NOISY TENSOR COMPLETION VIA ORIENTATION INVARIANT TUBAL NUCLEAR NORM*

ANDONG WANG†, GUOXU ZHOU‡, ZHONG JIN AND QIBIN ZHAO

Abstract: Noisy tensor completion is considered for reconstructing multi-way data from a small fraction of noisy observations caused by many practical reasons such as sensor dead pixels, communication loss, electromagnetic interferences, etc., and can provide a powerful pre-processing tool for subsequent tasks like classification, unmixing, and target detection in many remote sensing applications. Thanks to the capability to simultaneously model multi-orientational spectral low-rankness, the recently proposed Orientation Invariant Tubal Nuclear Norm (OITNN) empirically outperforms traditional tensor nuclear norms in tensor recovery tasks. However, its theoretical and empirical performance for noisy tensor completion is still insufficiently explored. To further unleash and understand the modeling potential of OITNN, this paper formulates an OITNN penalized least squares estimator for noisy tensor completion and studies its statistical and empirical performance. We compute the proposed estimator using an ADMM-based algorithm and rigorously characterize the statistical performance by establishing an upper bound on the estimation error. Simulation studies on nine different types of remote sensing data demonstrate the effectiveness of the OITNN for noisy tensor completion.

Key words: *tensor completion, tensor SVD, estimation error, ADMM*

Mathematics Subject Classification: *15A69, 90C25, 47N10, 90C59*

1 Introduction

Benefiting from the rapid development of modern sensor technology, there are emerging more and more tensor (i.e., multi-dimensional array) data, such as visible light images, multispectral images, hyperspectral data, thermal infrared data, Light Detection and Ranging (LiDAR), Synthetic Aperture Radar (SAR), etc. [16, 46, 49, 36, 35]. However, due to many reasons like sensor failures, communication losses, occlusion by objects, shortage of instruments, electromagnetic interferences, and so on in some empirical cases, the tensor data may be incomplete and noisy, leading to reduced accuracy of the subsequent data processing in applications such as classification, unmixing, and target detection [45, 23, 44, 8]. Thus, it

is necessary to study the noisy tensor completion which aims to reconstruct a data tensor from a fraction of noisy observations.

It is generally impossible to recover the unknown entries of a tensor from only a small number of noisy observations without any assumptions on the underlying tensor. To make the problem of noisy tensor completion well-posed, the mostly adopted assumption is that the underlying tensor has some low-rankness, such that it can be represented by a relatively small number of latent parameters. By estimating the latent parameters of a low-rank tensor from limited noisy observations, the unknown entries then can be well reconstructed. Recently, the low-rank tensor models have become a paradigm in visual data recovery [13, 14, 49].

As a higher-order generalization of matrix low-rankness [10], tensor low-rankness is often characterized by different tensor rank functions [22], e.g. the CANDECOMP/PARAFAC (CP) rank [3], Tucker rank [40], Tensor Train (TT) rank [32], Tensor Ring (TR) rank [51], etc. From a signal processing perspective, all the above exampled tensor rank functions are defined in the original domain of the tensor signal. A typically different example is the recently proposed tensor tubal rank [50, 18] which measures low-rankness in the frequency domain defined via Discrete Fourier Transform (DFT). It is advocated in [24, 23, 43, 45] that the tensor tubal rank can simultaneously reflect both low-rank and smooth structures of the signal tensor, and is thus ideal for the analysis and processing of various visual data which often possess both low-rankness and smoothness at the same time [52, 13].

Although the low-tubal-rank tensor model is broadly used in imagery data recovery, a typical defect of it is the orientation sensitivity owing to low-rankness strictly defined along the tubal orientation which makes it fail to simultaneously exploit frequency low-rankness in multiple orientations. To address this issue, Wang et al. [43] proposed the so called Orientation Invariant Tubal Nuclear Norm (OITNN) to exploit multi-orientational frequency low-rankness for an arbitrary K-way ($K \geq 3$) tensors, which achieve higher accuracy in robust tensor decomposition over traditional tensor low-rank models.

Motivated by the promising performance of OITNN, we introduce it to solve the problem of noisy tensor completion in this paper (see Fig. 1). Specifically, the contributions of this work are three-fold:

1). First, we formulate an OITNN-based estimator to estimate the underlying tensor from partial noisy observations. Benefiting from the orientational invariance of OITNN, the proposed estimator can simultaneously exploit the spectral low-rankness of the underlying tensor data along all orientations.

2). Algorithmically, we propose an algorithm based on Alternating Direction Method of Multipliers (ADMM) [12, 34] to compute the estimator and evaluate its effectiveness on nine different types of remote sensing data.

3). Statistically, the statistical performance of the proposed estimator is analyzed by establishing an upper bound on the estimation error.

The rest of this paper is organized as follows. First, the notations and preliminaries of tensor Singular Value Decomposition (t-SVD) and OITNN are introduced in Sec. 2. Then, the proposed estimator for noisy tensor completion is formulated in Sec. 3. To compute the proposed estimator, we design an ADMM-based algorithm in Sec. 4. To understand the statistical behavior of the estimator, we establish an upper bound on the estimation error in Sec. 5. Extensive experimental results on nine different types of remote sensing data are reported in Sec. 6. This work is briefly summarized in Sec. 7. Proofs for the theoretical results are given in the appendix.
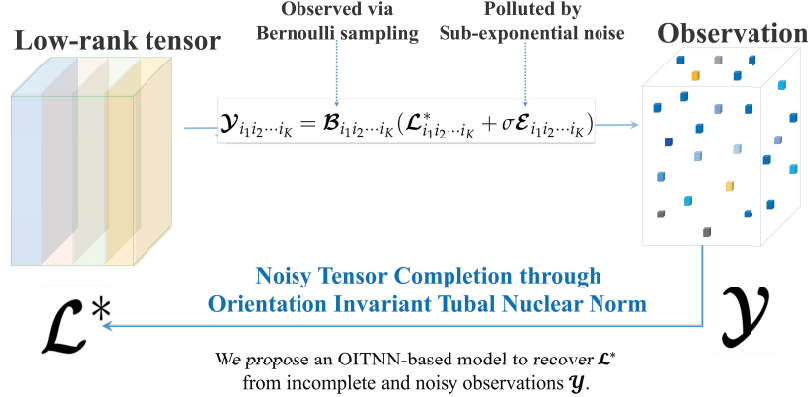
Low-rank tensor   Observed via Bernoulli sampling   Polluted by Sub-exponential noise   Observation

$$\boldsymbol{\mathcal{Y}}_{i_1 i_2 \cdots i_K} = \boldsymbol{\mathcal{B}}_{i_1 i_2 \cdots i_K}(\boldsymbol{\mathcal{L}}^*_{i_1 i_2 \cdots i_K} + \sigma \boldsymbol{\mathcal{E}}_{i_1 i_2 \cdots i_K})$$

Noisy Tensor Completion through Orientation Invariant Tubal Nuclear Norm

$\boldsymbol{\mathcal{L}}^*$ $\boldsymbol{\mathcal{Y}}$

We propose an OITNN-based model to recover $\boldsymbol{\mathcal{L}}^*$ from incomplete and noisy observations $\boldsymbol{\mathcal{Y}}$.

Figure 1: An illustration of this work.

2 Notations and Preliminaries

Notations. We use lowercase boldface, uppercase boldface, and calligraphy letters to denote vectors (e.g., $\boldsymbol{v}$), matrices (e.g., $\boldsymbol{M}$), and tensors (e.g., $\mathcal{T}$), respectively. For any positive integer n, let $[n] := \{1, \ldots, n\}$ for notational simplicity. For any real numbers a, b, let $a \vee b = \max\{a, b\}$, and $a \wedge b = \min\{a, b\}$. 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 $\mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$. We use c, c', c_1 etc. to denote constants whose values can vary from line to line. 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 a matrix $\boldsymbol{M} \in \mathbb{R}^{d_1 \times d_2}$, its nuclear norm and spectral norm are defined as $\|\boldsymbol{M}\|_* := \sum_i \sigma_i$ and $\|\boldsymbol{M}\| := \max_i \sigma_i$ respectively, where $\{\sigma_i \mid i \in [d_1 \wedge d_2]\}$ are its singular values. 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

$$\|\mathcal{T}\|_0 := \|\mathrm{vec}(\mathcal{T})\|_0, \ \|\mathcal{T}\|_1 := \|\mathrm{vec}(\mathcal{T})\|_1, \ \|\mathcal{T}\|_{\mathrm{F}} := \|\mathrm{vec}(\mathcal{T})\|_2, \ \|\mathcal{T}\|_\infty := \|\mathrm{vec}(\mathcal{T})\|_\infty,$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operation of a tensor [21]. Given $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, let $\boldsymbol{T}^{(i)} := \mathcal{T}(:,:,i)$ denotes its ith frontal slice. Other notations are introduced at their first appearance.

2.1 Tensor Singular Value Decomposition

We briefly recall the tensor singular value decomposition.

Definition 2.1 (t-product [18]). 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 $(i,j)^{\mathrm{th}}$ 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.

Definition 2.2 (Tensor transpose [18]). 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.
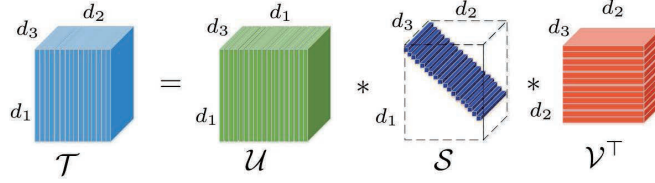
Figure 2: Illustration of t-SVD [44].

Definition 2.3 (Identity tensor [18])**.** 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 2.4 (f-diagonal tensor [18])**.** A tensor is called f-diagonal if each frontal slice of the tensor is a diagonal matrix.

Definition 2.5 (Orthogonal tensor [18])**.** 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 are further defined for the convenience of analysis.

Definition 2.6 (Block-diagonal matrix [18])**.** Let $\overline{\boldsymbol{T}}$ (or $\overline{\mathcal{T}}$) denote the block-diagonal matrix of the tensor $\widetilde{\mathcal{T}}$ in the Fourier domain, i.e.,

$$\overline{\boldsymbol{T}} := \begin{bmatrix} \widetilde{\boldsymbol{T}}^{(1)} & & \\ & \ddots & \\ & & \widetilde{\boldsymbol{T}}^{(d_3)} \end{bmatrix} \in \mathbb{C}^{d_1 d_3 \times d_2 d_3} \tag{2.1}$$

Then, t-SVD can be defined as follows (see Fig. 2).

Definition 2.7 (t-SVD, tubal rank [18])**.** Any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has a tensor singular value decomposition as

$$\mathcal{T} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top, \tag{2.2}$$

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 $\mathcal{S}$:

$$r_{\text{tb}}(\mathcal{T}) := \#\big\{i \,\big|\, \mathcal{S}(i,i,:) \neq \boldsymbol{0}\big\}. \tag{2.3}$$

Definition 2.8 (Average rank, tubal nuclear norm, tensor spectral norm [26])**.** 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 $\text{rank}_{\text{avg}}(\cdot)$, tubal nuclear norm $\|\cdot\|_\star$ of $\mathcal{T}$ are defined as the averaged rank and nuclear norm of frontal slices of $\widetilde{\boldsymbol{T}}$:

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

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

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

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

$$\mathcal{T} \in \Re^{d_1 \times \cdots \times d_K}$$

mode 1
mode 2
mode K
mode k
mode t

mode-(k, t)

3D-unfolding

Permute

Reshape

mode k
mode t
mode $k-1$
mode K
mode 1

$$\mathcal{T}_{[k,t]} \in \Re^{d_k \times \frac{D}{d_k d_t} \times d_t}$$
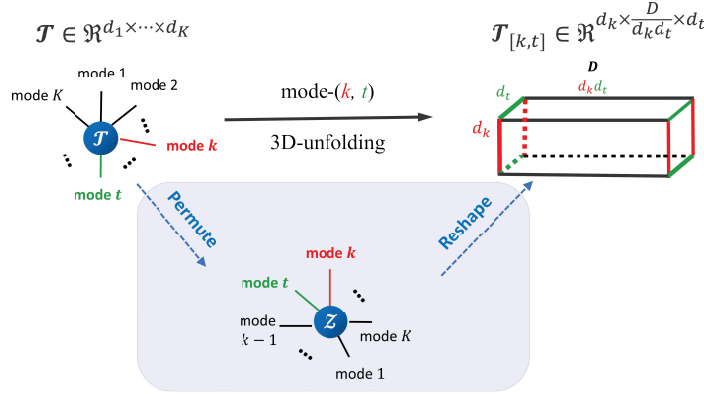
D

d_t $d_k d_t$

d_k

Figure 3: Illustration of 3D-unfolding.

It has been shown in [26] that just like the relationship between the matrix nuclear norm and matrix rank, TNN is the tightest convex relaxation of tensor average rank for tensors within the unit ball of tensor spectral norm. Thus according to Definition 2.8, TNN indeed imposes a low-average-rank structure which captures low-rankness in the spectral domain. By viewing low-rankness as a special type of sparsity, and recalling the wisdom from signal processing that "sparsity in the spectral domain usually means smoothness in the original domain", it can be said that a low-average-rank structure may simultaneously model both low-rankness and smoothness of the tensor data. This can explain the superior performance of TNN in visual data (like images and videos) restoration to some extent since many visual data are both low-rank and smooth. However, as pointed out in [43], TNN is orientation sensitive in the sense that just mode-3 fibers are chosen to perform DFT, and thus only spectral low-rankness along orientation of mode-3 can be exploited, leading to very limited representation ability for general higher-way tensors with spectral low-rankness along multiple orientations.

2.2 Orientation Invariant TNN

To overcome the orientation sensitivity of TNN, a new tensor norm named orientation invariant TNN is proposed in [43] which heavily relies on a novel tensor 3D-unfolding operation as follows.

Definition 2.9 (mode-(k, t) 3D-unfolding [43]). For different $k, t \in [K]$, 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 $\mathcal{T}_{[k,t]} \in \mathbb{R}^{d_k \times (D/(d_k d_t)) \times d_t}$ obtained by the following two steps (See Fig. 3).

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 kth 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

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

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

Inspired the spirit of t-product which treats a 3-way tensor as a "2-way array" whose entries are tubes, we can also take a K-way ($K \geq 3$) tensor $\mathcal{T}$ as a "(K-1)-way array $\mathscr{T}$" of

size $d_1 \times d_2 \times \cdots \times d_{t-1} \times d_{t+1} \times \cdots \times d_K$ whose entries are mode-t fibers. In this manner, the mode-(k, t) 3D-unfolding $\mathcal{T}_{[k,t]}$ of $\mathcal{T}$ can also be analogously seen as a "mode-k unfolding" of $\mathscr{T}$ whose entries are mode-t fibers.

The mode-(k, t) 3d-unfolding is defined to transform an original K-way tensor to 3-way, where the mode-k is re-indexed to the 1-st mode of the resulted tensor, while the mode-t is re-indexed to the 3-rd mode. Thus, the average rank (as well as the TNN) of the resulted 3-way tensor can measure spectral low-rankness along the t-th orientation of the original tensor. Then, if we fix the mode-t, we can in general choose any $k \neq t$ to obtain $K - 1$ different 3-way tensors, whose average ranks (and TNNs) all measures low-rankness along the same (i.e., the t-th) orientation. Although it is considerable and feasible for us to simultaneously consider the $K - 1$ resulted 3-way tensors, this would significantly bring with more complexities of notation, exposition, and computation. For simplicity, we follow the idea of [43] which straightforwardly lets $k = t - 1$, and considers the mode-$(t-1, t)$ 3d-unfolding, or equivalently the mode-$(k, k+1)$ 3d-unfolding.

