilde{\mathcal{K}} = \mathcal{K}^t + \frac{\mathcal{W}^t}{\rho}, \tilde{\mathcal{K}}_k = \mathcal{K}_k^t + \frac{\mathcal{Y}^t}{\rho}, \text{ and } \tilde{\mathcal{T}} = \mathcal{T}^t + \frac{\mathcal{Z}^t}{\rho}.$$

By solving matrix equation group, we get the closed-form solution of $\mathcal{L}^{t+1}$:

$$(\mathcal{L}^{(k)})^{t+1} = \rho^{-1}(\rho\tilde{\mathcal{K}} + \sum_k \tilde{\mathcal{K}} + \mathcal{Y} - (1+\rho)\mathcal{M} - \mathcal{S}^{t+1})$$

with

$$\mathcal{S}^{t+1} = \frac{(1+K)\mathcal{Y} + (K+\rho+K\rho)\tilde{\mathcal{T}} - K\tilde{\mathcal{K}} - \sum_k \tilde{\mathcal{K}}_k}{(1+K)(1+\rho)+K},$$

where

$$\mathcal{M} = \frac{K(1+\rho)\tilde{\mathcal{K}} + (1+\rho)\sum_k \tilde{\mathcal{K}}_k + K\mathcal{Y} - K\tilde{\mathcal{T}}}{(1+K)(1+\rho)+K}.$$

Update $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$ Fixing $(\{\mathcal{L}^{(k)}\}, \mathcal{S})$, we update $\{\mathcal{K}_k\}_k, \mathcal{T}$, and $\mathcal{K}$ by minimizing the augmented Lagrangian L_ρ^{II} with respect to $(\{\mathcal{K}_k\}, \mathcal{T}, \mathcal{K})$. The problem can be solved separately as follows:

$$\begin{aligned}\mathcal{K}_k^{t+1} &= \min_{\mathcal{K}_k}\ \ \lambda_\iota v_k\|\mathcal{K}_k\|_\star + \frac{\rho}{2}\left\|\mathcal{K}_k - \mathfrak{F}_k((\mathcal{L}^{(k)})^{t+1}) + \frac{\mathcal{Y}_k^t}{\rho}\right\|_{\mathrm{F}}^2 \\ &= \operatorname{Prox}_{\lambda_\iota v_k/\rho}^{\|\cdot\|_\star}(\mathfrak{F}_k((\mathcal{L}^{(k)})^{t+1}) - \frac{\mathcal{Y}_k^t}{\rho}),\end{aligned}$$

$$\begin{aligned}\mathcal{T}^{t+1} &= \underset{\mathcal{T}}{\operatorname{argmin}}\ \ \mu_\iota\|\mathcal{T}\|_{l_1} + \frac{\rho}{2}\left\|\mathcal{T} - \left(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}\right)\right\|_{\mathrm{F}}^2 \\ &= \operatorname{Prox}_{\mu_\iota/\rho}^{\|\cdot\|_{l_1}}(\mathcal{S}^{t+1} - \frac{\mathcal{Z}^t}{\rho}),\end{aligned}$$

$$\begin{aligned}\mathcal{K}^{t+1} &= \underset{\mathcal{K}}{\operatorname{argmin}}\ \delta_\alpha^{l_\infty}(\mathcal{K}) + \frac{\rho}{2}\left\|\mathcal{K} - \sum_k (\mathcal{L}^{(k)})^{t+1} + \frac{\mathcal{W}^t}{\rho}\right\|_{\mathrm{F}}^2 \\ &= \operatorname{Proj}_\alpha^{\|\cdot\|_{l_\infty}}(\sum_k (\mathcal{L}^{(k)})^{t+1} - \frac{\mathcal{W}^t}{\rho}).\end{aligned}$$

6.3 Computational complexity

6.3.1 One iteration computational complexity

In each single iteration of the proposed algorithms, the main time cost comes from updating the low tubal rank components by the proximal operator of TNN which involves FFT, IFFT, and n_3 SVDs of $n_1 \times n_2$ matrices for tensors of size $n_1 \times n_2 \times n_3$. Hence for tensors of size $d_1 \times d_2 \times \cdots \times d_K$ in the proposed models, both Algorithms 1 and 2 have per-iteration computational complexity

$$O\Big(KD\log D + D\sum\nolimits_{k=1}^{K}\min(d_k, d_k^{-1}d_{k+1}^{-1}D)\Big). \tag{38}$$

Note that when $K = 3$, the cost in eq. (38) is significantly higher than vanilla TNN [6] due to the need of computing TNNs in all orientations. The complexity in eq. (38) is of the same order as ADMM-based algorithms for SNN [5] and LatentNN [20]. We also observed agreements with eq. (38) in the results of running time comparison in Sect. 7.2: the proposed OITNNs (especially OITNN-L) are much more costly than TNN, and comparable to (or slightly more time-consuming than) SNN and LatentNN.

6.3.2 *Convergence analysis*

According to the analysis in ref. [40], the convergence rate of general ADMM-based algorithms is $O(1/T)$, where T denotes the iteration number. We analyze the convergence behaviors of Algorithms 1 and 2 in Theorems 5 and 6, respectively.

Theorem 5 (Convergence behavior of Algorithm 1)**.** For any positive constant ρ, if the unaugmented Lagrangian function $L_0^{\mathrm{I}}(\mathcal{L}, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}_k, \{\mathcal{Y}_k\}_k, \mathcal{Z}, \mathcal{W})$ has a saddle point, then the iterations $(\mathcal{L}^t, \mathcal{S}^t, \mathcal{T}^t, \mathcal{K}^t, \{\mathcal{K}_k^t\}_k, \{\mathcal{Y}_k^t\}_k, \mathcal{Z}^t, \mathcal{W}^t)$ in Algorithm 1 satisfy the residual convergence, objective convergence and dual variable convergence (defined in ref. [41]) of Problem (31) as $t \to \infty$.

