Robust Tensor Decomposition via Orientation Invariant Tubal Nuclear Norms

Andong Wang,[1,2,3] **Chao Li,**[2] **Zhong Jin,**[*1,4] **Qibin Zhao**[*2,3]
[1]School of Computer Science & Engineering, NJUST [2]Tensor Learning Unit, RIKEN AIP [3] School of Automatation, GDUT
[4]Key Laboratory of Intelligent Perception and System for High-Dimensional Information of Ministry of Education, NJUST
Corresponding authors: Zhong Jin (zhongjin@njust.edu.cn), Qibin Zhao (qibin.zhao@riken.jp)

Abstract

Low-rank tensor recovery has been widely applied to computer vision and machine learning. Recently, tubal nuclear norm (TNN) based optimization is proposed with superior performance as compared to other tensor nuclear norms. 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 robust tensor decomposition 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.

Introduction

Tensor decomposition has become a paradigm in modern multi-way data analysis. Due to various reasons like sensor failures, occlusion in videos, or abnormalities, the multi-way data are often corrupted by noises and gross corruptions. For example, the embedded noises in hyper-spectral image is probably a mixture of small dense noises and sparse gross corruptions (Zhao et al. 2015a). To tackle both small noises and gross corruptions, the robust tensor decomposition (RTD) (Gu, Gui, and Han 2014) is studied to robustify traditional tensor decompositions like CANDECOMP/PARAFAC (CP) decomposition (Harshman 1970) and Tucker decomposition (Tucker 1966) 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 (Harshman 1970), Tucker rank (Tucker 1966), TT rank (Oseledets 2011) and Tubal rank (Kilmer et al. 2013), 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) (Liu et al. 2013). Unfortunately, RMP is NP-hard in general for matrices (2-way tensors) (Candès and Tao 2010) and even harder for higher-way tensors (Hillar and Lim 2009). In low-rank matrix estimation, matrix nuclear norm is proposed as the convex envelop of rank function (Fazel 2002) for tractable algorithms. Motivated by the great success of matrix nuclear norm, its tensor extensions have been extensively studied, like tensor trace norm (Yuan and Zhang 2016), overlapped Schatten-1 norm (SNN) (Tomioka et al. 2011; Liu et al. 2013), latent Schatten norm (LatentNN) (Tomioka and Suzuki 2013), squared nuclear norm (SqNN) (Mu et al. 2014) and tubal nuclear norm (TNN) (Zhang et al. 2014). Among existing tensor nuclear norms[1], TNN is induced by the tensor singular value decomposition (t-SVD) (Kilmer et al. 2013) and has shown superior performance in various applications, such as image/video inpainting/denoising (Zhou and Feng 2017; Zhang and Aeron 2017; Lu et al. 2019), clustering (Xie et al. 2017) and WiFi fingerprint-based indoor localization (Liu et al. 2016b).

In real multi-way data like images and videos, there is an ubiquitous "spatial-shifting" correlation making such data spatial-temporally smooth (Liu et al. 2016a). From a signal processing standpoint, smoothness in original domain often reflects the existence of some simple patterns in spectral domain (Yokota et al. 2018). 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-order 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:

1) We propose two tensor norms via a novel 3d unfolding operation on K-way tensors, which are orientation invariant, thus can be exploited for the multi-orientational spectral low-rankness.
2) The new norms are employed to formulate RTD as two

[1]In this paper, "tensor nuclear norms" refer to tensor extensions of matrix nuclear norm instead of the nuclear norm induced by CP decomposition (Yuan and Zhang 2016).

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.

Notations and Preliminaries

Notations. Matrices and tensors are denoted by uppercase boldface, and calligraphy letters, respectively. Let $[n] := \{1, \cdots, n\}, \forall n \in \mathbb{N}_+$. 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 $\mathbb{R}^{d_1 \times \cdots \times d_K}$. For a matrix $\mathbf{M}$ with singular values σ_i's, define its nuclear norm $\|\mathbf{M}\|_* := \sum_i \sigma_i$ and spectral norm $\|\mathbf{M}\| := \max_i \sigma_i$. Given $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$, define its l_0-norm $\|\mathcal{T}\|_{l_0} := \|\mathrm{vec}(\mathcal{T})\|_0$, l_1-norm $\|\mathcal{T}\|_{l_1} := \|\mathrm{vec}(\mathcal{T})\|_1$, F-norm $\|\mathcal{T}\|_{\mathrm{F}} := \|\mathrm{vec}(\mathcal{T})\|_2$, and l_∞-norm $\|\mathcal{T}\|_{l_\infty} := \|\mathrm{vec}(\mathcal{T})\|_\infty$, where $\mathrm{vec}(\cdot)$ denotes the vectorization (Kolda and Bader 2009). For notational simplicity, let $d_{K+1} = d_1$, $D = \prod_{k\in[K]} d_k$, $d_{\backslash k} = D/(d_k d_{k+1})$, $\tilde{d}_k = \sqrt{d_{k+1}}\big(\sqrt{d_k} + \sqrt{d_{\backslash k}}\big)$, $\forall k \in [K]$. Given $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, let $\mathbf{T}^{(i)} := \mathcal{T}(:,:,i)$ denotes its i^{th} frontal slice. We use c, c', c_1 etc. to denote constants whose values can vary from line to line. Other notations are introduced when they first appear.