For the ease of presentation, let

$$\mathcal{T}_{[k]} := \mathcal{T}_{[k,k+1]}, \quad k = 1, 2, \ldots, K$$

and name it the *mode-k 3D-unfolding* of $\mathcal{T}$. For the ease of notations, 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 follows

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

Based on the tensor 3D-unfolding, we are now ready to introduce two new tensor ranks [43].

Definition 2.10 (Orientation invariant tubal rank, Orientation invariant average rank [43]). For any tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, its Orientation Invariant Tubal Rank (OITR) $\vec{\mathfrak{r}}_{\mathrm{t}}$ and Orientation Invariant Average Rank (OIAR) $\vec{\mathfrak{r}}_{\mathrm{a}}$ 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:

$$
\begin{aligned}
\vec{\mathfrak{r}}_{\mathrm{t}}(\mathcal{T}) &:= \big(r_{\mathrm{tb}}(\mathcal{T}_{[1]}), \ldots, r_{\mathrm{tb}}(\mathcal{T}_{[K]})\big)^\top \in \mathbb{R}^K, \\
\vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{T}) &:= \big(\mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[1]}), \ldots, \mathrm{rank}_{\mathrm{avg}}(\mathcal{T}_{[K]})\big)^\top \in \mathbb{R}^K.
\end{aligned}
\tag{2.4}
$$

Thus, OITR is a complexity measure in the original domain, whereas the OIAR measures low-rankness in the spectral domain as shown in Eq. (2.4). According to Lemma 1 in [43], the following relationship between the OITR, OIAR, and the classical Tucker rank $\vec{\mathfrak{r}}_{\mathrm{Tucker}}$ holds

$$\vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{T}) \leq \min\{\vec{\mathfrak{r}}_{\mathrm{t}}(\mathcal{T}), \vec{\mathfrak{r}}_{\mathrm{Tucker}}(\mathcal{T})\}, \tag{2.5}$$

where the partial order "$\leq$" is defined entry-wisely. Eq. (2.5) shows that if a tensor has a low OITR or Tucker rank, then it will have a low OIAR, which indicates the low OIAR assumption is weaker than the popular low Tucker rank assumption [43].

Motivated by the definition of TNN, it is natural to relax the average ranks in OIAR to their tightest convex envelops in all orientations and obtain the following orientation invariant norm:

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

According to the definitions of OITR and OIAR, either of them has K elements which represent spectral low-rankness along K different orientations, and thus the dimensionality K could not be reduced from the standpoint of signal modeling.

Definition 2.11 (Overlapped orientation invariant tubal nuclear norm [43]). 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 follows:

$$\|\mathcal{T}\|_{\star \mathrm{o}} := \sum_{k=1}^{K} w_k \|\mathcal{T}_{[k]}\|_{\star}, \tag{2.6}$$

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

As discussed in [43], OITNN-O indeed imposes a low OIAR structure, which exploits low-rankness in the spectral domain of all orientations. Thus in the original domain, it models a data tensor as simultaneously low tubal rank in all orientations.

3 Noisy Tensor Completion via OITNN

In this section, we first specify the observation model in the noisy tensor completion setting and then formulate the OITNN-based model.

3.1 The Observation Model

Let $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ be the unknown tensor to be estimated. Suppose it's partial noisy observation $\mathcal{Y} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is measured through a noisy channel specified by the following observation model:

$$\mathcal{Y}_{i_1 i_2 \ldots i_K} = \mathcal{B}_{i_1 i_2 \ldots i_K}(\mathcal{L}^*_{i_1 i_2 \ldots i_K} + \sigma \mathcal{E}_{i_1 i_2 \ldots i_K}), \quad \forall (i_1, i_2, \ldots, i_K) \in [d_1] \times [d_2] \times \ldots [d_K] \tag{3.1}$$

where $\mathcal{B} = (\mathcal{B}_{i_1 i_2 \ldots i_K}) \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is *the sampling tensor* whose entries are independent Bernoulli random variables, the positive constant σ is the known standard deviation of random noises, and $\mathcal{E}_{i_1 i_2 \ldots i_K} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ represents *the noise tensor* whose entries are independent random variables with normalized variance. For simplicity of exposition, we make the following assumptions on the sampling scheme and the noise setting:

Assumption 3.1 (Uniform sampling). All the entries of the underlying tensor $\mathcal{L}^*$ are sampled independently with the same probability p. Equivalently, all the entries of the sampling tensor $\mathcal{B}$ are *i.i.d.* sampled from Bernoulli distribution of parameter $p \in (0, 1]$, i.e.,

$$\mathcal{B}_{i_1 i_2 \ldots i_K} = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p. \end{cases} \tag{3.2}$$

Assumption 3.2 (Sub-exponential noises). The noise variables $\mathcal{E}_{i_1 i_2 \ldots i_K}$ are sub-exponential random variables with zero expectation and unit variance, i.e.,

$$\mathbb{E}[\mathcal{B}_{i_1 i_2 \ldots i_K}] = 0, \text{ and } \operatorname{var}(\mathcal{B}_{i_1 i_2 \ldots i_K}) = 1$$

and there exists a positive constant ϱ as the smallest number such that the following inequality holds

$$\mathbb{E}[\exp(\varrho^{-1}|\mathcal{B}_{i_1 i_2 \ldots i_K}|)] \leq +\infty, \quad \forall (i_1, i_2, \ldots, i_K) \in [d_1] \times [d_2] \times \ldots [d_K] \tag{3.3}$$

Remark 3.1. In this paper, we assume uniform sampling only for the ease of exposition. Both the proposed estimator in Eq. (3.5) and Algorithm 1 are not limited to uniform sampling. One can easily adopts a K-way generalization of the Assumption 1 in [44] and uses very similar arguments to extend the estimation error in Sec. 5 to a more generalized sampling schemes.

3.2 The Proposed Model for Noisy Tensor Completion

The goal of noisy tensor completion is to recover the unknown tensor $\mathcal{L}^*$ from the incomplete noisy observation $\mathcal{Y}$ satisfying the observation model (3.1). Our idea is to seek an estimation of the underlying tensor by using a least squares estimator penalized by the OIAR defined in (2.4) to exploit spectral low-rankness along all orientations:

$$\min_{\mathcal{L}} \ \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_{\mathrm{F}}^2 + \lambda_0\|\vec{\mathfrak{r}}_{\mathrm{a}}(\mathcal{L})\|_1 \tag{3.4}$$

where $\odot$ represents the entry-wise multiplication between tensors, and λ_0 is a tunable regularization parameter which balances the low-rankness and the fidelity term.

However, general rank minimization is NP-hard [10, 2], making it extremely hard to soundly solve Problem (3.4). For tractable low-OIAR optimization, we follow the most common idea to relax the non-convex function OIAR to its convex surrogate (i.e., the overlapped OITNN $\|\cdot\|_{\star\mathrm{o}}$), and obtain the following estimator:

$$\hat{\mathcal{L}} \in \mathrm{argmin}_{\|\mathcal{L}\|_\infty \leq \mathbf{a}} \ \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_{\mathrm{F}}^2 + \lambda\|\mathcal{L}\|_{\star\mathrm{o}} \tag{3.5}$$

where $\mathbf{a} > 0$ is a known constant constraining the magnitude of entries in $\mathcal{L}^*$. The additional constraint $\|\mathcal{L}\|_\infty \leq \mathbf{a}$ is very mild since most signals are of limited energy in real applications. It can also provide a theoretical benefit to exclude the "spiky" tensors, which is important in controlling the identifiability of $\mathcal{L}^*$. Such "non-spiky" constraints are also imposed in previous literatures [30, 19, 14, 49], playing a key role in bounding the estimation error.

Then, it is natural to ask the following questions:

Q1: *How to compute the proposed estimator?*

Q2: *How well can the proposed estimator estimate $\mathcal{L}^*$?*

We first discuss **Q1** in Sec. 4 and then answer **Q2** in Sec. 5.

4 Optimization Algorithm

In this section, we answer **Q1** by designing an algorithm based on ADMM to compute the proposed estimator.

For notational simplicity, we recall the definition 3D-unfolding operator for $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ as $\mathfrak{F}_k(\mathcal{T}) := \mathcal{T}_{[k]}$ and its inverse $\mathfrak{F}_k^{-1}(\cdot)$ such that $\mathfrak{F}_k^{-1}(\mathcal{T}_{[k]}) = \mathcal{T}$. To solve Problem (3.5), the first step is to introduce auxiliary variables $\{\mathcal{K}^{(k)}\}$ as follows:

$$\min_{\mathcal{L},\{\mathcal{K}^{(k)}\}} \ \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_{\mathrm{F}}^2 + \lambda\sum_{k=1}^{K} w_k\|\mathcal{K}_{[k]}^{(k)}\|_\star + \boldsymbol{\delta}_{\mathbf{a}}^\infty(\mathcal{L}),$$
$$\text{s.t.} \quad \mathcal{L} = \mathcal{K}^{(k)}, \forall k \in [K] \tag{4.1}$$

where $\boldsymbol{\delta}_{\mathbf{a}}^\infty(\mathcal{L})$ is the indicator function of tensor l_∞-norm ball defined as

$$\boldsymbol{\delta}_{\mathbf{a}}^\infty(\mathcal{L}) = \begin{cases} 0 & \|\mathcal{L}\|_\infty \leq \mathbf{a} \\ +\infty & \|\mathcal{L}\|_\infty > \mathbf{a} \end{cases}$$

The augmented Lagrangian of Problem (4.1) is given by

$$L_\rho(\mathcal{L}, \{\mathcal{K}^{(k)}, \{\mathcal{W}^{(k)}\}\}) = \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_F^2 + \lambda \sum_{k=1}^{K} w_k \|\mathcal{K}_{[k]}^{(k)}\|_\star + \boldsymbol{\delta}_\mathbf{a}^\infty(\mathcal{L})$$
$$+ \sum_{k=1}^{K} \left(\langle \mathcal{W}^{(k)}, \mathcal{L} - \mathcal{K}^{(k)} \rangle + \frac{\rho}{2}\|\mathcal{L} - \mathcal{K}^{(k)}\|_F^2 \right) \tag{4.2}$$

where $\{\mathcal{W}^{(k)}\}$ are Lagrangian multipliers and $\rho > 0$ is a penalty parameter.

According to the framework of ADMM [1, 12], we update the variables alternatively after the t-th iteration as follows:

Update $\mathcal{L}$: We update $\mathcal{L}$ by keeping the other variables fixed:

$$\mathcal{L}_{t+1} = \operatorname{argmin}_\mathcal{L} L_\rho(\mathcal{L}, \{\mathcal{K}_t^{(k)}, \{\mathcal{W}_t^{(k)}\}\})$$
$$= \operatorname{argmin}_\mathcal{L} \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_F^2 + \boldsymbol{\delta}_\mathbf{a}^\infty(\mathcal{L}) + \sum_{k=1}^{K} \frac{\rho}{2}\|\mathcal{L} - (\mathcal{K}_t^{(k)} - \rho^{-1}\mathcal{W}_t^{(k)})\|_F^2$$
$$= \operatorname{Proj}_\mathbf{a}^{\|\cdot\|_\infty} \left(\left(\mathcal{B} \odot \mathcal{Y} + \rho \sum_{k=1}^{K} \mathcal{K}_t^{(k)} - \sum_{k=1}^{K} \mathcal{W}_t^{(k)}\right) \oslash (\mathcal{B} + K\rho\mathbf{1}) \right) \tag{4.3}$$

where $\oslash$ denotes entry-wise division, $\mathbf{1} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is the tensor with all entires being 1, and $\operatorname{Proj}_\mathbf{a}^{\|\cdot\|_\infty}(\cdot)$ is the projector onto the tensor l_∞-norm ball of radius $\mathbf{a}$ which is given by $\operatorname{Proj}_\mathbf{a}^{\|\cdot\|_\infty}(\mathcal{L}) = \operatorname{sign}(\mathcal{L}) \odot \min(|\mathcal{L}|, \mathbf{a})$ [43].

Update $\{\mathcal{K}^{(k)}\}$: We update each $\{\mathcal{K}^{(k)}\}$ separately with $\mathcal{L}$ and $\{\mathcal{W}^{(k)}\}$ fixed as follows:

$$\mathcal{K}_{t+1}^{(k)} = \operatorname{argmin}_{\mathcal{K}^{(k)}} L_\rho(\mathcal{L}_{t+1}, \{\mathcal{K}^{(k)}, \{\mathcal{W}_t^{(k)}\}\})$$
$$= \operatorname{argmin}_{\mathcal{K}^{(k)}} \lambda w_k \|\mathcal{K}_{[k]}^{(k)}\|_\star + \frac{\rho}{2}\|\mathcal{K}^{(k)} - (\mathcal{L}_{t+1} + \rho^{-1}\mathcal{W}_t^{(k)})\|_F^2 \tag{4.4}$$
$$= \mathfrak{F}_k^{-1} \left(\operatorname{Prox}_{\rho^{-1}\lambda w_k}^{\|\cdot\|_\star}(\mathfrak{F}_k(\mathcal{L}_{t+1} + \rho^{-1}\mathcal{W}_t^{(k)})) \right)$$

where $\operatorname{Prox}_\tau^{\|\cdot\|_\star}(\mathcal{K})$ is the proximal operator of TNN at point $\mathcal{K}$ given in the following lemma:

Lemma 4.1 (Proximal operator of TNN [44]). *Let tensor $\mathcal{T}_0 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ with t-SVD $\mathcal{T}_0 = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, 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 $\|\cdot\|_\star$ at point $\mathcal{T}_0$ with parameter τ, can be computed as follows:*

$$\operatorname{Prox}_\tau^{\|\cdot\|_\star}(\mathcal{K}_0) := \operatorname{argmin}_\mathcal{T} \frac{1}{2}\|\mathcal{T}_0 - \mathcal{T}\|_F^2 + \tau\|\mathcal{T}\|_\star$$
$$= \mathcal{U} * \operatorname{ifft}_3(\max(\operatorname{fft}_3(\mathcal{S}) - \tau, 0)) * \mathcal{V}^\top$$

Update $\{\mathcal{W}^{(k)}\}$: Following [1], we update the dual variables $\{\mathcal{W}^{(k)}\}$ by dual ascent as follows:

$$\mathcal{W}_{t+1}^{(k)} = \mathcal{W}_t^{(k)} + \rho(\mathcal{L}_{t+1} - \mathcal{K}_t^{(k)}), \quad k \in [K] \tag{4.5}$$

Algorithm 1 ADMM for Problem (4.1)

Require: Observation $\mathcal{Y}$, parameters $\lambda \geq 0, \{w_k\} \geq 0$.
1: Let $\rho = 1$, $\epsilon = 10^{-6}$, and initialize $\mathcal{L}_0 = \mathbf{0}$, $\mathcal{K}_0^{(k)} = \mathcal{W}_0^{(k)} = \mathbf{0}, \forall k \in [K]$.
2: **while** not converged **do**
3:     Update $\mathcal{L}$ by Eq. (4.3);
4:     Update $\{\mathcal{K}^{(k)}\}$ by Eq. (4.4);
5:     Update $\{\mathcal{W}^{(k)}\}$ by Eq. (4.5);
6:     Check the convergence criteria as follows:
        Convergence in primal variables: $\|\mathcal{X}_{t+1} - \mathcal{X}_t\|_\infty \leq \epsilon, \ \forall \mathcal{X} \in \{\mathcal{L}\} \bigcup \{\mathcal{K}^{(k)}\}_k$;
        Convergence in constraints: $\|\mathcal{L}_{t+1} - \mathcal{K}_{t+1}^{(k)}\|_\infty \leq \epsilon, \forall k \in [K]$;
7:     $t = t + 1$.
8: **end while**

We then summarize the algorithm in Algorithm 1.