Theorem 6 (Convergence behavior of Algorithm 2)**.** For any positive constant ρ, if the unaugmented Lagrangian function $L_0^{\mathrm{II}}(\{\mathcal{L}^{(k)}\}_k, \mathcal{S}, \mathcal{T}, \mathcal{K}, \{\mathcal{K}_k\}_k, \{\mathcal{Y}_k\}_k, \mathcal{Z}, \mathcal{W})$ has a saddle point, then the iterate $(\{\mathcal{L}^{(k),t}\}_k, \mathcal{S}^t, \mathcal{T}^t, \mathcal{K}^t, \{\mathcal{K}_k^t\}_k, \{\mathcal{Y}_k^t\}_k, \mathcal{Z}^t, \mathcal{W}^t)$ in Algorithm 2 satisfy the residual convergence, objective convergence, and dual variable convergence (defined in ref. [41]) of Problem (35) as $t \to \infty$.

Since both Theorems 5 and 6 can be proven similar to Theorem 4 in ref. [6] by straightforwardly reformulating Problems (25) and (26) to the standard form of the two-block ADMM framework in eq. (3.1) of ref. [41], we simply omit the proof.

7 Experiments

In this section, we first verify the correctness of the proposed error bounds through experiments on synthetic data, and then evaluate the effectiveness of the proposed norms on real datasets. The code is written in Matlab language, and all experiments are performed on a Windows 10 laptop with AMD Ryzen 3.0GHz CPU and 12GB RAM.

7.1 Correctness of the proposed error bounds

To validate the correctness of Theorems 3 and 4, we conduct simulations to check whether the proposed upper bounds in eq. (27) can predict the scaling behavior of the estimation error.

Generation of $\mathcal{L}^*$. We generate the low-rank tensor $\mathcal{L}^* \in \mathbb{R}^{d_1\times d_2\times\cdots\times d_K}$ in the following manner. Given K random integers $p_k < d_k, \forall k \in [K]$, we first generate a standard Gaussian tensor (i.e., tensors with *i.i.d.* $N(0,1)$ entries) $\mathcal{G}_0 \in \mathbb{R}^{p_1\times\cdots\times p_K}$. Then, we repeat the recursive operation $\mathcal{G}_k = \mathfrak{F}_k^{-1}(\mathcal{U}_k * \mathfrak{F}_k(\mathcal{G}_{k-1}))$, $\forall k \in [K-1]$, where $\mathcal{U}_k \in \mathbb{R}^{d_k\times p_k\times p_{k+1}}$ are also standard Gaussian tensors and $\mathcal{G}_k \in \mathbb{R}^{n_1\times\cdots\times n_k\times p_{k+1}\times\cdots\times p_K}$. We further generate $\mathcal{G}_K = \mathfrak{F}_K^{-1}(\mathcal{U}_K * \mathfrak{F}_K(\mathcal{G}_{K-1}))$ with standard Gaussian tensor $\mathcal{U}_K \in \mathbb{R}^{d_K\times p_K\times d_1}$. Finally, we let $\mathcal{L}^* = \mathcal{G}_K/\|\mathcal{G}_K\|_{l_\infty}$.

We form the sparse corruption tensor $\mathcal{S}^*$ by choosing its support uniformly at random according to ref. [24]. We generate the noise tensor $\mathcal{E}$ with entries drawing *i.i.d.* from $\mathcal{N}(0, \sigma^2)$ with $\sigma = c\|\mathcal{L}^*\|_{\mathrm{F}}/\sqrt{D}$ to keep a constant signal noise ratio. For simplicity, we consider cubical tensors, i.e., $d_1 = \cdots = d_K = d$. We test tensors of size $40\times 40\times 40$. We randomly choose $p_k \in \{2, 3, \cdots, 10\}$ to generate $\mathcal{L}^*$. We generate the corruption tensor $\mathcal{S}^*$ with sparsity $s = \mathfrak{s}D$ where $\mathfrak{s} \in \{0.02 : 0.02 : 0.3\}$ and form the noise tensor $\mathcal{E}$ with noise level $c = 0.1$. We run the proposed Algorithms 1 and 2 and then compute the estimation errors $\mathfrak{E}^{\mathrm{o}}/D$ and $\mathfrak{E}^{\iota}/D$ for 500 random choices of p_k's. We computed the OITR $(\underline{r}_1^{\mathrm{o}}, \cdots, \underline{r}_K^{\mathrm{o}})$ of $\mathcal{L}^*$, since it is not equal to $(p_1, \cdots, p_K)$ in general. We then compute $\mathfrak{r}_{\mathrm{o}}$ and $\mathfrak{r}_\iota$ in eq. (27). We will check whether the errors $\mathfrak{E}^{\mathrm{o}}/D$ and $\mathfrak{E}^{\iota}/D$ scale like $a_1\mathfrak{r}_{\mathrm{o}} + b_1\mathfrak{s}$ and $a_2\mathfrak{r}_\iota + b_2\mathfrak{s}$, respectively, with some constants a_1, a_2, b_1, b_2.

Figure 8 shows the results of $\mathfrak{E}^{\mathrm{o}}/D$ versus $\mathfrak{r}_{\mathrm{o}}$ and $\mathfrak{s}$, and $\mathfrak{E}^{\iota}/D$ versus $\mathfrak{r}_\iota$ and $\mathfrak{s}$ by keeping other variables fixed. From Figure 8, we can see that the errors $\mathfrak{E}^{\mathrm{o}}/D$ and $\mathfrak{E}^{\iota}/D$ have approximately linear scaling behavior with respect to $\mathfrak{r}_{\mathrm{o}}$ and $\mathfrak{s}$,