Tensor singular value decomposition

We briefly recall the tensor singular value decomposition.

Definition 1. (Kilmer et al. 2013). *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 $(i,j)^{\text{th}}$ tube $\mathcal{T}(i,j,:) = \sum_{k=1}^{d_2} \mathcal{T}_1(i,k,:) \bullet \mathcal{T}_2(k,j,:)$, where $\bullet$ is the circular convolution.*

With notions like tensor transpose, orthogonal tensor, and f-diagonal tensor (Kilmer et al. 2013), t-SVD can be defined.

Definition 2. (Kilmer et al. 2013). *Any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has a tensor singular value decomposition as*

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \tag{1}$$

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 $\mathcal{S}$:

$$\mathrm{rank}_{\mathrm{tb}}(\mathcal{T}) := \#\big\{i \,\big|\mathcal{S}(i,i,:) \neq \mathbf{0}\big\}. \tag{2}$$

Definition 3. (Lu et al. 2016). *Given $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, let $\widetilde{\mathcal{T}}$ be its Fourier version[2] in $\mathbb{C}^{d_1 \times d_2 \times d_3}$. The tensor average rank $\mathrm{rank}_{\mathrm{avg}}(\cdot)$, tubal nuclear norm $\|\cdot\|_\star$ of $\mathcal{T}$ are defined as the averaged rank and nuclear norm of frontal slices of $\widetilde{\mathcal{T}}$:*

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

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

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

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

As proved in (Lu et al. 2019), 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 3. 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.

Orientation Invariant TNNs

To overcome the drawbacks of TNN, we first propose a new tensor 3d-unfolding operation as follows.

Definition 4 (mode-(k,t) 3d-unfolding)**.** *For different $k, t \in [K]$, the mode-(k,t) 3d-unfolding of $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is a 3-way tensor $\mathcal{T}_{[k,t]} \in \mathbb{R}^{d_k \times (D/(d_k d_t)) \times d_t}$ obtained by the following two steps (see Fig. 1.).*

First, permute $\mathcal{T}$ to $\mathcal{Z} \in \mathbb{R}^{d'_1 \times d'_2 \times \cdots \times d'_K}$ whose 1st and K^{th} modes are respectively the k^{th} and t^{th} modes of $\mathcal{T}$, with the rest modes permuted circularly. Second, reshape $\mathcal{Z}$ to $\mathcal{T}_{[k,t]} \in \mathbb{R}^{d_k \times (D d_k^{-1} d_t^{-1}) \times d_t}$ 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 $J_l = \sum_{m=2}^{l-1} d'_m$.

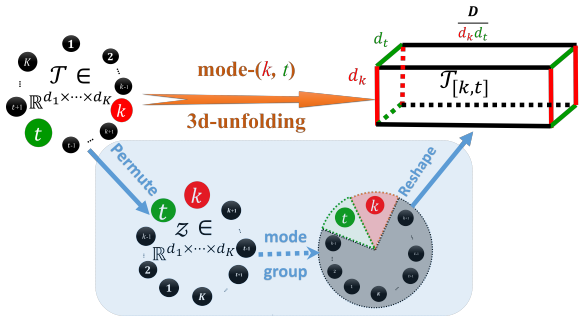


Figure 1: Illustration of 3d-unfolding.

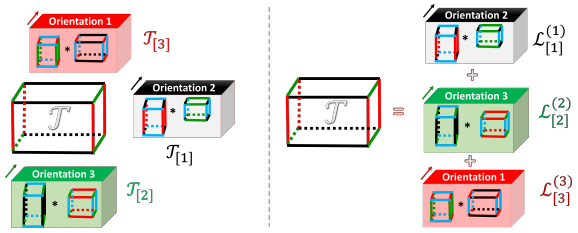
Intuitively, by viewing a K-way tensor $\mathcal{T}$ as a "$(K$-1)-way array $\mathbf{T}$" of size $d_1 \times d_2 \times \cdots \times d_{t-1} \times d_{t+1} \times \cdots d_K$ whose entries are mode-t tubes, the mode-(k,t) 3d-unfolding $\mathcal{T}_{[k,t]}$ can also be analogously viewed as a "mode-k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

Based on 3d-unfolding, two tensor ranks are defined.

Definition 5. *The Orientation Invariant Tubal Rank (OITR) $\vec{\mathfrak{r}}_{\mathrm{t}}$ and Orientation Invariant Average Rank (OIAR) $\vec{\mathfrak{r}}_{\mathrm{a}}$ of any $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ are defined as the following vectors:*

$$\begin{aligned}\vec{\mathfrak{r}}_{\mathrm{t}}(\mathcal{T}) &:= \big(\mathrm{rank}_{\mathrm{tb}}(\mathcal{T}_{[1]}), \cdots, \mathrm{rank}_{\mathrm{tb}}(\mathcal{T}_{[K]})\big)^\top \in \mathbb{R}^K,\\ \vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{T}) &:= \big(\mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[1]}), \cdots, \mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[K]})\big)^\top \in \mathbb{R}^K.\end{aligned} \tag{3}$$

[3]Using circular order of modes, let $d_{K+1} = d_{(K+1) \bmod K} = d_1$.



(a) OITNN-O $\|\mathcal{T}\|_{\star o}$ (b) OITNN-L $\|\mathcal{T}\|_{\star \iota}$

Figure 2: 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 $\{\mathcal{L}^{(k)}\}$.