Complexity analysis: In each single iteration of Algorithms 1, 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 Algorithm 1 has per-iteration complexity

$$O\big(KD\log D + D\sum_k \min(d_k, d_k^{-1}d_{k+1}^{-1}D)\big), \tag{4.6}$$

which is the same order as ADMM-based algorithms for SNN [11].

Convergence analysis: We then discuss the convergence of Algorithm 1 as follows.

Theorem 4.2 (Convergence of Algorithm 1). *For any positive constant ρ, if the unaugmented Lagrangian function $L_0(\mathcal{L}, \{\mathcal{K}^{(k)}\}, \{\mathcal{W}^{(k)}\})$ has a saddle point, then the iterations $(\mathcal{L}_t, \{\mathcal{K}_t^{(k)}\}, \{\mathcal{W}_t^{(k)}\})$ in Algorithm 1 satisfy the residual convergence, objective convergence and dual variable convergence (defined in [1]) of Problem (4.1) as $t \to \infty$.*

The proof of Theorem 4.2 is given in Appendix B. Since there are only equality constraints in the convex Problem (4.1), strong duality holds naturally as a corollary of Slater's condition [37], which further indicates that the unaugmented Lagrangian $L_0(\mathcal{L}, \{\mathcal{K}^{(k)}\}, \{\mathcal{W}^{(k)}\})$ has a saddle point. Moreover, according to the analysis in [12], the convergence rate of general ADMM-based algorithms is $O(1/T)$, where T denotes the iteration number. In this way, the convergence behavior of Algorithm 1 is analyzed.

5 Statistical Performance of the Proposed Estimator

We answer **Q2** by establishing an upper bound on the estimation error of any estimator $\hat{\mathcal{L}}$ in the solution set of Problem (3.5) to the underlying tensor $\mathcal{L}^*$. Let $\Delta := \hat{\mathcal{L}} - \mathcal{L}^*$ be the error tensor whose F-norm will be upper bounded in this section.

Lemma 5.1 (Error bound on the observed entries). *By setting the regularization parameter*

$$\lambda \geq 2\sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^* \tag{5.1}$$

then we obtain

$$\|\Delta\|_{\star\mathrm{o}} \ \leq \ \sum_k w_k\sqrt{32r_k^*}\|\Delta\|_\mathrm{F} \tag{5.2}$$

and

$$\frac{1}{2}\|\mathcal{B} \odot \Delta\|_\mathrm{F}^2 \leq \frac{3\lambda}{2}\sum_k w_k\sqrt{2r_k^*}\|\Delta\|_\mathrm{F} \tag{5.3}$$

The proof can be found in Appendix A.1. Lemma 5.1 gives an upper bound on the "observed part" of the estimation error $\|\mathcal{B} \odot \Delta\|_F$. As our goal is to establish a bound on $\|\Delta\|_F$, we then connect $\|\mathcal{B} \odot \Delta\|_F$ with $\|\Delta\|_F$ by quantifying the probability of the following so-called Restricted Strong Convexity (RSC) property of the sampling operator defined by $\mathcal{B}$:

$$\|\mathcal{B} \odot \Delta\|_F^2 \geq \frac{p}{2}\|\Delta\|_F^2 \ - \ \text{an intercept term} \tag{5.4}$$

when the error tensor Δ belongs to some set $\mathsf{C}(\beta, \boldsymbol{r})$ defined as

$$\mathsf{C}(\beta, \boldsymbol{r}) := \left\{ \Delta \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K} \ \Big| \ \|\Delta\|_\infty \leq 1, \ p\|\Delta\|_F^2 \geq \beta, \ \|\Delta\|_{\star\mathrm{o}} \leq \sum_k w_k \sqrt{r_k} \|\Delta\|_F \right\} \tag{5.5}$$

where β is an F-norm tolerance parameter and $\boldsymbol{r} = (r_1, r_2, \ldots, r_K)$ is a rank parameter whose values will be specified in the sequel.

Lemma 5.2 (Restricted Strong Convexity). *For any $\Delta \in \mathsf{C}(\beta, \boldsymbol{r})$, it holds with probability at least $1 - 1/\tilde{d}$ that*

$$\|\mathcal{B} \odot \Delta\|_F^2 \geq \frac{p}{2}\|\Delta\|_F^2 - 64ep^{-1}\Big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} + 2\Big)^2 \tag{5.6}$$

where e is the base of the natural logarithm, and the entries $\mathcal{R}_{i_1 i_2 \ldots i_K}$ of tensor $\mathcal{R}$ are i.i.d. Rademacher random variables.

The proof of Lemma 5.2 can be found in Appendix A.1. According to Lemma 5.1 and Lemma 5.2, it remains to bound $\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*$ and $\mathbb{E}[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*]$. The following lemmas upper bound them respectively.

Lemma 5.3. *Let $p \geq \max_k(d_k \wedge d_{\setminus k})^{-1}$. Then it holds with probability at least $1 - \sum_k \tilde{d}_k^{-1}$ that*

$$\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^* \leq \frac{C_\varrho}{K^2} \sum_{k=1}^{K} w_k^{-1} \sqrt{pD(d_k \wedge d_{\setminus k})^{-1} \log \tilde{d}_k} \tag{5.7}$$

Lemma 5.4. *Let $p \geq \max_k(d_k \wedge d_{\setminus k})^{-1}$ and $pD(d_k \wedge d_{\setminus k})^{-1} \geq 2\log^2(d_k \wedge d_{\setminus k})\log \tilde{d}_k, \forall k \in [K]$. Then, the following inequality holds*

$$\mathbb{E}[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*] \leq \frac{C_2}{K^2} \sum_{k=1}^{K} w_k^{-1} \sqrt{pD(d_k \wedge d_{\setminus k})^{-1} \log \tilde{d}_k} \tag{5.8}$$

For the proof of Lemma 5.3 and Lemma 5.4, please see Appendix A.1.

Then, combining Lemmas 5.1 to 5.4, we arrive at an upper bound on the estimation error in the follow theorem.

Theorem 5.5. *Suppose Assumption 3.1 and Assumption 3.2 are satisfied and $\|\mathcal{L}^*\|_\infty \leq \mathbf{a}$. Let $p \geq \max_k(d_k \wedge d_{\setminus k})^{-1}$ and $pD(d_k \wedge d_{\setminus k})^{-1} \geq 2\log^2(d_k \wedge d_{\setminus k})\log \tilde{d}_k, \forall k \in [K]$. By setting*

$$\lambda = \frac{c_1'\sigma}{K^2} \sum_{k=1}^{K} w_k^{-1} \sqrt{pD(d_k \wedge d_{\setminus k})^{-1} \log \tilde{d}_k} \tag{5.9}$$

the estimation error of any estimator $\hat{\mathcal{L}}$ defined in Problem (3.5) can be upper bounded as follows

$$\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\mathrm{F}}^2 \leq \frac{C}{p^2} \left(\frac{pD(\sigma \vee \mathbf{a})^2}{K^4} \Big(\sum_{k=1}^{K} w_k^{-1} \sqrt{(d_k \wedge d_{\setminus k})^{-1} \log \tilde{d}_k} \Big)^2 \Big(\sum_{k=1}^{K} w_k \sqrt{r_k^*} \Big)^2 + \mathbf{a}^2 + \log \tilde{d} \right)$$
(5.10)

with probability at least $1 - \tilde{d}^{-1} - \sum_k \tilde{d}_k^{-1}$.

The proof of Theorem 5.5 can be found in Appendix A.2. To understand the proposed bound, we consider the K-way cubical tensor $\mathcal{L}^* \in \mathbb{R}^{d \times d \times \cdots \times d}$ with regularization weights $w_1 = w_2 = \cdots = w_K = 1/K$ and sampling ratio $p \geq 1/d$. Then the bound in Eq. (5.10) is simplified to the following element-wise error:

$$\frac{\|\hat{\mathcal{L}} - \mathcal{L}^*\|_{\mathrm{F}}^2}{d^K} \leq O \left(\frac{1}{p} \cdot (\sigma \vee \mathbf{a})^2 \cdot \frac{1}{K} \left(\sum_{k=1}^{K} \sqrt{\frac{r_k^*}{d}} \right)^2 \log d \right)$$
(5.11)

which means the *estimation error is controlled by the averaged low-tubal-rank ratio in all orientations of the underlying tensor* $\mathcal{L}^*$. Eq. (5.11) also indicates that as the sampling ratio

$$p \geq \Omega \left(\frac{1}{K} \left(\sum_{k=1}^{K} \sqrt{\frac{r_k^*}{d}} \right)^2 \log d \right)$$
(5.12)

the estimation error will be small. Equivalently, the expected observation number $N = pd^K$ for approximate tensor completion is

$$N \geq \Omega \left(\frac{1}{K} \Big(\sum_{k=1}^{K} \sqrt{r_k^*} \Big)^2 d^{K-1} \log d \right)$$
(5.13)

It can be seen from Eq. (5.11) and Eq. (5.13) that both the estimation error and the sample complexity are *controlled by the low-tubal-rankness along all orientations*, which is in consistence with the results for sum-of-norms-based estimators in [33, 39, 43, 29]. A more general analysis in [29] and [33] indicates that the performance of sum-of-norms-based estimators are determined by all the structural complexities for a simultaneously structured signal, just as the proposed bounds show. However, a straightforward corollary of the results in [29] and [33] also shows that for tensor recovery from a small number of noisy observations, OITNN *cannot provide essentially lower estimation error or sample complexity than TNN*. Fortunately, despite this general and unavoidable limitation of sum-of-norms-based estimators [33, 39, 43, 29], one can still find a significant gain in empirical performance of OITNN over TNN in Sec. 6.

$\boxed{6}$ Experiments

In this section, we evaluate the effectiveness of OITNN in comparison with traditional matrix/tensor nuclear norms by conducting simulated noisy tensor completion on nine different types of datasets collected from several remote sensing related applications from §§6.1 to §§6.9.

Sampling schemes: In the simulations, we consider two sampling schemes:

1) *Uniform sampling with ratio p*: As described in Assumption 3.1, each element of the binary sampling tensor $\mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ is drawn *i.i.d.* from Bernoulli distribution with parameter $p \in (0, 1]$.

2) *Non-uniform sampling with proxy ratio $\mathfrak{p}$*: This is a more challenging setting where there exists totally missing rows, columns, tubes and possibly higher-order fibers. Specifically, given the proxy ratio $\mathfrak{p} \in (0, 1]$, we let $q = (1 - \mathfrak{p})^{1/K}$ and construct the sampling tensor $\mathcal{B}$ as follows:
   Step 1: Initialize the sampling tensor $\mathcal{B}^0$ to all ones, i.e., $\mathcal{B}^0 \leftarrow \text{ones}([d_1, \ldots, d_K])$ in Matlab.
   Step 2: For each k from 1 to K, we generate the tensor $\mathcal{B}^k$ by specifying its mode-k matricization $\boldsymbol{B}^k_{(k)}$ as the matrix obtained by randomly setting each column of $\boldsymbol{B}^{k-1}_{(k)}$ to zero with probability q.
   Step 3: We finally construct the sampling tensor by $\mathcal{B} \leftarrow \mathcal{B}^K$.
   According to the above steps, we can see that the sampling tensor $\mathcal{B}$ may process totally zero rows, columns, tubes, and higher-order fiber. We call the parameter $\mathfrak{p}$ *the proxy sampling ratio* because the expected sample number, i.e., the expectation of the number of non-zero entries in $\mathcal{B}$ is $\mathfrak{p}D$.

Observation model: We conduct noisy tensor completion on the datasets . First we rescale the signal tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ by $\|\mathcal{L}^*\|_\infty = 1$, and then we generate the noise tensor $\mathcal{E} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$ with *i.i.d.* standard Gaussian entries. Next, the standard deviation parameter σ of the random noises is chosen by $\sigma = c\|\mathcal{L}^*\|_\mathrm{F}/\sqrt{D}$ for a constant Signal to Noise Ratio (SNR) $c = 0.05$. Further, the binary sampling tensor $\mathcal{B}$ is generated according to the above mentioned uniform or non-uniform sampling schemes. Finally, we obtain the observation tensor $\mathcal{Y} = \mathcal{B} \odot (\mathcal{L}^* + \sigma\mathcal{E})$ according to the observation model (3.1).

Competitor nuclear norms and performance metrics: We evaluate the effectiveness of OITNN in comparison with mostly adopted matrix/tensor nuclear norms, including matrix Nuclear Norm (NN) [19, 31], the Sum of Nuclear Norms (SNN) [39], the vanilla Tubal Nuclear Norm (TNN-DFT) [42], the Tubal Nuclear Norm defined with Discrete Fourier Transform (TNN-DCT) [27, 45], and the Square Nuclear Norm (SqNN) [29] (only for 4-way tensors in §§6.9). The effectiveness of norms are measured by the Peaks Signal Noise Ratio (PSNR) and structural similarity (SSIM) [47]. Larger PSNR and SSIM values indicate higher quality of the estimator $\hat{\mathcal{L}}$.

6.1 Experiments on An Urban Area Imagery Dataset

This section uses the well known area imagery dataset *UCMerced* land use dataset [48], which is composed of 256×256 pixels RGB images. The 85-th image of all the 21 classes are chosen in this experiment and we obtain 21 images as shown in Fig. 4.

We consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d \times d \times 3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of an image are missing. For NN [19, 31], we set the regularization parameter $\lambda = \lambda_\iota \sigma \sqrt{pd \log(2d)}$. For SNN [39], we set the regularization parameter $\lambda = \lambda_\iota$ and choose the weight $\boldsymbol{\alpha}$ by $\alpha_1 : \alpha_2 : \alpha_3 = 1 : 1 : 0.01$. For TNN-DFT [42] and TNN-DCT [27], we set $\lambda = \lambda_\iota \sqrt{pd \log(6d)}$. For the proposed OITNN-O, we set the regularization parameter $\lambda = \lambda_\iota \sqrt{pd \log(6d)}$ and choose the weight $\boldsymbol{w}$ by $w_1 : w_2 : w_3 = 1 : 1 : 2.5$. The factor λ_ι is then tuned in $\{10^{-3}, 10^{-2}, \ldots, 10^3\}$ for each norm, and we choose the one with highest PSNRs in most cases in the parameter tuning phase. In each setting, we test each image for 10 trials and report the averaged PSNR (in db) and SSIM values. We present the PSNR and SSIM values for two sampling schemes in Fig. 5 and Fig. 6 respectively for quantitative comparison. Several visual examples are shown in Fig. 7. It can be found from Fig. 5 and Fig. 6 that

> The non-uniform sampling setting is harder than the uniform setting with equal (proxy) sampling ratio because the non-uniform setting involves totally missing rows, columns, and tubes which will be automatically filled with zeros by directly minimizing traditional nuclear norms.

> 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 (See Fig. 7): 1) First, NN and SNN 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-DFT and TNN-DCT prevents them from recovering the totally missing tubes, since the direction minimization of TNN leads to a preference to filling in the unknown tubes with zeros. 3) Thanks to its capability to exploit multi-orientational spectral low-rankness, our OITNN-O can introduce information of observed rows/columns/tubes to reconstruct the unseen ones, leading to improved recovery.