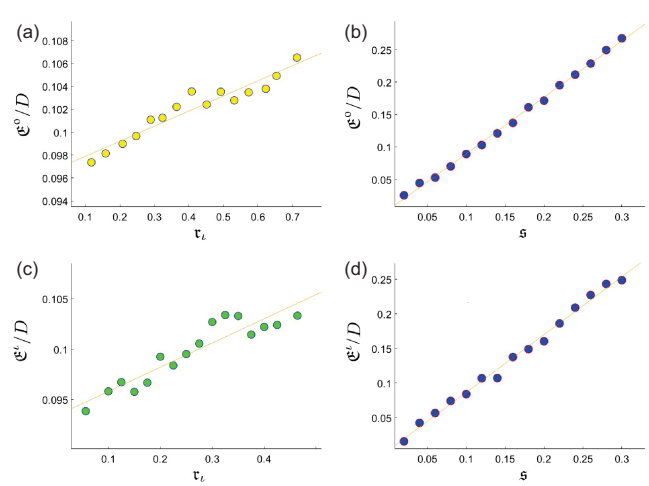

Figure 8 (Color online) The element-wise estimation errors versus the rank complexity and sparse ratio for tensors in $\mathbb{R}^{40\times 40\times 40}$. (a) $\mathfrak{E}^{\mathrm{o}}$ vs $\mathfrak{r}_{\mathrm{o}}$ with $\mathfrak{s} = 0.1$; (b) $\mathfrak{E}^{\mathrm{o}}$ vs $\mathfrak{s}$ with $\mathfrak{r}_{\mathrm{o}} = 0.4478$; (c) $\mathfrak{E}^{\iota}$ vs $\mathfrak{r}_\iota$ with $\mathfrak{s} = 0.1$; (d) $\mathfrak{E}^{\iota}$ vs $\mathfrak{s}$ with $\mathfrak{r}_\iota = 0.275$.

and $\mathfrak{r}_\iota$ and $\mathfrak{s}$, respectively. Thus, it can be said that the proposed bounds can approximately predict the scaling behavior of the estimation error.

7.2 Effectiveness of the proposed OITNNs

7.2.1 Color images

We first evaluate effectiveness of the proposed norms in comparison with other nuclear norm-based models on nine color images of size $256 \times 256 \times 3$ (see Figure 9). The competitor norms include SNN [14], LatentNN [20], SqNN [21], TNN [23], twist TNN (t-TNN) [42] and matrix nuclear norm (NN) [15]. We first conduct RTD on color images and then carry out color image inpainting to further demonstrate the power of the proposed norms. RTD models based on the aforementioned norms are formulated by replacing OITNN-O in Problem (25) and the corresponding optimization problems are solved by ADMM via our own implementations in Matlab. We use the peak signal noise ratio (PSNR) to measure the recovery quality. The higher the PSNR is, the better the recovery.

Robust image recovery Given a color image $\mathcal{M} \in \mathbb{R}^{d_1 \times d_2 \times 3}$, we first normalize it such that $\|\mathcal{M}\|_{l_\infty} = 1$. Then, we select the support set Ω (with size $|\Omega| = \mathfrak{s}D$) of $\mathcal{S}^*$ uniformly at random with corruption ratio $\mathfrak{s}$, and add independent zero-mean Gaussian noises with standard deviation $\sigma = c\sigma_0$, where $\sigma_0 = \|\mathcal{M}\|_{\mathrm{F}}/\sqrt{D}$ is the normalized signal-noise-ratio [6]. Specifically, we follow the setting in ref. [43] to generate the outlier tensor $\mathcal{S}^*$, that is, for all $(i, j, k) \in \Omega$, let $\mathcal{S}^*_{ijk} = \mathcal{B}_{ijk}$, where $\mathcal{B}$ is a tensor with independent Bernoulli ± 1 entries.

The weight parameters for the proposed norms OITNN-O and OITNN-L are set as $w_1 : w_2 : w_3 = 1 : 10 : 1$ and $v_1 : v_2 : v_3 = 1 : 0.0288 : 1$, respectively, since they perform better in most cases in the parameter tuning phase. The weight parameters $\boldsymbol{\alpha}$ of SNN in eq. (2) are chosen to satisfy $\alpha_1 : \alpha_2 : \alpha_3 = 1 : 1 : 0.01$ as suggested in ref. [14]. For LatentNN, we adopt its scaled version in eq. (4) which achieves better performance. The "sparse/low-rank" parameter ratio μ/λ of NN is $1/\sqrt{\max\{d_1, d_2\}}$ [44], and $1/\sqrt{3\max\{d_1, d_2\}}$ for SqNN, TNN and t-TNN [6]. We suggest the "sparse/low-rank ratio" parameter for OITNN-O to be $1/\sqrt{3\max\{d_1, d_2\}}$. The "sparse/low-rank ratio" parameter for OITNN-L is much more difficult to select, and we find 3.33×10^{-4} performs better in most cases on the tested color images[8]. We first test two cases by setting the pair of corruption ratio and noise level as $(\mathfrak{s}, c) = (0.05, 0.1)$ and $(0.15, 0.15)$, and then consider a more difficult setting where 30% entries are corrupted by outliers (i.e., $\mathfrak{s} = 0.3$) and the R, G, B channels are polluted by dense Gaussian noises with noise level $c = 0.1, 0.2$, and 0.3, respectively. Given a color image and a corruption-noise setting, we test 10 times and report the averaged PSNR and running time (in seconds).

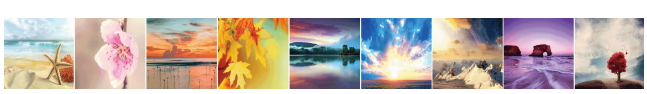

Figure 9 (Color online) Nine test images.

For quantitative comparison, we report the PSNR and running time on the nine images in Figures 10 and 11. An visual example is shown in Row 1 of Figure 12 for qualitative evaluation. As illustrated in Figure 12, the proposed norms obtain higher visual quality than the competitor norms. According to Figures 10 and 11, we can find that the tensorial norms outperform matrix nuclear norm since more structural information is exploited by tensorial treatments, and the proposed OITNN-O and OITNN-L can achieve better performance than other tensorial norms in most cases since the multi-orientational spectral low-rankness is considered. As regards to the computational complexity, it can be seen that the proposed OITNNs (especially OITNN-L) are much more costly than NN, SqNN, TNN, and t-TNN, and slightly more time-consuming than SNN and LatentNN, which may potentially limit the application of OITNNs on large scale datasets.