As shown in Eq. (3), OITR serves as a complexity measure in the original domain, whereas the OIAR measures low-rankness in the spectral domain. They have the following relationship with the classical Tucker rank $\vec{\mathfrak{r}}_{\text{Tucker}}$.

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

$$\vec{\mathfrak{r}}_{\text{a}}(\mathcal{T}) \leq \min\{\vec{\mathfrak{r}}_{\text{t}}(\mathcal{T}), \vec{\mathfrak{r}}_{\text{Tucker}}(\mathcal{T})\}, \tag{4}$$

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

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. By relaxing average rank to its convex envelop in each orientation, we naturally define the following norm.

Definition 6. *The Overlapped Orientation Invariant Tubal Nuclear Norm (OITNN-O) of $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined as:*

$$\|\mathcal{T}\|_{\star o} := \sum_{k=1}^{K} w_k \|\mathcal{T}_{[k]}\|_\star, \tag{5}$$

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 Fig. 2(a)). It differs from SNN (Liu et al. 2013) which only considers low Tucker rank. As special cases, if $K = 3$, OITNN-O degenerates to triple TNN (Wei et al. 2018); If $K = 3, (w_1, w_3) \to \mathbf{0}$, then it approximates TNN.

Although the assumption of low OIAR is weaker than low Tucker rank, it may still be strict for some real data tensors. In (Tomioka and Suzuki 2013), it is observed that a latent Schatten norm (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 7. *The Latent Orientation Invariant Tubal Nuclear Norm (OITNN-L) of $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ is defined as*

$$\|\mathcal{T}\|_{\star \iota} := \inf_{\sum_k \mathcal{L}^{(k)} = \mathcal{T}} \sum_{k=1}^{K} v_k \|\mathcal{L}^{(k)}_{[k]}\|_\star, \tag{6}$$

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 Fig. 2(b)). According to Definitions 6 and 7, 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 (Zhang et al. 2014; Lu et al. 2019), 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 $\|\cdot\|_{\star \iota}$, denoted by $\|\cdot\|^*_{\star o}$ and $\|\cdot\|^*_{\star \iota}$ respectively, are given as follows:*

$$\begin{aligned} \|\mathcal{T}\|^*_{\star o} &:= \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \max_k \left\{ w_k^{-1} \|\mathcal{X}^{(k)}_{[k]}\| \right\}, \\ \|\mathcal{T}\|^*_{\star \iota} &:= \sum_{k=1}^{K} v_k^{-1} \|\mathcal{T}_{[k]}\|. \end{aligned} \tag{7}$$

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

Robust Tensor Decomposition via OITNNs

Observation model

Suppose we observe $\mathcal{Y} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ by the following model

$$\mathcal{Y} = \mathcal{L}^* + \mathcal{S}^* + \mathcal{E}, \tag{8}$$

where $\mathcal{L}^*$ is the true low-rank tensor, $\mathcal{S}^*$ stores entry-wisely sparse corruptions, and tensor $\mathcal{E}$ represents dense small noise (see Fig. 3 (Zhao et al. 2015b)). The goal of RTD is to recover $\mathcal{L}^*$ and $\mathcal{S}^*$ from $\mathcal{Y}$. If $\mathcal{E} = \mathbf{0}$, RTD degenerates to the TRPCA (Lu et al. 2019); If $\mathcal{S}^* = \mathbf{0}$, it becomes the noisy tensor decomposition (Tomioka and Suzuki 2013).

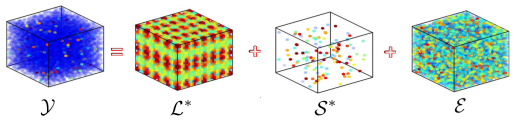


Figure 3: Observation model of RTD.

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 (Agarwal, Negahban, and Wainwright 2012; Klopp, Lounici, and Tsybakov 2017):

$$\|\mathcal{L}^*\|_{l_\infty} \leq \alpha. \tag{9}$$

Second, let $\{\mathcal{L}^{(k)*}\}$ be any latent components obtained while computing $\|\mathcal{L}^*\|_{\star \iota}$ in Eq. (6). Then, the signal $\mathcal{L}^*$ can be written as

$$\mathcal{L}^* = \sum_{k=1}^{K} \mathcal{L}^{(k)*}. \tag{10}$$

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

$$\|\mathcal{L}^{(k)*}_{[l]}\| \leq \beta \tilde{d}_l, \ \forall l \neq k \in [K], \tag{11}$$

where $\tilde{d}_l = \sqrt{d_{l+1}}\big(\sqrt{d_l}+\sqrt{d_{\backslash l}}\big)$. The motivation of Eq. (11) 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[4] in any mode-l 3d-unfolding ($l \neq k$).