6.2 Experiments on Gray Image Sequences

Time sequences of images are broadly used in many computer vision based remote sensing applications. We experiment on a $180 \times 320 \times 30$ tensor which consists of the first 30 gray image frames of the *Sky* dataset.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. For NN [19, 31], we set the parameter $\lambda = \lambda_\iota \sigma \sqrt{p(d_1 \vee d_2) \log(d_1 + d_2)}$. For SNN [39], we set the regularization parameter $\lambda = \lambda_\iota$ and chose the weight $\boldsymbol{\alpha}$ by $\alpha_1 : \alpha_2 : \alpha_3 = 1 : 1 : 1$. For TNN-DFT [42], we set $\lambda = \lambda_\iota \sigma \sqrt{pd_3(d_1 \vee d_2) \log(d_1 d_3 + d_2 d_3)}$. For TNN-DCT [27], we set $\lambda = \lambda_\iota \sigma \sqrt{p(d_1 \vee d_2) \log(d_1 d_3 + d_2 d_3)}$. For the proposed OITNN-O, we set the regularization parameter $\lambda = \lambda_\iota \sigma \sqrt{pd_3(d_1 \vee d_2) \log(d_1 d_3 + d_2 d_3)}$ and choose the weight $\boldsymbol{w}$ by
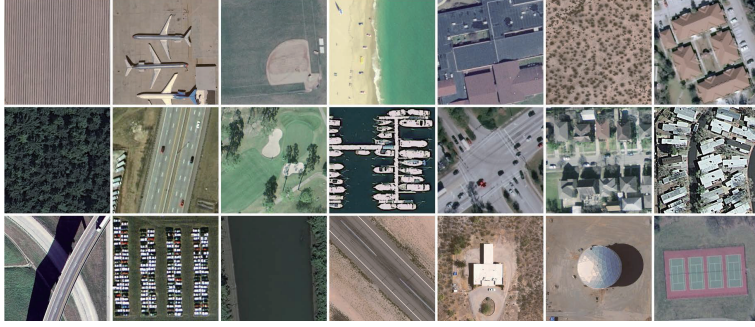
Figure 4: The dataset consists of the 85-th frame of all the 21 classes in the *UCMerced* dataset [46].

$w_1 : w_2 : w_3 = 1 : 1 : 1$. The factor λ_ι is then tuned in $\{10^{-3}, 10^{-2}, \ldots, 10^3\}$ for each norm, and we chose the one with highest PSNRs in most cases in the parameter tuning phase. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

We present the PSNR and SSIM values for two sampling schemes in Table 1 for quantitative comparison. The results are so similar to §§6.1: First, the non-uniform sampling setting is harder than the uniform setting with equal (proxy) sampling ratio. Second, the tensorial norms (SNN, TNN-DFT, TNN-DCT, and OITNN-O) outperform vanilla matrix NN, and our OITNN-O performs best in the tensorial norms, owing to its capability in modeling spectral low-rankness in multiple orientations.

6.3 Experiments on Hyperspectral Data

Benefit from its fine spectral and spatial resolutions, hyperspectral image processing has been extensively adopted in many remote sensing applications [13]. In this section, we conduct noisy tensor completion on subsets of the two representative hyperspectral datasets described as follows:

Indian Pines: This dataset was collected by AVIRIS sensor in 1992 over the Indian Pines test site in North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands. We use the first 30 bands in the experiments due to the trade-off between the limitation of computing resources and the efforts for parameter tuning.

Salinas A: The data were acquired by AVIRIS sensor over the Salinas Valley, California in 1998, and consists of 224 bands over a spectrum range of 400–2500nm. This dataset has a spatial extent of 86×83 pixels with a resolution of 3.7m. We use the first 30 bands in the experiments too.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values. We present the PSNR and SSIM values for two sampling schemes in Table 2 for quantitative comparison, with visual examples shown in Fig. 8. The experimental results are also very similar to §§6.1: Our

(a) $p = 0.05$

(b) $p = 0.1$

(c) $p = 0.15$

Figure 5: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *UCMerced* dataset in the uniform sampling setting with ratio $p \in \{0.05, 0.1, 0.15\}$.

OITNN-O performs best in the tensorial norms thanks to its capability in modeling spectral low-rankness in multiple orientations.

6.4 Experiments on Multispectral Images

Multispectral imaging captures image data within specific wavelength ranges across the electromagnetic spectrum, and has become one of the most widely utilized datatype in remote sensing. This section presents simulated experiments on multi-spectral images. The original data are two multispectral images *Cloth, Hair, Jelly Beans*, and *Oil Painting* from the Columbia MSI Database containing scenes of a variety of real-world objects. Each MSI is of size $512 \times 512 \times 31$ with intensity range scaled to $[0, 1]$.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

We present the PSNR and SSIM values for two sampling schemes in Table 3 and Table 4 respectively for quantitative comparison, with visual examples shown in Fig. 9. The

---

Available at http: //www1.cs.columbia.edu/CAVE/databases/multispectral

(a) $\mathfrak{p} = 0.05$

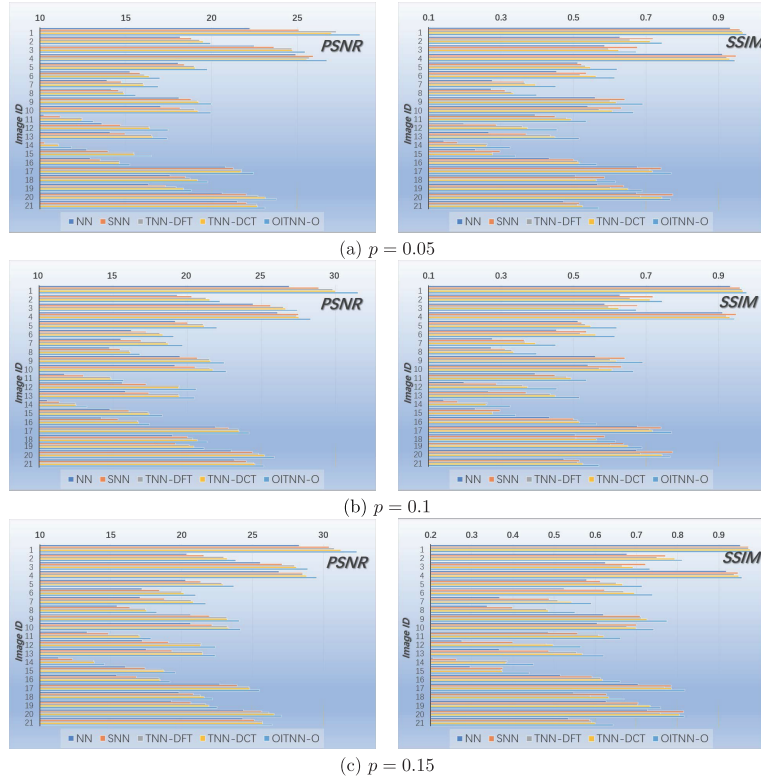(b) $\mathfrak{p} = 0.1$
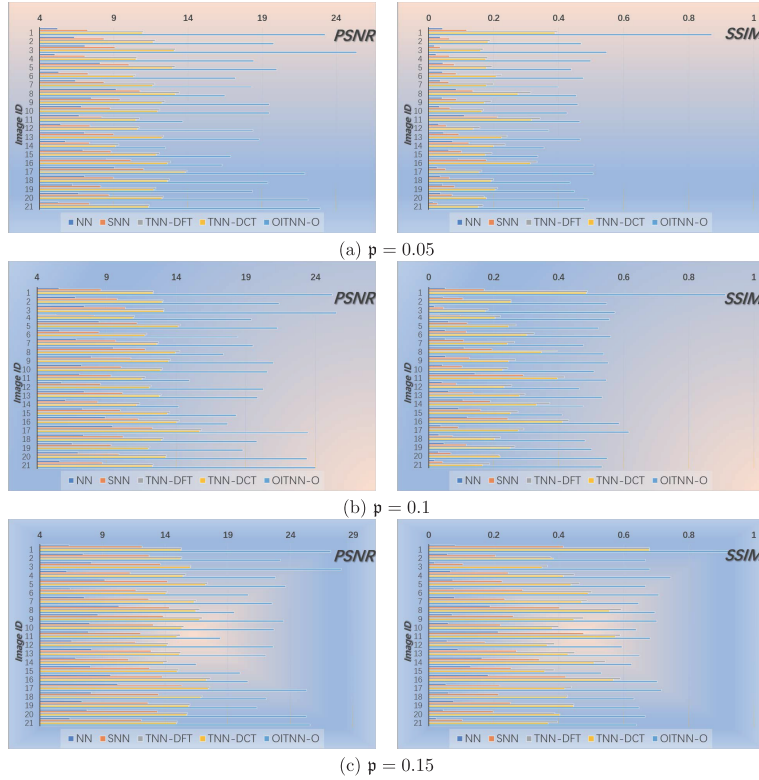

(c) $\mathfrak{p} = 0.15$

Figure 6: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *UCMerced* dataset in the non-uniform sampling setting with proxy ratio $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$.

experimental results are also very similar to §§6.1: thanks to its capability in modeling spectral low-rankness in multiple orientations, our OITNN-O outperforms the other competitor norms.

6.5 | Experiments on Point Could Data

With the rapid advances of sensor technology, the emerging point cloud data provide better performance than 2-D images in many remote sensing applications due to its flexible and scalable geometric representation [28]. In this section, we also conduct experiments on a dataset for Unmanned Ground Vehicle (UGV). The dataset contains a sequence of point cloud data acquired from a Velodyne HDL-64E LiDAR. We select 30 frames (Frame Nos. 65-94) from the data sequence. The point cloud data is formatted into two tensors sized $64 \times 870 \times 30$ representing the distance data (named *SenerioB Distance*) and the intensity data (named *SenerioB Intensity*), respectively.

We also consider two sampling schemes, i.e., *uniform sampling with ratio $p \in \{0.05, 0.1, 0.15\}$* and *non-uniform sampling with proxy ratio $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$* for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials

---

Scenario B from http://www.mrt.kit.edu/z/publ/download/velodynetracking/dataset.html

(a) Orignal   (b) Observation   (c) NN   (d) SNN   (e) TNN(DFT)   (f) TNN(DCT)   (g) OITNN-O
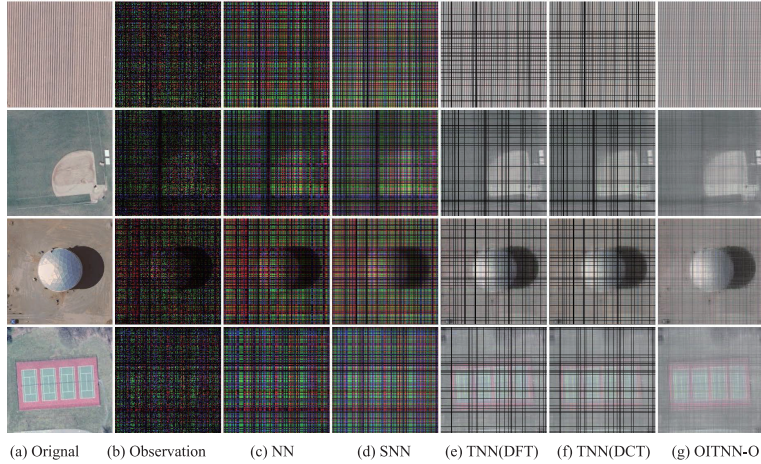
Figure 7: Visual results obtained by five norms for noisy tensor completion on the *UCMerced* dataset in the non-uniform sampling setting with approximately 95% missing entries. The sub-plots from (a) to (g): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) recovered by SNN [39]; (e) image recovered by the vanilla TNN-DFT [42]; (f) image recovered by the TNN-DCT [27]; (g) image recovered by our OITNN-O.

and report the averaged PSNR (in db) and SSIM values.

We present the PSNR and SSIM values for two sampling schemes in Table 5 for quantitative comparison. The experimental results are also very similar to §§6.1: our OITNN-O performs better than the other norms thanks to its capability in modeling spectral low-rankness in multiple orientations.

6.6 Experiments on SAR Data

Polarimetric synthetic aperture radar (PolSAR) has attracted lots of attention from remote sensing scientists because of its various advantages, e.g., all-weather, all-time, penetrating capability, and multi-polarimetry. In this section, we adopt the *PolSAR UAVSAR Change*

(a) Orignal   (b) Observation   (c) NN   (d) SNN   (e) TNN(DFT)   (f) TNN(DCT)   (g) OITNN-O

Figure 8: Visual results obtained by five norms for noisy tensor completion on the 21-th bound of hyper-spectral dataset in the non-uniform sampling setting with approximately 95% missing entries. The first row corresponds to the *Indian Pines* dataset with the second corresponding to *Salinas A*. The sub-plots from (a) to (g): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) recovered by SNN [39]; (e) image recovered by the vanilla TNN-DFT [42]; (f) image recovered by the TNN-DCT [27]; (g) image recovered by our OITNN-O.

(a) Orignal   (b) Observation   (c) NN   (d) SNN   (e) TNN(DFT)   (f) TNN(DCT)   (g) OITNN-O
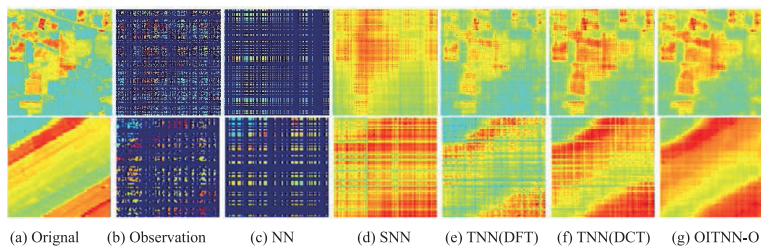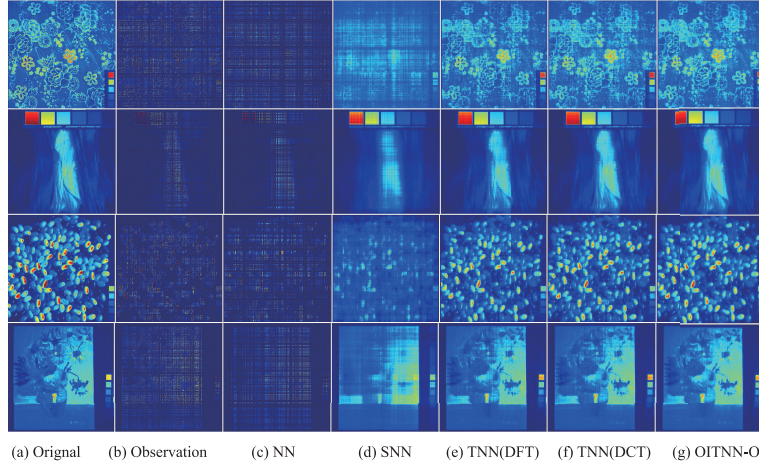
Figure 9: Visual results obtained by five norms for noisy tensor completion on the 21-th bound of multi-spectral dataset in the non-uniform sampling setting with approximately 95% missing entries. The first, second, third, and fourth row correspond to the *Cloth*, *Hair*, *Jelly Beans*, and *Oil Painting*, respectively. The sub-plots from (a) to (g): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) recovered by SNN [39]; (e) image recovered by the vanilla TNN-DFT [42]; (f) image recovered by the TNN-DCT [27]; (g) image recovered by our OITNN-O.

Detection Images dataset. It is a dataset of single-look quad-polarimetric SAR images acquired by the UAVSAR airborne sensor in L-band over an urban area in San Francisco city on 18 September 2009, and May 11, 2015. The dataset #1 have length and width of 200 pixels, and we use the first 30 bands.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

We present the PSNR and SSIM values for two sampling schemes in Table 6 for quantitative comparison. The experimental results are also very similar to §§6.1: thanks to its capability in modeling spectral low-rankness in multiple orientations, our OITNN-O achieves better performance than the other competitors.