Image inpainting To further show the effectiveness of the proposed OITNNs, we also apply them to the classical image inpainting problem. Specifically, we consider two settings of missing patterns on the nine test images in Figure 10. In Setting I, we adopt the uniform sampling strategy where 90% of the entries are missing uniformly at random, whereas in Setting II, we consider non-uniform sampling under which some rows are first missing and columns of the rest rows are then randomly sampled with a total missing ratio of 85%. Note that Setting II is very challenging since all the three matricizations of an input image suffer from missing columns which can hardly be recovered by matrix low-rankness in original domain.

For qualitative comparison, inpainting examples in Settings I and II are shown in Rows 2 and 3 of Figure 12, respectively. The quantitative comparison in PSNR and running time (in seconds) is presented in Figure 13.

The following can be found from Figures 12 and 13.

- Setting II is harder than Setting I because the non-uniform sampling in Setting II involves totally missing rows, columns, and tubes which will be automatically filled with zeros by directly minimizing traditional nuclear norms.

8) It is of theoretical significance to derive "optimal values" of the parameters for OITNN-L by adopting a technology road-map similar to but more challenging than the famous robust PCA [44].

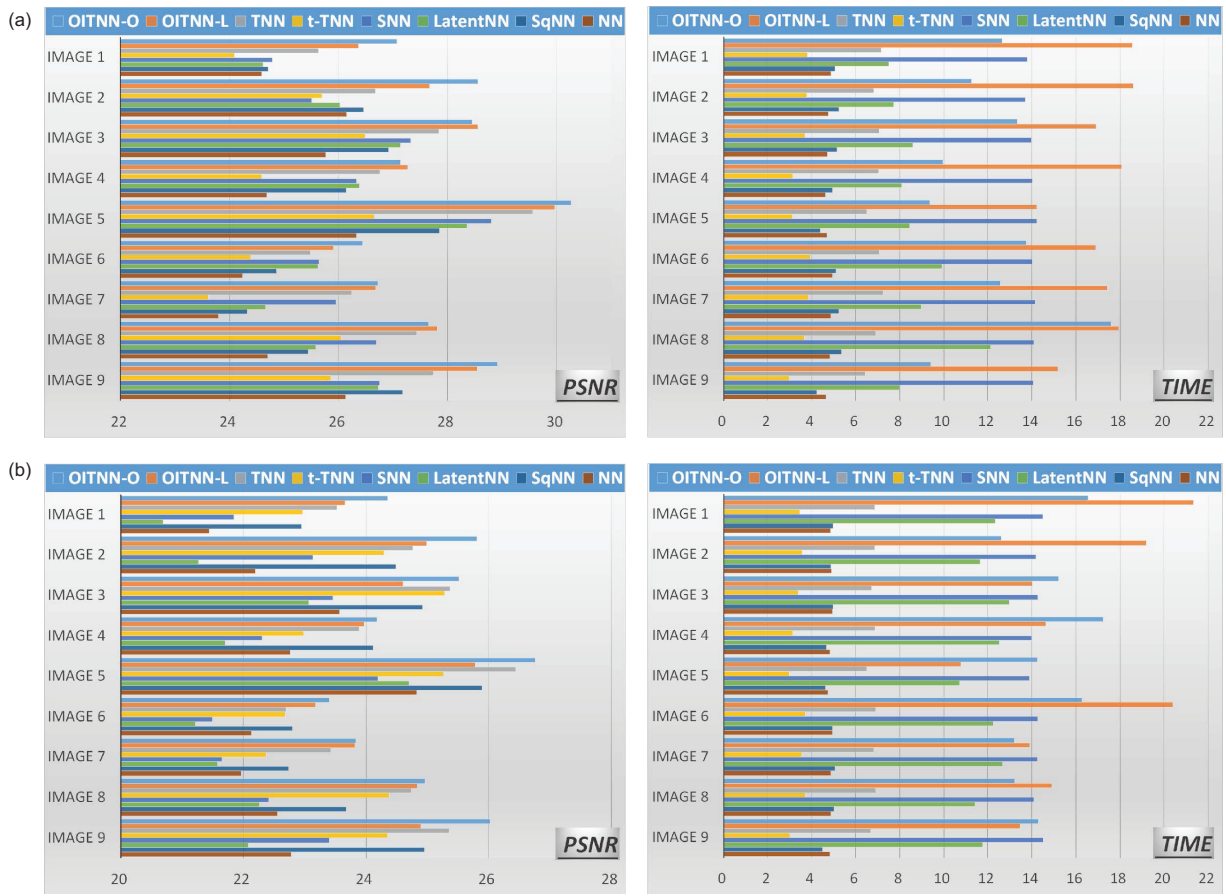

Figure 10 (Color online) Quantitative comparison in PSNR values and running time (in seconds) of RTD models based on eight different norms on color images. Subplot (a) shows the results for the setting $(\mathfrak{s}, c) = (0.05, 0.1)$ with subplot (b) for $(\mathfrak{s}, c) = (0.15, 0.15)$.

