remote sensing

MDPI

Article

Guaranteed Robust Tensor Completion via $*_L$-SVD with Applications to Remote Sensing Data

Andong Wang [1,2,3], **Guoxu Zhou** [1,3] and **Qibin Zhao** [1,2,*]

[1] School of Automation, Guangdong University of Technology, Guangzhou 510006, China; w.a.d@gdut.edu.cn (A.W.); gx.zhou@gdut.edu.cn (G.Z.)
[2] Tensor Learning Team, RIKEN AIP, Tokyo 103-0027, Japan
[3] Key Laboratory of Intelligent Detection and The Internet of Things in Manufacturing, Ministry of Education, Guangzhou 510006, China
[*] Correspondence: qibin.zhao@riken.jp

Abstract: This paper conducts a rigorous analysis for the problem of robust tensor completion, which aims at recovering an unknown three-way tensor from incomplete observations corrupted by gross sparse outliers and small dense noises simultaneously due to various reasons such as sensor dead pixels, communication loss, electromagnetic interferences, cloud shadows, etc. To estimate the underlying tensor, a new penalized least squares estimator is first formulated by exploiting the low rankness of the signal tensor within the framework of tensor $*_L$-Singular Value Decomposition ($*_L$-SVD) and leveraging the sparse structure of the outlier tensor. Then, an algorithm based on the Alternating Direction Method of Multipliers (ADMM) is designed to compute the estimator in an efficient way. Statistically, the non-asymptotic upper bound on the estimation error is established and further proved to be optimal (up to a log factor) in a minimax sense. Simulation studies on synthetic data demonstrate that the proposed error bound can predict the scaling behavior of the estimation error with problem parameters (i.e., tubal rank of the underlying tensor, sparsity of the outliers, and the number of uncorrupted observations). Both the effectiveness and efficiency of the proposed algorithm are evaluated through experiments for robust completion on seven different types of remote sensing data.

Keywords: remote sensing data restoration; robust tensor completion; tensor SVD; statistical performance; ADMM

check for updates

Citation: Wang, A.; Zhou, G.; Zhao Q. Guaranteed Robust Tensor Completion via $*_L$-SVD with Applications to Remote Sensing Data. *Remote Sens.* **2021**, *13*, 3671. https://doi.org/10.3390/rs13183671

Academic Editors: Karen Egiazarian, Aleksandra Pizurica and Vladimir Lukin

Received: 30 July 2021
Accepted: 9 September 2021
Published: 14 September 2021

1. Introduction

Despite the broad adoption of advanced sensors in various remote sensing tasks, the quality of data remains a critical issue and can significantly influence the actual performances of the backend applications. Many types of modern remote sensing data in the modality of optical, hyperspectral, multispectral, thermal, Light Detection and Ranging (LiDAR), Synthetic Aperture Radar (SAR), etc., are typically multi-way and can be readily stored, analyzed, and processed by tensor-based models [1–7]. In some extreme circumstances, the data tensor may encounter missing entries, gross sparse outliers, and small dense noises at the same time, as a result of partial sensor failures, communication errors, occlusion by obstacles, and so on [8,9]. To robustly complete a partially observed data tensor corrupted by outliers and noises, the problem of robust tensor completion arises.

When only a fraction of partially corrupted observations are available, the crucial point of robust tensor completion lies in the assumption that the underlying data tensor is highly redundant such that the main components of it remain only slightly suppressed by missing information, outliers, and noises, and thus can be effectively reconstructed by exploiting the intrinsic redundancy. The tensor low-rankness is an ideal tool to model the redundancy of tensor data, and has gained extensive attention in remote sensing data restoration [5,10,11].

As higher-order extensions of low-rank matrix models [12], low-rank tensor models are typically formulated as minimization problems of the tensor rank function [13]. However, there are multiple definitions of tensor ranks, such as the CP rank [14], Tucker rank [15], TT rank [16], TR rank [17], etc., which focus on low rank structures in the original domains (like the pixel domain of optimal images) [18,19]. Recently, a remarkably different example named the low-tubal-rank tensor model [20,21] was proposed within the algebraic framework of tensor Singular Value Decomposition (t-SVD) [20,22], which captures low-rankness in the frequency domain defined via Discrete Fourier Transform (DFT). As discussed in [18,19,21,23], the low-tubal-rank tensor models are capable to exploit both low-rankness and smoothness of the tensor data, making it quite suitable to analyze and process diverse remote sensing imagery data which are often simultaneously low-rank and smooth [5,10].

Motivated by the advantages of low-tubal-rankness in modeling remote sensing data, we resolve the robust tensor completion problem by utilizing a generalized low-tubal-rank model based on the tensor $*_L$-Singular Value Decomposition ($*_L$-SVD) [24], which leverages low-rankness in more general transformed domains rather than DFT. What needs to be pointed out is that the $*_L$-SVD has become a research focus in tensor-based signal processing, computer vision, and machine learning very recently [18,23,25,26]. Regarding the preference of theory in this paper, we only introduce several typical works with statistical analysis as follows. For tensor completion in the noiseless settings, Lu et al. [26] proposed a $*_L$-SVD-based model which can exactly recover the underlying tensor under mild conditions. For tensor completion from partial observations corrupted by sparse outliers, Song et al. [27] designed a $*_L$-SVD-based algorithm with exact recovery guarantee. Zhang et al. [25] developed a theoretically guaranteed approach via the $*_L$-SVD to for tensor completion from Poisson noises. The problem of tensor recovery from noisy linear observations is studied in [18] based the $*_L$-SVD with guaranteed statistical performance.

In this paper, we focus on statistical guaranteed approaches in a more challenging setting than the aforementioned $*_L$-SVD-based models, where the underlying signal tensor suffers from missing entries, sparse outliers, and small dense noises simultaneously. Specifically, we resolve the problem of robust tensor completion by formulating a $*_L$-SVD-based estimator whose estimation error is established and further proved to be minimax optimal (up to a log factor). We propose an algorithm based on Alternating Direction Method of Multipliers (ADMM) [28,29] to compute the estimator and evaluate both the effectiveness and efficiency on seven different types of remote sensing data.

The remainder of this paper proceeds as follows. We first introduce some notation and preliminaries in Section 2. Then, the proposed estimator for robust tensor completion is formulated in Section 3. We compute the estimator by using an ADMM-based algorithm described in Section 4. The statistical performance of the proposed estimator is analyzed in Section 5. Experimental results on both synthetic and real datasets are reported in Section 7. We summarize this paper and discuss future directions briefly in Section 8. The proofs of the theoretical results are given in Appendix A.

2. Preliminaries

In this section, we first introduce some notations and then give a brief introduction to the $*_L$-SVD framework.

2.1. Notations

Main notations are listed in Table 1. Let $[d] := \{1, \ldots, d\}$, $\forall d \in \mathbb{N}_+$. Let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$, $\forall a, b \in \mathbb{R}$. For $i \in [d], \mathbf{e}_i \in \mathbb{R}^d$ denotes the standard vector basis whose i_{th} entry is 1 with the others 0. For $(i, j, k) \in [d_1] \times [d_2] \times [d_3]$, the outer product $\mathbf{e}_i \circ \mathbf{e}_j \circ \mathbf{e}_k$ denotes a standard tensor basis in $\mathbb{R}^{d_1 \times d_2 \times d_3}$, whose $(i, j, k)_{th}$ entry is 1 with the others 0. For a 3-way tensor, a tube is a vector defined by fixing indices of the first two modes and varying the third one; A slice is a matrix defined by fixing all but two indices. For any set Θ, $|\Theta|$ denotes its cardinality and $\Theta^{\perp}$ its complement.

Absolute positive constants are denoted by $C, c, c_0,$ etc whose values may vary from line to line. When the field and size of a tensor are not shown explicitly, it is defaulted to be in $\mathbb{R}^{d_1 \times d_2 \times d_3}$. The spectral norm $\|\cdot\|$ and nuclear norm $\|\cdot\|_*$ of a matrix are the maximum and the sum of the singular values, respectively.

Table 1. List of notations.

Notations	Descriptions	Notations	Descriptions
t	a scaler	$\mathbf{T}$	a matrix
$\mathbf{t}$	a vector	$\boldsymbol{\mathcal{T}}$	a tensor
$\boldsymbol{\mathcal{L}}^*$	the true low-rank tensor	$\hat{\boldsymbol{\mathcal{L}}}$	the estimator of $\boldsymbol{\mathcal{L}}^*$
$\boldsymbol{\mathcal{S}}^*$	the true sparse tensor	$\hat{\boldsymbol{\mathcal{S}}}$	the estimator of $\boldsymbol{\mathcal{S}}^*$
y_i	a scalar observation	ξ_i	Gaussian noise
$\boldsymbol{\mathcal{X}}_i$	a design tensor	N	number of observations
N_ι	number of uncorrupted observations	N_s	$N - N_\iota$
Θ_s	support of corruption tensor $\boldsymbol{\mathcal{S}}^*$	$\boldsymbol{\Theta}_s^\perp$	complement of Θ_s
$\mathfrak{X}(\cdot)$	design operator	$\mathfrak{X}^*(\cdot)$	adjoint operator of $\mathfrak{X}(\cdot)$
$\mathbf{L}$	an orthogonal matrix in $\mathbb{R}^{d_3 \times d_3}$	$L(\boldsymbol{\mathcal{T}}) := \boldsymbol{\mathcal{T}} \times_3 \mathbf{L}$	tensor L-transform
$\overline{\mathbf{T}}$	block-diagonal matrix of $L(\boldsymbol{\mathcal{T}})$	$\lVert\boldsymbol{\mathcal{T}}\rVert_{\mathrm{sp}} := \lVert\overline{\mathbf{T}}\rVert$	tensor spectral norm
$\boldsymbol{\mathcal{T}}_{ijk}$	$(i,j,k)_{th}$ entry of $\boldsymbol{\mathcal{T}}$	$\lVert\boldsymbol{\mathcal{T}}\rVert_\star := \lVert\overline{\mathbf{T}}\rVert_*$	tubal nuclear norm
$\boldsymbol{\mathcal{T}}(i,j,:)$	$(i,j)_{th}$ tube of $\boldsymbol{\mathcal{T}}$	$\lVert\boldsymbol{\mathcal{T}}\rVert_1 := \sum_{ijk} \lvert\boldsymbol{\mathcal{T}}_{ijk}\rvert$	tensor l_1-norm
$\boldsymbol{\mathcal{T}}(:,:,k)$	k_{th} frontal slice of $\boldsymbol{\mathcal{T}}$	$\lVert\boldsymbol{\mathcal{T}}\rVert_{\mathrm{F}} := \sqrt{\sum_{ijk} \boldsymbol{\mathcal{T}}_{ijk}^2}$	tensor F-norm
$\mathbf{T}^{(k)}$	$\boldsymbol{\mathcal{T}}(:,:,k)$	$\lVert\boldsymbol{\mathcal{T}}\rVert_\infty := \max_{ijk} \lvert\boldsymbol{\mathcal{T}}_{ijk}\rvert$	tensor l_∞-norm
$\mathbf{T}_{(k)}$	mode-k unfolding of $\boldsymbol{\mathcal{T}}$	$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle := \sum_{ijk} \boldsymbol{\mathcal{A}}_{ijk} \boldsymbol{\mathcal{B}}_{ijk}$	tensor inner product

2.2. *Tensor $*_L$-Singular Value Decomposition*

The tensor $*_L$-SVD is a generalization of the t-SVD [22]. To get a better understanding of $*_L$-SVD, we first introduce several basic notions of t-SVD as follows. For any tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, its block circulant matrix $\texttt{bcirc}(\boldsymbol{\mathcal{T}})$ is defined as

$$\texttt{bcirc}(\boldsymbol{\mathcal{T}}) := \begin{bmatrix} \boldsymbol{\mathcal{T}}^{(1)} & \boldsymbol{\mathcal{T}}^{(d_3)} & \cdots & \boldsymbol{\mathcal{T}}^{(2)} \\ \boldsymbol{\mathcal{T}}^{(2)} & \boldsymbol{\mathcal{T}}^{(1)} & \cdots & \boldsymbol{\mathcal{T}}^{(3)} \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{\mathcal{T}}^{(d_3)} & \boldsymbol{\mathcal{T}}^{(d_3-1)} & \cdots & \boldsymbol{\mathcal{T}}^{(1)} \end{bmatrix}$$

We also define the block vectorization operator and its inverse operator for any $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ by:

$$\texttt{bvec}(\boldsymbol{\mathcal{T}}) := \begin{bmatrix} \boldsymbol{\mathcal{T}}^{(1)} \\ \boldsymbol{\mathcal{T}}^{(2)} \\ \vdots \\ \boldsymbol{\mathcal{T}}^{(d_3)} \end{bmatrix}, \quad \texttt{bvfold}(\texttt{bvec}(\boldsymbol{\mathcal{T}})) = \boldsymbol{\mathcal{T}}$$

Then, based on the operators defined above, we are able to give the definition of the tensor t-product.

Definition 1 (T-product [22])**.** *For any tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{d_2 \times d_4 \times d_3}$, their t-product is a tensor $\boldsymbol{\mathcal{C}}$ of size $d_1 \times d_4 \times d_3$ computed as follows:*

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} := \texttt{bvfold}\big(\texttt{bcirc}(\boldsymbol{\mathcal{A}})\texttt{bvec}(\boldsymbol{\mathcal{B}})\big)$$

If we view the 3-way tensor $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{d_1 \times d_4 \times d_3}$ as a d_1-by-d_4 "matrix" $\mathscr{C}$ of tubes $\boldsymbol{\mathcal{C}}(i,j,:) \in \mathbb{R}^{d_3}$, then the t-product can be analogously conducted like the matrix multiplication by changing scalar multiplication by the circular convolution between the tubes (i.e., vectors), as follows:

$$\boldsymbol{\mathcal{C}}(i,j,:) = \sum_{k=1}^{d_2} \boldsymbol{\mathcal{A}}(i,k,:) \circledast \boldsymbol{\mathcal{B}}(k,j,:) \tag{1}$$

where the symbol $\circledast$ denotes the circular convolution of two tubes $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_3}$ defined as follows [22]:

$$(\mathbf{a} \circledast \mathbf{b})_j = \sum_{k=1}^{d_3} \mathbf{a}_k \mathbf{b}_{1+(j-k)\texttt{mod} d_3}$$

where $\texttt{mod}(\cdot)$ is the modulus operator. According to the well-known relationship between circular convolution and DFT, the t-product is equivalent to matrix multiplication between all the frontal slices in the Fourier domain [22], i.e.,

$$\overline{\boldsymbol{\mathcal{C}}} = \overline{\boldsymbol{\mathcal{A}}} \odot \overline{\boldsymbol{\mathcal{B}}} \tag{2}$$

where $\overline{\boldsymbol{\mathcal{T}}}$ denotes the tensor obtained by conducting DFT on all the mode-3 fibers of any tensor $\boldsymbol{\mathcal{T}}$, i.e.,

$$\overline{\boldsymbol{\mathcal{T}}} = \boldsymbol{\mathcal{T}} \times_3 \mathbf{F}_{d_3} \tag{3}$$

where $\mathbf{F}_{d_3}$ is the transform matrix of DFT [22], and $\times_3$ denotes the tensor mode-3 product [30].

In [24], Kernfeld et al. extended the t-product to the tensor $*_L$-product by replacing DFT by any invertible linear transform $L(\cdot)$ induced by a non-singular transformation matrix $\mathbf{L}$, and established the framework of $*_L$-SVD. In the latest studies, the transformation matrix $\mathbf{L}$ defining the transform L is restricted to be orthogonal [18,26,31,32] (unitary in [25,27]) for better properties, which is also followed in this paper.

Given any orthogonal matrix $\mathbf{L} \in \mathbb{R}^{d_3 \times d_3}$ (though we restrict $\mathbf{L}$ to be orthogonal for simplicy, our analysis still holds with simple extensions for unitary $\mathbf{L}$ [27]), define the associated linear transform $L(\cdot)$ with inverse $L^{-1}(\cdot)$ on any $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ as

$$\overline{\boldsymbol{\mathcal{T}}} = L(\boldsymbol{\mathcal{T}}) := \boldsymbol{\mathcal{T}} \times_3 \mathbf{L}, \text{ and } \quad L^{-1}(\boldsymbol{\mathcal{T}}) := \boldsymbol{\mathcal{T}} \times_3 \mathbf{L}^{-1} \tag{4}$$

Definition 2 (Tensor $*_L$-product [24])**.** *The $*_L$–product of any $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{d_2 \times d_4 \times d_3}$ under the invertible linear transform L in Equation (4), denoted by $\boldsymbol{\mathcal{A}} *_L \boldsymbol{\mathcal{B}}$, is defined as the tensor $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{d_1 \times d_4 \times d_3}$ such that $L(\boldsymbol{\mathcal{C}}) = L(\boldsymbol{\mathcal{A}}) \odot L(\boldsymbol{\mathcal{B}})$.*

Definition 3 ($*_L$–block-diagonal matrix [18])**.** *For any $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, its $*_L$–block-diagonal matrix, denoted by $\overline{\mathbf{T}}$, is defined as the block diagonal matrix whose i-th diagonal block is the i-th frontal slice $\overline{\mathbf{T}}^{(i)}$ of $\overline{\boldsymbol{\mathcal{T}}} = L(\boldsymbol{\mathcal{T}})$, i.e.,*

$$\overline{\mathbf{T}} := \texttt{bdiag}(\overline{\boldsymbol{\mathcal{T}}}) := \begin{bmatrix} \overline{\mathbf{T}}^{(1)} & & \\ & \ddots & \\ & & \overline{\mathbf{T}}^{(d_3)} \end{bmatrix} \in \mathbb{R}^{d_1 d_3 \times d_2 d_3}$$

Based on the notions of tensor $*_L$–transpose, $*_L$-identity tensor, $*_L$-orthogonal tensor, and f-diagonal tensor [24], the $*_L$–SVD (illustrated in Figure 1) is given.

Theorem 1 (Tensor $*_L$–SVD, $*_L$-tubal rank [24])**.** *Any $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ has a tensor $*_L$–Singular Value Decomposition ($*_L$–SVD) under any L in Equation (4), given as follows*

$$\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{U}} *_L \boldsymbol{\mathcal{D}} *_L \boldsymbol{\mathcal{V}}^\top \tag{5}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{d_1 \times d_1 \times d_3}$, $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{d_2 \times d_2 \times d_3}$ are $*_L$-orthogonal, and $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is f-diagonal.