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

$$\begin{aligned} &\min_{\mathcal{L},\mathcal{S}}\ l(\mathcal{L},\mathcal{S}) + \lambda_{\mathrm{o}}\|\mathcal{L}\|_{\star\mathrm{o}} + \mu_{\mathrm{o}}\|\mathcal{S}\|_{l_1}, \\ &\text{s.t.}\ \ \|\mathcal{L}\|_{l_\infty} \leq \alpha. \end{aligned} \tag{12}$$

Model II: RTD-OITNN-L

$$\begin{aligned} &\min_{\{\mathcal{L}^{(k)}\},\mathcal{S}}\ l\big(\sum\nolimits_k \mathcal{L}^{(k)},\mathcal{S}\big) + \lambda_\iota \sum\nolimits_k v_k\|\mathcal{L}^{(k)}_{[k]}\|_\star + \mu_\iota\|\mathcal{S}\|_{l_1} \\ &\text{s.t.}\ \ \big\|\sum\nolimits_k \mathcal{L}^{(k)}\big\|_{l_\infty} \leq \alpha,\ \|\mathcal{L}^{(l)}_{[k]}\| \leq \beta\tilde{d}_k,\ \forall l \neq k, \end{aligned} \tag{13}$$

where $\lambda_{\mathrm{o}}, \mu_{\mathrm{o}}, \lambda_\iota, \mu_\iota$ denote regularization parameters, square loss $l(\mathcal{L},\mathcal{S}) = \|\mathcal{Y}-\mathcal{L}-\mathcal{S}\|_{\mathrm{F}}^2/2$ 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 $\{\mathcal{L}^{(k)}\}$.

Statistical Performance

We analyze statistical performance of the proposed models. Let $(\hat{\mathcal{L}}_{\mathrm{o}}, \hat{\mathcal{S}}_{\mathrm{o}})$ and $(\{\hat{\mathcal{L}}^{(k)}\}, \hat{\mathcal{S}}_\iota)$ be any solution to Problem (12) and Problems (13), respectively. We establish both deterministic and non-asymptotic bounds on the estimation errors, i.e., $\mathfrak{E}^{\mathrm{o}}$, $\mathfrak{E}^\iota$, $\mathfrak{E}^\iota_{\mathrm{com}}$ (defined in Table 1), of the low-rank component $\mathcal{L}^*$ and the sparse component $\mathcal{S}^*$ in their sum.

Table 1: List of some notations

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

Deterministic bounds

When $\mathcal{E}$ in the observation model (8) represents any noise, we bound the estimation error in the following theorems where the dual norms in Lemma 2 are used.

Theorem 1. *If $\lambda_{\mathrm{o}} \geq 2\|\mathcal{E}\|^*_{\star\mathrm{o}}$ and $\mu_{\mathrm{o}} \geq 2(\|\mathcal{E}\|_{l_\infty} + 2\alpha)$ in Problem (12), then any solution $(\hat{\mathcal{L}}_{\mathrm{o}}, \hat{\mathcal{S}}_{\mathrm{o}})$ satisfies:*

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

[4]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 8 in the supplementary material).

Theorem 1 indicates that once parameters $(\lambda_{\mathrm{o}}, \mu_{\mathrm{o}})$ 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 $\lambda_\iota \geq 2\max_k\{\|\mathcal{E}\|^*_{\star\iota} + v_k^{-1}(K-1)\beta\tilde{d}_k\}$ and $\mu_\iota \geq 2(\|\mathcal{E}\|_{l_\infty} + 2\alpha)$ in Problem (13), then it holds that:*

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

Theorem 2 shows that when $(\lambda_\iota, \mu_\iota)$ exceed some thresholds in terms of $\mathcal{E}$, estimation error $\mathfrak{E}^\iota_{\mathrm{com}}$ 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 $\mathcal{S}^*$.

Non-asymptotic bounds

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

Theorem 3. *If parameters $\lambda_{\mathrm{o}} = 2\sigma K^{-2}\sum_k(\tilde{d}_k/w_k)$ and $\mu_{\mathrm{o}} = 8\sigma\sqrt{\log D} + 16\alpha$ in Problem (12), then with high probability it holds that*

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

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 (13), then with high probability it holds that:*

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

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

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 $(\underline{r}^{\mathrm{o}}_1, \cdots, \underline{r}^{\mathrm{o}}_K)$. 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:*

$$\frac{\mathfrak{E}^{\mathrm{o}}}{D} \precsim \sigma^2(\mathfrak{r}_{\mathrm{o}} + \mathfrak{s}\log D), \text{and}\ \ \frac{\mathfrak{E}^\iota}{D} \precsim \sigma^2(\mathfrak{r}_\iota + \mathfrak{s}\log D), \tag{14}$$

where $\mathfrak{r}_{\mathrm{o}} = \big(K^{-1}\sum_k\sqrt{\underline{r}^{\mathrm{o}}_k/d}\big)^2$ and $\mathfrak{r}_\iota = \min_k\{\overline{r}^\iota_k/d\}$ 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 (Huang et al. 2015; Lu et al. 2019) which intrinsically ensure separability between low-rank and sparse components.*

Optimization Algorithms

We develop two algorithms to solve Model I and Model II respectively. By adding auxiliary variables to Problem (12), we obtain

$$\min_{\mathcal{L},\mathcal{S},\mathcal{T},\mathcal{K},\{\mathcal{K}_k\}} l(\mathcal{L},\mathcal{S}) + \lambda_{\mathrm{o}} \sum_k w_k \|\mathcal{K}_k\|_\star + \mu_{\mathrm{o}} \|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K})$$
$$\text{s.t.} \quad \mathcal{K}_k = \mathcal{L}_{[k]}, \forall k; \mathcal{T} = \mathcal{S}; \mathcal{K} = \mathcal{L},$$

where $\delta_\alpha^{l_\infty}(\mathcal{K})$ is the indicator function of tensor l_∞-norm ball whose value is 0 if $\|\mathcal{K}\|_{l_\infty} \leq \alpha$, and $+\infty$ otherwise.