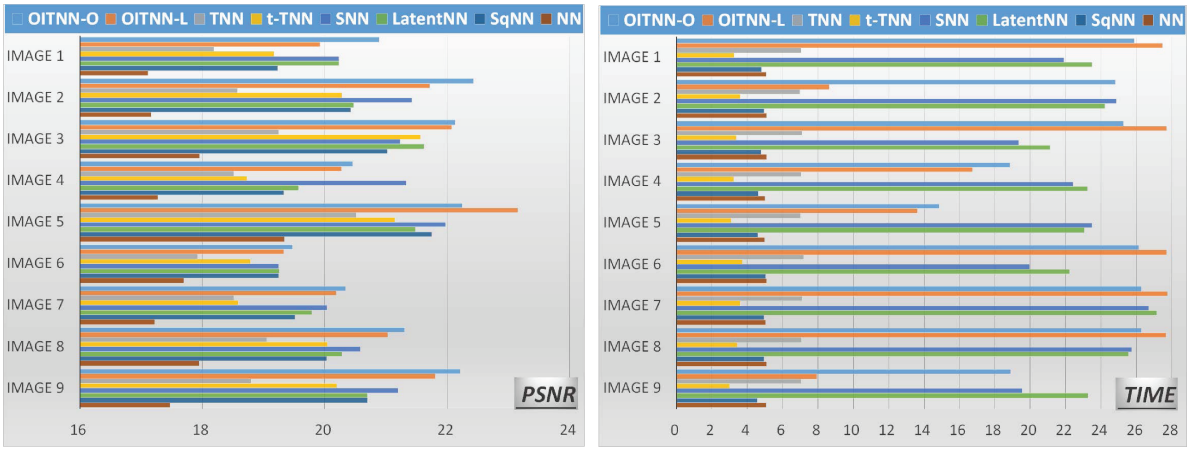

Figure 11 (Color online) Quantitative comparison in PSNR values and running time (in seconds) of RTD models based on eight different norms on color images for the setting where 30% entries are corrupted by outliers, and the R, G, B channels are respectively polluted by dense noises with noise level $c = 0.1, 0.2,$ and 0.3.

• Our OITNN-O performs best in the tensorial norms, and we owe the gain in performance to its capability to model spectral low-rankness in multiple orientations. The effect is particularly significant in dealing with non-uniform sampling: (1) First, matrix nuclear norm based models fail to recover totally missing rows, columns, and tubes since the minimization of nuclear norm simply fills the void with zeros. (2) The orientation sensitivity of TNN prevents them from recovering the totally missing tubes, since the direct minimization of TNN leads to a preference to fill in the unknown tubes with zeros. (3) Thanks to its capability to exploit multi-orientational spectral low-rankness, our OITNNs

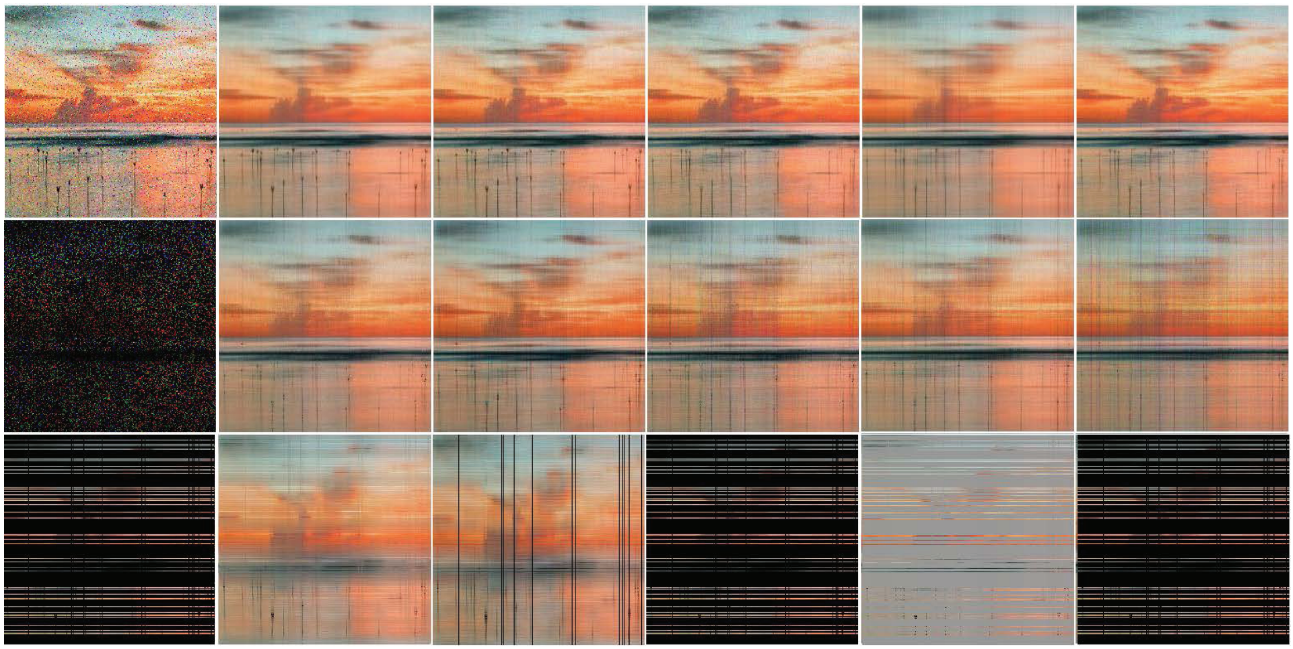

Figure 12 (Color online) Visual performances of the proposed OITNN-O and OITNN-L compared with tightly related TNN, SNN, and LatentNN. Row 1: robust image recovery with corruption ratio $\mathfrak{s} = 0.05$ and noise level $c = 0.1$. Row 2: image inpainting with 90% random missing entries. Row 3: image inpainting with missing columns and rows (total missing ratio 85%).

can introduce information of observed rows/columns/tubes to reconstruct the unseen ones, leading to improved recovery.

• Despite the remarkable recovery performance, the proposed OITNNs (especially OITNN-L) consume more running time than TNN, t-TNN, NN, and SqNN, and have a comparable computational complexity to SNN and LatentNN. Although this kind of trade-off in performance and complexity is reasonable, the high computational cost may be unaffordable for high-dimensional data.