The $*_L$-tubal rank of $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ is defined as the number of non-zero tubes of $\mathcal{D}$ in its $*_L$–SVD in Equation (5) i.e.,

$$r_{\text{tb}}(\mathcal{T}) := \#\{i \mid \mathcal{D}(i,i,:) \neq \mathbf{0}, i \in [d_1 \wedge d_2]\}$$

where # counts the number of elements of a given set.

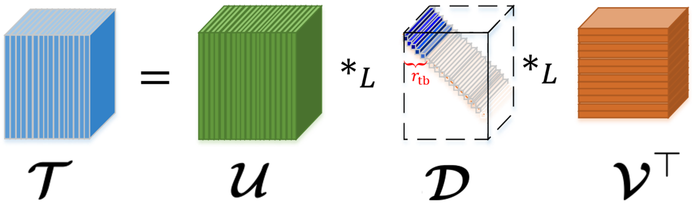


Figure 1. An illustration of $*_L$–SVD [18].

For any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, we have the following equivalence between its $*_L$-SVD and the matrix SVD of its $*_L$–block-diagonal matrix $\overline{\mathbf{T}}$:

$$\mathcal{T} = \mathcal{U} *_L \mathcal{D} *_L \mathcal{V}^\top \quad \Leftrightarrow \quad \overline{\mathbf{T}} = \overline{\mathbf{U}} \cdot \overline{\mathbf{D}} \cdot \overline{\mathbf{V}}^\top.$$

Considering the block diagonal structure of $\overline{\mathbf{T}}$, we define the tensor $*_L$-multi-rank on its diagonal blocks $\overline{\mathbf{T}}^{(i)}$:

Definition 4 (Tensor $*_L$–nuclear norm, tensor $*_L$-spectral norm [26]). *The tensor $*_L$–nuclear norm ($*_L$-TNN) and $*_L$-spectral norm of any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ under any L in Equation (4) are defined as the matrix nuclear norm and matrix spectral norm of $\overline{\mathbf{T}}$, respectively, i.e.,*

$$\|\mathcal{T}\|_\star := \|\overline{\mathbf{T}}\|_*, \quad \|\mathcal{T}\|_{\text{sp}} := \|\overline{\mathbf{T}}\|.$$

As proved in [26,27], $*_L$–TNN is the convex envelop of the l_1-norm of the $*_L$–multi-rank in unit tensor $*_L$-spectral norm ball. Thus, $*_L$–TNN encourages a low $*_L$–multi-rank structure which means low-rankness in spectral domain. When the linear transform L represents the DFT (although we restrict the $\mathbf{L}$ in Equation (4) to be orthogonal, we still consider TNN as a special case of $*_L$–TNN up to constants and real/complex domain) along the 3-rd mode, $*_L$–TNN and tensor $*_L$-spectral norm degenerate to the Tubal Nuclear Norm (TNN) and the tensor spectral norm, respectively, up to a constant factor d_3^{-1} [26,33].

3. Robust Tensor Completion

In this section, we will formulate the robust tensor completion problem. The observation model will be shown first.

3.1. The Observation Model

Consider an underling signal tensor $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ which possesses intrinsically low-dimensionality structure characterized by low-tubal-rankness, that is $r_{\text{tb}}(\mathcal{L}^*) \ll d_1 \wedge d_2$. Suppose we obtain N scalar observations y_i of $\mathcal{L}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ from the noisy observation model:

$$y_i = \langle \mathcal{L}^* + \mathcal{S}^*, \mathcal{X}_i \rangle + \xi_i, \quad \forall i \in [N], \tag{6}$$

where the tensor $\mathcal{S}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ represents some gross corruptions (e.g., outliers, errors, etc.) additive to the signal $\mathcal{L}^*$ which is element-wisely sparse (the presented theoretical analysis and optimization algorithm can be generalized to more sparsity settings of corruptions (e.g. the tube-wise sparsity [20,34], and slice-wise sparsity [34,35]) by using the tools developed for robust matrix completion in [36] and robust tensor decomposition in [34]; for simplicity, we only consider the most common element-wisely sparse case), ξ_i's are random noises sampled i.i.d. from Gaussian distribution $\mathcal{N}(0,\sigma^2)$, and $\boldsymbol{\mathcal{X}}_i$'s are known random design tensors in $\mathbb{R}^{d_1\times d_2\times d_3}$ satisfying the following assumptions:

Assumption 1. *We make two natural assumptions on the design tensors:*

I. *All the corrupted positions of $\boldsymbol{\mathcal{L}}^*$ are observed, that is, the (unknown) support $\Theta_s = \mathtt{supp}(\boldsymbol{\mathcal{S}}^*) := \{(i,j,k) \mid \boldsymbol{\mathcal{S}}^*_{ijk} \neq 0\}$ of the corruption tensor $\boldsymbol{\mathcal{S}}^*$ is fully observed. Formally speaking, there exists an unknown subset $\mathbf{X}_s \subset \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N$ drawn from an (unknown) distribution $\boldsymbol{\Pi}_{\Theta_s}$ on the set $\mathbf{X}_{\Theta_s} := \{\mathbf{e}_j \circ \mathbf{e}_k \circ \mathbf{e}_l, \forall (j,k,l) \in \Theta_s\}$, such that each element in $\mathbf{X}_{\Theta_s}$ is sampled at least once.*

II. *All uncorrupted positions of $\boldsymbol{\mathcal{L}}^*$ are sampled uniformly with replacement for simplicity of exposition. Formally speaking, each element of the set $\mathbf{X}_s^\perp := \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N \backslash \mathbf{X}_s$ is sampled i.i.d. from an uniform distribution $\boldsymbol{\Pi}_{\boldsymbol{\Theta}_s^\perp}$ on the set $\mathbf{X}_{\boldsymbol{\Theta}_s^\perp} := \{\mathbf{e}_j \circ \mathbf{e}_k \circ \mathbf{e}_l, \forall (j,k,l) \in \boldsymbol{\Theta}_s^\perp\}$.*

According to the observation model (6), the true tensor $\boldsymbol{\mathcal{L}}^*$ is first corrupted by a sparse tensor $\boldsymbol{\mathcal{S}}^*$ and then sampled to N scalars $\{y_i\}$ with additive Gaussian noises $\{\xi_i\}$ (see Figure 2). The corrupted positions of $\boldsymbol{\mathcal{L}}^*$ are further assumed in Assumption 1 to be totally observed with design tensors in $\mathbf{X}_s \subset \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N$, and the remaining uncorrupted positions are sampled uniformly through design tensors in $\mathbf{X}_s^\perp = \{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N \backslash \mathbf{X}_s$.

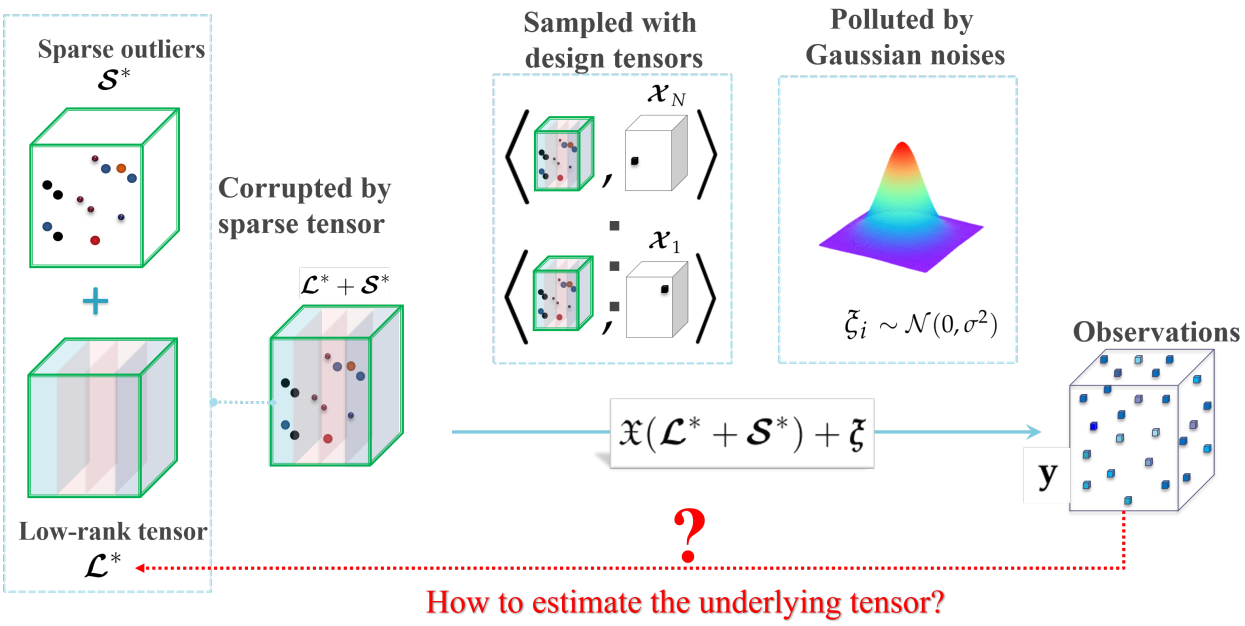

Sparse outliers
$\boldsymbol{\mathcal{S}}^*$
+
Low-rank tensor
$\boldsymbol{\mathcal{L}}^*$
Corrupted by
sparse tensor
$\boldsymbol{\mathcal{L}}^* + \boldsymbol{\mathcal{S}}^*$
Sampled with
design tensors
$\boldsymbol{\mathcal{X}}_N$
$\boldsymbol{\mathcal{X}}_1$
Polluted by
Gaussian noises
$\xi_i \sim \mathcal{N}(0,\sigma^2)$
$\mathfrak{X}(\boldsymbol{\mathcal{L}}^* + \boldsymbol{\mathcal{S}}^*) + \boldsymbol{\xi}$
Observations
y
?
How to estimate the underlying tensor?

Figure 2. An illustration of the robust tensor completion problem.

Let $\mathbf{y} = (y_1,\ldots,y_N)^\top \in \mathbb{R}^N$ and $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_N)^\top \in \mathbb{R}^N$ be the vector of observations and noises, respectively. Define the design operator $\mathfrak{X} : \mathbb{R}^{d_1\times d_2\times d_3} \to \mathbb{R}^N$ as $\mathfrak{X}(\cdot) := (\langle\cdot,\boldsymbol{\mathcal{X}}_1\rangle,\ldots,\langle\cdot,\boldsymbol{\mathcal{X}}_N\rangle)^\top$, and its adjoint operator $\mathfrak{X}^*(\mathbf{z}) := \sum_{i=1}^N z_i\boldsymbol{\mathcal{X}}_i$ for any $\mathbf{z} \in \mathbb{R}^N$. Then the observation model (6) can be rewritten in a compact form

$$\mathbf{y} = \mathfrak{X}(\boldsymbol{\mathcal{L}}^* + \boldsymbol{\mathcal{S}}^*) + \boldsymbol{\xi}.$$

3.2. *The Proposed Estimator*

The aim of robust tensor completion is to reconstruct the unknown low-rank $\boldsymbol{\mathcal{L}}^*$ and sparse $\boldsymbol{\mathcal{S}}^*$ from incomplete and noisy measurements $\{(\boldsymbol{\mathcal{X}}_i, y_i)\}_{i=1}^N$ generated by the observation model (6). It can be treated as a robust extension of tensor completion in [33], and a noisy partial variant of tensor robust PCA [37].

To reconstruct the underlying low-rank tensor $\boldsymbol{\mathcal{L}}^*$ and sparse tensor $\boldsymbol{\mathcal{S}}^*$, it is natural to consider the following minimization model:

$$\min_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}} \frac{1}{2N}\|\mathbf{y}-\mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}})\|_2^2+\lambda_\iota r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}})+\lambda_s\|\boldsymbol{\mathcal{S}}\|_1, \tag{7}$$

where we use least squares as the fidelity term for Gaussian noises, the tubal rank as the regularization to impose low-rank structure in $\boldsymbol{\mathcal{L}}$, the tensor l_0-(pseudo)norm to regularize $\boldsymbol{\mathcal{S}}$ for sparsity, $\lambda_\iota, \lambda_s \geq 0$ are tunable regularization parameters, balancing the regularizations and the fidelity term.

However, general rank and l_0-norm minimization is NP-hard [12,38], making it extremely hard to soundly solve Problem (7). For tractable low-rank and sparse optimization, we follow the most common idea to relax the non-convex functions $r_{\mathrm{tb}}(\cdot)$ and $\|\cdot\|_0$ to their convex surrogates, i.e., the $*_L$–tubal nuclear norm $\|\cdot\|_\star$ and the tensor l_1-norm $\|\cdot\|_1$, respectively. Specifically, the following estimator is defined:

$$(\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{S}}}) := \operatorname*{argmin}_{\|\boldsymbol{\mathcal{L}}\|_\infty\leq a,\|\boldsymbol{\mathcal{S}}\|_\infty\leq a} \frac{1}{2N}\|\mathbf{y}-\mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}})\|_2^2+\lambda_\iota\|\boldsymbol{\mathcal{L}}\|_\star+\lambda_s\|\boldsymbol{\mathcal{S}}\|_1, \tag{8}$$

where $a > 0$ is a known constant constraining the magnitude of entries in $\boldsymbol{\mathcal{L}}^*$ and $\boldsymbol{\mathcal{S}}^*$. The additional constraint $\|\boldsymbol{\mathcal{L}}\|_\infty \leq a$ and $\|\boldsymbol{\mathcal{S}}\|_\infty \leq a$ is very mild since most signals and corruptions 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 separability of $\boldsymbol{\mathcal{L}}^*$ and $\boldsymbol{\mathcal{S}}^*$. Such "non-spiky" constraints are also imposed in previous literatures [36,39,40], 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 $\boldsymbol{\mathcal{L}}^*$ and $\boldsymbol{\mathcal{S}}^*$?

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

4. Algorithm

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

To solve Problem (8), the first step is to introduce auxiliary variables $\mathbf{g}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}$ to deal with the complex couplings between $\mathfrak{X}(\cdot)$, $\|\cdot\|_2$ $\|\cdot\|_\star$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ as follows:

$$\begin{aligned} \min_{\mathbf{g},\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}}} \quad & \frac{1}{2N}\|\mathbf{g}\|_2^2+\lambda_\iota\|\boldsymbol{\mathcal{K}}\|_\star+\lambda_s\|\boldsymbol{\mathcal{T}}\|_1+\delta_a^\infty(\boldsymbol{\mathcal{M}})+\delta_a^\infty(\boldsymbol{\mathcal{N}}), \\ \text{s.t.} \quad & \mathbf{g}=\mathbf{y}-\mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}), \boldsymbol{\mathcal{L}}=\boldsymbol{\mathcal{K}}=\boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{S}}=\boldsymbol{\mathcal{T}}=\boldsymbol{\mathcal{N}} \end{aligned} \tag{9}$$

where $\delta_a^\infty(\cdot)$ is the indicator function of tensor l_∞-norm ball defined as follows

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

We then give the augmented Lagrangian of Equation (9) with Lagrangian multipliers $\mathbf{z}$ and $\{\mathbf{Z}_i\}_{i=1}^4$ and penalty parameter $\rho > 0$:

$$\begin{aligned}
&L_\rho(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}},\mathbf{g},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}},\mathbf{z},\boldsymbol{\mathcal{Z}}_1,\boldsymbol{\mathcal{Z}}_2,\boldsymbol{\mathcal{Z}}_3,\boldsymbol{\mathcal{Z}}_4)\\
&= \frac{1}{2N}\|\mathbf{g}\|_2^2 + \lambda_\iota\|\boldsymbol{\mathcal{K}}\|_\star + \lambda_s\|\boldsymbol{\mathcal{T}}\|_1 + \delta_a^\infty(\boldsymbol{\mathcal{M}}) + \delta_a^\infty(\boldsymbol{\mathcal{N}})\\
&\quad + \langle \mathbf{z}, \mathbf{g} + \mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) - \mathbf{y}\rangle + \frac{\rho}{2}\|\mathbf{g} + \mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) - \mathbf{y}\|_2^2\\
&\quad + \langle \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}}\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}}\|_\mathrm{F}^2 + \langle \boldsymbol{\mathcal{Z}}_2, \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{M}}\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{M}}\|_\mathrm{F}^2\\
&\quad + \langle \boldsymbol{\mathcal{Z}}_3, \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{T}}\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{T}}\|_\mathrm{F}^2 + \langle \boldsymbol{\mathcal{Z}}_4, \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{N}}\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{N}}\|_\mathrm{F}^2
\end{aligned} \tag{10}$$

Following the framework of standard two-block ADMM [41], we separate the primal variables into two blocks $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})$ and $(\mathbf{g},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}})$, and update them alternatively as follows:

Update the first block $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})$: After the t-th iteration, we first update $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})$ by keeping the other variables fixed as follows:

$$\begin{aligned}
&(\boldsymbol{\mathcal{L}}^{t+1},\boldsymbol{\mathcal{S}}^{t+1})\\
&= \operatorname*{argmin}_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}} L_\rho(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}},\mathbf{g}^t,\boldsymbol{\mathcal{K}}^t,\boldsymbol{\mathcal{T}}^t,\boldsymbol{\mathcal{M}}^t,\boldsymbol{\mathcal{N}}^t,\mathbf{z}^t,\boldsymbol{\mathcal{Z}}_1^t,\boldsymbol{\mathcal{Z}}_2^t,\boldsymbol{\mathcal{Z}}_3^t,\boldsymbol{\mathcal{Z}}_4^t)\\
&= \operatorname*{argmin}_{\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}} \ \langle \mathbf{z}^t, \mathbf{g}^t + \mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) - \mathbf{y}\rangle + \frac{\rho}{2}\|\mathbf{g}^t + \mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) - \mathbf{y}\|_2^2\\
&\quad + \langle \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}}^t\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}}^t\|_\mathrm{F}^2 + \langle \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{M}}^t\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{M}}^t\|_\mathrm{F}^2\\
&\quad + \langle \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{T}}^t\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{T}}^t\|_\mathrm{F}^2 + \langle \boldsymbol{\mathcal{Z}}_4^t, \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{N}}^t\rangle + \frac{\rho}{2}\|\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{N}}^t\|_\mathrm{F}^2
\end{aligned} \tag{11}$$

By taking derivatives, respectively, to $\boldsymbol{\mathcal{L}}$ and $\boldsymbol{\mathcal{S}}$ and setting them to zero, we obtain the following system of equations:

$$\begin{aligned}
\mathfrak{X}^*\mathbf{z}^t + \rho\mathfrak{X}^*\big(\mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) + \mathbf{g}^t - \mathbf{y}\big) + \boldsymbol{\mathcal{Z}}_1^t + \rho(\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{K}}^t) + \boldsymbol{\mathcal{Z}}_2^t + \rho(\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{M}}^t) = \mathbf{0}\\
\mathfrak{X}^*\mathbf{z}^t + \rho\mathfrak{X}^*\big(\mathfrak{X}(\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}) + \mathbf{g}^t - \mathbf{y}\big) + \boldsymbol{\mathcal{Z}}_3^t + \rho(\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{T}}^t) + \boldsymbol{\mathcal{Z}}_4^t + \rho(\boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{N}}^t) = \mathbf{0}
\end{aligned} \tag{12}$$

Through solving the system of equations in Equation (12), we obtain

$$\begin{aligned}
\boldsymbol{\mathcal{L}}^{t+1} &= \frac{1}{4\rho}\Big(\mathfrak{X}^*\mathfrak{X}(\mathfrak{X}^*\mathfrak{X} + \mathbb{I})^{-1}(2\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}_\iota + \boldsymbol{\mathcal{B}}_s) - 2(\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}_\iota)\Big)\\
\boldsymbol{\mathcal{S}}^{t+1} &= \frac{1}{4\rho}\Big(\mathfrak{X}^*\mathfrak{X}(\mathfrak{X}^*\mathfrak{X} + \mathbb{I})^{-1}(2\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}_\iota + \boldsymbol{\mathcal{B}}_s) - 2(\boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}_s)\Big)
\end{aligned} \tag{13}$$

where $\mathbb{I}$ denotes the identity operator, and the intermediate tensors are given by $\boldsymbol{\mathcal{A}} = \mathfrak{X}^*(\mathbf{z}^t + \rho\mathbf{g}^t - \mathbf{y})$, $\boldsymbol{\mathcal{B}}_\iota = \boldsymbol{\mathcal{Z}}_1^t + \boldsymbol{\mathcal{Z}}_2^t - \rho(\boldsymbol{\mathcal{K}}^t + \boldsymbol{\mathcal{M}}^t)$, and $\boldsymbol{\mathcal{B}}_s = \boldsymbol{\mathcal{Z}}_3^t + \boldsymbol{\mathcal{Z}}_4^t - \rho(\boldsymbol{\mathcal{T}}^t + \boldsymbol{\mathcal{N}}^t)$.