6.7 Experiments on Seismic Data

With the rapid development of GIS and remote sensing technologies, valuable progress has been made in the earthquake-related research field. We use the seismic data tensor of size $512 \times 512 \times 3$ which is abstracted from the test data "seismic.mat" of a toolbox for seismic data processing from Center of Geopyhsics, Harbin Institute of Technology, China.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ images, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

Available at https://github.com/sevenysw/MathGeo2018

(a) Orignal   (b) Observation   (c) NN   (d) SNN   (e) TNN(DFT)   (f) TNN(DCT)   (g) OITNN-O
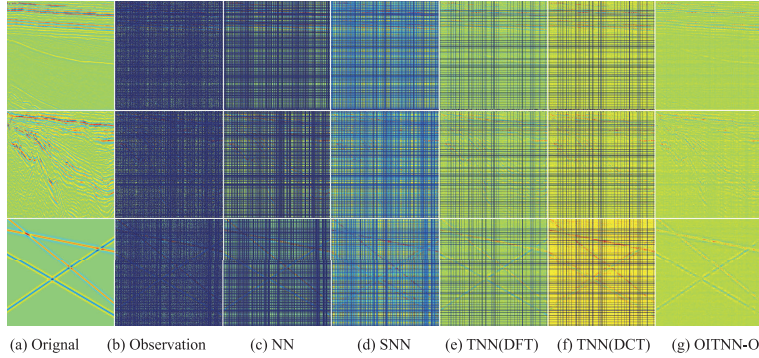
Figure 10: Visual results obtained by five norms for noisy tensor completion on the *Seismic* dataset in the non-uniform sampling setting with approximately 85% missing entries. The first, second, and third rows correspond to the first, second, and third frontal slice of the tensor. The sub-plots from (a) to (g): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) recovered by SNN [39]; (e) image recovered by the vanilla TNN-DFT [42]; (f) image recovered by the TNN-DCT [27]; (g) image recovered by our OITNN-O.

We present the PSNR and SSIM values for two sampling schemes in Table 7 for quantitative comparison, with visual examples shown in Fig. 10. The experimental results are also very similar to §§6.1: our OITNN-O outperforms the other competitor norms benefiting from its capability in modeling spectral low-rankness in multiple orientations.

6.8 Experiments on Thermal Imaging Data

Thermal infrared data can provide important measurements of surface energy fluxes and temperatures in various remote sensing applications. In this section, we experiment on two infrared datasets as follows:

> The *OSU Thermal Database* [9]: The sequences were recorded on the Ohio State University campus during the months of February and March 2005, and show several people, some in groups, moving through the scene. We use the first 30 frames of Sequences 1 and form a tensor of size $320 \times 240 \times 30$.

> The *Infraed Detection* dataset [15]: This dataset is collected for infrared detection and tracking of dim-small aircraft targets under ground/air background. It consists of 22 subsets of infrared image sequences of all aircraft targets. We use the first 30 frames of data3.zip to form a $256 \times 256 \times 30$ tensor due to the trade-off between the limitation of computing resources and the efforts for parameter tuning.

We also consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1, 0.15\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$ for the $d_1 \times d_2 \times d_3$ tensors, that is, $\{95\%, 90\%, 85\%\}$ entries of a tensor are missing. We use the same parameter tuning scheme with §§6.2. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values.

Available at http://www.csdata.org/p/387/

(a) Orignal　(b) Observation　(c) NN　(d) SNN　(e) TNN(DFT)　(f) TNN(DCT)　(g) OITNN-O
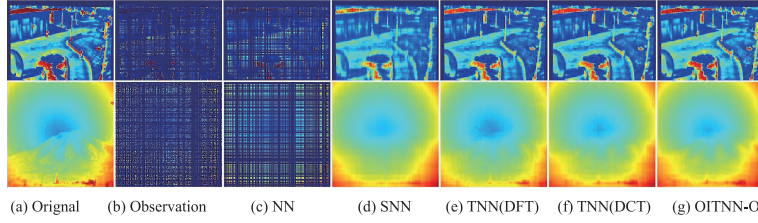
Figure 11: Visual results obtained by five norms for noisy tensor completion on the 21-th bound of the thermal imaging dataset in the non-uniform sampling setting with approximately 95% missing entries. The first and second rows correspond to the *OSU Thermal Database* and *Infraed Detection* dataset, respectively. The sub-plots from (a) to (g): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) recovered by SNN [39]; (e) image recovered by the vanilla TNN-DFT [42]; (f) image recovered by the TNN-DCT [27]; (g) image recovered by our OITNN-O.

We present the PSNR and SSIM values for two sampling schemes in Table 8 for quantitative comparison, with visual examples shown in Fig. 11. The experimental results are also very similar to §§6.1: thanks to its capability in modeling spectral low-rankness in multiple orientations, our OITNN-O outperforms the other tensorial norms.

6.9　Experiments on RGB Video Data

RGB aerial videos are broadly used in many computer vision based remote sensing tasks. We experiment on a $216 \times 288 \times 3 \times 30$ tensor which consists of the first 30 frames of the *Ground* dataset for small object detection [25].

Since the data tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}$ is 4-way, we evaluate NN by first unfolding the $\mathcal{L}^*$ along all the four modes to get four matrices $\boldsymbol{L}^*_{(k)}, k \in [4]$, and then reporting the result with highest PSNR values. We adopt TNN-DFT and TNN-DCT separately on three $d_1 \times d_2 \times d4$ gray videos. Thus, we can use the same parameter tuning scheme for NN, TNN-DFT, and TNN-DCT as §§6.2. We also consider the Square Nuclear Norm (SqNN) [29] for which we consider the general unfolding with mode partitions $\{1, 4\}$ and $\{2, 3\}$ and we set the regularization parameter $\lambda = \lambda_\iota \sigma \sqrt{p d_1 d_4 \log(d_1 d_4 + d_2 d_3)}$. For SNN [39], we set the regularization parameter $\lambda = \lambda_\iota$ and chose the weight $\boldsymbol{\alpha}$ by $\alpha_1 : \alpha_2 : \alpha_3 : \alpha_4 = 1 : 1 : 0.01 : 1$. For the proposed OITNN-O, we set the regularization parameter $\lambda = \lambda_\iota \sigma \sqrt{p d_1 d_3 d_4 \log(d_1 d_3 d_4 + d_2 d_3)}$ and choose the weight $\boldsymbol{w}$ by $w_1 : w_2 : w_3 : w_4 = 1 : 100 : 1 : 1$. The factor λ_ι is then tuned in $\{10^{-3}, 10^{-2}, \dots, 10^3\}$ for each norm, and we chose the one with highest PSNRs in most cases in the parameter tuning phase.

We consider two sampling schemes, i.e., *uniform sampling with ratio* $p \in \{0.05, 0.1\}$ and *non-uniform sampling with proxy ratio* $\mathfrak{p} \in \{0.05, 0.1, 0.15\}$, that is, $\{95\%, 90\%, 85\%\}$ entries of the video are missing. In each setting, we test each tensor for 10 trials and report the averaged PSNR (in db) and SSIM values. We present the PSNR and SSIM values for two sampling schemes in Table 9 for quantitative comparison with visual examples shown in Fig. 12. The results are so similar to §§6.1: First, the non-uniform sampling setting is harder than the uniform setting with equal (proxy) sampling ratios. Second, the tensorial norms (SqNN, SNN, TNN-DFT, TNN-DCT, and OITNN-O) outperform vanilla matrix NN, and our OITNN-O performs best in the tensorial norms, owing to its capability in modeling spectral low-rankness in multiple orientations.

Available at http://www.loujing.com/rss-small-target

(a) Orignal   (b) Observation   (c) NN   (d) SqNN   (e) SNN   (f) TNN(DFT)   (g) TNN(DCT)   (h) OITNN-O
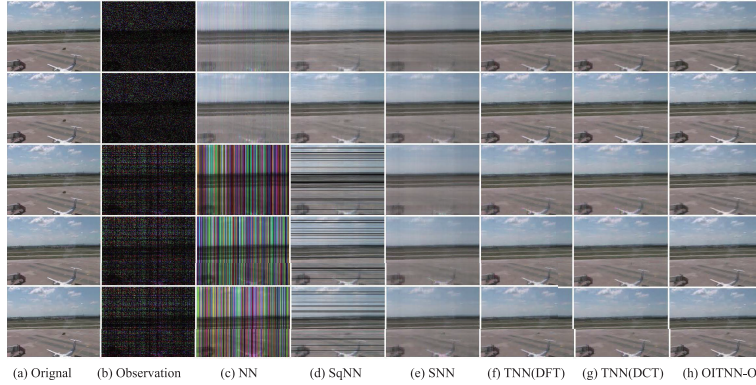
Figure 12: Visual results obtained by five norms for noisy tensor completion on the 21-th frame of *Ground* dataset in the uniform and non-uniform sampling settings. The first and second rows correspond to the uniform sampling with ratio $p = 0.05$ and 0.1 respectively, with the third, fourth, and fifth corresponding to non-uniform sampling with proxy ratio $\mathfrak{p} = 0.05$, 0.1, and 0.15, respectively. The sub-plots from (a) to (h): (a) the original image; (b) the observed image; (c) image recovered by NN [19]; (d) image recovered by SqNN [29]; (e) image recovered by SNN [39]; (f) image recovered by the vanilla TNN-DFT [42]; (g) image recovered by the TNN-DCT [27]; (h) image recovered by our OITNN-O.

7 Conclusion and Discussion

To reconstruct an unknown tensor data from partial noisy observations, we first formulated an OITNN penalized least squares estimator for noisy tensor completion and then explored its statistical and empirical performance. The proposed estimator can be tractably computed via an ADMM-based algorithm and its statistical performance is analyzed through establishing an upper bound on the estimation error. The effectiveness of OITNN is validated through extensive simulation studies on nine different types of remote sensing data.

However, generally speaking, the proposed estimator has the following two drawbacks due to the adoption of OITNN:

1) Computational inefficiency: Compared to TNN, it is more time-consuming since it involves computing TNNs along all orientations.

2) Sample inefficiency: The analysis of [29] and [33] indicates that for tensor recovery from a small number of observations, OITNN cannot provide essentially lower sample complexity than TNN.

Thus, one future direction is to consider more efficient solvers of Problem (3.5) using factorization strategy or sampling-based methods. Another direction is to study more general transforms (linear or nonlinear) to define the tensor low-rankness like [17, 27, 45, 7] and discuss whether lower sample complexity can be obtained. It is also interesting to consider smoothness [4, 36] by constructing a hyper-graph-lile structure [5] in the proposed models for more sound tensor modeling.

Acknowledgment

References

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

[2] E. Candès, X. Li, Y. Ma, and J. Wright, Robust principal component analysis? *J. ACM* 58 (2011) 11.

[3] J. Carroll and J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition, *Psychometrika* 35 (1970) 283–319.

[4] J. Chang, Y. Chen, L. Qi, and H. Yan, Hypergraph clustering using a new laplacian tensor with applications in image processing, *SIAM J. Imaging Sci.* 13 (2020) 1157–1178.

[5] J. Chang, W. Ding, L. Qi, and H. Yan, Computing the p-spectral radii of uniform hypergraphs with applications, *J. Sci. Comput.* 75 (2018) 1–25.

[6] S. Chatterjee, Matrix estimation by universal singular value thresholding, *Ann. Statist.* 43 (2015) 177–214.

[7] C. Cui, K. Zhang, T. Daulbaev, J. Gusak, I. Oseledets, and Z. Zhang, Active subspace of neural networks: Structural analysis and universal attacks, *SIAM J. Math. Data Sci.* 2 (2020) 1096–1122.

[8] C. Cui and Z. Zhang, High-dimensional uncertainty quantification of electronic and photonic ic with non-gaussian correlated process variations, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39 (2019) 1649–1661.

[9] J. Davis and V. Sharma, Background-subtraction using contour-based fusion of thermal and visible imagery, *Computer Vision and Image Understanding*, 106 (2007) 162–182.

[10] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[11] Q. Gu, H. Gui, and J. Han, Robust tensor decomposition with gross corruption, In *Proceedings of Advances in Neural Information Processing Systems*, 2014, pp. 1422–1430.

[12] B. He and X. Yuan, On the o(1/n) convergence rate of the douglas–rachford alternating direction method, *SIAM J. Numer. Anal.* 50 (2012) 700–709.

[13] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, Non-local meets global: An integrated paradigm for hyperspectral image restoration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44 (2022) 2089–2107.

[14] J. Hou, F. Zhang, H. Qiu, J. Wang, Y. Wang, and D. Meng, Robust low-tubal-rank tensor recovery from binary measurements, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2021.

[15] B. Hui, Z. Song, and H. Fan, A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background, *China Scientific Data* 5 (2020) 291–302.

[16] Q. Jiang and M. Ng, Robust low-tubal-rank tensor completion via convex optimization, In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 2649–2655.

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

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

[19] O. Klopp, Noisy low-rank matrix completion with general sampling distribution, *Bernoulli* 20 (2014) 282–303.

[20] O. Klopp, Matrix completion by singular value thresholding: Sharp bounds, *Electron. J. Stat.* 9 (2015) 2348–2369.

[21] T. Kolda and B. Bader, Tensor decompositions and applications, *SIAM Rev.* 51 (2009) 455–500.

[22] J. Liu, P. Musialski, P. Wonka, and J. Ye, Tensor completion for estimating missing values in visual data, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013) 208–220.

[23] X. Liu, S. Aeron, V. Aggarwal, and X. Wang, Low-tubal-rank tensor completion using alternating minimization, *IEEE Trans. Inform. Theory* 66 (2020) 1714–1737.

[24] X. Liu and X. Wang, Fourth-order tensors with multidimensional discrete transforms, *arXiv preprint arXiv:1705.01576* 2017.

[25] J. Lou, W. Zhu, H. Wang, and M. Ren, Small target detection combining regional stability and saliency in a color image, *Multimedia Tools and Applications* 76 (2017) 14781–14798.

[26] C. Lu, J. Feng, W. Liu, Z. Lin, and S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (2020) 925 – 938.

[27] C. Lu, X. Peng, and Y. Wei, Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms, In *the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5996–6004.

[28] A. Mayalu, K. Kochersberger, B. Jenkins, and F. Malassenet, Lidar data reduction for unmanned systems navigation in urban canyon, *Remote Sensing* 12 (2020): 1724.

[29] C. Mu, B. Huang, J. Wright, and D. Goldfarb, Square deal: Lower bounds and improved relaxations for tensor recovery, In *International Conference on Machine Learning*, 2014, pp. 73–81.

[30] S. Negahban and M. Wainwright, Estimation of (near) low-rank matrices with noise and high-dimensional scaling, *Ann. Statist.* 39 (2011) 1069–1097.

[31] S. Negahban and M. Wainwright, Restricted strong convexity and weighted matrix completion: Optimal bounds with noise, *J. Mach. Learn. Res.* 13 (2012) 1665–1697.

[32] I. Oseledets, Tensor-train decomposition, *SIAM J. Sci. Comput.* 33 (2011) 2295–2317.

[33] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi, Simultaneously structured models with application to sparse and low-rank matrices, *IEEE Trans. Inform. Theory* 61 (2015) 2886–2908.

[34] N. Parikh and S. Boyd, Proximal algorithms, *Foundations and Trends® in Optimization* 1 (2014) 127–239.

[35] Y. Qiu, G. Zhou, X. Chen, D. Zhang, X. Zhao, and Q. Zhao, Semi-supervised non-negative tucker decomposition for tensor data representation, *Science China Technological Sciences* 64 (2021) 1881–1892.

[36] Y. Qiu, G. Zhou, Y. Wang, Y. Zhang, and S. Xie, A generalized graph regularized non-negative tucker decomposition framework for tensor data representation, *IEEE Transactions on Cybernetics* 52 (2020) 594 – 607.

[37] R. T. Rockerfeller, Convex analysis, 1970.

[38] M. Talagrand, A new look at independence, *Ann. Probab.* 24 (1996) 1–34.