7.2.2 *YUV videos*

The efficacy of the proposed OITNNs are also evaluated through experiments on the widely used eleven YUV videos[9]: akiyo, bridge-far, silent, carphone, claire, coastguard, container, foreman, salesman, grandma, and hall. We simply use the first n frames of each video, resulting in eleven 4-way tensors in $\mathbb{R}^{144\times176\times3\times n}$. We conduct video inpainting with 85% random missing.

The proposed norm is compared with TNN, SNN, and LatentNN. To use OITNN-O and OITNN-L, we first permute the tensors[10] to $144 \times n \times 3 \times 176$. TNN is tested separately on three (i.e., the channel number) tensors in $\mathbb{R}^{144\times n\times176}$. For LatentNN, we also adopt its scaled version in eq. (4), which achieves better performance. Parameters for OITNN-O are set as $w_1 : w_2 : w_3 : w_4 = 1 : 100 : 1 : 1$, and for OITNN-L $v_1 : v_2 : v_3 : v_4 = 1 : 0.001 : 1 : 1$. We simply set $n = 8$ due to the trade-off between the parameter tuning and computational cost. We tune other parameters for better performances in most cases.

We show the PSNR values and running time (in seconds) in Figure 14. It can be seen that the proposed OITNNs have better performances thanks to their flexibility in exploiting the multi-orientation correlations in color videos. According to the results of running time comparison, the proposed ADMM-based algorithms for OITNN minimization unavoidably increase the computational burden in comparison with ADMM-based TNN minimization due to the need of computing K TNNs by definition, but the computational complexity is comparable to (or only slightly higher) the ADMM-based SNN and LatentNN minimization.

8 Conclusions

To improve the representation ability and flexibility of the recently proposed TNN in modeling multi-orientational correlations, we defined two new tensor norms for general K-way ($K \geq 3$) tensors, dubbed OITNN-O and OITNN-L, which can simultaneously exploit the low-rankness in spectral domain for all orientations. Then, we adopted them to RTD and rigorously established upper bounds on the estimation error. Correctness of the error bounds was verified through simulation study on synthetic datasets. Experiments on real datasets

9) The videos are available at https://media.xiph.org/video/derf/.

10) It is observed that this way of tensor construction performs better, despite one has some other ways. Similar tricks are also used in refs. [24, 45].

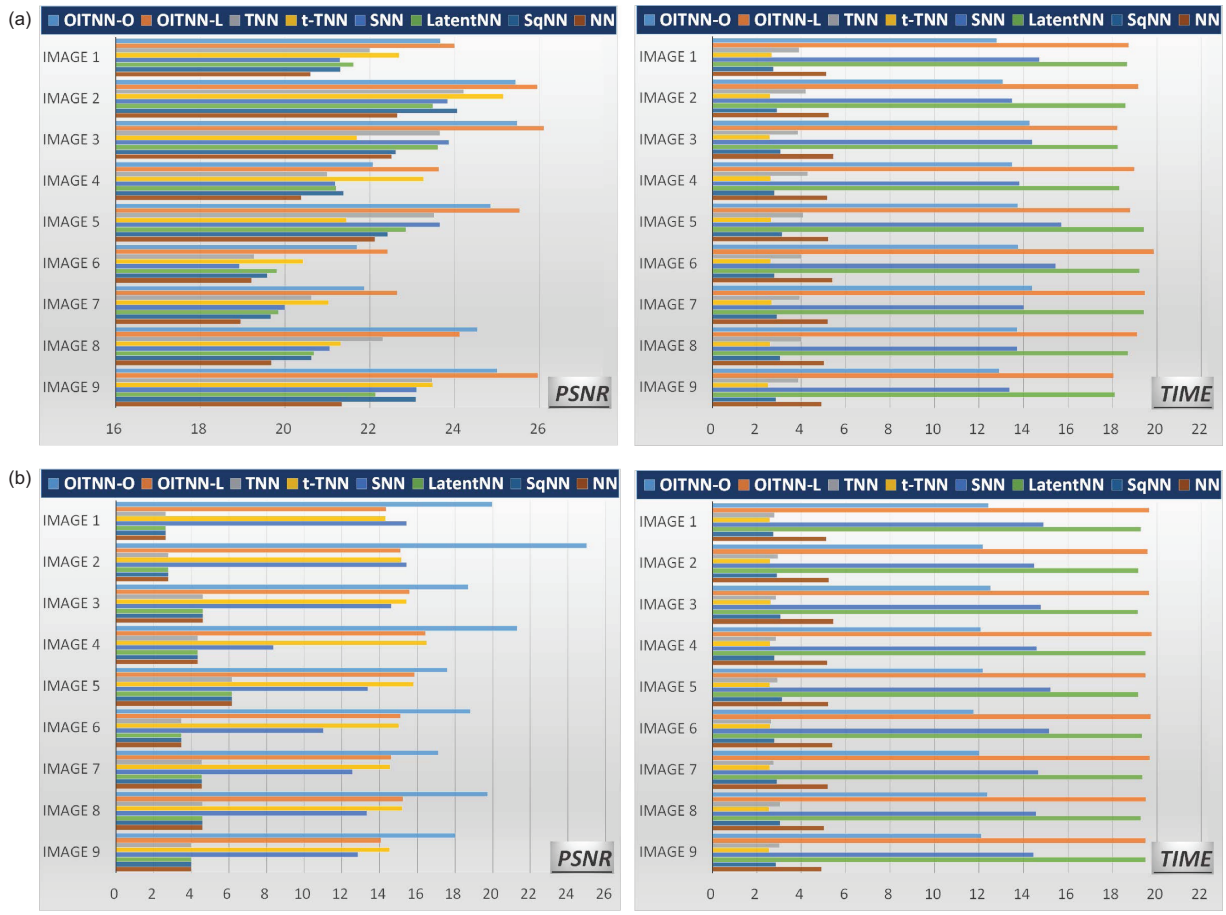

Figure 13 (Color online) Quantitative comparison in PSNR values and running time (in seconds) of image inpainting methods based on eight different norms. Subplot (a) show the results for Setting I where approximately 90% of the entries are missing uniformly at random, and subplot (b) is for Setting II where there exist simultaneously missing columns, rows, and tubes with approximately 85% missing entries.