Update the second block $(\mathbf{g},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}})$: According to the special form of the Lagrangian in Equation (10), the variables $\mathbf{g},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}}$ in the second block can be updated separately as follows.

We first update $\mathbf{g}$ with fixed $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})$:

$$\begin{aligned}
\mathbf{g}^{t+1} &= \operatorname*{argmin}_{\mathbf{g}} L_\rho(\boldsymbol{\mathcal{L}}^{t+1},\boldsymbol{\mathcal{S}}^{t+1},\mathbf{g},\boldsymbol{\mathcal{K}},\boldsymbol{\mathcal{T}},\boldsymbol{\mathcal{M}},\boldsymbol{\mathcal{N}},\mathbf{z}^t,\boldsymbol{\mathcal{Z}}_1^t,\boldsymbol{\mathcal{Z}}_2^t,\boldsymbol{\mathcal{Z}}_3^t,\boldsymbol{\mathcal{Z}}_4^t)\\
&= \operatorname*{argmin}_{\mathbf{g}} \ \frac{1}{2N}\|\mathbf{g}\|_2^2 + \frac{\rho}{2}\|\mathbf{g} + \mathfrak{X}(\boldsymbol{\mathcal{L}}^{t+1} + \boldsymbol{\mathcal{S}}^{t+1}) - \mathbf{y} + \rho^{-1}\mathbf{z}^t\|_2^2\\
&= \frac{N\rho}{1 + N\rho}\Big(\mathbf{y} - \mathfrak{X}(\boldsymbol{\mathcal{L}}^{t+1} + \boldsymbol{\mathcal{S}}^{t+1}) - \rho^{-1}\mathbf{z}^t\Big)
\end{aligned} \tag{14}$$

We then update $\boldsymbol{\mathcal{K}}$ with fixed $(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}})$:

$$\begin{aligned}\boldsymbol{\mathcal{K}}^{t+1} &= \operatorname*{argmin}_{\boldsymbol{\mathcal{K}}} L_\rho(\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1}, \mathbf{g}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}^t, \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{Z}}_4^t) \\ &= \operatorname*{argmin}_{\boldsymbol{\mathcal{K}}} \lambda_\iota \|\boldsymbol{\mathcal{K}}\|_\star + \left\langle \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{K}} \right\rangle + \frac{\rho}{2} \|\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{K}}\|_{\mathrm{F}}^2 \\ &= \mathtt{Prox}^{\|\cdot\|_\star}_{\lambda_\iota \rho^{-1}}(\boldsymbol{\mathcal{L}}^{t+1} + \rho^{-1} \boldsymbol{\mathcal{Z}}_1^t),\end{aligned} \tag{15}$$

where $\mathtt{Prox}^{\|\cdot\|_\star}_{\rho^{-1}}(\cdot)$ is the proximality operator of $*_L$–TNN given in the following lemma.

Lemma 1 (A modified version of Theorem 3.2 in [26]). *Let $\boldsymbol{\mathcal{L}}_0 \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ be any tensor with $*_L$–SVD $\boldsymbol{\mathcal{L}}_0 = \boldsymbol{\mathcal{U}} *_L \boldsymbol{\mathcal{D}} *_L \boldsymbol{\mathcal{V}}^\top$. Then the proximality operator of $*_L$–TNN at $\boldsymbol{\mathcal{L}}_0$ with constant $\tau > 0$, defined as $\mathtt{Prox}^{\|\cdot\|_\star}_{\tau}(\boldsymbol{\mathcal{L}}_0) := \operatorname{argmin}_{\boldsymbol{\mathcal{L}}} \tau \|\boldsymbol{\mathcal{L}}\|_\star + \frac{1}{2}\|\boldsymbol{\mathcal{L}} - \boldsymbol{\mathcal{L}}_0\|_F$, can be computed by*

$$\mathtt{Prox}^{\|\cdot\|_\star}_{\tau}(\boldsymbol{\mathcal{L}}_0) = \boldsymbol{\mathcal{U}} *_L \boldsymbol{\mathcal{D}}_\tau *_L \boldsymbol{\mathcal{V}}^\top \tag{16}$$

where

$$\boldsymbol{\mathcal{D}}_\tau = L^{-1}(L(\boldsymbol{\mathcal{D}}) - \tau)_+). \tag{17}$$

where t_+ denotes the positive part of t, i.e., $t_+ = \max(t, 0)$.

We update $\boldsymbol{\mathcal{T}}$ with fixed $(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}})$:

$$\begin{aligned}\boldsymbol{\mathcal{T}}^{t+1} &= \operatorname*{argmin}_{\boldsymbol{\mathcal{T}}} L_\rho(\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1}, \mathbf{g}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}^t, \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{Z}}_4^t) \\ &= \operatorname*{argmin}_{\boldsymbol{\mathcal{T}}} \lambda_s \|\boldsymbol{\mathcal{T}}\|_1 + \left\langle \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{T}} \right\rangle + \frac{\rho}{2} \|\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{T}}\|_{\mathrm{F}}^2 \\ &= \mathtt{Prox}^{\|\cdot\|_1}_{\lambda_s \rho^{-1}}(\boldsymbol{\mathcal{S}}^{t+1} + \rho^{-1} \boldsymbol{\mathcal{Z}}_3^t),\end{aligned} \tag{18}$$

where $\mathtt{Prox}^{\|\cdot\|_1}_{\tau}(\boldsymbol{\mathcal{T}})$ is the proximality operator [19] of the tensor l_1-norm at point $\boldsymbol{\mathcal{T}}$ given as $\mathtt{Prox}^{\|\cdot\|_1}_{\tau}(\boldsymbol{\mathcal{T}}) = \mathtt{sign}(\boldsymbol{\mathcal{T}}) \odot (|\boldsymbol{\mathcal{T}}| - \tau)_+$, where $\odot$ denotes the element-wise product.

We then update $\boldsymbol{\mathcal{M}}$ with fixed $(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}})$:

$$\begin{aligned}\boldsymbol{\mathcal{M}}^{t+1} &= \operatorname*{argmin}_{\boldsymbol{\mathcal{M}}} L_\rho(\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1}, \mathbf{g}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}^t, \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{Z}}_4^t) \\ &= \operatorname*{argmin}_{\boldsymbol{\mathcal{M}}} \delta_a^\infty(\boldsymbol{\mathcal{M}}) + \left\langle \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{M}} \right\rangle + \frac{\rho}{2} \|\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{M}}\|_{\mathrm{F}}^2 \\ &= \mathtt{Proj}^{\|\cdot\|_\infty}_{a}(\boldsymbol{\mathcal{L}}^{t+1} + \rho^{-1} \boldsymbol{\mathcal{Z}}_2^t),\end{aligned} \tag{19}$$

where $\mathtt{Proj}^{\|\cdot\|_\infty}_{a}(\cdot)$ is the projector onto the tensor l_∞-norm ball of radius a, which is given by $\mathtt{Proj}^{\|\cdot\|_\infty}_{a}(\boldsymbol{\mathcal{M}}) = \mathtt{sign}(\boldsymbol{\mathcal{M}}) \odot \min(|\boldsymbol{\mathcal{M}}|, a)$ [19].

Similarly, we update $\boldsymbol{\mathcal{N}}$ as follows:

$$\begin{aligned}\boldsymbol{\mathcal{N}}^{t+1} &= \operatorname*{argmin}_{\boldsymbol{\mathcal{N}}} L_\rho(\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1}, \mathbf{g}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}^t, \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{Z}}_4^t) \\ &= \operatorname*{argmin}_{\boldsymbol{\mathcal{N}}} \delta_a^\infty(\boldsymbol{\mathcal{N}}) + \left\langle \boldsymbol{\mathcal{Z}}_4^t, \boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{N}} \right\rangle + \frac{\rho}{2} \|\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{N}}\|_{\mathrm{F}}^2 \\ &= \mathtt{Proj}^{\|\cdot\|_\infty}_{a}(\boldsymbol{\mathcal{S}}^{t+1} + \rho^{-1} \boldsymbol{\mathcal{Z}}_4^t).\end{aligned} \tag{20}$$

Update the dual variables $\mathbf{z}$ and $\{\boldsymbol{\mathcal{Z}}_i\}$: According to the update strategy of dual variables in ADMM [41], the variables $\mathbf{z}$ and $\{\boldsymbol{\mathcal{Z}}_i\}$ can be updated using dual ascent as follows:

$$\begin{aligned}
\mathbf{z}^{t+1} &= \mathbf{z}^{t} + \rho(\mathbf{g}^{t+1} + \mathfrak{X}(\boldsymbol{\mathcal{L}}^{t+1} + \boldsymbol{\mathcal{S}}^{t+1}) - \mathbf{y}) \\
\boldsymbol{\mathcal{Z}}_1^{t+1} &= \boldsymbol{\mathcal{Z}}_1^{t+1} + \rho(\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{K}}^{t+1}) \\
\boldsymbol{\mathcal{Z}}_2^{t+1} &= \boldsymbol{\mathcal{Z}}_2^{t+1} + \rho(\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{M}}^{t+1}) \\
\boldsymbol{\mathcal{Z}}_3^{t+1} &= \boldsymbol{\mathcal{Z}}_3^{t+1} + \rho(\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{T}}^{t+1}) \\
\boldsymbol{\mathcal{Z}}_4^{t+1} &= \boldsymbol{\mathcal{Z}}_4^{t+1} + \rho(\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{N}}^{t+1})
\end{aligned} \tag{21}$$

The algorithm for solving Problem (8) is summarized in Algorithm 1.

Algorithm 1 Solving Problem (8) using ADMM.

```
Input: The design tensors $\{\boldsymbol{\mathcal{X}}_i\}$ and observations $\{y_i\}$, the regularization parameters
    $\lambda_l, \lambda_s$, the $l_1$-norm bound $a$, the penalty parameter $\rho$ of the Lagrangian, the convergence
    tolerance $\delta$, the maximum iteration number $T_{\max}$.
Initialize $t = 0$, $\mathbf{g}^0 = \mathbf{z}^0 = \mathbf{0} \in \mathbb{R}^N$, $\boldsymbol{\mathcal{L}}^0 = \boldsymbol{\mathcal{S}}^0 = \boldsymbol{\mathcal{K}}^0 = \boldsymbol{\mathcal{T}}^0 = \boldsymbol{\mathcal{M}}^0 = \boldsymbol{\mathcal{N}}^0 = \boldsymbol{\mathcal{Z}}_1^0 = \boldsymbol{\mathcal{Z}}_2^0 =$
    $\boldsymbol{\mathcal{Z}}_3^0 = \boldsymbol{\mathcal{Z}}_4^0 = \mathbf{0} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$
for $t = 0, \cdots, T_{\max}$ do
    Update $(\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1})$ by Equation (13);
    Update $(\mathbf{g}^{t+1}, \boldsymbol{\mathcal{K}}^{t+1}, \boldsymbol{\mathcal{T}}^{t+1}, \boldsymbol{\mathcal{M}}^{t+1}, \boldsymbol{\mathcal{N}}^{t+1})$ by Equations (14)–(20), respectively;
    Update $(\mathbf{z}^{t+1}, \boldsymbol{\mathcal{Z}}_1^{t+1}, \boldsymbol{\mathcal{Z}}_2^{t+1}, \boldsymbol{\mathcal{Z}}_3^{t+1}, \boldsymbol{\mathcal{Z}}_4^{t+1})$ by Equation (21);
    Check the convergence criteria:
    (i) convergence of primal variables:
        $\|\boldsymbol{\mathcal{A}}^{t+1} - \boldsymbol{\mathcal{A}}^{t}\|_\infty \le \delta, \ \forall \boldsymbol{\mathcal{A}} \in \{\mathbf{g}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}\}$
    (ii) convergence of constraints:
        $\max\{\|\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{K}}^{t+1}\|_\infty, \|\boldsymbol{\mathcal{L}}^{t+1} - \boldsymbol{\mathcal{M}}^{t+1}\|_\infty\} \le \delta$
        $\max\{\|\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{T}}^{t+1}\|_\infty, \|\boldsymbol{\mathcal{S}}^{t+1} - \boldsymbol{\mathcal{N}}^{t+1}\|_\infty\} \le \delta$
        $\|\mathbf{g}^{t+1} + \mathfrak{X}(\boldsymbol{\mathcal{L}}^{t+1} + \boldsymbol{\mathcal{S}}^{t+1}) - \mathbf{y}\|_\infty \le \delta$
end for
Output: $(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}}) = (\boldsymbol{\mathcal{L}}^{t+1}, \boldsymbol{\mathcal{S}}^{t+1})$.
```

Complexity Analysis: The time complexity of Algorithm 1 is analyzed as follows. Due to the special structures of design tensors $\{\boldsymbol{\mathcal{X}}_i\}$, the operators $\mathfrak{X}$ and $(\mathfrak{X}^*\mathfrak{X}\mathfrak{X}^*\mathfrak{X} + \mathbb{I})^{-1}$ can be implemented with time cost $O(N)$ and $O(d_1d_2d_3 + N)$, respectively. The cost of updating $\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}$ and $\{\boldsymbol{\mathcal{Z}}_i\}$ is $O(d_1d_2d_3)$. The main time cost in Algorithm 1 lies in the update of $\boldsymbol{\mathcal{K}}$ which needs the $*_L$–SVD on $d_1 \times d_2 \times d_3$ tensors, involving the $*_L$-transform (costing $O(d_1d_2d_3^2)$ in general), and d_3 matrix SVDs on $d_1 \times d_2$ matrices (costing $O(d_1d_2d_3(d_1 \wedge d_2))$). Thus, the one-iteration cost of Algorithm 1 is

$$O(d_1d_2d_3((d_1 \wedge d_2) + d_3)) \tag{22}$$

in general, and can be reduced to $O(d_1d_2d_3((d_1 \wedge d_2) + \log d_3))$ for some linear transforms L which have fast implementations (like DFT and DCT).

Convergence Analysis: According to [28], the convergence rate of general ADMM-based algorithms is $O(1/t)$, where t is the iteration number. The convergence analysis of Algorithm 1 is established in Theorem 2.

Theorem 2 (Convergence of Algorithm 1). *For any positive constant ρ, if the unaugmented Lagrangian function $L_0(\mathbf{g}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}, \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\mathcal{Z}}_2, \boldsymbol{\mathcal{Z}}_3, \boldsymbol{\mathcal{Z}}_4)$ has a saddle point, then the iterations $(\mathbf{g}^t, \boldsymbol{\mathcal{L}}^t, \boldsymbol{\mathcal{S}}^t, \boldsymbol{\mathcal{K}}^t, \boldsymbol{\mathcal{T}}^t, \boldsymbol{\mathcal{M}}^t, \boldsymbol{\mathcal{N}}^t, \mathbf{z}^t, \boldsymbol{\mathcal{Z}}_1^t, \boldsymbol{\mathcal{Z}}_2^t, \boldsymbol{\mathcal{Z}}_3^t, \boldsymbol{\mathcal{Z}}_4^t)$ in Algorithm 1 satisfy the residual convergence, objective convergence and dual variable convergence (defined in [41]) of Problem (9) as $t \to \infty$.*

Proof. The key idea is to rewrite Problem (9) into a standard two-block ADMM problem. For notational simplicity, let

$$f(\mathbf{u}) = 0, \qquad g(\mathbf{v}) = \frac{1}{2N}\|\mathbf{g}\|_2^2 + \lambda_\iota \|\boldsymbol{\mathcal{K}}\|_\star + \lambda_s \|\boldsymbol{\mathcal{S}}\|_1 + \delta_a^\infty(\boldsymbol{\mathcal{M}}) + \delta_a^\infty(\boldsymbol{\mathcal{N}}),$$

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

$$\mathbf{u} = \begin{bmatrix} \mathtt{vec}(\boldsymbol{\mathcal{L}}) \\ \mathtt{vec}(\boldsymbol{\mathcal{S}}) \end{bmatrix} \in \mathbb{R}^{2d_1d_2d_3}, \qquad \mathbf{v} = \begin{bmatrix} \mathbf{g} \\ \mathtt{vec}(\boldsymbol{\mathcal{K}}) \\ \mathtt{vec}(\boldsymbol{\mathcal{T}}) \\ \mathtt{vec}(\boldsymbol{\mathcal{M}}) \\ \mathtt{vec}(\boldsymbol{\mathcal{N}}) \end{bmatrix} \in \mathbb{R}^{N+4d_1d_2d_3},$$

$$\mathbf{w} = \begin{bmatrix} \mathbf{z} \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_1) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_2) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_3) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_4) \end{bmatrix} \in \mathbb{R}^{N+4d_1d_2d_3}, \qquad \mathbf{c} = \begin{bmatrix} -\mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{N+4d_1d_2d_3},$$

and

$$\mathbf{A} = \begin{bmatrix} -\mathbf{X} & -\mathbf{X} \\ \mathbf{I}_{d_1d_2d_3} & \mathbf{0} \\ \mathbf{I}_{d_1d_2d_3} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_1d_2d_3} \\ \mathbf{0} & \mathbf{I}_{d_1d_2d_3} \end{bmatrix} \in \mathbb{R}^{(N+4d_1d_2d_3)\times(2d_1d_2d_3)}, \text{ with } \mathbf{X} = \begin{bmatrix} \mathtt{vec}(\boldsymbol{\mathcal{X}}_1)^\top \\ \mathtt{vec}(\boldsymbol{\mathcal{X}}_2)^\top \\ \vdots \\ \mathtt{vec}(\boldsymbol{\mathcal{X}}_N)^\top \end{bmatrix} \in \mathbb{R}^{N\times(2d_1d_2d_3)}, \tag{23}$$

where $\mathtt{vec}(\cdot)$ denotes the operation of tensor vectorization (see [30]).