[39] R. Tomioka, T. Suzuki, K. Hayashi, and H. Kashima, Statistical performance of convex tensor decomposition, In *Proceedings of Annual Conference on Neural Information Processing Systems*, 2011, pp. 972–980.

[40] L. Tucker, Some mathematical notes on three-mode factor analysis, *Psychometrika* 31 (1966) 279–311.

[41] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*, volume 47, Cambridge University Press, 2018.

[42] A. Wang, Z. Lai, and Z. Jin, Noisy low-tubal-rank tensor completion, *Neurocomputing* 330 (2019) 267–279.

[43] A. Wang, C. Li, Z. Jin, and Q. Zhao, Robust tensor decomposition via orientation invariant tubal nuclear norms, In *The AAAI Conference on Artificial Intelligence*, 2020, pp. 6102–6109.

[44] A. Wang, D. Wei, B. Wang, and Z. Jin, Noisy low-tubal-rank tensor completion through iterative singular tube thresholding, *IEEE Access* 6 (2018) 35112–35128.

[45] A. Wang, G. Zhou, Z. Jin, and Q. Zhao, Tensor recovery via $*_L$-spectral k-support norm, *IEEE Journal of Selected Topics in Signal Processing* 15 (2021) 522–534.

[46] A. Wang, G. Zhou, and Q. Zhao. Guaranteed robust tensor completion via $*_L$-SVD with applications to remote sensing data, *Remote Sensing* 13 (2021) 3671.

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

[48] Y. Yang and S. Newsam, Bag-of-visual-words and spatial extensions for land-use classi-fication, In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010, pp. 270–279.

[49] X. Zhang and M. Ng. Low rank tensor completion with Poisson observations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2021.

[50] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3842–3849.

[51] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, Tensor ring decomposition, *arXiv preprint arXiv:1606.05535* 2016.

[52] Y. Zheng, T. Huang, X. Zhao, Y. Chen, and W. He, Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image, *IEEE Transactions on Geoscience and Remote Sensing* 58 (2020) 8450–8464.

$\boxed{\text{A}}$ Proofs of the Theoretical Results in Sec. 5

In the section, we provide technical proof of the theoretical results in Sec. 5. Before giving detailed proofs, we first introduce some preliminaries basis for further analysis.

Suppose any tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with tubal rank r^* has reduced t-SVD as follows

$$\mathcal{X} = \mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top, \tag{A.1}$$

where $\mathcal{U}_\mathcal{X} \in \mathbb{R}^{n_1 \times r^* \times n_3}$ and $\mathcal{V}_\mathcal{X} \in \mathbb{R}^{n_2 \times r^* \times n_3}$ are orthogonal and $\mathcal{S}_\mathcal{X} \in \mathbb{R}^{r^* \times r^* \times n_3}$ is f-diagonal.

Define a tensor space T as:

$$T := \left\{ \mathcal{U}_\mathcal{X} * \mathcal{A} + \mathcal{B} * \mathcal{V}_\mathcal{X}^\top \ : \ \text{where } \mathcal{A} \in \mathbb{R}^{r^* \times n_2 \times n_3}, \mathcal{B} \in \mathbb{R}^{n_1 \times r^* \times n_3} \right\}.$$

We further define the projectors to T and $T^\perp$ as $P_T : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $P_{T^\perp} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^{n_1 \times n_2 \times n_3}$ respectively

$$\begin{aligned} P_T(\mathcal{T}) &= \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * \mathcal{T} + \mathcal{T} * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top - \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * \mathcal{T} * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top, \\ P_{T^\perp}(\mathcal{T}) &= (\mathcal{I} - \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top) * \mathcal{T} * (\mathcal{I} - \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top). \end{aligned} \tag{A.2}$$

Thus, we have

$$\begin{aligned} r_{\text{tb}}(P_T(\mathcal{T})) &\leq r_{\text{tb}}(\mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * \mathcal{T}) + r_{\text{tb}}((\mathcal{I} - \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top) * \mathcal{T} * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top) \\ &\leq 2 r_{\text{tb}}(\mathcal{X}). \end{aligned} \tag{A.3}$$

Equipped with the decomposibility of TNN in Lemma 2 in the supplementary material of [42], we will present an inequality frequently used in the analysis as follows.

Lemma A.1. *Given* $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *we have*

$$\|\mathcal{X} + P_{T^\perp}(\mathcal{T})\|_\star = \|\mathcal{X}\|_\star + \|P_{T^\perp}(\mathcal{T})\|_\star. \tag{A.4}$$

A.1 **Proofs for Lemma 5.1 to Lemma 5.4**

Proof of Lemma 5.1. Let $\Delta := \hat{\mathcal{L}} - \mathcal{L}^*$ be the tensor of estimation error. Let $\mathscr{L}(\mathcal{L}) = \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_{\mathrm{F}}^2 + \lambda\|\mathcal{L}\|_{\star\mathrm{o}}$. Then according to the optimality of $\hat{\mathcal{L}}$ to Problem (3.5), we have

$$\mathscr{L}(\hat{\mathcal{L}}) \leq \mathscr{L}(\mathcal{L}^*) \tag{A.5}$$

which means

$$\frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - (\mathcal{L}^* + \Delta))\|_{\mathrm{F}}^2 + \lambda \sum_{k=1}^{K} w_k\|\mathcal{L}_{[k]}^* + \Delta_{[k]}\|_\star \leq \frac{1}{2}\|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L}^*)\|_{\mathrm{F}}^2 + \lambda \sum_{k=1}^{K} w_k\|\mathcal{L}_{[k]}^*\|_\star$$

$$\Rightarrow \frac{1}{2}\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2 - \sigma\langle\mathcal{B} \odot \mathcal{E}, \Delta\rangle \leq \lambda(\sum_{k=1}^{K} w_k\|\mathcal{L}_{[k]}^*\|_\star - \sum_{k=1}^{K} w_k\|\mathcal{L}_{[k]}^* + \Delta_{[k]}\|_\star)$$

Let $\mathcal{P}_k(\cdot) = \mathcal{P}_{\mathfrak{F}_k(\mathcal{L}^*)}(\cdot)$ (see the definition of P_T in Eq. (A.2)). For any tensor $\Delta \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, define

$$\Delta'_k = \mathcal{P}_k(\Delta_{[k]}), \text{ and } \Delta''_k = \Delta_{[k]} - \Delta'_k.$$

Using Lemma A.1 directly yields

$$\|\mathcal{L}_{[k]}^* - \Delta''_k\|_\star = \|\mathcal{L}_{[k]}^*\|_\star + \|\Delta''_k\|_\star.$$

It leads to

$$\begin{aligned}\|\mathcal{L}_{[k]}^* - \Delta_{[k]}\|_\star &= \|(\mathcal{L}_{[k]}^* - \Delta''_k) - \Delta'_k\|_\star \\ &\geq \|\mathcal{L}_{[k]}^* - \Delta''_k\|_\star - \|\Delta'_k\|_\star \\ &= \|\mathcal{L}_{[k]}^*\|_\star + \|\Delta''_k\|_\star - \|\Delta'_k\|_\star.\end{aligned}$$

which further indicates

$$\begin{aligned}\sum_k w_k\left(\|\mathcal{L}_{[k]}^*\|_\star - \|(\mathcal{L}^* - \Delta)_{[k]}\|_\star\right) &\leq \sum_k w_k\left(\|\mathcal{L}_{[k]}^*\|_\star - (\|\mathcal{L}_{[k]}^*\|_\star + \|\Delta''_k\|_\star - \|\Delta'_k\|_\star)\right) \\ &= \sum_k w_k\|\Delta'_k\|_\star - \sum_k w_k\|\Delta''_k\|_\star.\end{aligned}$$

Thus, we have

$$\frac{1}{2}\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2$$
$$\leq \lambda(\sum_k w_k\|\Delta'_k\|_\star - \sum_k w_k\|\Delta''_k\|_\star) + \sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*\|\Delta\|_{\star\mathrm{o}}$$
$$\leq \lambda(\sum_k w_k\|\Delta'_k\|_\star - \sum_k w_k\|\Delta''_k\|_\star) + \sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*(\sum_k w_k\|\Delta'_k\|_\star + \sum_k w_k\|\Delta''_k\|_\star) \tag{A.6}$$
$$\leq (\lambda + \sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*)\sum_k w_k\|\Delta'_k\|_\star - (\lambda - \sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*)\sum_k w_k\|\Delta''_k\|_\star$$

Then by setting

$$\lambda \geq 2\sigma\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^* \tag{A.7}$$

we obtain

$$\frac{1}{2}\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2 + \frac{\lambda}{2}\sum_k w_k\|\Delta''_k\|_\star \leq \frac{3\lambda}{2}\sum_k w_k\|\Delta'_k\|_\star$$

which means

$$\sum_k w_k \|\Delta_k''\|_\star \le 3 \sum_k w_k \|\Delta_k'\|_\star \tag{A.8}$$

further leading to

$$
\begin{aligned}
\|\Delta\|_{\star\mathrm{o}} &\le \sum_k w_k \|\Delta_k'\|_\star + \sum_k w_k \|\Delta_k''\|_\star \\
&\le 4 \sum_k w_k \|\Delta_k'\|_\star \\
&\le 4 \sum_k w_k \sqrt{2r_k^*} \|\Delta_k'\|_\mathrm{F} \\
&\le 4 \sum_k w_k \sqrt{2r_k^*} \|\Delta_{[k]}\|_\mathrm{F} \\
&= \sum_k w_k \sqrt{32 r_k^*} \|\Delta\|_\mathrm{F}
\end{aligned}
\tag{A.9}
$$

According to Eq. (A.6), we also have

$$\frac{1}{2}\|\mathcal{B}\odot\Delta\|_\mathrm{F}^2 \le \frac{3\lambda}{2}\sum_k w_k\|\Delta_k'\|_\star \le \frac{3\lambda}{2}\sum_k w_k\sqrt{2r_k^*}\|\Delta_k'\|_\mathrm{F} \le \frac{3\lambda}{2}\sum_k w_k\sqrt{2r_k^*}\|\Delta\|_\mathrm{F} \tag{A.10}$$

$\square$

Proof of Lemma 5.2. We define the following event and upper bound the probability of it:

$$\mathfrak{E} := \Big\{\exists \Delta \in \mathsf{C}(\beta, \boldsymbol{r}) \text{ such that}$$

$$\|\mathcal{B}\odot\Delta\|_\mathrm{F}^2 \le \frac{p}{2}\|\Delta\|_\mathrm{F}^2 - 64p^{-1}\alpha\big(\mathbb{E}\big[\|\mathcal{B}\odot\mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{r_k} + 2\big)^2\Big\}$$

We also define sub-events $\mathfrak{E}_l$ with $l \in \mathbb{N}_+$:

$$\mathfrak{E}_l := \Big\{\exists\Delta \in \mathsf{C}'(t) \text{ with } t \in [\alpha^{l-1}\beta, \alpha^l\beta],$$

$$\text{such that } \|\mathcal{B}\odot\Delta\|_\mathrm{F}^2 \le \frac{p}{2}\|\Delta\|_\mathrm{F}^2 - 64p^{-1}\alpha\big(\mathbb{E}\big[\|\mathcal{B}\odot\mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{r_k} + 2\big)^2\Big\}$$

It holds according to Lemma A.2 that

$$\mathbb{P}\big[\mathfrak{E}_l\big] \le \mathbb{P}\Big[Z_t \ge \frac{1}{2\alpha}\alpha^l\beta + 64p^{-1}\alpha\big(\mathbb{E}\big[\|\mathcal{B}\odot\mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{r_k} + 2\big)^2\Big] \le 4\exp\big(-\frac{p\alpha^l\beta}{32\alpha^2}\big)$$

Then, we have

$$
\begin{aligned}
\mathbb{P}\big[\mathfrak{E}\big] \le \mathbb{P}\big[\bigcup\mathfrak{E}_l\big] \le \sum_{l=1}^{+\infty}\mathbb{P}\big[\mathfrak{E}_l\big] &\le \sum_{l=1}^{+\infty}4\exp\big(-\frac{p\alpha^l\beta}{32\alpha^2}\big) \\
&\le \sum_{l=1}^{+\infty}4\exp\big(-\frac{p\beta l\log\alpha}{32\alpha^2}\big) \\
&= \frac{4\exp\big(-\frac{p\beta\log\alpha}{32\alpha^2}\big)}{1 - \exp\big(-\frac{p\beta\log\alpha}{32\alpha^2}\big)}
\end{aligned}
$$

Letting $\alpha = e$ and $\beta = 32e^2\log 8\tilde{d}/p$, then we have $\mathbb{P}[\mathfrak{E}] \le 1/\tilde{d}$. $\square$

Proof of Lemma 5.3. According to Lemma A.5, we have

$$\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^* \leq \frac{1}{K^2} \sum_{k=1}^{K} w_k^{-1} \|(\mathcal{B} \odot \mathcal{E})_{[k]}\|. \tag{A.11}$$

For a fixed $k \in [K]$, suppose $p \geq (d_k \wedge d_{\backslash k})^{-1}$. Then according to Lemma 5 in [44], it holds with probability at least $1 - \tilde{d}_k^{-1}$ that

$$\|(\mathcal{B} \odot \mathcal{E})_{[k]}\| \leq C_\varrho \sqrt{pD(d_k \wedge d_{\backslash k})^{-1} \log \tilde{d}_k} \tag{A.12}$$

Then by taking union bound $\forall k \in [K]$ and letting $p \geq \max_k (d_k \wedge d_{\backslash k})^{-1}$, we obtain

$$\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^* \leq \frac{C_\varrho}{K^2} \sum_{k=1}^{K} w_k^{-1} \sqrt{pD(d_k \wedge d_{\backslash k})^{-1} \log \tilde{d}_k} \tag{A.13}$$

with probability at least $1 - \sum_k \tilde{d}_k^{-1}$. $\qquad\square$

Proof of Lemma 5.4. According to Lemma A.5, we have

$$\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^* \leq \frac{1}{K^2} \sum_{k=1}^{K} w_k^{-1} \|(\mathcal{B} \odot \mathcal{R})_{[k]}\|. \tag{A.14}$$

For a fixed $k \in [K]$, suppose $p \geq (d_k \wedge d_{\backslash k})^{-1}$ and $pD(d_k \wedge d_{\backslash k})^{-1} \geq 2\log^2(d_k \wedge d_{\backslash k}) \log \tilde{d}_k$. Then according to Lemma 6 in [44], it holds that

$$\mathbb{E}[\|(\mathcal{B} \odot \mathcal{R})_{[k]}\|] \leq C' \sqrt{pD(d_k \wedge d_{\backslash k})^{-1} \log \tilde{d}_k} \tag{A.15}$$

Then by adding them together and letting $p \geq (d_k \wedge d_{\backslash k})^{-1}$ and $pD(d_k \wedge d_{\backslash k})^{-1} \geq 2\log^2(d_k \wedge d_{\backslash k}) \log \tilde{d}_k$, we obtain

$$\mathbb{E}[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*] \leq \frac{C_2}{K^2} \sum_{k=1}^{K} w_k^{-1} \sqrt{pD(d_k \wedge d_{\backslash k})^{-1} \log \tilde{d}_k} \tag{A.16}$$

$$\square$$

A.2 Proof for Theorem 5.5

Proof of Theorem 5.5. Let

$$\beta = \frac{32e^2 \log(8\tilde{d})}{p} \tag{A.17}$$

and consider two cases:

Case 1: If $p\|\Delta\|_{\mathrm{F}}^2 < 128e^2\mathbf{a}^2 \log(8\tilde{d})/p = 4\mathbf{a}^2\beta$, then we obtain

$$\|\Delta\|_{\mathrm{F}}^2 < 4\mathbf{a}^2\beta = \frac{128e^2\mathbf{a}^2 \log(8\tilde{d})}{p^2} \tag{A.18}$$

and thus Bound (5.10) in Theorem 5.5 is satisfied.