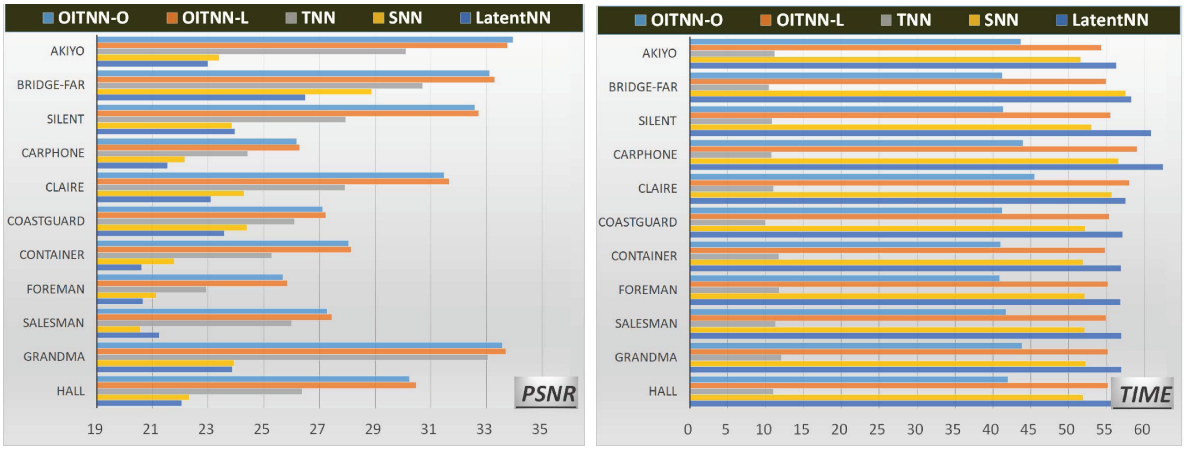

Figure 14 (Color online) Quantitative comparison in PSNR values and running time (in seconds) for video recovery.

demonstrated the effectiveness of the proposed norms.

However, despite the promising empirical performance, the proposed OITNNs have the following two drawbacks.

- (Parameter selection) A main drawback of OITNNs is that we cannot automatically determine the optimal tuning parameters. Although we have suggested the parameter settings for OITNN-O and OITNN-L on the tested datasets of color images and videos, one still has to tune the parameters on different datasets (e.g., the hyperspectral/multispectral images [7]).

• (Computational complexity) According to the experimental results of running time, the proposed OITNNs are more time-consuming than TNN. Although it seems unavoidable since OITNNs have to compute TNNs in all orientations to model the multi-orientational correlations, the high computational burden may severely limit practical applications of OITNNs.

Thus, there is a need to develop suitable tools for parameter tuning (like the variational Bayesian treatments [46]), to derive "optimal values" of the parameters using a similar technique road-map with ref. [44], or to design adaptive re-weighting schemes like ref. [47] for OITNNs in future work. Future research directions also include developing fast algorithms for the proposed models using techniques like [6], and considering more general transformations rather than DFT [48–50] or structures of the underlying tensor [51]. Another interesting direction is to exploit smoothness [2,31] via a hyper-graph structure [52] for more sound tensor modeling.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61872188, 62103110, 62073087, 62071132, 61903095, U191140003, and 61973090), the China Postdoctoral Science Foundation (Grant No. 2020M672536), the Natural Science Foundation of Guangdong Province (Grant Nos. 2020A1515010671, 2019B010154002, and 2019B010118001), the Guangdong Provincial Key Laboratory of Electronic Information Products Reliability Technology (Grant No. 2017B030314151). The first author would like to thank Shasha Jin for her kind understanding in these months.

Supporting Information

The supporting information is available online at tech.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