It can be verified that $f(\cdot)$ and $g(\cdot)$ are closed, proper convex functions. Then, Problem (9) can be re-written as follows:

$$\begin{aligned} \min_{\mathbf{u},\mathbf{v}} \quad & f(\mathbf{u}) + g(\mathbf{v}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{u} - \mathbf{v} = \mathbf{c}. \end{aligned}$$

According to the convergence analysis in [41], we have:

$$\begin{aligned} &\text{objective convergence:} && \lim_{t\to\infty} f(\mathbf{u}^t) + g(\mathbf{v}^t) = f^\star + g^\star, \\ &\text{dual variable convergence:} && \lim_{t\to\infty} \mathbf{w}^t = \mathbf{w}^\star, \\ &\text{constraint convergence:} && \lim_{t\to\infty} \mathbf{A}\mathbf{u}^t - \mathbf{v}^t = \mathbf{c}, \end{aligned}$$

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

$$\mathbf{w}^\star = \mathbf{w} = \begin{bmatrix} \mathbf{z}^\star \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_1^\star) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_2^\star) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_3^\star) \\ \mathtt{vec}(\boldsymbol{\mathcal{Z}}_4^\star) \end{bmatrix}$$

where $(\mathbf{z}^\star, \boldsymbol{\mathcal{Z}}_1^\star, \boldsymbol{\mathcal{Z}}_2^\star, \boldsymbol{\mathcal{Z}}_3^\star, \boldsymbol{\mathcal{Z}}_4^\star)$ is the component of dual variables in a saddle point $(\mathbf{g}^\star, \boldsymbol{\mathcal{L}}^\star, \boldsymbol{\mathcal{S}}^\star, \boldsymbol{\mathcal{K}}^\star, \boldsymbol{\mathcal{T}}^\star, \boldsymbol{\mathcal{M}}^\star, \boldsymbol{\mathcal{N}}^\star, \mathbf{z}^\star, \boldsymbol{\mathcal{Z}}_1^\star, \boldsymbol{\mathcal{Z}}_2^\star, \boldsymbol{\mathcal{Z}}_3^\star, \boldsymbol{\mathcal{Z}}_4^\star)$ of the unaugmented Lagrangian $L_0(\mathbf{g}, \boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{K}}, \boldsymbol{\mathcal{T}}, \boldsymbol{\mathcal{M}}, \boldsymbol{\mathcal{N}}, \mathbf{z}, \boldsymbol{\mathcal{Z}}_1, \boldsymbol{\mathcal{Z}}_2, \boldsymbol{\mathcal{Z}}_3, \boldsymbol{\mathcal{Z}}_4)$. □

5. Statistical Performance

In this section, we answer **Q2** by studying the statistical performances of the proposed estimator $(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}})$. Specifically, the goal is to upper bound the squared F-norm error $\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_{\mathrm{F}}^2 + \|\hat{\boldsymbol{\mathcal{S}}} - \boldsymbol{\mathcal{S}}^*\|_{\mathrm{F}}^2$. We will first give an upper bound on the estimation error in a non-asymptotic manner, and then prove that the upper bound is minimax optimal up to a logarithm factor.

5.1. Upper Bound on the Estimation Error

We establish an upper bounds on the estimation error in the following theorem. For notational simplicity, let $N_s = |\mathbf{X}_s|$ and $N_\iota = |\mathbf{X}_s^\perp|$ denote the number of corrupted and uncorrupted observations of $\boldsymbol{\mathcal{L}}^*$ in the observation model (6), respectively.

Theorem 3 (Upper bounds on the estimation error)**.** *If the number of uncorrupted observations in the observation model (6) satisfy*

$$N_\iota \ge c_1 d_1 d_3 \log(d_1 d_3 + d_2 d_3) \log^2(d_1 + d_2) \tag{24}$$

and regularization parameters in Problem (8) are set by

$$\lambda_\iota = c_2(\sigma \vee a)\sqrt{\frac{\log(d_1 d_3 + d_2 d_3)}{d_1 \wedge d_2}}, \quad \text{and} \qquad \lambda_s = c_3(\sigma \vee a)\frac{log(d_1 d_3 + d_2 d_3)}{N}, \tag{25}$$

then it holds with probability at least $1 - c_5(d_1 d_3 + d_2 d_3)^{-1}$ *that:*

$$\begin{aligned} &\frac{\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_F^2 + \|\hat{\boldsymbol{\mathcal{S}}} - \boldsymbol{\mathcal{S}}^*\|_F^2}{d_1 d_2 d_3} \\ &\le C\left(r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*) \cdot \frac{(\sigma^2 \vee a^2)(d_1 \vee d_2) d_3 \log \tilde{d}}{N_\iota} + \frac{N_s \log(d_1 d_3 + d_2 d_3)}{N_\iota} + \|\boldsymbol{\mathcal{S}}^*\|_0 \cdot \frac{a^2}{d_1 d_2 d_3}\right). \end{aligned} \tag{26}$$

Theorem 3 implies that, if the noise level σ and spikiness level a are fixed, and all the corrupted positions are observed exactly only once (i.e., the number of corrupted observations $N_s = \|\boldsymbol{\mathcal{S}}^*\|_0$), then the estimation error in Equation (26) would be bounded by

$$O\Big(r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*) \cdot \frac{(d_1 \vee d_2) d_3 \log(d_1 d_3 + d_2 d_3)}{N_\iota} + \|\boldsymbol{\mathcal{S}}^*\|_0 \cdot \big(\frac{\log(d_1 d_3 + d_2 d_3)}{N_\iota} + \frac{1}{d_1 d_2 d_3}\big)\Big). \tag{27}$$

Note that, the bound in Equation (27) is intuition-consistent: if the underlying tensor $\boldsymbol{\mathcal{L}}^*$ gets more complex (i.e., with higher tubal rank), then the estimation error will be larger; if the corruption tensor $\boldsymbol{\mathcal{S}}^*$ gets denser, then the estimation error will also become larger; if the number of uncorrupted observations N_ι gets larger, then the estimation error will decrease. The scaling behavior of the estimation error in Equation (27) will be verified through experiments on synthetic data in Section 7.1.

Remark 1 (Consistence with prior models for robust low-tubal-rank tensor completion)**.** *According to Equation (27), our $*_L$–SVD-based estimator in Equation (8) allows the tubal rank $r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*)$ to take the order $O(d_2/\log(d_1 d_3 + d_2 d_3))$, and the corruption ratio $\|\boldsymbol{\mathcal{S}}^*\|_0/(d_1 d_2 d_3)$ to be $O(1)$ for approximate estimation with small error. It is slightly better with a logarithm factor than the results for t-SVD-based tensor robust completion model in [8] which allows $r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*) = O(d_2/\log^2(d_1 d_3 + d_2 d_3))$ and $\|\boldsymbol{\mathcal{S}}^*\|_0/(d_1 d_2 d_3) = O(1)$.*

Remark 2 (Consistence with prior models for noisy low-tubal-rank tensor completion)**.** *If $\|\boldsymbol{\mathcal{S}}^*\|_0 = 0$, i.e., the corruption $\boldsymbol{\mathcal{S}}^*$ vanishes, then we obtain*

$$\frac{\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_F^2}{d_1 d_2 d_3} = O\big(\frac{r_{\rm tb}(\boldsymbol{\mathcal{L}}^*)(d_1 \vee d_2) d_3 \log(d_1 d_3 + d_2 d_3)}{N}\big)$$

which is consistent with the error bound for t-SVD-based noisy tensor completion [42–44], and $*_L$–SVD-based tensor Dantzig Selector in [18].

Remark 3 (Consistence with prior models for robust low-tubal-rank tensor decomposition)**.** *In the setting of Robust Tensor Decomposition (RTD) [34], the fully observed model instead of our estimation model in Equation (6) is considered. For the RTD problem, our error bound in Equation (27) is consistent with the t-SVD-based bound for RTD [34] (up to a logarithm factor).*

Remark 4 (No exact recovery guarantee)**.** *According to Theorem 3, when $\sigma = 0$ and $\|\boldsymbol{\mathcal{S}}^*\|_0 = 0$, i.e., in the noiseless case, the estimation error is upper bounded by $O(a(d_1 \vee d_2) d_3 r_{\rm tb}(\boldsymbol{\mathcal{L}}^*) \log \tilde{d}/N)$ which is not zero. Thus, no exact recovery is guaranteed by Theorem 3. It can be seen as a trade-off that we do not assume the low-tubal-rank tensor $\boldsymbol{\mathcal{L}}^*$ to satisfy the tensor incoherent conditions [8,35,37] which essentially ensures the separability between $\boldsymbol{\mathcal{L}}^*$ and $\boldsymbol{\mathcal{S}}^*$.*

5.2. A Minimax Lower Bound for the Estimation Error

In Theorem 3, we established the estimation error for Model (8). Then one may ask the complementary questions: how tight is the upper bound? Are there fundamental (model-independent) limits of estimation error in robust tensor completion? In this section, we will answer the questions.

To analyze the optimality of the proposed upper bound in Theorem 3, the minimax lower bounds of the estimation error is established for the tensor pair $(\boldsymbol{\mathcal{L}}^*, \boldsymbol{\mathcal{S}}^*)$ belonging to the class $\mathbf{A}(r,s,a)$ of tensor pairs defined as:

$$\mathbf{A}(r,s,a) := \big\{(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}) \;\big|\; r_{\rm tb}(\boldsymbol{\mathcal{L}}) \le r,\ \|\boldsymbol{\mathcal{S}}\|_0 \le s,\ \|\boldsymbol{\mathcal{L}}\|_\infty < a, \|\boldsymbol{\mathcal{S}}\|_\infty \le a\big\} \tag{28}$$

We then define the associated element-wise minimax error as follows

$$\mathscr{M}(\mathbf{A}(r,s,a)) := \inf_{(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}})} \sup_{(\boldsymbol{\mathcal{L}}^*, \boldsymbol{\mathcal{S}}^*) \in \mathbf{A}(r,s,a)} \mathbb{E}\left[\frac{\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_{\rm F}^2 + \|\hat{\boldsymbol{\mathcal{S}}} - \boldsymbol{\mathcal{S}}^*\|_{\rm F}^2}{d_1 d_2 d_3}\right], \tag{29}$$

where the infimum ranges over all pairs of estimators $(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}})$, the supremum ranges over all pairs of underlying tensors $(\boldsymbol{\mathcal{L}}^*, \boldsymbol{\mathcal{S}}^*)$ in the given tensor class $\mathbf{A}(r,s,a)$, and the expectation is taken over the design tensors $\{\boldsymbol{\mathcal{X}}_i\}$ and i.i.d. Gaussian noises $\{\xi_i\}$ in the observation model (6). We come up with the following theorem.

Theorem 4 (Minimax lower bound)**.** *Suppose the dimensionality $d_1, d_2 \ge 2$, the rank and sparsity parameters $r \in [d_1 \vee d_2]$, $s \le d_1 d_2 d_3 / 2$, the number of uncorrupted entries $N_\iota \ge r d_1 d_3$, and the number of corrupted entries $N_s \le \tau r \tilde{d}$ with a constant $\tau > 0$. Then, under Assumption 1, there exist absolute constants $b \in (0,1)$ and $c > 0$, such that*

$$\mathscr{M}(\mathbf{A}(r,s,a)) \ge b\phi(N,r,s) \tag{30}$$

where

$$\phi(N,r,s) := (\sigma \wedge a)^2 \left(\frac{r(d_1 + d_2) d_3 + N_s}{N_\iota} + \frac{s}{d_1 d_2 d_3}\right). \tag{31}$$

The lower bound given in Equation (30) implies that the proposed upper bound in Theorem 3 is optimal (up to a log factor) in the minimax sense for tensors belonging to the set $\mathbf{A}(r,s,a)$. That is to say no estimator can obtain more accurate estimation than our estimator in Equation (8) (up to a log factor) for $(\mathcal{L}^*, \mathcal{S}^*) \in \mathbf{A}(r, s, a)$, thereby showing the optimality of the proposed estimator.

6. Connections and Differences with Previous Works

In this section, we discuss the connections and differences with existing nuclear norm based robust matrix/tensor completion models, where the underlying matrix/tensor suffers from missing values, gross sparse outliers, and small dense noises at the same time.

First, we briefly introduce and analyze the two most related models, i.e., the matrix nuclear norm based model [36] and the sum of mode-wise matrix nuclear norms based model [45] as follows.

(1) The matrix Nuclear Norm (NN) based model [36]: If the underlying tensor is of 2-way, i.e., a matrix, then the observation model in Equation (6) becomes the setting for robust matrix completion, and the proposed estimator in Equation (8) degenerates to the matrix nuclear norm based estimator in [36]. In both model formulation and statistical analysis, this work can be seen as a 3-way generalization of [36].
Moreover, by conducting robust matrix completion on each frontal slice of a 3-way tensor, we can obtain the matrix nuclear norm based robust tensor completion model as follows:

$$\min_{\mathcal{L},\mathcal{S}} \frac{1}{2N} \|\mathbf{y} - \mathfrak{X}(\mathcal{L} + \mathcal{S})\|_2^2 + \lambda_\iota \sum_{k=1}^{d_3} (\|\mathbf{L}^{(k)}\|_* + \lambda_s \|\mathbf{S}^{(k)}\|_1) \tag{32}$$

(2) The Sum of mode-wise matrix Nuclear Norms (SNN) based model [45]: Huang et al. [45] proposed a robust tensor completion model based on the sum of mode-wise nuclear norms deduced by the Tucker decomposition as follows

$$\min_{\mathcal{L},\mathcal{S}} \frac{1}{2N} \|\mathbf{y} - \mathfrak{X}(\mathcal{L} + \mathcal{S})\|_2^2 + \sum_{k=1}^{3} \alpha_k \|\mathbf{L}_{(k)}\|_* + \lambda_s \|\mathcal{S}\|_1, \tag{33}$$

where $\mathbf{L}_{(k)} \in \mathbb{R}^{d_i \times \prod_{j \neq k} d_j}$ is the mode-k matriculation of tensor $\mathcal{L} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, for all $i = 1, 2, 3$.
The main differences between SNN and this work are two-fold: (i) SNN is based on the Tucker decomposition [15], whereas this work is based on the recently proposed tensor $*_L$-SVD [24]; (ii) the theoretical analysis for SNN cannot guarantee the minimax optimality of the model in [45], whereas this works rigorously proof of the minimax optimality of the proposed estimator is established in Section 5.

Then, we discuss the following related works which can be seen as special cases of this work.

(1) The robust tensor completion model based on t-SVD [46]: In a short conference presentation [46] (whose first author is the same as this paper), the t-SVD-based robust tensor completion model is studied. As t-SVD can be viewed as a special case of the $*_L$-SVD (when DFT is used as the transform L), the model in [46] can be a special case of ours.

(2) The robust tensor recovery models with missing values and sparse outliers [8,27]: In [8,27], the authors considered the robust reconstruction of incomplete tensor polluted by sparse outliers, and proposed t-SVD (or $*_L$-SVD) based models with theoretical guarantees for exact recovery. As they did not consider small dense noises, their settings are indeed a special case of our observation model (6) when $\mathcal{E} = 0$.

(3) The robust tensor decomposition based on t-SVD [34]: In [34], the authors studied the t-SVD-based robust tensor decomposition, which aims at recovering a tensor corrupted by both gross sparse outliers and small dense noises. Comparing with this work, Ref. [34] can be seen as a special case when there are no missing values.

7. Experiments

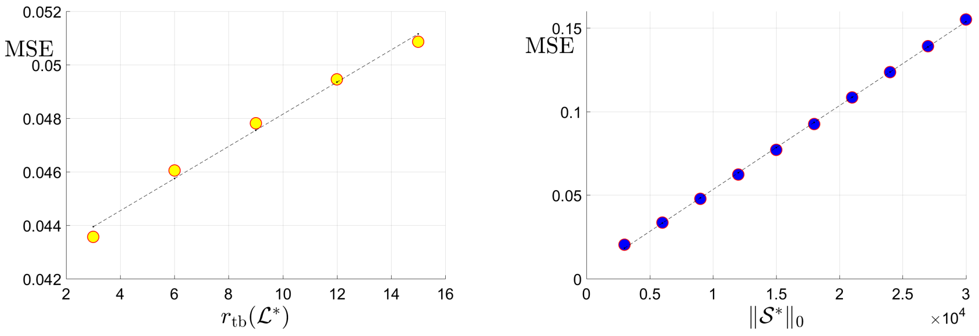
In this section, experiments on synthetic datasets will be first conducted to validate the sharpness of the proposed upper bounds in Theorem 3. Then, both effectiveness and efficiency of the proposed algorithm will be demonstrated through experiments on seven different types of remote sensing datasets. All codes are written in Matlab, and all experiments are performed on a Windows 10 laptop with AMD Ryzen 3.0 GHz CPU and 8 GB RAM.

7.1. Sharpness of the Proposed Upper Bound

Sharpness of the proposed upper bounds in Theorem 3 will be validated. Specifically, we will check whether the upper bounds in Equation (27) can reflect the true scaling behavior of the estimation error. As predicted in Equation (27), if the upper bound is "sharp", then it is expected that the Mean Square Errors (MSE) $(\|\hat{\boldsymbol{\mathcal{L}}}-\boldsymbol{\mathcal{L}}^*\|_{\mathrm{F}}^2+\|\hat{\boldsymbol{\mathcal{S}}}-\boldsymbol{\mathcal{S}}^*\|_{\mathrm{F}}^2)/(d_1d_2d_3)$ will possess a scaling behavior very similar to the upper bound: approximately linear w.r.t the tubal rank of the underlying tensor $\boldsymbol{\mathcal{L}}^*$, the l_0-norm of the corruption tensor $\boldsymbol{\mathcal{S}}^*$, and the reciprocal of uncorrupted observation number N_ι. We will examine whether this expectation will happen in simulation studies on synthetic datasets.