Case 2: If $p\|\Delta\|_{\mathrm{F}}^2 \geq 4\mathbf{a}^2\beta$, according to the feasibility of $\mathcal{L}^*$ and $\hat{\mathcal{L}}$ to Problem (3.5) and the triangular inequality, we have

$$\|\Delta\|_\infty = \|\hat{\mathcal{L}} - \mathcal{L}^*\|_\infty \leq \|\hat{\mathcal{L}}\|_\infty - \|\mathcal{L}^*\|_\infty \leq 2\mathbf{a} \tag{A.19}$$

At the same time, Eq. (A.9) indicates that

$$\|\Delta\|_{\star\mathrm{o}} \leq \sum_k w_k \sqrt{32 r_k^*}\|\Delta\|_{\mathrm{F}} \tag{A.20}$$

Combining Eqs. (A.19) and (A.20) with the definition of set $\mathsf{C}(\beta, \boldsymbol{r})$ yields

$$\frac{\Delta}{2\mathbf{a}} \in \mathsf{C}(\beta, 32\boldsymbol{r}^*) \tag{A.21}$$

Then according to Lemma 5.2, it holds with probability at least $1 - 1/\tilde{d}$ that

$$\frac{1}{4\mathbf{a}^2}\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2 \geq \frac{p}{8\mathbf{a}^2}\|\Delta\|_{\mathrm{F}}^2 - 256ep^{-1}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k \sqrt{2r_k} + 2\big)^2 \tag{A.22}$$

Combining Eq. (A.22) and Lemma 5.1 gives

$$\frac{p}{4}\|\Delta\|_{\mathrm{F}}^2 \leq \frac{3\lambda}{2}\sum_k w_k\sqrt{2r_k^*}\|\Delta\|_{\mathrm{F}} + 1024e\mathbf{a}^2p^{-1}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{2r_k} + 2\big)^2$$

According to $2\sqrt{xy} \leq xc + y/c$, we have

$$\frac{3\lambda}{2}\sum_k w_k\sqrt{2r_k^*}\|\Delta\|_{\mathrm{F}} = 2 \cdot \frac{3\lambda}{4}\sum_k w_k\sqrt{2r_k^*}\sqrt{8p^{-1}} \cdot \sqrt{\frac{p}{8}\|\Delta\|_{\mathrm{F}}^2}$$

$$\leq 9\lambda p^{-1}(\sum_k w_k\sqrt{r_k^*})^2 + \frac{p}{8}\|\Delta\|_{\mathrm{F}}^2$$

Thus, we obtain a bound on $\|\Delta\|_{\mathrm{F}}^2$ as follows

$$\frac{p}{8}\|\Delta\|_{\mathrm{F}}^2 \leq 9\lambda p^{-1}(\sum_k w_k\sqrt{r_k^*})^2 + 1024e\mathbf{a}^2p^{-1}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{2r_k} + 2\big)^2 \tag{A.23}$$

Combing $\lambda = 2\|\mathcal{B} \odot \mathcal{E}\|_{\star\mathrm{o}}^*$, Lemma 5.3 and Lemma 5.4 with Eq. (A.23) directly yields Bound (5.10) in Theorem 5.5.

$\square$

A.3 Other Important Lemmas

Lemma A.2. *Consider the set*

$$\mathsf{C}'(t) := \Big\{\Delta \ \Big| \ p\|\Delta\|_{\mathrm{F}}^2 \leq t\Big\}\bigcap\mathsf{C}(\beta, \boldsymbol{r}) \tag{A.24}$$

and define the following quantity

$$Z_t := \sup_{\Delta\in\mathsf{C}'(t)}\ \big\{\big|\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2 - \mathbb{E}[\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2]\big|\big\} \tag{A.25}$$

Then, it holds with any constant $\alpha > 1$ that

$$\mathbb{P}\Big[Z_t \geq \frac{1}{2\alpha} + 64p^{-1}\alpha\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^*\big]\sum_k w_k\sqrt{r_k} + 2\big)^2\Big] \leq 4\exp\big(-\frac{pt}{32\alpha^2}\big)$$

Lemma A.3 (Talagrand's concentration inequality [38] modified by [6]). *Suppose that $f :$ $[-1,1]^N \to \mathbb{R}$ is a convex ρ-Lipschitz function. Let $X_1, \ldots, X_N$ be independent random variables taking values in [-1,1]. Let $Z = f(X_1, \ldots, X_N)$. Then it holds for any $t \geq 0$,*

$$\mathbb{P}\left[|Z - \mathbb{E}[Z]| \geq 16\rho + t\right] \leq 4\exp\left(-\frac{t^2}{2\rho^2}\right)$$

Proof of Lemma A.2. The explicit form of Z_t is given as follows:

$$Z_t = \sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2 - \mathbb{E}[\|\mathcal{B} \odot \Delta\|_{\mathrm{F}}^2] \right| \right\}$$

$$= \sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \sum_{i_1 i_2 \ldots i_K} \left(\mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2 - \mathbb{E}[\mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2] \right) \right| \right\}$$

We study the concentration behavior of random variable Z_t by first upper bounding its expectation and then specifying its tail behavior.

First, we bound the expectation of Z_t. By the symmetrization principle [41], we have

$$\mathbb{E}[Z_t] = \mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \sum_{i_1 i_2 \ldots i_K} \left(\mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2 - \mathbb{E}[\mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2] \right) \right| \right\} \right]$$

$$= 2\mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \sum_{i_1 i_2 \ldots i_K} \mathcal{R}_{i_1 i_2 \ldots i_K} \mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2 \right| \right\} \right]$$

where $\mathcal{R}_{i_1 i_2 \ldots i_K}$ are *i.i.d.* Rademacher random variables.

Further, according to the contraction trick [41], it holds that

$$\mathbb{E}[Z_t] \leq 8\mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \sum_{i_1 i_2 \ldots i_K} \mathcal{R}_{i_1 i_2 \ldots i_K} \mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K} \right| \right\} \right]$$

$$= 8\mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \langle \mathcal{R} \odot \mathcal{B}, \Delta \rangle \right| \right\} \right]$$

$$\leq 8\mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^* \|\Delta\|_{\star\mathrm{o}} \right\} \right]$$

$$\leq 8\mathbb{E}\left[\sup_{\Delta \in \mathsf{C}'(t)} \left\{ \|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^* \sum_k w_k \sqrt{r_k} \|\Delta\|_{\mathrm{F}} \right\} \right]$$

$$\leq 8\mathbb{E}\left[\|\mathcal{B} \odot \mathcal{R}\|_{\star\mathrm{o}}^* \right] \sum_k w_k \sqrt{r_k} \sqrt{t/p}$$

Then, we bound the tail behavior of Z_t by using Lemma A.3. By letting

$$f(\mathcal{B}) := Z_t = \sup_{\Delta \in \mathsf{C}'(t)} \left\{ \left| \sum_{i_1 i_2 \ldots i_K} \left(\mathcal{B}_{i_1 i_2 \ldots i_K} \Delta_{i_1 i_2 \ldots i_K}^2 - p\Delta_{i_1 i_2 \ldots i_K}^2 \right) \right| \right\}$$

we can verify that $f(\cdot)$ is a Lipschitz function with parameter x. Indeed following the

proof of Lemma 10 in [20], we have

$$
\begin{aligned}
&f(\mathcal{B}) - f(\mathcal{B}') \\
&= \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 \big| \big\} - \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}'_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 \big| \big\} \\
&\leq \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 \big| - \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}'_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 \big| \big\} \\
&\leq \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 - (\mathcal{B}'_{i_1 i_2 \dots i_K} - p)\Delta_{i_1 i_2 \dots i_K}^2 \big| \big\} \\
&\leq \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \big| \sum_{i_1 i_2 \dots i_K} (\mathcal{B}_{i_1 i_2 \dots i_K} - \mathcal{B}'_{i_1 i_2 \dots i_K})\Delta_{i_1 i_2 \dots i_K}^2 \big| \big\} \\
&\leq \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \|\mathcal{B} - \mathcal{B}'\|_{\mathrm{F}} \sqrt{\sum_{i_1 i_2 \dots i_K} \Delta_{i_1 i_2 \dots i_K}^4} \big\} \\
&\leq \sup_{\Delta \in \mathsf{C}'(t)} \big\{ \|\mathcal{B} - \mathcal{B}'\|_{\mathrm{F}} \sqrt{\sum_{i_1 i_2 \dots i_K} \Delta_{i_1 i_2 \dots i_K}^2} \big\} \leq \|\mathcal{B} - \mathcal{B}'\|_{\mathrm{F}} \sqrt{t/p}
\end{aligned}
$$

Then, according to Lemma A.3, we have

$$
\mathbb{P}[|Z_t - \mathbb{E}[Z_t]| \geq 16\sqrt{t/p} + s] \leq 4\exp(-\frac{ps^2}{2t})
$$

which further leads to

$$
\mathbb{P}[Z_t \geq 8\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star \mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} \sqrt{t/p} + 16\sqrt{t/p} + s] \leq 4\exp(-\frac{ps^2}{2t}) \qquad (\mathrm{A.26})
$$

Using the inequality $2\sqrt{xy} \leq cx + y/c$ yields

$$
\begin{aligned}
&8\sqrt{t/p}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star \mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} + 2\big) \\
&= 2 \cdot \sqrt{ct} \cdot \sqrt{16(cp)^{-1}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star \mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} + 2\big)^2} \qquad (\mathrm{A.27}) \\
&\leq ct + 16(cp)^{-1}\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star \mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} + 2\big)^2
\end{aligned}
$$

By letting $c = 1/(4\alpha), s = ct$ with a constant $\alpha > 1$, we have

$$
\mathbb{P}\Big[Z_t \geq \frac{1}{2\alpha} + 64p^{-1}\alpha\big(\mathbb{E}\big[\|\mathcal{B} \odot \mathcal{R}\|_{\star \mathrm{o}}^*\big] \sum_k w_k \sqrt{r_k} + 2\big)^2\Big] \leq 4\exp\big(-\frac{pt}{32\alpha^2}\big)
$$

$\square$

Lemma A.4 (Dual norm of OITNN-O [43]). *The dual norms of $\|\cdot\|_{\star \mathrm{o}}$, denoted by $\|\cdot\|_{\star \mathrm{o}}^*$ is given as follows:*

$$
\|\mathcal{T}\|_{\star \mathrm{o}}^* := \sup_{\|\mathcal{R}\|_{\star \mathrm{o}} \leq 1} \langle \mathcal{T}, \mathcal{R} \rangle = \inf_{\sum_k \mathcal{X}^{(k)} = \mathcal{T}} \max_k \big\{ w_k^{-1}\|\mathcal{X}_{[k]}^{(k)}\| \big\} \qquad (\mathrm{A.28})
$$

Lemma A.5. *For any K-way ($K \geq 3$) tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}$, the following inequality holds:*

$$\|\mathcal{T}\|_{\star o}^* \leq \frac{1}{K^2} \sum_k w_k^{-1} \|\mathcal{T}_{[k]}\|. \tag{A.29}$$

Proof of Lemma A.5. Recall the formulation of $\|\mathcal{T}\|_{\star o}^*$ in Lemma A.4 as follows

$$\|\mathcal{T}\|_{\star o}^* := \inf_{\sum_k \mathcal{T}^{(k)} = \mathcal{T}} \max_k \{ w_k^{-1} \|\mathcal{T}_{[k]}^{(k)}\| \}. \tag{A.30}$$

Letting

$$\mathcal{T}^{(k)} = \frac{w_k \|\mathcal{T}_{[k]}^{(k)}\|^{-1}}{\sum_k w_k \|\mathcal{T}_{[k]}^{(k)}\|^{-1}} \mathcal{T}, \tag{A.31}$$

then for any $k \in [K]$,

$$\begin{aligned}
w_k^{-1} \|\mathcal{T}_{[k]}^{(k)}\| &= w_k^{-1} \frac{w_k \|\mathcal{T}_{[k]}^{(k)}\|^{-1}}{\sum_k w_k \|\mathcal{T}_{[k]}^{(k)}\|^{-1}} \|\mathcal{T}_{[k]}^{(k)}\| \\
&= \frac{1}{\sum_k w_k \|\mathcal{T}_{[k]}^{(k)}\|^{-1}} \\
&\leq \frac{1}{K^2} \sum_k w_k^{-1} \|\mathcal{T}_{[k]}^{(k)}\|,
\end{aligned} \tag{A.32}$$

where the last inequality holds because the "harmonic mean" is no larger than the "arithmetic mean". In this way, the lemma is proved. $\qquad\square$

B | Proofs of the Theoretical Results in Sec. 4

Proof of Theorem 4.2. The key idea is to rewrite Problem (4.1) into a standard two-block ADMM problem. For notational simplicity, let

$$f(\boldsymbol{u}) = \frac{1}{2} \|\mathcal{B} \odot (\mathcal{Y} - \mathcal{L})\|_{\mathrm{F}}^2 + \boldsymbol{\delta}_{\mathbf{a}}^{\infty}(\mathcal{L}), \qquad g(\boldsymbol{v}) = \lambda \sum_{k=1}^{K} w_k \|\mathcal{K}_{[k]}^{(k)}\|_{\star},$$

with $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ and $\boldsymbol{A}$ defined as follows

$$\boldsymbol{u} := \mathrm{vec}(\mathcal{L}) \in \mathbb{R}^D, \qquad \boldsymbol{v} := \begin{bmatrix} \mathrm{vec}(\mathcal{K}^{(1)}) \\ \mathrm{vec}(\mathcal{K}^{(2)}) \\ \vdots \\ \mathrm{vec}(\mathcal{K}^{(K)}) \end{bmatrix} \in \mathbb{R}^{KD},$$

$$\boldsymbol{w} := \begin{bmatrix} \mathrm{vec}(\mathcal{W}^{(1)}) \\ \mathrm{vec}(\mathcal{W}^{(2)}) \\ \vdots \\ \mathrm{vec}(\mathcal{W}^{(K)}) \end{bmatrix} \in \mathbb{R}^{KD}, \qquad \boldsymbol{A} := \begin{bmatrix} \boldsymbol{I}_D \\ \boldsymbol{I}_D \\ \vdots \\ \boldsymbol{I}_D \end{bmatrix} \in \mathbb{R}^{KD \times D}$$

where $\mathrm{vec}(\cdot)$ denotes the operation of tensor vectorization (see [21]).