1 Liu X Y, Aeron S, Aggarwal V. Low-tubal-rank tensor completion using alternating minimization. IEEE Trans Inform Theor, 2020, 66: 1714–1737
2 Chang J, Chen Y, Qi L. Hypergraph clustering using a new laplacian tensor with applications in image processing. SIAM J Imag Sci, 2020, 13: 1157–1178
3 Zhang X, Ng M K P. Low rank tensor completion with poisson observations. IEEE Trans Pattern Anal Mach Intell, 2021, doi: 10.1109/TPAMI.2021.3059299
4 Qiu Y N, Zhou G X, Chen X Q. Semi-supervised non-negative Tucker decomposition for tensor data representation. Sci China Tech Sci, 2021, 64: 1881–1892
5 Gu Q, Gui H, Han J. Robust tensor decomposition with gross corruption. In: Proceedings of Advances in Neural Information Processing Systems. Montral: MIT Press, 2014. 1422–1430
6 Wang A, Jin Z, Tang G. Robust tensor decomposition via t-SVD: Near-optimal statistical guarantee and scalable algorithms. Signal Process, 2020, 167: 107319
7 Wang A, Zhou G, Zhao Q. Guaranteed robust tensor completion via *L-SVD with applications to remote sensing data. Remote Sens, 2021, 13: 3671
8 Cui C, Zhang Z. High-dimensional uncertainty quantification of electronic and photonic IC with non-gaussian correlated process variations. IEEE Trans Comput-Aided Des Integr Circuits Syst, 2020, 39: 1649–1661
9 Zhao Q, Meng D, Kong X, et al. A novel sparsity measure for tensor recovery. In: Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 271–279
10 Harshman R A. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. 1970
11 Tucker L R. Some mathematical notes on three-mode factor analysis. Psychometrika, 1966, 31: 279–311
12 Oseledets I V. Tensor-train decomposition. SIAM J Sci Comput, 2011, 33: 2295–2317
13 Kilmer M E, Braman K, Hao N. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM J Matrix Anal Appl, 2013, 34: 148–172
14 Liu J, Musialski P, Wonka P. Tensor completion for estimating missing values in visual data. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 208–220
15 Candes E J, Tao T. The power of convex relaxation: Near-optimal matrix completion. IEEE Trans Inform Theor, 2010, 56: 2053–2080
16 Hillar C J, Lim L H. Most tensor problems are np-hard. J ACM, 2013, 60: 1–39
17 Fazel M. Matrix rank minimization with applications. Dissertation for Doctoral Degree. Stanford: Stanford University, 2002
18 Yuan M, Zhang C H. On tensor completion via nuclear norm minimization. Found Comput Math, 2016, 16: 1031–1068
19 Tomioka R, Suzuki T, Hayashi K, et al. Statistical performance of convex tensor decomposition. In: Proceedings of Annual Conference on Neural Information Processing Systems. Granada: MIT Press, 2011. 972–980
20 Tomioka R, Suzuki T. Convex tensor decomposition via structured schatten norm regularization. In: Proceedings of Annual Conference on Neural Information Processing Systems. Lake Tahoe: MIT Press, 2013. 1331–1339
21 Mu C, Huang B, Wright J, et al. Square deal: Lower bounds and improved relaxations for tensor recovery. In: Proceedings of International Conference on Machine Learning. Beijing: ACM, 2014. 73–81
22 Zhang Z, Ely G, Aeron S, et al. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014. 3842–3849
23 Zhang Z, Aeron S. Exact tensor completion using t-SVD. IEEE Trans Signal Process, 2017, 65: 1511–1526
24 Lu C, Feng J, Chen Y. Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans Pattern Anal Mach Intell, 2020, 42: 925–938
25 Zhang L, Song L, Du B. Nonlocal low-rank tensor completion for visual data. IEEE Trans Cybern, 2021, 51: 673–685
26 Yokota T, Erem B, et al. Missing slice recovery for tensors using a low-rank model in embedded space. arXiv: 1804.01736
27 Wang A, Li C, Jin Z, et al. Robust tensor decomposition via orientation invariant tubal nuclear norms. In: Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 6102–6109
28 Kolda T G, Bader B W. Tensor decompositions and applications. SIAM Rev, 2009, 51: 455–500
29 Huang B, Mu C, Goldfarb D, et al. Provable low-rank tensor recovery. Optim-Online, 2014, 4252: 455–500
30 Li X, Wang A, Lu J. Statistical performance of convex low-rank and sparse tensor recovery. Pattern Recognition, 2019, 93: 193–203
31 Qiu Y, Zhou G, Wang Y. A generalized graph regularized non-negative tucker decomposition framework for tensor data representation. IEEE Trans Cybern, 2022, 52: 594–607
32 Goldfarb D, Qin Z T. Robust low-rank tensor recovery: Models and algorithms. SIAM J Matrix Anal Appl, 2014, 35: 225–253
33 Wimalawarne K, Sugiyama M, Tomioka R. Multitask learning meets tensor factorization: Task imputation via convex optimization. Proc

Adv Neural Inf Proc Syst, 2014, 27: 2825–2833
34 Rockerfeller R. Convex Analysis. Princeton: Princeton University Press, 1970
35 Wei D, Wang A, Feng X. Tensor completion based on triple tubal nuclear norm. Algorithms, 2018, 11: 94
36 Zheng Y B, Huang T Z, Zhao X L. Tensor N-tubal rank and its convex relaxation for low-rank tensor recovery. Inf Sci, 2020, 532: 170–189
37 Klopp O, Lounici K, Tsybakov A B. Robust matrix completion. Probab Theor Relat Fields, 2017, 169: 523–564
38 Huang B, Mu C, Goldfarb D, et al. Provable models for robust low-rank tensor completion. Pacific J Optim, 2015, 11: 339–364
39 Wang A, Wei D, Wang B. Noisy low-tubal-rank tensor completion through iterative singular tube thresholding. IEEE Access, 2018, 6: 35112–35128
40 He B, Yuan X. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. SIAM J Numer Anal, 2012, 50: 700–709
41 Boyd S. Distributed optimization and statistical learning via the alternating direction method of multipliers. FNT Machine Learn, 2010, 3: 1–122
42 Hu W, Tao D, Zhang W. The twist tensor nuclear norm for video completion. IEEE Trans Neural Netw Learn Syst, 2017, 28: 2961–2973
43 Lu C. Transforms based tensor robust pca: Corrupted low-rank tensors recovery via convex optimization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2021. 1145–1152
44 Candès E J, Li X, Ma Y, et al. Robust principal component analysis? J ACM, 2011, 58: 11
45 Lu C, Peng X, Wei Y. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 5996–6004
46 Zhao Q, Zhou G, Zhang L. Bayesian robust tensor factorization for incomplete multiway data. IEEE Trans Neural Netw Learn Syst, 2016, 27: 736–748
47 Ng M K P, Yuan Q, Yan L. An adaptive weighted tensor completion method for the recovery of remote sensing images with missing data. IEEE Trans Geosci Remote Sens, 2017, 55: 3367–3381
48 Li C, Khan M E, Sun Z, et al. Beyond unfolding: Exact recovery of latent convex tensor decomposition under reshuffling. In: Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 4602–4609
49 Wang A, Zhou G, Jin Z. Tensor recovery via $*_L$-spectral k-support norm. IEEE J Sel Top Signal Process, 2021, 15: 522–534
50 Cui C, Zhang K, Daulbaev T. Active subspace of neural networks: Structural analysis and universal attacks. SIAM J Math Data Sci, 2020, 2: 1096–1122
51 Li C, Sun Z. Evolutionary topology search for tensor network decomposition. In: Proceedings of the International Conference on Machine Learning. ACM, 2020. 5947–5957
52 Chang J, Ding W, Qi L. Computing the p-spectral radii of uniform hypergraphs with applications. J Sci Comput, 2018, 75: 1–25