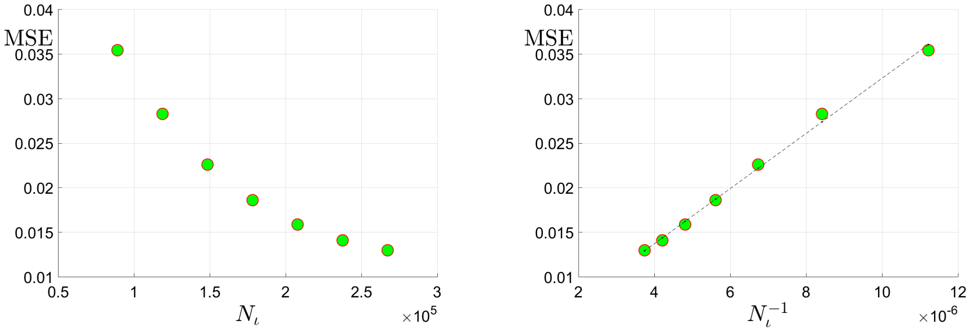
The synthetic datasets are generated as follows. Similar to [26], we consider three cases of linear transform L with orthogonal matrix $\mathbf{L}$: (1) Discrete Fourier Transform (DFT); (2) Discrete Cosine Transform (DCT) [24]; (3) Random Orthogonal Matrix (ROM) [26]. The underlying low-rank tensor $\boldsymbol{\mathcal{L}}^* \in \mathbb{R}^{d_1\times d_2\times d_3}$ with $*_L$-tubal rank r^* is generated by $\boldsymbol{\mathcal{L}}^* = \boldsymbol{\mathcal{P}} *_L \boldsymbol{\mathcal{Q}}$, where $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{d_1\times r^*\times d_3}$ and $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{r^*\times d_2\times d_3}$ are i.i.d. sampled from $\mathcal{N}(0,1)$. $\boldsymbol{\mathcal{L}}^*$ is then normalized such that $\|\boldsymbol{\mathcal{L}}^*\|_\infty = 1$. Second, to generate the sparse corruption tensor $\boldsymbol{\mathcal{S}}^*$, we first form $\boldsymbol{\mathcal{S}}_0$ with i.i.d. uniform distribution $\mathrm{Uni}(0,1)$ and then uniformly select $\gamma d_1d_2d_3$ entries. Thus the number of corrupted entries $\|\boldsymbol{\mathcal{S}}^*\|_0 = \gamma d_1d_2d_3$. Third, we uniformly select N_ι elements from the uncorrupted positions of $(\boldsymbol{\mathcal{L}}^*+\boldsymbol{\mathcal{S}}^*)$. Finally, the noise $\{\xi_i\}$ are sampled from i.i.d. Gaussian $N(0,\sigma^2)$ with $\sigma = 0.1\|\boldsymbol{\mathcal{L}}^*\|_{\mathrm{F}}/\sqrt{d_1d_2d_3}$. We consider f-diagonal tensors with $d_1 = d_2 = d \in \{80,100,120\}$, $d_3 = 30$ and tubal rank $r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*) \in \{3,6,9,12,15\}$. We choose corruption ratio $\gamma \in \{0.01 : 0.01 : 0.1\}$ and uncorrupted observation ratio $N_\iota/(d_1d_2d_3 - N_s) \in \{0.4 : 0.1 : 0.9\}$. In each setting, the MSE averaged over 30 trials is reported.

In Figure 3, we report the results for 100 × 100 × 30 tensors when the DFT is adopted as the linear transform L in Equation (4). According to sub-plots (a), (b), and (d) in Figure 3, it can be seen that the MSE scales approximately linearly w.r.t. $r_{\mathrm{tb}}(\boldsymbol{\mathcal{L}}^*)$, $\|\boldsymbol{\mathcal{S}}^*\|_0$, and N_ι^{-1}. There results accord well with our expectation for the size 100 × 100 × 30 and linear transform $L =$ DFT. As very similar phenomena are also observed in all the other settings where $d \in \{80,120\}$ and $L \in$ {DCT, ROM}, we simply omit them.Thus, it can be verified that the scaling behavior of the estimation error can be approximately predicted by the proposed upper bound in Equation (27).



(**a**) (**b**)

Figure 3. *Cont.*



(**c**) (**d**)

Figure 3. Plots of the MSE versus the tubal rank $r_{\text{tb}}(\boldsymbol{\mathcal{L}}^*)$ of the underlying tensor, the number of corruptions $\|\boldsymbol{\mathcal{S}}^*\|_0$, the number of uncorrupted observations N_ι and its inversion N_ι^{-1} : (**a**) MSE vs. the tubal rank $r_{\text{tb}}(\boldsymbol{\mathcal{L}}^*)$ with fixed corruption level $\|\boldsymbol{\mathcal{S}}^*\|_0 = 0.03d_1d_2d_3$ and number of uncorrupted observations $N_\iota = 0.7d_1d_2d_3 - \|\boldsymbol{\mathcal{S}}^*\|_0$; (**b**) MSE vs. the number of corruptions $\|\boldsymbol{\mathcal{S}}^*\|_0$ with fixed tubal rank 9 and total observation number $0.7d_1d_2d_3$; (**c**) MSE vs. the number of uncorrupted observation N_ι with $r_{\text{tb}}(\boldsymbol{\mathcal{L}}^*) = 3$ and corruption level $\|\boldsymbol{\mathcal{S}}^*\|_0 = 0.01d_1d_2d_3$; (**d**) MSE vs. N_ι^{-1} with $r_{\text{tb}}(\boldsymbol{\mathcal{L}}^*) = 3$ and $\|\boldsymbol{\mathcal{S}}^*\|_0 = 0.01d_1d_2d_3$.

7.2. Effectiveness and Efficient of the Proposed Algorithm

In this section, we evaluate both the effectiveness and efficiency of the proposed Algorithm 1 by conducting robust tensor completion on seven different types of datasets collected from several remote sensing related applications from Sections 7.2.1– 7.2.7.

Following [25], we adopted three different transformations L in Equation (4) to define the $*_L$–TNN: the first two transformations are DFT and DCF (denoted by TNN (DFT) and TNN (DCT), respectively), and the third one named TNN (Data) depends on the given data motived by [27,31]. We first perform SVD on the mode-3 unfolding matrix of $\boldsymbol{\mathcal{L}}^*$ as $\mathbf{L}^*_{(3)} = \mathbf{USV}^\top$, and then use $\mathbf{U}^\top$ as the desired transform matrix in the $*_L$–product (4). The proposed algorithm is compared with the aforementioned models NN [36] in Equation (32) and SNN [45] in Equation (33) in Section 6. Both Model (32) and Model (33) are solved by using ADMM with implementations by ourselves in Matlab language.

We conduct robust tensor completion on the datasets in Figure 4 with a similar settings as [47]. For a $d_1 \times d_2 \times d_3$ tensor data $\boldsymbol{\mathcal{L}}^*$ re-scaled by $\|\boldsymbol{\mathcal{L}}^*\|_\infty = 1$, we choose its support uniformly at random with ratio ρ_{s} and fill in the values with i.i.d. standard Gaussian variables to generate the corruption $\boldsymbol{\mathcal{S}}^*$. Then, we randomly sample the entries of $\boldsymbol{\mathcal{L}}^* + \boldsymbol{\mathcal{S}}^*$ uniformly with observation ratio ρ_{obs}. The noises $\{\xi_i\}$ are further generated with i.i.d. zero-mean Gaussian entries whose standard deviation is given by $\sigma = 0.05\|\boldsymbol{\mathcal{L}}^*\|_{\text{F}}/\sqrt{d_1d_2d_3}$ to generate the observations $\{y_i\}$. The goal in the experiments is to estimate the underlying signal $\boldsymbol{\mathcal{L}}^*$ from $\{y_i\}$. The effectiveness of algorithms are measured by the Peaks Signal Noise Ratio (PSNR) and structural similarity (SSIM) [48]. Specifically, the PSNR of an estimator $\hat{\boldsymbol{\mathcal{L}}}$ is defined as

$$\texttt{PSNR} := 10\log_{10}\left(\frac{d_1d_2d_3\|\boldsymbol{\mathcal{L}}^*\|_\infty^2}{\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_{\text{F}}^2}\right),$$

for the underlying tensor $\boldsymbol{\mathcal{L}}^* \in \mathbb{R}^{d_1 \times d_2 \times d_3}$. The SSIM is computed via

$$\texttt{SSIM} := \frac{(2\mu_{\boldsymbol{\mathcal{L}}^*}\mu_{\hat{\boldsymbol{\mathcal{L}}}} + (0.01\bar{\omega})^2)(2\sigma_{\boldsymbol{\mathcal{L}}^*,\hat{\boldsymbol{\mathcal{L}}}} + (0.03\bar{\omega})^2)}{(\mu^2_{\boldsymbol{\mathcal{L}}^*} + \mu^2_{\hat{\boldsymbol{\mathcal{L}}}} + (0.01\bar{\omega})^2)(\sigma^2_{\boldsymbol{\mathcal{L}}^*} + \sigma^2_{\hat{\boldsymbol{\mathcal{L}}}} + (0.03\bar{\omega})^2)},$$

where $\mu_{\boldsymbol{\mathcal{L}}^*}, \mu_{\hat{\boldsymbol{\mathcal{L}}}}, \sigma_{\boldsymbol{\mathcal{L}}^*}, \sigma_{\hat{\boldsymbol{\mathcal{L}}}}, \sigma_{\boldsymbol{\mathcal{L}}^*,\hat{\boldsymbol{\mathcal{L}}}}$ and $\bar{\omega}$ denotes the local means, standard deviation, cross-covariance, and dynamic range of the magnitude of tensors $\boldsymbol{\mathcal{L}}^*$ and $\hat{\boldsymbol{\mathcal{L}}}$. Larger PSNR and SSIM values indicate higher quality of the estimator $\hat{\boldsymbol{\mathcal{L}}}$.

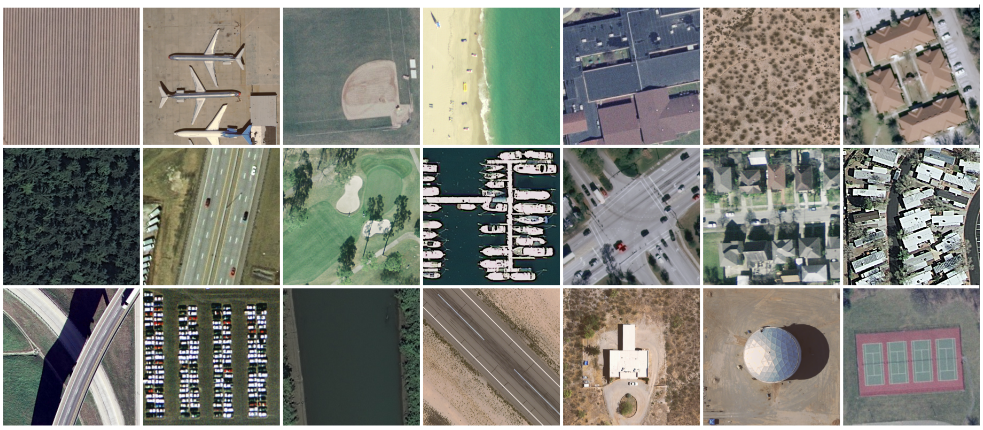


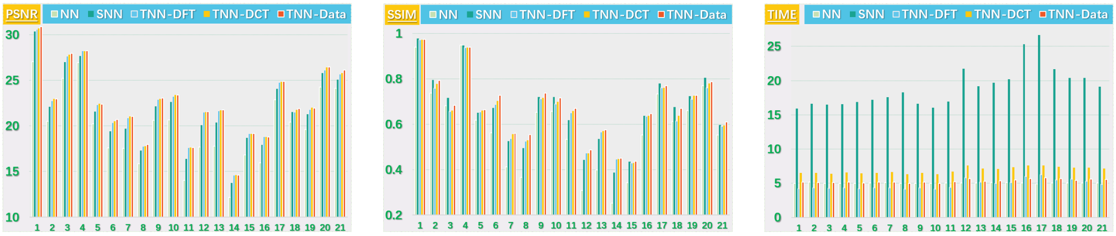
Figure 4. The dataset consists of the 85-th frame of all the 21 classes in the UCMerced dataset.

7.2.1. Experiments on an Urban Area Imagery Dataset

Area imagery data processing plays a key role in many remote sensing applications, such as land-use mapping [49]. We adopt the popular area imagery dataset UCMerced [50], which is a 21 class land use image dataset meant for research purposes. The images were manually extracted from large images from the USGS National Map Urban Area Imagery collection for various urban areas around the country. The pixel resolution of this public domain imagery is 1 foot, and each RGB image measures 256 × 256 pixels. There are 100 images for each class, and we chose the 85-th image to form a dataset of 21 images as shown in Figure 4.

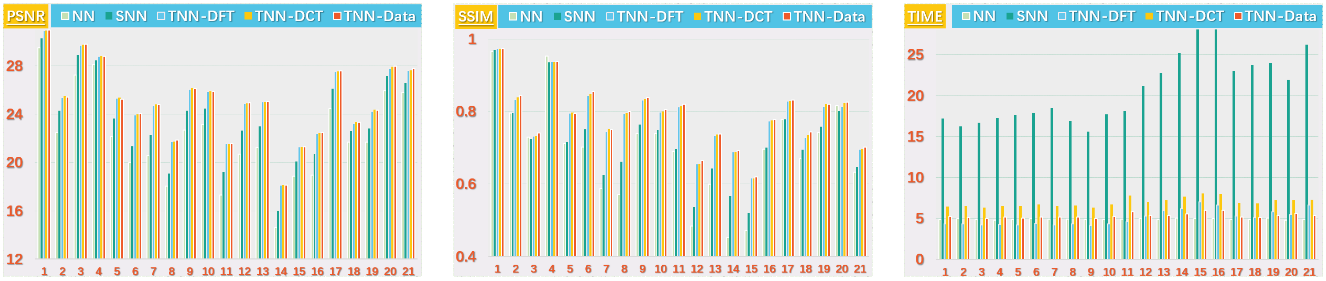
We consider two scenarios by setting $(\rho_{\text{obs}}, \rho_{\text{s}}) \in \{(0.3, 0.2), (0.8, 0.3)\}$ for the $d \times d \times 3$ images. For NN (Model (32)), we set the regularization parameters $\lambda_s = \lambda_\iota / \sqrt{d\rho_{\text{obs}}}$ (suggested by [38]), and tune the parameter λ_ι around $6.5\sigma\sqrt{\rho_{\text{obs}} d \log(6d)}$ (suggested by [51]). For SNN, the parameter λ_s is tuned in $\{0.01, 0.05, 0.1, 1\}$ for better performance in most cases, and the weight α is set by $\alpha_1 = \alpha_2 = \lambda_s\sqrt{3d\rho_{\text{obs}}}, \alpha_3 = 0.01\lambda_s\sqrt{3d\rho_{\text{obs}}}$. For Algorithm 1, we tune λ_ι around $2\sigma\sqrt{3\rho_{\text{obs}} d \log(6d)}$, and let $\lambda_s = \lambda_\iota / \sqrt{3d\rho_{\text{obs}}}$ for TNN (DFT) and $\lambda_s = \lambda_\iota / \sqrt{d\rho_{\text{obs}}}$ for TNN (DCT) and TNN (Data). In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds).

We present the PSNR, SSIM values and running time in Figures 5 and 6 for settings of $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.3, 0.2)$ and $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.8, 0.3)$, respectively, for quantitative evalution, with visual examples shown in Figures 7 and 8. It can seen that from Figures 5–8 that the proposed TNN (Data) has the highest recovery quality in most cases, and posses a comparative running time as NN. We attribute the promising performance of the proposed algorithm to the extraordinary representation power of the low-tubl-rank models: low-tubal-rankness can exploit both low-rankness and smoothness simultaneously, whereas traditional models like NN and SNN can only exploit low-rankness in the original domain [18].



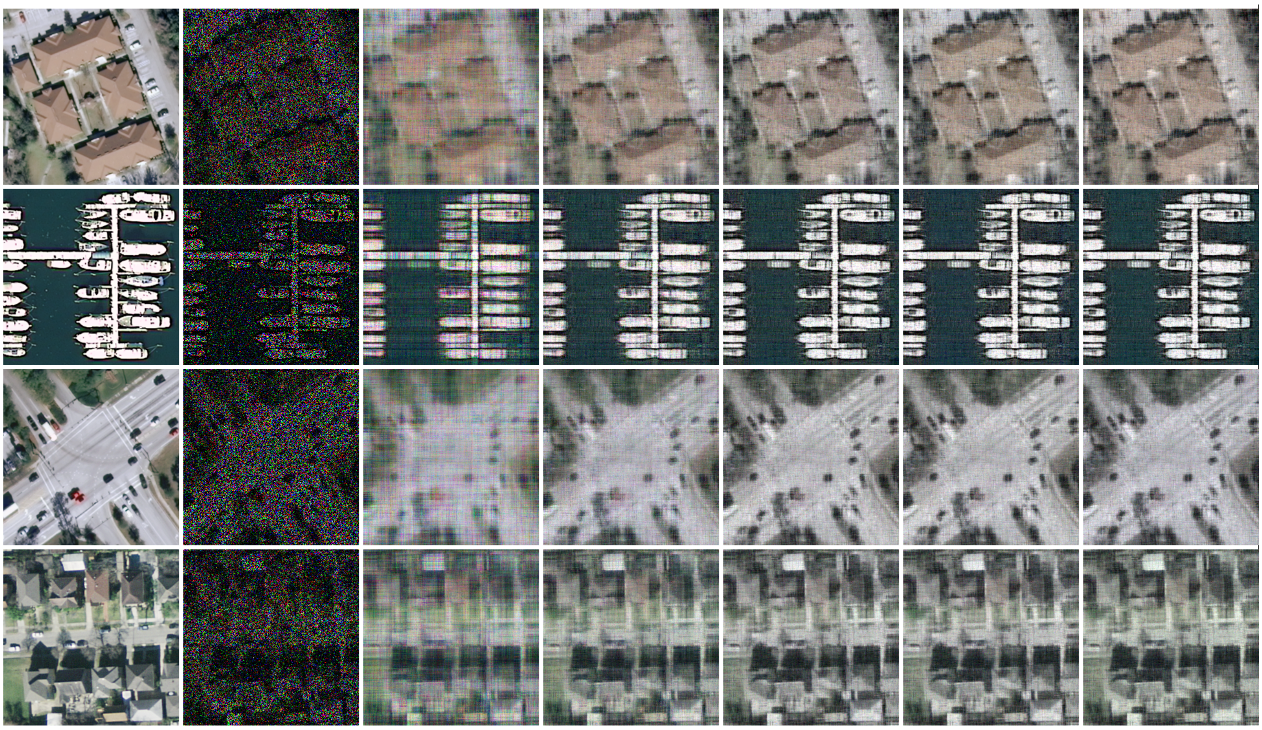
(**a**) PSNR (**b**) SSIM (**c**) TIME

Figure 5. The PSNR, SSIM values and running time (in seconds) on the UCMerced dataset for the setting $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.3, 0.2)$.