Adding auxiliary variables to Problem (13) [5] also yields

$$\min_{\substack{\mathcal{S},\mathcal{T},\mathcal{K},\\ \{\mathcal{L}^{(k)}\},\{\mathcal{K}_k\}}} l(\sum\nolimits_k \mathcal{L}^{(k)}, \mathcal{S}) + \lambda_\iota \sum\nolimits_k v_k \|\mathcal{K}_k\|_\star + \mu_{\mathrm{o}} \|\mathcal{T}\|_{l_1} + \delta_\alpha^{l_\infty}(\mathcal{K})$$
$$\text{s.t.} \quad \mathcal{K}_k = \mathcal{L}_{[k]}^{(k)}, \forall k; \mathcal{T} = \mathcal{S}; \mathcal{K} = \sum\nolimits_k \mathcal{L}^{(k)}.$$

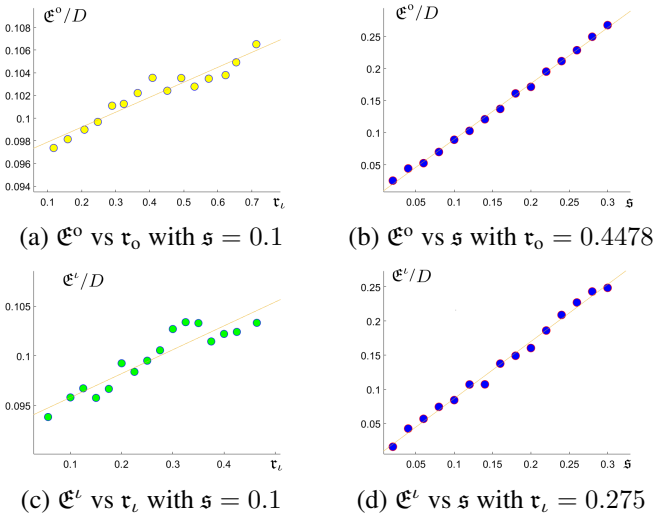
Then we continue solving Model I and Model II by ADMM presented in Algorithm 1 and Algorithm 2 respectively, where all the sub-problems can be solved in closed forms[6]. In each single iteration of Algorithms 1 and 2, the main cost comes from updating the low tubal rank components which involves FFT, IFFT and d_3 SVDs of $d_1 \times d_2$ matrices for tensors of size $d_1 \times d_2 \times d_3$. Hence Algorithms 1 and 2 have per-iteration complexity $O\big(KD \log D + D \sum_k \min(d_k, d_k^{-1} d_{k+1}^{-1} D)\big)$, which is the same order as ADMM-based algorithms for SNN (Gu, Gui, and Han 2014) and LatentNN (Tomioka and Suzuki 2013). The convergence of Algorithms 1 and 2 naturally holds since Problems (12) and (13) can be reformulated as the standard form of the two-block ADMM framework (Boyd et al. 2011).

Experiments

Correctness of the proposed error bounds

To validate the correctness of Theorem 3 and Theorem 4, we conduct simulations to check whether the proposed upper bounds in Eq. (14) can predict the scaling behavior of the estimation error.

Generation of $\mathcal{L}^*$. We generate the low-rank tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ 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.* $\mathcal{N}(0,1)$ 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 $\mathcal{L}^* = \mathcal{G}_K / \|\mathcal{G}_K\|_{l_\infty}$.



(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 $\mathfrak{r}_{\iota} = 0.275$

Figure 4: The element-wise estimation errors versus the rank complexity and sparse ratio for tensors in $\mathbb{R}^{40 \times 40 \times 40}$.

We form the sparse corruption tensor $\mathcal{S}^*$ by choosing its support uniformly at random according to (Lu et al. 2019). 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 Algorithm 1 and Algorithm 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. (14). *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.*

Fig. 4 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 Fig. 4, 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 $\mathfrak{s}$, 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.

Effectiveness of the proposed OITNNs

We evaluate effectiveness of the proposed norms in comparison with other nuclear norm-based models on real datasets. The competitor norms include SNN (Liu et al. 2013), LatentNN (Tomioka and Suzuki 2013), SqNN (Mu et al. 2014),

[5] The constraints on tensor spectral norm in Problem (13) are removed to get rid of $K(K-1)$ auxiliary variables in decoupling the constraints, and $K(K-1)$ projections on tensor spectral norm ball, which is rather space and time-consuming.

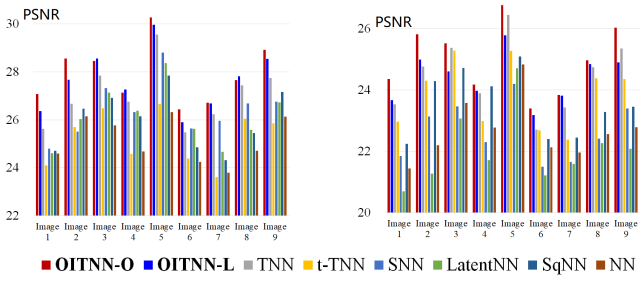
[6] Due to space limitation, the description of Algorithms 1 and 2 are shown in the supplementary material. Matlab implementations can be found in https://qibinzhao.github.io

TNN (Zhang and Aeron 2017), twist TNN (t-TNN) (Hu et al. 2017) and matrix nuclear norm (NN) (Candès and Tao 2010). We first conduct RTD on color images and color videos, 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 (12) 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.