It can be verified that $f(\cdot)$ and $g(\cdot)$ are closed, proper convex functions. Then, Problem (4.1) can be re-written as follows:

$$\min_{\boldsymbol{u},\boldsymbol{v}} \quad f(\boldsymbol{u}) + g(\boldsymbol{v})$$

$$\text{s.t.} \quad \boldsymbol{Au} - \boldsymbol{v} = \boldsymbol{0}.$$

According to the convergence analysis in [1], we have:

objective convergence: $\qquad \lim_{t\to\infty} f(\boldsymbol{u}^t) + g(\boldsymbol{v}^t) = f^\star + g^\star,$

dual variable convergence: $\qquad \lim_{t\to\infty} \boldsymbol{w}^t = \boldsymbol{w}^\star,$

constraint convergence: $\qquad \lim_{t\to\infty} \boldsymbol{Au}^t - \boldsymbol{v}^t = \boldsymbol{0},$

where $f^\star, g^\star$ are the optimal values of $f(\boldsymbol{u})$, $g(\boldsymbol{v})$, respectively. Variable $\boldsymbol{w}^\star$ is a dual optimal point defined as:

$$\boldsymbol{w}^\star = \begin{bmatrix} \text{vec}(\mathcal{W}^{(1)\star}) \\ \text{vec}(\mathcal{W}^{(2)\star}) \\ \vdots \\ \text{vec}(\mathcal{W}^{(K)\star}) \end{bmatrix}$$

where $(\{\mathcal{W}^{(k)\star}\}_k)$ are the dual variables in a saddle point $(\mathcal{L}^\star, \{\mathcal{K}^{(k)\star}\}_k, \{\mathcal{W}^{(k)\star}\}_k)$ of the unaugmented Lagrangian $L_0(\mathcal{L}, \{\mathcal{K}^{(k)}\}_k, \{\mathcal{W}^{(k)}\}_k)$. $\qquad\square$

ANDONG WANG
School of Automation, Guangdong University of Technology
Guangzhou, 510006, China
Tensor Learning Team, RIKEN AIP, Tokyo, 103-0027, Japan
E-mail address: w.a.d@outlook.com

GUOXU ZHOU
School of Automation, Guangdong University of Technology
Guangzhou, 510006, China
E-mail address: gx.zhou@gdut.edu.cn

ZHONG JIN
School of Computer Science and Engineering
Nanjing University of Science and Technology
Nanjing, 200094, China
E-mail address: zhongjin@njust.edu.cn

QIBIN ZHAO
School of Automation, Guangdong University of Technology
Guangzhou, 510006, China
Tensor Learning Team, RIKEN AIP, Tokyo, 103-0027, Japan
E-mail address: qibin.zhao@riken.jp

Table 1: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *Sky* dataset in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	22.57	25.23	27.52	27.66	**29.25**
	SSIM	0.3893	0.7712	0.6226	0.6937	**0.7953**
Uniform $p = 0.1$	PSNR	24.53	27.15	28.56	28.79	**30.61**
	SSIM	0.4826	0.7964	0.6572	0.7395	**0.8196**
Uniform $p = 0.15$	PSNR	25.68	28.38	29.28	29.52	**31.32**
	SSIM	0.5257	0.8097	0.6826	0.7679	**0.8276**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	3.58	23.84	27.22	27.22	**28.14**
	SSIM	0.0046	0.7075	0.6135	0.6776	**0.758**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	3.99	26.28	28.16	28.29	**29.56**
	SSIM	0.0057	0.7676	0.645	0.7219	**0.8095**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	4.38	27.6	28.82	29.01	**30.52**
	SSIM	0.0066	0.7872	0.6649	0.7473	**0.8361**

Table 2: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the hyperspectral datasets in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

(a) *Indian Pines*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	20.44	22.01	25.68	26.26	**28.64**
	SSIM	0.3895	0.6359	0.6293	0.6727	**0.7985**
Uniform $p = 0.1$	PSNR	22.23	24.94	27.45	28.4	**30.41**
	SSIM	0.4836	0.7171	0.7226	0.7744	**0.8484**
Uniform $p = 0.15$	PSNR	23.52	26.61	28.54	29.52	**31.24**
	SSIM	0.5438	0.7668	0.7713	0.8177	**0.8687**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	6.63	21.17	24.07	24.32	**26.58**
	SSIM	0.012	0.575	0.5277	0.5479	**0.7307**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	7.04	23.95	25.56	26.03	**28.2**
	SSIM	0.016	0.6759	0.6226	0.6535	**0.7933**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	7.43	25.67	26.62	27.06	**29.27**
	SSIM	0.0222	0.7328	0.6798	0.712	**0.8285**

(b) *Salinas A*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	15.21	20.79	22.55	26.52	**30.55**
	SSIM	0.2594	0.7547	0.5667	0.7384	**0.8756**
Uniform $p = 0.1$	PSNR	20.62	25.56	25.72	29.61	**32.64**
	SSIM	0.4775	0.8284	0.7027	0.8403	**0.9075**
Uniform $p = 0.15$	PSNR	23.09	27.99	28.06	31.32	**34.02**
	SSIM	0.5643	0.8622	0.7804	0.8798	**0.9246**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	6.48	18.11	19.08	21.99	**25.44**
	SSIM	0.0174	0.5246	0.383	0.5264	**0.7895**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	6.92	23.63	21.04	25.01	**27.98**
	SSIM	0.0264	0.7537	0.5257	0.6962	**0.8463**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	7.31	26.43	22.42	27.19	**30.31**
	SSIM	0.0358	0.8232	0.5947	0.7763	**0.889**

Table 3: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the multi-spectral datasets in the uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

(a)   *Cloth*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	20.1	20.95	25	26.09	**28.72**
	SSIM	0.3762	0.5096	0.6773	0.7283	**0.8674**
Uniform $p = 0.1$	PSNR	21.14	22.72	28	29.24	**32.13**
	SSIM	0.4341	0.5983	0.8132	0.854	**0.9324**
Uniform $p = 0.15$	PSNR	22.05	24.18	30.03	31.36	**34.2**
	SSIM	0.4889	0.6783	0.8722	0.9054	**0.954**

(b)   *Hair*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	25.33	30.09	33.16	35.31	**36.16**
	SSIM	0.7147	0.8631	0.8917	0.9248	**0.9533**
Uniform $p = 0.1$	PSNR	29.52	33.35	36.22	38.18	**38.88**
	SSIM	0.8008	0.9122	0.9292	0.9535	**0.9673**
Uniform $p = 0.15$	PSNR	31.12	35.24	38	39.88	**40.44**
	SSIM	0.8364	0.9336	0.9449	0.965	**0.9725**

(c)   *Jelly Beans*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	16.33	18.21	25.43	26.47	**29.6**
	SSIM	0.2397	0.4942	0.6726	0.7223	**0.8755**
Uniform $p = 0.1$	PSNR	18.12	22.11	28.5	30.14	**33.26**
	SSIM	0.3169	0.6629	0.79	0.8518	**0.9344**
Uniform $p = 0.15$	PSNR	19.92	24.67	30.51	32.33	**35.37**
	SSIM	0.4053	0.7592	0.8489	0.903	**0.9534**

(d)   *Oil Painting*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	22.46	24.98	29.74	30.26	**33.09**
	SSIM	0.5542	0.7363	0.8146	0.8327	**0.9129**
Uniform $p = 0.1$	PSNR	24.78	27.76	32.31	32.87	**36**
	SSIM	0.6466	0.8213	0.8784	0.8963	**0.9479**
Uniform $p = 0.15$	PSNR	26.51	29.66	34.18	34.73	**37.85**
	SSIM	0.7079	0.8641	0.9105	0.9267	**0.9612**

Table 4: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the multi-spectral datasets in the non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

(a) *Cloth*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Nonuniform $\mathfrak{p} = 0.05$	PSNR	13.67	19.53	22.17	23.3	**24.65**
	SSIM	0.0958	0.4461	0.5514	0.5882	**0.7028**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	14.07	21.73	24.05	25.63	**26.65**
	SSIM	0.1373	0.5554	0.6675	0.7156	**0.7943**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	14.46	23.24	25.32	27.02	**28.1**
	SSIM	0.1728	0.6346	0.7336	0.7805	**0.844**

(b) *Hair*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Nonuniform p=0.05	PSNR	15.48	26.64	28.94	31.87	**32.18**
	SSIM	0.2823	0.7933	0.8293	0.8733	**0.9107**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	15.92	30.81	30.54	33.62	**34.21**
	SSIM	0.3392	0.8748	0.8701	0.9116	**0.9366**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	16.24	33.09	31.5	34.83	**35.61**
	SSIM	0.3833	0.9068	0.8886	0.9285	**0.9481**

(c) *Jelly Beans*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Nonuniform $\mathfrak{p} = 0.05$	PSNR	13.78	17.18	23.4	23.93	**25.41**
	SSIM	0.1854	0.4429	0.5877	0.6065	**0.7614**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	14.2	20.32	25.68	26.59	**27.89**
	SSIM	0.2351	0.5798	0.7046	0.7396	**0.8437**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	14.59	22.7	27.21	28.3	**29.39**
	SSIM	0.2762	0.6784	0.7681	0.8059	**0.8799**

(d) *Oil Painting*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Nonuniform $\mathfrak{p} = 0.05$	PSNR	14.99	23.21	28.15	28.62	**29.84**
	SSIM	0.1857	0.6633	0.7618	0.7798	**0.8445**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	15.4	26.35	30.08	30.77	**32.02**
	SSIM	0.2319	0.7685	0.8243	0.8461	**0.8933**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	15.78	28.25	31.4	32.19	**33.41**
	SSIM	0.2687	0.825	0.8583	0.8824	**0.9174**

Table 5: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the LiDAR point cloud datasets in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

(a) *ScenerioB Distance*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	15.94	19.95	23.08	23.14	**23.16**
	SSIM	0.305	0.6253	0.7752	0.7818	**0.788**
Uniform $p = 0.1$	PSNR	17.4	22.11	24.71	24.78	**25.11**
	SSIM	0.4203	0.7164	0.8425	0.8514	**0.8568**
Uniform $p = 0.15$	PSNR	18.43	23.54	25.73	25.84	**26.18**
	SSIM	0.4966	0.7664	0.8733	0.8832	**0.8846**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	13.63	17.25	21.01	21.07	**21.22**
	SSIM	0.1655	0.4905	0.6889	0.6926	**0.7105**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	14	19.44	22.16	22.19	**22.45**
	SSIM	0.2027	0.6209	0.7472	0.7559	**0.7659**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	14.39	20.82	22.86	22.88	**23.17**
	SSIM	0.2325	0.675	0.7798	0.7877	**0.7958**

(b) *ScenerioB Intensity*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	14.81	17.59	20.76	20.8	**20.9**
	SSIM	0.1461	0.3301	0.5462	0.5487	**0.5558**
Uniform $p = 0.1$	PSNR	16.12	19.39	22.15	22.23	**22.5**
	SSIM	0.2311	0.4573	0.6628	0.6678	**0.6913**
Uniform $p = 0.15$	PSNR	17.04	20.59	22.98	23.09	**23.42**
	SSIM	0.2996	0.5458	0.7221	0.7293	**0.7511**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	12.55	15.94	19.27	19.28	**19.44**
	SSIM	0.1478	0.2453	0.4374	0.4384	**0.421**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	12.95	17.88	20.29	20.34	**20.57**
	SSIM	0.18	0.3719	0.5204	0.5229	**0.5222**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	13.29	18.94	20.9	20.95	**21.26**
	SSIM	0.215	0.4497	0.5741	0.5768	**0.5856**

Table 6: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *PolSAR* dataset in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	26.24	26.08	26.48	26.33	**28.16**
	SSIM	0.8740	0.8355	0.8007	0.7850	**0.9193**
Uniform $p = 0.1$	PSNR	27.72	28.28	28.22	28.40	**30.27**
	SSIM	0.9007	0.8904	0.8429	0.8529	**0.9419**
Uniform $p = 0.15$	PSNR	28.89	29.71	29.56	29.90	**32.09**
	SSIM	0.9162	0.9150	0.8643	0.8864	**0.9563**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	13.63	17.25	21.01	21.08	**21.22**
	SSIM	0.1655	0.4905	0.6889	0.6926	**0.7105**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	14	19.44	22.16	22.19	**22.45**
	SSIM	0.2027	0.6209	0.7472	0.7559	**0.7659**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	14.39	20.82	22.86	22.88	**23.17**
	SSIM	0.2325	0.675	0.7798	0.7877	**0.7958**

Table 7: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *Seismic* dataset in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	22.25	22.28	22.49	22.62	**23.17**
	SSIM	0.4369	0.4906	0.3928	0.4168	**0.5206**
Uniform $p = 0.1$	PSNR	23.58	23.53	23.48	23.61	**24.17**
	SSIM	0.5462	0.5740	0.5004	0.5265	**0.6163**
Uniform $p = 0.15$	PSNR	24.74	24.66	24.51	24.61	**25.42**
	SSIM	0.6266	0.6552	0.5898	0.6148	**0.7044**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	6.14	6.67	8.87	8.81	**22.24**
	SSIM	0.0115	0.0165	0.0272	0.0276	**0.4341**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	6.55	7.83	10.64	10.62	**23.23**
	SSIM	0.0151	0.0254	0.0699	0.0646	**0.512**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	6.94	9.19	11.97	11.94	**23.93**
	SSIM	0.0213	0.047	0.127	0.1163	**0.5768**

OISY TENSOR COMPLETION VIA OITNN   313

Table 8: The PSNR and SSIM values obtained by five norms (NN [19], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the thermal imaging data in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

(a) The *OSU Thermal Database*

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	13.19	15.83	28.06	27.99	**28.79**
	SSIM	0.1848	0.4759	0.8584	0.8707	**0.8913**
Uniform $p = 0.1$	PSNR	14.67	19.75	31.30	31.62	**32.14**
	SSIM	0.2509	0.6594	0.9151	0.9326	**0.9316**
Uniform $p = 0.15$	PSNR	16.27	22.52	33.02	33.51	**33.88**
	SSIM	0.3273	0.7621	0.9315	0.9509	**0.9428**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	9.41	14.77	22.47	22.38	**23.59**
	SSIM	0.1331	0.4102	0.6776	0.6855	**0.7691**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	9.83	18.12	24.75	24.72	**25.9**
	SSIM	0.1567	0.5783	0.7746	0.7874	**0.8474**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	10.2	20.56	26.78	26.84	**27.83**
	SSIM	0.1809	0.6823	0.8273	0.8427	**0.8877**

(b) The *Infraed Detection* dataset

Settings		NN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	27.50	29.52	31.37	31.86	**32.54**
	SSIM	0.5961	0.8106	0.7517	0.8042	**0.8278**
Uniform $p = 0.1$	PSNR	29.13	31.24	32.03	32.71	**33.50**
	SSIM	0.6510	0.8358	0.7610	0.8215	**0.8415**
Uniform $p = 0.15$	PSNR	30.00	32.24	32.56	33.27	**34.10**
	SSIM	0.6729	0.8486	0.7721	0.8344	**0.8481**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	6.42	28.26	31.25	31.58	**31.95**
	SSIM	0.0098	0.7715	0.7542	0.8003	**0.8207**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	6.83	30.5	31.78	32.32	**32.84**
	SSIM	0.0115	0.8165	0.7596	0.8157	**0.8384**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	7.23	31.64	32.23	32.87	**33.66**
	SSIM	0.0126	0.8377	0.7675	0.8285	**0.8553**

Table 9: The PSNR and SSIM values obtained by six norms (NN [19], SqNN [29], SNN [39], TNN-DFT [42], TNN-DCT [27], and our OITNN-O) for noisy tensor completion on the *Ground* datasets in the uniform and non-uniform sampling settings. The highest PSNR and SSIM values are highlighted in **bold**.

Settings		NN	SqNN	SNN	TNN-DFT	TNN-DCT	OITNN-O
Uniform $p = 0.05$	PSNR	24.83	26.80	25.03	30.90	30.94	**32.85**
	SSIM	0.7697	0.8011	0.8055	0.8955	0.8997	**0.9306**
Uniform $p = 0.1$	PSNR	27.03	29.10	27.36	32.27	32.49	**34.25**
	SSIM	0.8395	0.8725	0.8559	0.9083	0.9149	**0.9382**
Nonuniform $\mathfrak{p} = 0.05$	PSNR	7.6	13.08	24.94	29.57	29.6	**32.18**
	SSIM	0.0444	0.2344	0.7923	0.8763	0.8801	**0.9239**
Nonuniform $\mathfrak{p} = 0.1$	PSNR	8.54	15.35	26.66	30.93	31.1	**33.84**
	SSIM	0.0593	0.3242	0.8348	0.8924	0.8987	**0.9312**
Nonuniform $\mathfrak{p} = 0.15$	PSNR	9.21	17.1	28.19	31.52	31.8	**33.99**
	SSIM	0.0711	0.5342	0.8644	0.8971	0.905	**0.9372**