(**a**) PSNR (**b**) SSIM (**c**) TIME

Figure 6. The PSNR, SSIM values and running time (in seconds) on the UCMerced dataset for the setting $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.8, 0.3)$.



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

Figure 7. The visual examples for five models on UCMerced dataset for the setting $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.3, 0.2)$. (**a**) The original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

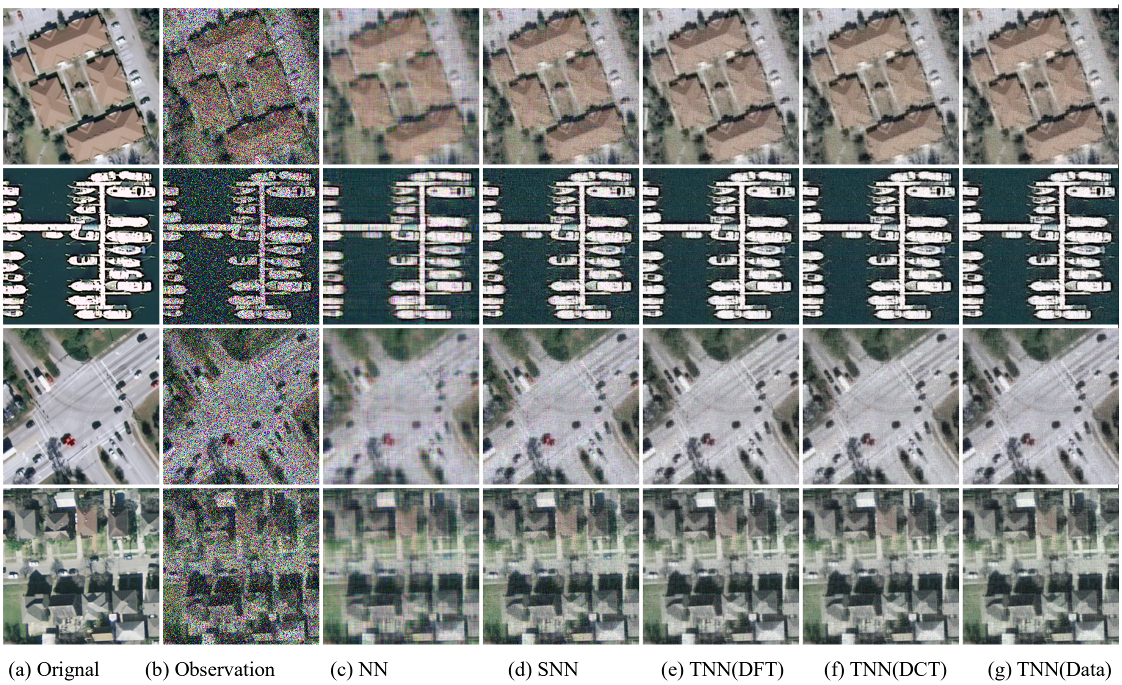


Figure 8. The visual examples for five models on UCMerced dataset for the setting $(\rho_{\text{obs}}, \rho_{\text{s}}) = (0.8, 0.3)$. (**a**) The original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

7.2.2. Experiments on Hyperspectral Data

Benefit from its fine spectral and spatial resolutions, hyperspectral image processing has been extensively adopted in many remote sensing applications [10,52]. In this section, we conduct robust 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 400–2500 nm. This dataset has a spatial extent of 86 × 83 pixels with a resolution of 3.7 m. We use the first 30 bands in the experiments too.

We consider three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3$, $\rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6$, $\rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8$, $\rho_{\text{s}} = 0.3$) for robust completion of hyper-spectral data. For NN, we set the regularization parameters $\lambda_s = \lambda_\iota / \sqrt{\rho_{\text{obs}}(d_1 \vee d_2)}$ (suggested by [38]), and tune the parameter λ_ι around $6.5\sigma\sqrt{\rho_{\text{obs}}(d_1 \vee d_2)\log(d_1 d_3 + d_2 d_3)}$ (suggested by [51]). For SNN, the parameter λ_s is tuned in $\{0.01, 0.05, 0.1, 1\}$ for better performance in most cases, and we chose the weight $\boldsymbol{\alpha}$ by $\alpha_1 = \alpha_2 = \alpha_3 = \lambda_s\sqrt{\rho_{\text{obs}}(d_1 \vee d_2)d_3}$ (suggested by [47]). For Algorithm 1, we tune the parameter λ_ι around $2\sigma\sqrt{\rho_{\text{obs}}(d_1 \vee d_2)d_3\log(d_1 d_3 + d_2 d_3)}$, and let $\lambda_s = \lambda_\iota / \sqrt{\rho_{\text{obs}}(d_1 \vee d_2)d_3}$ for TNN (DFT) and $\lambda_s = \lambda_\iota / \sqrt{\rho_{\text{obs}}(d_1 \vee d_2)}$ for TNN (DCT) and TNN (Data). In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds).

For quantitative evalution, we report the PSNR, SSIM values and running time in Tables 2 and 3 for the Indian Pines and Salinas A datasets, respectively. The visual examples are, respectively, shown in Figures 9 and 10. It can seen that the proposed TNN (Data) has the highest recovery quality in most cases, and has a comparative running time as NN, indicating the effectiveness and efficiency of low-tubal-rank models in comparison with original domain-based models NN and SNN.

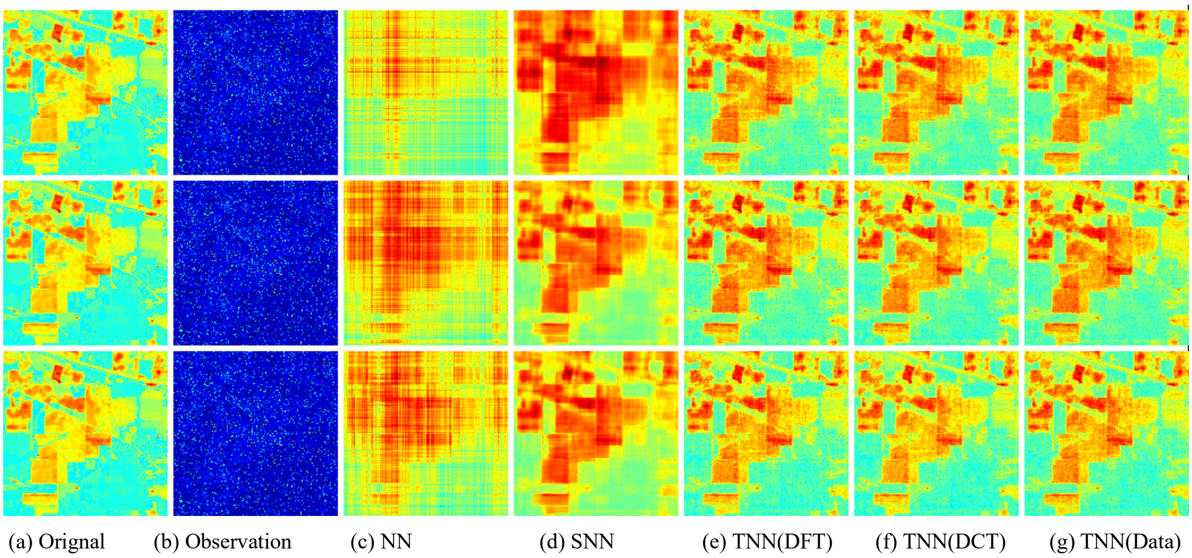


Figure 9. Visual results of robust tensor completion for five models on the 21st bound of Indian Pines dataset. The top, middle, and bottum row corresponds to the Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), respectively. The sub-plots from (**a**) to (**g**): (**a**) the original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

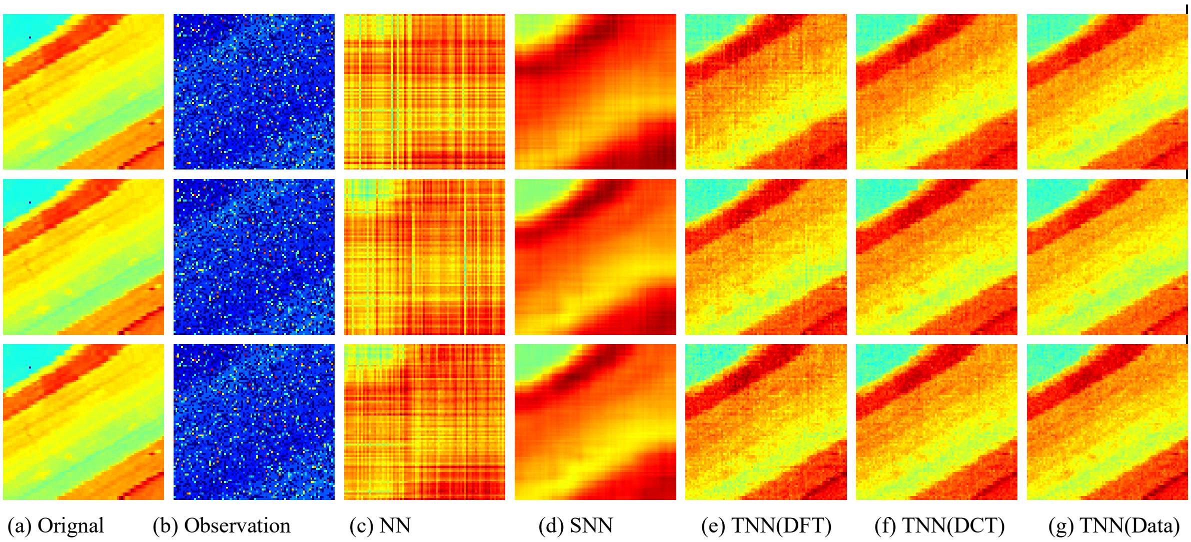


Figure 10. Visual results of robust tensor completion for five models on the 21st bound of Salinas A dataset. The top, middle, and bottum row corresponds to the Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), respectively. The sub-plots from (**a**) to (**g**): (**a**) the original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

Table 2. Quantitative evaluation on the Indian Pines dataset in PSNR, SSIM, and running time of five models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	20.63	25.46	28.49	29.33	**30.08**
	SSIM	0.4842	0.7275	0.7619	0.7872	**0.8181**
	TIME	14.77	40.1	**11.17**	15.53	13.54
Setting II	PSNR	21.95	27.66	29.49	30.17	**30.61**
	SSIM	0.5454	0.7864	0.7912	0.8073	**0.8296**
	TIME	14.18	39.76	**11.04**	15.23	13.29
Setting III	PSNR	22.43	28.22	29.64	30.31	**30.87**
	SSIM	0.5534	0.8051	0.7971	0.8139	**0.8345**
	TIME	14.21	38.88	**11.05**	15.27	13.43

Table 3. Quantitative evaluation on the Salinas A dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	19.01	26.18	27.1	30.99	**32.69**
	SSIM	0.4918	0.837	0.7501	0.8350	**0.8774**
	TIME	5.57	11.53	**4.07**	5.73	4.9
Setting II	PSNR	20.97	28.79	29.4	32.26	**33.3**
	SSIM	0.5806	0.8675	0.8117	0.8645	**0.8714**
	TIME	5.41	11.36	**4.02**	5.67	4.79
Setting III	PSNR	21.54	29.5	29.73	32.38	**33.54**
	SSIM	0.5914	0.8772	0.8208	0.8683	**0.8848**
	TIME	5.34	11.01	**3.98**	5.59	4.91

7.2.3. 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 multispectral images. The original data are two multispectral images Beads and Cloth from the Columbia MSI Database (available at http://www1.cs.columbia.edu/CAVE/databases/multispectral accessed on 28 July 2021) containing scenes of a variety of real-world objects. Each MSI is of size 512 × 512 × 31 with intensity range scaled to [0, 1].

We also consider three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6$, $\rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$) for robust completion of multi-spectral data. We tune the parameters in the same way as Section 7.2.2. In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds).

For quantitative evalution, we report the PSNR, SSIM values and running time in Tables 4 and 5 for the Beads and Cloth datasets, respectively. The visual examples for the Cloth dataset is shown in Figure 11. We can also find that the proposed TNN (Data) achieves the highest accuracy in most cases, and has a comparative running time as NN, which demonstrates both the effectiveness and efficiency of low-tubal-rank models.

Table 4. Quantitative evaluation on the Beads dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	18.58	18.71	25.11	25.18	**27.05**
	SSIM	0.448	0.6208	0.804	0.8203	**0.8673**
	TIME	309.55	933.12	280	260.46	**241.65**
Setting II	PSNR	20.4	21.35	27.31	27.46	**28.9**
	SSIM	0.5406	0.7603	0.8754	0.8894	**0.9143**
	TIME	302.95	915.57	276.2	268.58	**244.5**
Setting III	PSNR	21.01	22.36	27.96	28.13	**29.4**
	SSIM	0.5531	0.7848	0.8803	0.8944	**0.9165**
	TIME	301.92	922.07	276.99	272.59	**244.02**

Table 5. Quantitative evaluation on the Cloth dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	21.5	22.79	29.7	**30.83**	30.77
	SSIM	0.5054	0.6333	0.8649	0.8883	**0.8941**
	TIME	308.29	915.43	281.4	264.84	**242.63**
Setting II	PSNR	22.63	24.94	32.32	33.57	**33.86**
	SSIM	0.5566	0.7355	0.916	0.9323	**0.9391**
	TIME	300.3	911.27	273.36	268.46	**243.37**
Setting III	PSNR	22.99	25.78	32.76	34.02	**34.39**
	SSIM	0.5652	0.7643	0.9183	0.9342	**0.941**
	TIME	297.94	910.64	280.51	268.25	**246.14**

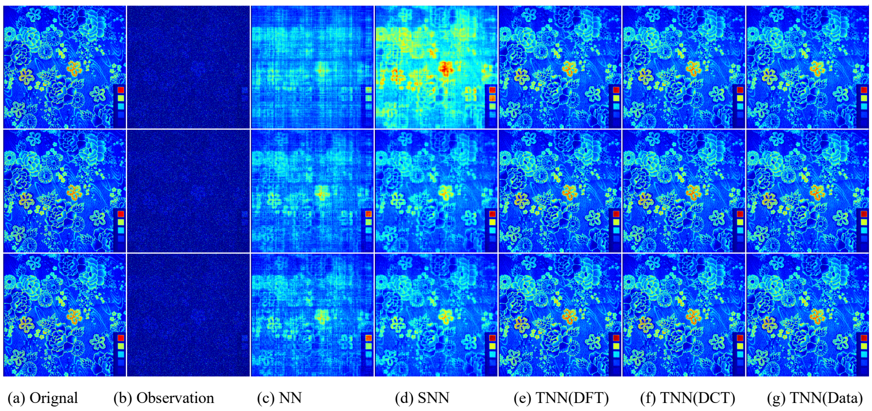


Figure 11. Visual results of robust tensor completion for five models on the 21st bound of Cloth dataset. The top, middle, and bottum row corresponds to the Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), respectively. The sub-plots from (**a**) to (**g**): (**a**) The original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

7.2.4. Experiments on Point Could Data

With the rapid advances of sensor technology, the emerging point cloud data provide better performance than 2D images in many remote sensing applications due to its flexible and scalable geometric representation [53]. In this section, we also conduct experiments on a dataset (scenario B from http://www.mrt.kit.edu/z/publ/download/velodynetracking/dataset.html, accessed on 28 July 2021) 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 × 870 × 30 representing the distance data (named SenerioB Distance) and the intensity data (named SenerioB Intensity), , respectively.

We also consider three settings, i.e., Setting I ($\rho_{\rm obs} = 0.3, \rho_{\rm s} = 0.2$), Setting II ($\rho_{\rm obs} = 0.6, \rho_{\rm s} = 0.25$), and Setting III ($\rho_{\rm obs} = 0.8, \rho_{\rm s} = 0.3$) for robust completion of point cloud data. We tune the parameters in the same way as Section 7.2.2. In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds). For quantitative evalution, we report the PSNR, SSIM values and running time in Tables 6 and 7 for the SenerioB Distance and SenerioB Intensity datasets, respectively. We can also find that the proposed TNN (Data) achieves the highest accuracy in most cases, and has a comparative running time as NN, which demonstrates both the effectiveness and efficiency of low-tubal-rank models.

Table 6. Quantitative evaluation on the SenerioB Distance dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\rm obs} = 0.3, \rho_{\rm s} = 0.2$), Setting II ($\rho_{\rm obs} = 0.6, \rho_{\rm s} = 0.25$), and Setting III ($\rho_{\rm obs} = 0.8, \rho_{\rm s} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	17.55	20.01	23.86	23.86	**23.87**
	SSIM	0.468	0.763	0.8732	0.8737	**0.8739**
	TIME	15.22	186.31	**14.49**	19.66	15.83
Setting II	PSNR	18.57	23.87	25.28	25.31	**25.34**
	SSIM	0.551	0.9055	0.9096	0.91	**0.9105**
	TIME	15.51	189.87	**15.55**	18.68	16.31
Setting III	PSNR	18.98	24.78	25.79	25.83	**25.87**
	SSIM	0.5678	**0.9197**	0.9179	0.9184	0.9189
	TIME	15.03	195	**14.8**	19.11	15.02

Table 7. Quantitative evaluation on the SenerioB Intensity dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\rm obs} = 0.3, \rho_{\rm s} = 0.2$), Setting II ($\rho_{\rm obs} = 0.6, \rho_{\rm s} = 0.25$), and Setting III ($\rho_{\rm obs} = 0.8, \rho_{\rm s} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	16.35	20.35	21.28	21.26	**21.31**
	SSIM	0.2588	0.7076	0.7114	0.7116	**0.7137**
	TIME	15.13	188.96	**14.18**	19.06	15.19
Setting II	PSNR	17.09	22.17	22.28	22.32	**22.45**
	SSIM	0.3149	**0.7889**	0.7708	0.7718	0.7781
	TIME	14.64	187.61	**14.09**	19.1	15.74
Setting III	PSNR	17.35	22.48	22.57	22.61	**22.79**
	SSIM	0.3331	**0.7985**	0.7828	0.7836	0.7914
	TIME	14.7	187.66	**14.23**	19	15.22

7.2.5. Experiments on Aerial Video Data

Aerial videos (or time sequences of images) are broadly used in many computer vision based remote sensing tasks [54]. We experiment on a 180 × 320 × 30 tensor which consists of the first 30 frames of the Sky dataset (available at http://www.loujing.com/rss-small-target, accessed on 28 July 2021) for small object detection [55].