Robust image recovery In this experiment, nine color images of size $256 \times 256 \times 3$ are tested (see Fig. 5(a)). Given a color image $\mathcal{M} \in \mathbb{R}^{d_1 \times d_2 \times 3}$, we randomly pick the support of $\mathcal{S}^*$ with probability $\mathfrak{s}$ and add *i.i.d.* Gaussian noise with standard deviation $\sigma = c\sigma_0$, where $\sigma_0 = \|\mathcal{M}\|_{\mathrm{F}}/\sqrt{3d_1d_2}$. We test two cases with $(\mathfrak{s}, c) = (0.05, 0.1)$ and $(0.15, 0.15)$. Given a color image and a corruption level, we test 10 times and report the averaged PSNR.

Parameters for OITNN-O are set as $w_1 : w_2 : w_3 = a_1 : 1 : a_1$, with $a_1 \in [0.1, 0.5]$, and OITNN-L $v_1 : v_2 : v_3 = 1 : a_2 : 1$ with $a_2 \in [0.025, 0.055]$. The weight parameters $\boldsymbol{\alpha}$ of SNN are chosen to satisfy $\alpha_1 : \alpha_2 : \alpha_3 = 1 : 1 : 0.01$ as suggested in (Liu et al. 2013). The "sparse/low-rank" parameter ratio μ/λ of NN is $1/\sqrt{d_2}$ (Candès et al. 2011), and $1/\sqrt{3d_2}$ for SqNN, TNN and t-TNN. We tune the "sparse/low-rank" parameter ratio for OITNN-O and OITNN-L. Other parameters of the algorithms are manually tuned.

(a) Nine test images.

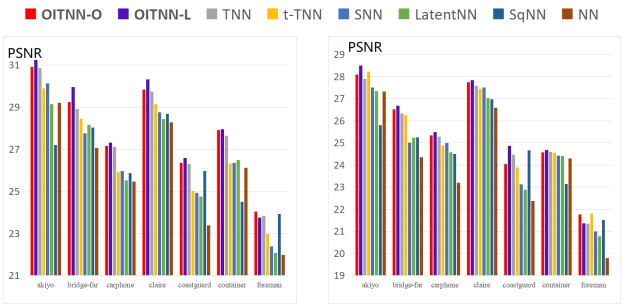


(b) $(\mathfrak{s}, c) = (0.05, 0.1)$ (c) $(\mathfrak{s}, c) = (0.15, 0.15)$

Figure 5: Test images and quantitative comparison of RTD models based on different norms on color images.

For quantitative comparison, PSNR values on the nine images are reported in Fig. 5(b)-(c). An visual example is shown in Row 1 of Fig. 7 for qualitative evaluation. As illustrated in Fig. 7, the proposed norms obtain higher visual quality than the competitor norms. According to Figs. 5 and 7, the proposed OITNN-O and OITNN-L have better performance in most cases on color images.

Robust video recovery The performance comparison is carried out on the widely used seven YUV videos[7]: akiyo, bridge-far, carphone, claire, coastguard, container and foreman. Due to computational limitation, we use the first 32 frames of each video, resulting in seven $144 \times 176 \times 3 \times 32$ tensors. We conduct robust video recovery against sparse corruptions and Gaussian noise with $(\mathfrak{s}, c) = (0.15, 0.15)$ and $(0.2, 0.2)$. NN is tested for matrices of size $(144 \cdot 172 \cdot 3) \times 32$; SqNN is tested for matrices of size $(144 \cdot 32) \times (176 \cdot 3)$; TNN and t-TNN are tested on 3 (the channel number) tensors of size $(144 \times 176 \times 32)$. The "sparse/low-rank" parameter ratio of NN, SqNN, TNN and t-TNN are set by the suggestion of RPCA and TRPCA (Candès et al. 2011; Lu et al. 2019). We tune other parameters for better performances in most cases. The PSNR values are reported in Fig. 6. It can be seen that the proposed OITNNs-based models have better performances thanks to their flexibility in exploiting the multi-orientation correlations in color videos.



(a) $(\mathfrak{s}, c) = (0.15, 0.15)$ (b) $(\mathfrak{s}, c) = (0.2, 0.2)$

Figure 6: Quantitative comparison on seven color videos.