We also consider three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). We tune the parameters in the same way as Section 7.2.2. In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds). For quantitative evalution, we report the PSNR, SSIM values and running time in Table 8. It is also found that the proposed TNN (Data) achieves the highest accuracy in most cases, and can run as fast as NN.

Table 8. Quantitative evaluation on the Sky dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	21.03	26.74	28.67	28.59	**29.74**
	SSIM	0.4875	0.7805	0.708	0.7076	**0.788**
	TIME	36.1	144.84	**28.1**	39.85	34.43
Setting II	PSNR	22.44	28.8	29.41	29.35	**30.48**
	SSIM	0.5715	**0.8026**	0.7155	0.7147	0.7814
	TIME	34.23	138.52	**27.11**	38.77	33.74
Setting III	PSNR	22.77	28.72	29.55	29.49	**30.59**
	SSIM	0.5786	0.7471	0.7324	0.7307	**0.7906**
	TIME	34.42	139.78	**26.97**	39.14	33.86

7.2.6. Experiments on Thermal Imaging Data

Thermal infrared data can provide important measurements of surface energy fluxes and temperatures in various remote sensing applications [7]. In this section, we experiment on two infrared datasets as follows:

- The Infraed Detection dataset [56]: this dataset is collected for infrared detection and tracking of dim-small aircraft targets under ground/air background (available at http://www.csdata.org/p/387/, accessed on 28 July 2021). 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 × 256 × 30 tensor due to the trade-off between the limitation of computing resources and the efforts for parameter tuning.
- The OSU Thermal Database [3]: 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 × 240 × 30.

Similiar to Section 7.2.2, we test in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), and use the same strategy for parameter tuning. In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds). For quantitative evalution, we report the PSNR, SSIM values and running time in Tables 9 and 10 for the Infraed Detection and OSU Thermal Database datasets, respectively. The visual examples are, respectively, shown in Figures 12 and 13. It can seen that the proposed TNN (Data) has the highest recovery quality in most cases, and has a comparative running time as NN, showing both effectiveness and efficiency of low-tubal-rank models in comparison with original domain-based models NN and SNN.

Table 9. Quantitative evaluation on the Infraed Detection dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	24.82	30.09	31.94	32.07	**32.84**
	SSIM	0.6021	**0.8231**	0.7408	0.7437	0.7768
	TIME	49.01	215.41	48.97	52.37	**46.59**
Setting II	PSNR	26.58	31.8	32.33	32.43	**33.11**
	SSIM	0.6679	**0.8414**	0.7428	0.7453	0.7724
	TIME	47.55	217.43	50.07	52.9	**47.18**
Setting III	PSNR	26.95	31.9	33.11	33.2	**33.81**
	SSIM	0.6682	**0.8454**	0.7237	0.7265	0.7525
	TIME	48.65	216.42	49.13	52.81	**46.94**

Table 10. Quantitative evaluation on the OSU Thermal Database in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	15.62	21.5	31.33	31.5	**31.51**
	SSIM	0.3402	0.8105	0.9345	0.9347	**0.935**
	TIME	49	222.49	**40.31**	49.45	42.22
Setting II	PSNR	17.47	29.48	32.88	33.19	**33.21**
	SSIM	0.4428	0.9057	0.9427	0.9431	**0.9433**
	TIME	46.18	197.9	**36.14**	47.19	41.79
Setting III	PSNR	18.17	30.83	33.31	33.71	**33.75**
	SSIM	0.468	0.9265	0.9495	0.95	**0.9507**
	TIME	45.85	200.39	**36.39**	46.96	41.36

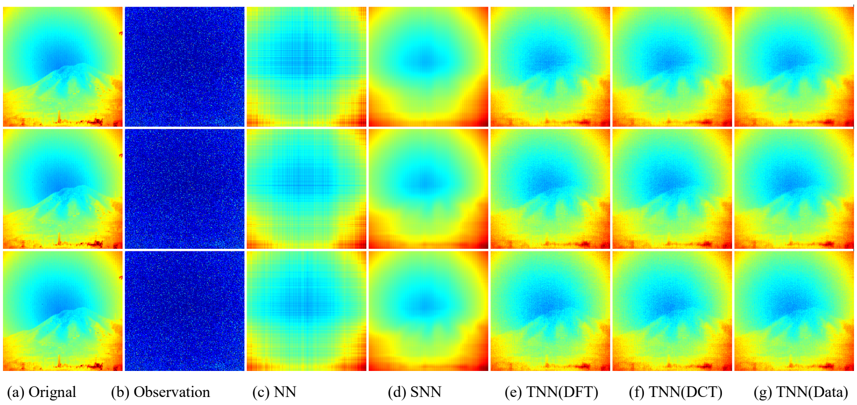

(a) Orignal (b) Observation (c) NN (d) SNN (e) TNN(DFT) (f) TNN(DCT) (g) TNN(Data)

Figure 12. Visual results of robust tensor completion for five models on the 21st bound of Infraed Detection dataset. The top, middle, and bottum row corresponds to the Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), respectively. The sub-plots from (**a**) to (**g**): (**a**) the original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

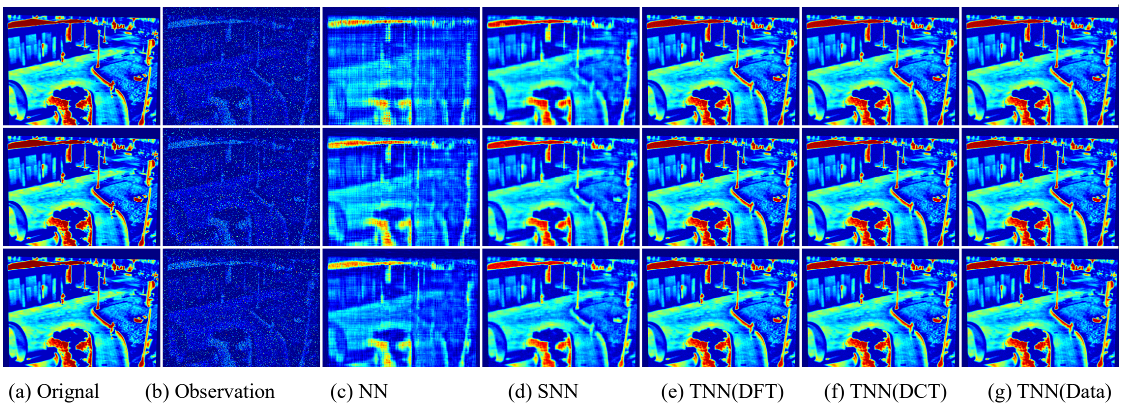


Figure 13. Visual results of robust tensor completion for five models on the 21st bound of OSU Thermal Database dataset. The top, middle, and bottum row corresponds to the Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$), respectively. The sub-plots from (**a**) to (**g**): (**a**) the original image; (**b**) the observed image; (**c**) image recovered by the matrix nuclear norm (NN) based Model (32); (**d**) recovered by the sum of mode-wise nuclear norms (SNN) based Model (33); (**e**) image recovered by TNN (DFT); (**f**) image recovered by TNN (DCT); (**g**) image recovered by TNN (Data).

7.2.7. 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 [57]. In this section, we adopt the PolSAR UAVSAR Change 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 three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6$, $\rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). We tune the parameters in the same way as Section 7.2.2. In each setting, we test each image for 10 trials and report the averaged PSNR (in db), SSIM and running time (in seconds) in Table 11. It is also found that the proposed TNN (Data) achieves the highest accuracy in most cases, and can run as fast as NN.

Table 11. Quantitative evaluation on the UAVSAR-Dataset1-2015 dataset in PSNR, SSIM, and running time of five tensor completion models for robust tensor completion in three settings, i.e., Setting I ($\rho_{\text{obs}} = 0.3, \rho_{\text{s}} = 0.2$), Setting II ($\rho_{\text{obs}} = 0.6, \rho_{\text{s}} = 0.25$), and Setting III ($\rho_{\text{obs}} = 0.8, \rho_{\text{s}} = 0.3$). The highest PSNR/SSIM, or lowest time (in seconds) is highlighted in **bold**.

Settings	Metrics	NN	SNN	TNN-DFT	TNN-DCT	TNN-Data
Setting I	PSNR	29.14	25.86	26.22	26.5	**31.62**
	SSIM	0.8748	0.8797	0.8868	0.8909	**0.9438**
	TIME	23.54	75.07	**17.46**	25.24	22.6
Setting II	PSNR	31.3	26.71	26.96	27.31	**34.28**
	SSIM	0.9044	0.8742	0.9018	0.9059	**0.9615**
	TIME	23.09	74.75	**17.6**	24.77	22.59
Setting III	PSNR	31.8	26.98	27.14	27.66	**35.03**
	SSIM	0.9118	0.8829	0.903	0.9092	**0.9649**
	TIME	22.88	73.03	**17.64**	25.22	22.54

8. Conclusions

In this paper, we resolve the challenging robust tensor completion problem by proposing a $*_L$-SVD-based estimator to robustly reconstruct a low-rank tensor in the presence of missing values, gross outliers, and small noises simultaneously. Specifically, this work can be concluded in the following three aspects:

(1) Algorithmically, we design an efficient algorithm within the framework of ADMM to efficiently compute the proposed estimator with guaranteed convergence behavior.

(2) Statistically, we analyze the statistical performance of the proposed estimator by establishing a non-asymptotic upper bound on the estimation error. The proposed upper bound is further proved to be minimax optimal (up to a log factor).

(3) Experimentally, the correctness of the upper bound is first validated through simulations on synthetic datasets. Then both effectiveness and efficiency of the proposed algorithm are demonstrated by extensive comparisons with state-of-the-art nuclear norm based models (i.e., NN and SNN) on seven different types of remote sensing data.

However, from a critical point of view, the proposed method has the following two limitations:

(1) The orientational sensitivity of $*_L$-SVD: Despite the promising empirical performance of the $*_L$-SVD-based estimator, 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 transformed low-rankness in multiple orientations [19,58].

(2) The difficulty in finding the optimal transform $L(\cdot)$ for $*_L$-SVD: Although a direct use of fixed transforms (like DFT and DCT) may produce fairish empirical performance, it is still unclear how to find the best optimal transformation $L(\cdot)$ for any certain tensor $\boldsymbol{\mathcal{L}}^*$ when only partial and corrupted observations are available.

According to the above limitations, it is interesting to consider higher-order extensions of the proposed model in an orientation invariant way like [19] and discuss the statistical performance. It is also interesting to consider the data-dependent transformation learning like [31,59]. Another future direction is to consider more efficient solvers of Problem (8) using the factorization strategy or Frank–Wolfe method [47,60–62].

Author Contributions: Conceptualization, A.W. and Q.Z.; Data curation, Q.Z.; Formal analysis, G.Z.; Funding acquisition, A.W., G.Z. and Q.Z.; Investigation, A.W.; Methodology, G.Z.; Project administration, A.W. and Q.Z.; Resources, Q.Z.; Software, A.W.; Supervision, G.Z. and Q.Z.; Validation, A.W., G.Z. and Q.Z.; Visualization, Q.Z.; Writing—original draft, A.W. and G.Z.; Writing—review & editing, A.W., G.Z. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 61872188, 62073087, 62071132, U191140003, 6197309, in part by the China Postdoctoral Science Foundation under Grant 2020M672536, in part by the Natural Science Foundation of Guangdong Province under Grants 2020A1515010671, 2019B010154002, 2019B010118001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: In this paper, all the data supporting our experimental results are publicly available with references or URL links.

Acknowledgments: The authors are grateful to the editor and reviewers for their valuable time in processing this manuscript. The first author would like to thank Shasha Jin for her kind understanding in these months, and Zhong Jin in Nanjing University of Science and Technology for his long time support. The authors are also grateful to Xiongjun Zhang for sharing the code for [25] and Olga Klopp for her excellent theoretical analysis in [36,51] which is so helpful.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theoretical Results

Appendix A.1. Additional Notations and Preliminaries

For the ease of exposition, we first list the additive notations often used in the proofs in Table A1.

Table A1. Additional notations in the proofs.

Notations	Descriptions	Notations	Descriptions
$\boldsymbol{\Delta}^{\iota} := \mathcal{L}^* - \hat{\mathcal{L}}$	estimation error of $\mathcal{L}^*$	$\boldsymbol{\Delta}^{s} := \mathcal{S}^* - \hat{\mathcal{S}}$	estimation error of $\mathcal{S}^*$
$r^* := r_{\rm tb}(\mathcal{L}^*)$	$*_L$-tubal-rank of $\mathcal{L}^*$	$s^* := \|\mathcal{S}^*\|_0$	sparsity of $\mathcal{S}^*$
$\varrho := N/N_\iota$	inverse uncorrupted ratio	a	l_∞-norm bound in Equation (8)
$\tilde{d}$	$(d_1 + d_3)d_3$		
$\Omega_\iota := \{i \in [N] \| \langle \mathcal{X}_i, \mathcal{S}^* \rangle = 0\}$	index set of design tensors $\{\mathcal{X}_i\}$ corresponding to uncorrupted entries		
$\Omega_s := \{i \in [N] \| \langle \mathcal{X}_i, \mathcal{S}^* \rangle \neq 0\}$	index set of design tensors $\{\mathcal{X}_i\}$ corresponding to corrupted entries		
$\mathcal{E} := \frac{1}{N}\sum_{i\in\Omega_\iota} \xi_i \mathcal{X}_i$	stochastic tensor defined to lower bound parameters λ_ι and λ_s		
$\mathcal{W} := \frac{1}{N}\sum_{i\in\Omega_\iota} \mathcal{X}_i$	stochastic tensor defined to lower bound parameter λ_s		
$\mathcal{R}_\Sigma := \frac{1}{N_\iota}\sum_{i\in\Omega_\iota} \varepsilon_i \mathcal{X}_i$	random tensor defined in bounding $\|\boldsymbol{\Delta}^\iota\|_{\rm F}$ with i.i.d. Rademacher $\{\varepsilon_i\}$		
$\|\mathcal{T}\|_\Pi := \sqrt{\mathbb{E}_{\mathcal{X}_i}\left[\left\langle \mathcal{X}_i, \mathcal{T}_{\boldsymbol{\Theta}_s^\perp} \right\rangle^2\right]}$	expectation of $\langle \mathcal{X}_i, \cdot \rangle^2$ for $i \in \Omega_\iota$ defined to establish the RSC condition		

We then introduce the decomposability of $*_L$-TNN and tensor l_1-norm which plays a key role in the analysis.

Decomposability of $*_L$-TNN. Suppose $\mathcal{L}^*$ has reduced $*_L$-SVD as $\mathcal{L}^* = \mathcal{U} *_L \mathcal{D} *_L \mathcal{V}^\top$, 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 and $\mathcal{D} \in \mathbb{R}^{r^* \times r^* \times d_3}$ is f-diagonal. Define projectors $\mathcal{P}^\star(\cdot)$ and $\mathcal{P}^\perp(\cdot)$ as follows:

$$\begin{aligned}
&\mathcal{P}^\star(\mathcal{T}) = \mathcal{U} *_L \mathcal{U}^\top *_L \mathcal{T} + \mathcal{T} *_L \mathcal{V} *_L \mathcal{V}^\top - \mathcal{U} *_L \mathcal{U}^\top *_L \mathcal{T} *_L \mathcal{V} *_L \mathcal{V}^\top, \\
&\mathcal{P}^\perp(\mathcal{T}) = (\mathcal{I} - \mathcal{U} *_L \mathcal{U}^\top) *_L \mathcal{T} *_L (\mathcal{I} - \mathcal{V} *_L \mathcal{V}^\top).
\end{aligned}$$

Then, it can be verified that:

(I). $\mathcal{T} = \mathcal{P}^\star(\mathcal{T}) + \mathcal{P}^\perp(\mathcal{T}), \forall \mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$;

(II). $\left\langle \mathcal{P}^\star(\mathcal{A}), \mathcal{P}^\perp(\mathcal{B}) \right\rangle = 0, \forall \mathcal{A}, \mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$.

(III). $r_{\rm tb}(\mathcal{P}^\star(\mathcal{T})) \leq 2 r_{\rm tb}(\mathcal{L}^*), \forall \mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$.

In the same way to the results in supplementary material of [43], it can also be shown that the following equations hold:

(I). (Decomposability of $*_L$–TNN) For any $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$ satisfying $\mathcal{A} *_L \mathcal{B}^\top = \mathbf{0}, \mathcal{A}^\top *_L \mathcal{B} = \mathbf{0}$,

$$\|\mathcal{A} + \mathcal{B}\|_\star = \|\mathcal{P}^\star(\mathcal{A})\|_\star + \|\mathcal{P}^\perp(\mathcal{B})\|_\star. \tag{A1}$$

(II). (Norm compatibility inequality) For any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$

$$\|\mathcal{T}\|_\star = \sqrt{d_3 r_{\rm tb}(\mathcal{T})} \|\mathcal{T}\|_{\rm F}. \tag{A2}$$

Decomposability of tensor l_1-norm [63]. Let $\Omega_{\rm e} \subset [d_1] \times [d_2] \times [d_3]$ denote any index set and $\Omega_{\rm e}^\perp$ its complement. Then for any tensor $\mathcal{T}$, define two tensors $\mathcal{T}_{\Omega_{\rm e}}$ and $\mathcal{T}_{\Omega_{\rm e}^\perp}$ as follows

$$\mathcal{T}_{\Omega_{\rm e}}(i,j,k) := \begin{cases} T_{ijk}, & (i,j,k) \in \Omega_{\rm e} \\ 0, & (i,j,k) \in \Omega_{\rm e}^\perp \end{cases}, \ \mathcal{T}_{\Omega_{\rm e}^\perp} := \mathcal{T} - \mathcal{T}_{\Omega_{\rm e}}. \tag{A3}$$

Then, one has

(I). (Decomposability of l_1-norm) For any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $\|\mathcal{T}\|_1 = \|\mathcal{T}_{\Omega_{\rm e}}\|_1 + \|\mathcal{T}_{\Omega_{\rm e}^\perp}\|_1$.