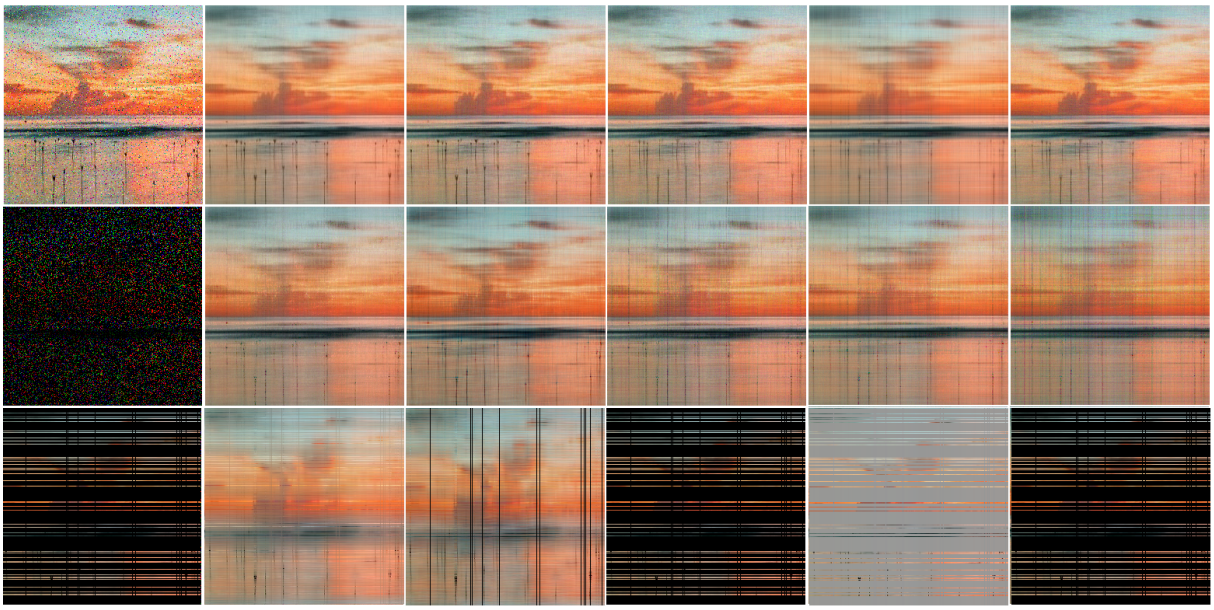
Robust recovery of UGV data Experiments on a dataset [8] for unmanned ground vehicle (UGV) are also conducted. This dataset contains a sequence of gray camera images and point cloud data acquired from a Velodyne HDL-64E LiDAR. A scenario containing 32 frames (Frame Nos. 65-96) is selected. The camera data is resized to a tensor of size $128 \times 256 \times 32$, and the point cloud data is formatted into two tensors (distance data tensor and intensity data tensor) sized $64 \times 436 \times 32$. We conduct RTD against sparse corruptions and Gaussian noise with $(\mathfrak{s}, c) = (0.15, 0.15)$ and $(0.3, 0.2)$. The parameters are tuned for better performances in most cases. The PSNR values are reported in Table 2. We can see that the proposed norms perform better on UGV data.

Table 2: The quantitative comparison on the dataset for UGV.

Norms	$(\mathfrak{s}, c) = (0.15, 0.15)$			$(\mathfrak{s}, c) = (0.3, 0.2)$		
	Camera	Distance	Intensity	Camera	Distance	Intensity
NN	23.12	24.49	21.61	21.43	23.41	20.91
SNN	24.29	25.32	22.21	22.55	24.49	21.56
TNN	25.54	25.86	22.88	23.36	25.08	22.08
t-TNN	25.19	25.92	23.13	23.22	25.37	22.35
SqNN	23.45	24.71	21.86	21.75	23.77	21.38
LatentNN	23.61	24.67	21.97	21.89	23.79	21.45
OITNN-O	**26.12**	25.97	**23.31**	**24.08**	25.42	22.41
OITNN-L	25.79	**26.35**	23.28	23.87	**25.79**	**22.45**

[7]The videos are available from https://sites.google.com/site/subudhibadri/fewhelpfuldownloads.

[8]*Scenario B* and *Scenario B-additional* dataset from http://www.mrt.kit.edu/z/publ/download/velodynetracking/dataset.html

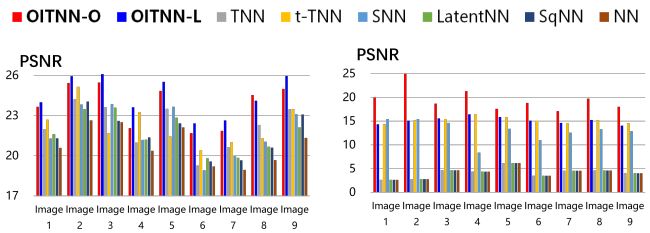


(a) Observation (b) OITNN-O (c) OITNN-L (d) TNN (e) SNN (f) LatentNN

Figure 7: 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%).

Color 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 Fig. 5(a). In Setting I, 90% of the entries are missing randomly, whereas in Setting II 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 Setting I and Setting II are shown in Row 2 and Row 3 of Fig. 7, respectively. The quantitative comparison in PSNR is presented in Fig. 8. Thank to their ability in exploiting multi-orientational low-rankness in spectral domain, the proposed OITNNs outperform the competitors in most cases, especially when the missing pattern conflicts with low-rankness in original domain (like Setting II).



(a) Setting I (b) Setting II

Figure 8: Quantitative comparison in image inpainting.

Conclusions

Two new tensor norms for general K-way ($K \geq 3$) tensors are first defined to exploit the low-rankness in spectral domain for all orientations. We then adopt them to robust tensor recovery and rigorously establish upper bounds on the estimation error. Correctness of the error bounds is verified through simulation study. Experiments on real datasets demonstrate the effectiveness of the proposed norms. A main drawback of the proposed models is that we cannot automatically determine the optimal tuning parameters, and it is interesting to develop a suitable tool for parameter tuning (like (Zhao et al. 2015b)) as future work. Future research directions also include developing fast algorithms for the proposed models using techniques like (Wang et al. 2019; Wang, Jin, and Tang 2020).

Acknowledgments

The authors thank Jin Wang, Bo Wang and Dongxu Wei for their long time support. This work is partially supported by National Natural Science Foundation of China [Grant Nos. 61872188, U1713208, 61972204, 61672287, 61861136011, 61773215, 61703209, 61972212]. This work is also partially supported by JSPS KAKENHI (Grant No. 17K00326), and the National Key Research and Development Program Intergovernmental International Science and Technology Innovation Cooperation Project (MOST-RIKEN) under Grant 2017YFE0116800.

References

Agarwal, A.; Negahban, S.; and Wainwright, M. J. 2012. Noisy matrix decomposition via convex relaxation: Optimal

rates in high dimensions. *The Annals of Statistics* 1171–1197.

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

Candès, E. J., and Tao, T. 2010. The power of convex relaxation: near-optimal matrix completion. *IEEE TIT* 56(5):2053–2080.

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):11.

Fazel, M. 2002. *Matrix rank minimization with applications*. Ph.D. Dissertation, PhD thesis, Stanford University.

Gu, Q.; Gui, H.; and Han, J. 2014. Robust tensor decomposition with gross corruption. In *NIPS 2014*, 1422–1430.

Harshman, R. A. 1970. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis.

Hillar, C. J., and Lim, L. 2009. Most tensor problems are np-hard. *Journal of the ACM* 60(6):45.

Hu, W.; Tao, D.; Zhang, W.; Xie, Y.; and Yang, Y. 2017. The twist tensor nuclear norm for video completion. *IEEE TNNLS* 28(12):2961–2973.

Huang, B.; Mu, C.; Goldfarb, D.; and Wright, J. 2015. Provable models for robust low-rank tensor completion. *Pacific Journal of Optimization* 11(2):339–364.

Kilmer, M. E.; Braman, K.; Hao, N.; and Hoover, R. C. 2013. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications* 34(1):148–172.

Klopp, O.; Lounici, K.; and Tsybakov, A. B. 2017. Robust matrix completion. *Probability Theory and Related Fields* 169(1-2):523–5642.

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

Liu, J.; Musialski, P.; Wonka, P.; and Ye, J. 2013. Tensor completion for estimating missing values in visual data. *IEEE TPAMI* 35(1):208–220.

Liu, X. Y.; Aeron, S.; Aggarwal, V.; and Wang, X. 2016a. Low-tubal-rank tensor completion using alternating minimization. *arXiv preprint arXiv:1610.01690*.

Liu, X.; Aeron, S.; Aggarwal, V.; Wang, X.; and Wu, M. 2016b. Adaptive sampling of rf fingerprints for fine-grained indoor localization. *IEEE Transactions on Mobile Computing* 15(10):2411–2423.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2016. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *CVPR 2016*, 5249–5257.

Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; and Yan, S. 2019. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE TPAMI* 1–1.

Mu, C.; Huang, B.; Wright, J.; and Goldfarb, D. 2014. Square deal: Lower bounds and improved relaxations for tensor recovery. In *ICML 2014*, 73–81.

Oseledets, I. V. 2011. Tensor-train decomposition. *SIAM Journal on Scientific Computing* 33(5):2295–2317.

Tomioka, R., and Suzuki, T. 2013. Convex tensor decomposition via structured schatten norm regularization. In *NIPS 2013*, 1331–1339.

Tomioka, R.; Suzuki, T.; Hayashi, K.; and Kashima, H. 2011. Statistical performance of convex tensor decomposition. In *NIPS 2011*, 972–980.

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

Wang, A.; Song, X.; Wu, X.; Lai, Z.; and Jin, Z. 2019. Latent schatten tt norm for tensor completion. In *ICASSP 2019*, 2922–2926. IEEE.

Wang, A.; Jin, Z.; and Tang, G. 2020. Robust tensor decomposition via t-svd: Near-optimal statistical guarantee and scalable algorithms. *Signal Processing* 167:107319.

Wei, D.; Wang, A.; Feng, X.; Wang, B.; and Wang, B. 2018. Tensor completion based on triple tubal nuclear norm. *Algorithms* 11(7).

Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2017. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *IJCV* (4):1–23.

Yokota, T.; Erem, B.; Guler, S.; Warfield, S. K.; and Hontani, H. 2018. Missing slice recovery for tensors using a low-rank model in embedded space. *arXiv preprint arXiv:1804.01736*.

Yuan, M., and Zhang, C. H. 2016. On tensor completion via nuclear norm minimization. *Foundations of Computational Mathematics* 16(4):1–38.

Zhang, Z., and Aeron, S. 2017. Exact tensor completion using t-svd. *IEEE TSP* 65(6):1511–1526.

Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; and Kilmer, M. 2014. Novel methods for multilinear data completion and de-noising based on tensor-svd. In *CVPR 2014*, 3842–3849.

Zhao, Q.; Meng, D.; Kong, X.; Xie, Q.; Cao, W.; Wang, Y.; and Xu, Z. 2015a. A novel sparsity measure for tensor recovery. In *ICCV 2015*, 271–279.

Zhao, Q.; Zhou, G.; Zhang, L.; Cichocki, A.; and Amari, S.-I. 2015b. Bayesian robust tensor factorization for incomplete multiway data. *IEEE TNNLS* 27(4):736–748.

Zhou, P., and Feng, J. 2017. Outlier-robust tensor pca. In *CVPR 2017*, 3938–3946.