(II). (Norm compatibility inequality) For any $\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, $\|\mathcal{T}_{\Omega_{\rm e}}\|_1 = \sqrt{|\Omega_{\rm e}|} \|\mathcal{T}_{\Omega_{\rm e}}\|_{\rm F}$.

Appendix A.2. The Proof for Theorem 3

The proof follows the lines of [34,36]. For notational simplicity, we define the following two sets

$$\Omega_\iota := \{i \in [N] \mid \langle \mathcal{X}_i, \mathcal{S}^* \rangle = 0\}, \ \Omega_s := \{i \in [N] \mid \langle \mathcal{X}_i, \mathcal{S}^* \rangle \neq 0\} \tag{A4}$$

which denote the index set of design tensors $\{\mathcal{X}_i\}$ corresponding to uncorrupted/corrupted entries, respectively.

Appendix A.2.1. Mainstream of Proving Theorem 3

Proof of Theorem 3. Let $\mathscr{F}(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \langle \boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{X}}_i\rangle)^2 + \lambda_\iota\|\boldsymbol{\mathcal{L}}\|_\star + \lambda_s\|\boldsymbol{\mathcal{S}}\|_1$ for simplicity. Then, according to the optimality of $(\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{S}}})$ to Problem (8), it holds that

$$\mathscr{F}(\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{S}}}) \le \mathscr{F}(\boldsymbol{\mathcal{L}}^*,\boldsymbol{\mathcal{S}}^*) \tag{A5}$$

and

$$\begin{aligned}
\|\boldsymbol{\Delta}^\iota\|_\infty &= \|\hat{\boldsymbol{\mathcal{L}}}-\boldsymbol{\mathcal{L}}^*\|_\infty \le \|\hat{\boldsymbol{\mathcal{L}}}\|_\infty + \|\boldsymbol{\mathcal{L}}^*\|_\infty \le 2a\\
\|\boldsymbol{\Delta}^s\|_\infty &= \|\hat{\boldsymbol{\mathcal{S}}}-\boldsymbol{\mathcal{S}}^*\|_\infty \le \|\hat{\boldsymbol{\mathcal{S}}}\|_\infty + \|\boldsymbol{\mathcal{S}}^*\|_\infty \le 2a
\end{aligned}\tag{A6}$$

Equation (A5) indicates that

$$\frac{1}{N}\sum_{i=1}^{N}(\xi_i + \langle\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s,\boldsymbol{\mathcal{X}}_i\rangle)^2 + \lambda_\iota\|\hat{\boldsymbol{\mathcal{L}}}\|_\star + \lambda_s\|\hat{\boldsymbol{\mathcal{S}}}\|_1 \le \frac{1}{N}\sum_{i=1}^{N}\xi_i^2 + \lambda_\iota\|\boldsymbol{\mathcal{L}}^*\|_\star + \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1 \tag{A7}$$

which leads to

$$\begin{aligned}
\frac{1}{N}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s,\boldsymbol{\mathcal{X}}_i\rangle^2 \le\ & \underbrace{\frac{2}{N}\sum_{i\in\Omega_s}|\langle\xi_i\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle| - \frac{1}{N}\sum_{i\in\Omega_s}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle^2}_{:=\mathbf{I}}\\
&+\underbrace{2|\langle\boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^\iota\rangle| + \lambda_\iota(\|\boldsymbol{\mathcal{L}}^*\|_\star - \|\hat{\boldsymbol{\mathcal{L}}}\|_\star)}_{:=\mathbf{II}}\\
&+\underbrace{2\left|\left\langle\boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\right\rangle\right| + \lambda_s(\|\boldsymbol{\mathcal{S}}^*\|_1 - \|\hat{\boldsymbol{\mathcal{S}}}\|_1)}_{:=\mathbf{III}}
\end{aligned}\tag{A8}$$

where $\boldsymbol{\mathcal{E}} := \frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\xi_i\boldsymbol{\mathcal{X}}_i$, and the equality $\langle\boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle = \left\langle\boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\right\rangle$ holds. Now each item in the right hand side of (A8) will be upper bounded separately as follows. Following the idea of [36], the upper bound will be analyzed upon the following event

$$\mathsf{E} := \left\{\max_{1\le i\le N}|\xi_i| \le C_*\sigma\sqrt{\log\tilde{d}}\right\} \tag{A9}$$

According to the tail behavior of the maximum in a sub-Gaussian sequence, it holds with an absolute constant $C_* > 0$ such that $\mathbb{P}[\mathsf{E}] \ge 1 - 1/(2\tilde{d})$.

Bound I. On the event E, we get

$$\mathbf{I} \le \frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2 \le \frac{C\sigma^2|\Theta_s|\log\tilde{d}}{N}. \tag{A10}$$

Bound II. Note that according to the properties of $*_L$-TNN, we have

$$\begin{aligned}
\|\boldsymbol{\mathcal{L}}^*\|_\star - \|\hat{\boldsymbol{\mathcal{L}}}\|_\star &= \|\boldsymbol{\mathcal{L}}^*\|_\star - \|\boldsymbol{\mathcal{L}}^* - \boldsymbol{\Delta}^\iota\|_\star\\
&= \|\boldsymbol{\mathcal{L}}^*\|_\star - \|\boldsymbol{\mathcal{L}}^* - \boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota) - \boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star\\
&\le \|\boldsymbol{\mathcal{L}}^*\|_\star - (\|\boldsymbol{\mathcal{L}}^* - \boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star - \|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star)\\
&= \|\boldsymbol{\mathcal{L}}^*\|_\star - (\|\boldsymbol{\mathcal{L}}^*\|_\star + \|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star - \|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star)\\
&= \|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star - \|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star
\end{aligned}$$

Thus, we can bound term **II** by

$$\mathbf{II} \le 2\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}}\|\boldsymbol{\Delta}^\iota\|_\star + \lambda_\iota\left(\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star - \|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star\right)$$

By letting $\lambda_\iota \geq 4\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}}$, it holds that

$$\mathbf{II} \leq \frac{3}{2}\lambda_\iota \|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star \leq \frac{3}{2}\lambda_\iota \sqrt{2r^* d_3}\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}. \tag{A11}$$

Bound III: Note that since $\boldsymbol{\mathcal{S}}^*_{\boldsymbol{\Theta}_s^\perp} = \mathbf{0}$, we have $\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp} = -\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}$, leading to

$$\mathbf{III} \leq 2\|\boldsymbol{\mathcal{E}}\|_\infty \|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1 + \lambda_s(\|\boldsymbol{\mathcal{S}}^*\|_1 - \|\hat{\boldsymbol{\mathcal{S}}}\|_1) \leq (2\|\boldsymbol{\mathcal{E}}\|_\infty - \lambda_s)\|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1 + \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1$$

Letting $\lambda_s \geq 4\|\boldsymbol{\mathcal{E}}\|_\infty$ yields

$$\mathbf{III} \leq \lambda_s \|\boldsymbol{\mathcal{S}}^*\|_1 \tag{A12}$$

Thus, putting Equations (A10)–(A12) together, we have the following inequality on the event $\mathbf{E}$:

$$\frac{1}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\rangle^2 \leq \frac{3\lambda_\iota\sqrt{r^* d_3}}{\sqrt{2}}\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}} + \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1 + \frac{C\sigma^2|\Theta_s|\log\tilde{d}}{N} \tag{A13}$$

Then, we follow the line of [36] to specify a kind of Restricted Strong Convexity (RSC) for the random sampling operator formed by the design tensors $\{\boldsymbol{\mathcal{X}}_i\}$ on a carefully chosen constrained set. The RSC will show that when the error tensors $(\boldsymbol{\Delta}^\iota, \boldsymbol{\Delta}^s)$ belong to the constrained set, the following relationship:

$$\frac{1}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\rangle^2 \geq \kappa_0\|\boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\|_{\Pi}^2 - \boldsymbol{\tau}, \tag{A14}$$

holds with an appropriate residual $\boldsymbol{\tau}$ with high probability.

Before explicitly defining the constrained set, we first consider the following set where $\boldsymbol{\Delta}^s$ should lie:

$$\mathbf{B}(\delta_1, \delta_2) := \{\boldsymbol{\mathcal{B}} \in \mathbb{R}^{d_1\times d_2\times d_3} \mid \|\boldsymbol{\mathcal{B}}\|_\Pi^2 \leq \delta_1^2,\ \|\boldsymbol{\mathcal{B}}\|_1 \leq \delta_2\} \tag{A15}$$

with two positive constants δ_1 and δ_2 whose values will be specified later. We also define the following set of tensor pairs:

$$\mathbf{D}(r,\kappa,\beta) := \left\{(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) \mid \|\boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\|_\Pi^2 \geq \beta, \|\boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\|_\infty \leq 1, \|\boldsymbol{\mathcal{A}}\|_\star \leq \sqrt{rd_3}\|\boldsymbol{\mathcal{A}}_{\boldsymbol{\Theta}_s^\perp}\|_{\mathrm{F}} + \kappa\right\} \tag{A16}$$

We then define the constrained set as the intersection:

$$\mathbf{D}(r,\kappa,\beta) \cap \{\mathbb{R}^{d_1\times d_2\times d_3} \times \mathbf{B}(\delta_1,\delta_2)\} \tag{A17}$$

To bound the estimation error in Equation (26), we will upper bound $\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}$ and $\|\boldsymbol{\Delta}^s\|_{\mathrm{F}}$ separately.

Note that

$$\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}^2 = \|\boldsymbol{\Delta}^\iota_{\Theta_s}\|_{\mathrm{F}}^2 + \|\boldsymbol{\Delta}^\iota_{\boldsymbol{\Theta}_s^\perp}\|_{\mathrm{F}}^2 \leq \|\boldsymbol{\Delta}^\iota_{\Theta_s}\|_{\mathrm{F}}^2 + 4a^2|\boldsymbol{\Theta}_s^\perp| = |\Theta_s|\|\boldsymbol{\Delta}^\iota_{\boldsymbol{\Theta}_s^\perp}\|_\Pi^2 + 4a^2|\boldsymbol{\Theta}_s^\perp| \tag{A18}$$

and similarly

$$\|\boldsymbol{\Delta}^s\|_{\mathrm{F}}^2 \leq |\Theta_s|\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_\Pi^2 + 4a^2|\boldsymbol{\Theta}_s^\perp| \tag{A19}$$

We will bound $\|\boldsymbol{\Delta}^\iota_{\boldsymbol{\Theta}_s^\perp}\|_\Pi^2$ and $\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_\Pi^2$ separately. We first bound $\|\boldsymbol{\Delta}^\iota_{\boldsymbol{\Theta}_s^\perp}\|_\Pi^2$ in what follows.

Case 1: If $\|\boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\|_\Pi^2 \leq 16a^2\sqrt{\frac{128\log\tilde{d}}{N_\iota}}$, then we use the following inequality

$$\|\boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\|_\Pi^2 \geq \frac{1}{2}\|\boldsymbol{\Delta}^\iota\|_\Pi^2 - \|\boldsymbol{\Delta}^s\|_\Pi^2 \tag{A20}$$

which holds due to $(x+y)^2 = x^2 + y^2 + 2xy \geq x^2 + y^2 + 2 \cdot x/\sqrt{2} \cdot \sqrt{2}y \geq x^2 + y^2 - (x^2/2 + 2y^2) = x^2/2 - y^2$.

Thus, we can upper bound $\|\boldsymbol{\Delta}^{\iota}\|_{\Pi}$ with an upper bound of $\|\boldsymbol{\Delta}^{s}\|_{\Pi}$ in Lemma A1

$$\|\boldsymbol{\Delta}^{\iota}\|_{\Pi}^2 \leq 16a^2\sqrt{\frac{128\log\tilde{d}}{N_{\iota}}} + \|\boldsymbol{\Delta}^{s}\|_{\Pi}^2 \tag{A21}$$

Case 2: Suppose $\|\boldsymbol{\Delta}^{\iota} + \boldsymbol{\Delta}^{s}\|_{\Pi}^2 \geq 16a^2\sqrt{\frac{128\log\tilde{d}}{N_{\iota}}}$. First, according to Lemma A3, it holds on the event $\mathbf{E}$ defined in Equation (A9) that

$$\begin{aligned}
\|\boldsymbol{\Delta}^{\iota}\|_{\star} &\overset{(i)}{\leq} \|\boldsymbol{\mathcal{P}}^{\perp}(\boldsymbol{\Delta}^{\iota})\|_{\star} + \|\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\Delta}^{\iota})\|_{\star} \\
&\overset{(ii)}{\leq} 4\|\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\Delta}^{\iota})\|_{\star} + \frac{2N_s}{N\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d}) \\
&\overset{(iii)}{\leq} 4\sqrt{2r^* d_3}\|\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\Delta}^{\iota})\|_{\mathrm{F}} + \frac{2N_s}{N\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d}) \\
&\overset{(iv)}{\leq} 4\sqrt{2r^* d_3}\|\boldsymbol{\Delta}^{\iota}\|_{\mathrm{F}} + \frac{2N_s}{N\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d}) \\
&\overset{(v)}{\leq} 4\sqrt{2r^* d_3}(\|\boldsymbol{\Delta}^{\iota}_{\boldsymbol{\Theta}_s^{\perp}}\|_{\mathrm{F}} + \|\boldsymbol{\Delta}^{\iota}_{\Theta_s}\|_{\mathrm{F}}) + \frac{2N_s}{N\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d}) \\
&\overset{(vi)}{\leq} \sqrt{32r^* d_3}\|\boldsymbol{\Delta}^{\iota}_{\boldsymbol{\Theta}_s^{\perp}}\|_{\mathrm{F}} + a\sqrt{128r^* d_3|\Theta_s|} + \frac{2N_s}{N\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d})
\end{aligned} \tag{A22}$$

where (i) holds due to the triangular inequality; (ii) is a direct consequence of Lemma A3, and the definition of event $\mathbf{E}$; (iii) holds because $r_{\mathrm{tb}}(\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\mathcal{T}})) \leq 2r^*$, and $\|\boldsymbol{\mathcal{T}}\|_{\star} \leq \sqrt{r_{\mathrm{tb}}(\boldsymbol{\mathcal{T}})d_3}\|\boldsymbol{\mathcal{T}}\|_{\mathrm{F}}$ for any $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times d_2\times d_3}$; (iv) stems from the inequality $\|\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\mathcal{T}})\|_{\mathrm{F}} \leq \|\boldsymbol{\mathcal{T}}\|_{\mathrm{F}} = \sqrt{\|\boldsymbol{\mathcal{P}}^{\star}(\boldsymbol{\mathcal{T}})\|_{\mathrm{F}}^2 + \|\boldsymbol{\mathcal{P}}^{\perp}(\boldsymbol{\mathcal{T}})\|_{\mathrm{F}}^2}$; (v) is due to the triangular inequality; (vi) holds since $\|\boldsymbol{\Delta}^{\iota}\|_{\infty} \leq 2a$.

Note that according to Lemma A1, we have with probability at least $1 - 2.5/\tilde{d}$:

$$\frac{\boldsymbol{\Delta}^{s}}{4a} \in \mathbf{B}(\delta_1, \delta_2) \ \text{ with } \ \delta_1 = \frac{\sqrt{\Delta_1}}{4a},\ \delta_2 = \frac{N_s}{4aN\lambda_s}(4a^2 + 8aN\lambda_s + C\sigma^2\log\tilde{d}) \tag{A23}$$

where Δ_1 is defined in Lemma A4.

Together with Equation (A22), we have

$$\frac{1}{4a}(\boldsymbol{\Delta}^{\iota}, \boldsymbol{\Delta}^{s}) \in \mathbf{D}(r, \kappa, \beta) \cap (\mathbb{R}^{d_1\times d_2\times d_3} \times \mathbf{B}(\delta_1, \delta_2)) \tag{A24}$$

with the following parameters

$$r = 32r^* \quad \text{and} \quad \kappa = 2a\sqrt{2r^* d_3|\Theta_s|} + \frac{N_s}{2aN\lambda_{\iota}}(aN\lambda_s + C\sigma^2\log\tilde{d}) \tag{A25}$$

Then, according to Lemma A6, it holds with probability at least $1 - 2/\tilde{d}$ that

$$\frac{1}{16a^2 N_{\iota}}\sum_{i\in\Omega_{\iota}}\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^{\iota} + \boldsymbol{\Delta}^{s}\rangle^2 \geq \frac{1}{32a^2}\|\boldsymbol{\Delta}^{\iota} + \boldsymbol{\Delta}^{s}\|_{\Pi}^2 - \boldsymbol{\tau}(r, \kappa, \delta_1, \delta_2) \tag{A26}$$

where $\boldsymbol{\tau}(r, \kappa, \delta_1, \delta_2)$ is defined in Equation (A64).

Recall that Equation (A13) writes:

$$\frac{1}{N}\sum_{i\in\Omega_{\iota}}\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^{\iota} + \boldsymbol{\Delta}^{s}\rangle^2 \leq \frac{3\lambda_{\iota}\sqrt{r^* d_3}}{\sqrt{2}}\|\boldsymbol{\Delta}^{\iota}\|_{\mathrm{F}} + N_s(a\lambda_s + \frac{C\sigma^2\log\tilde{d}}{N}) \tag{A27}$$

Letting $\varrho := N/N_\iota$, we further obtain

$$\begin{aligned}\frac{1}{2}\|\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\|_\Pi^2 &\le \frac{3\varrho\lambda_\iota}{\sqrt{2}}\sqrt{r^*d_3}\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}+C\boldsymbol{\tau}' \\ &\le \frac{9}{8}\varrho^2\lambda_\iota^2 r^* d_3\cdot 4d_1d_2d_3+\frac{\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}^2}{4d_1d_2d_3}+C\boldsymbol{\tau}' \\ &\le \frac{9}{2}\varrho^2\lambda_\iota^2 r^* d_1d_2d_3^2+\frac{\|\boldsymbol{\Delta}^\iota_{\Theta_s^\perp}\|_{\mathrm{F}}^2}{4d_1d_2d_3}+\frac{a^2|\Theta_s|}{d_1d_2d_3}+C\boldsymbol{\tau}'\end{aligned} \tag{A28}$$

Thus, by using Equation (A20) and Lemma A1, we have

$$\frac{\|\boldsymbol{\Delta}^\iota_{\Theta_s^\perp}\|_{\mathrm{F}}^2}{d_1d_2d_3}\le C\left(\varrho^2\lambda_\iota^2r^*d_1d_2d_3^2+\frac{a^2|\Theta_s|}{d_1d_2d_3}+\boldsymbol{\tau}'\right) \tag{A29}$$

where

$$\boldsymbol{\tau}'=\varrho N_s(a\lambda_s+\frac{C\sigma^2\log\tilde{d}}{N})+16a^2\boldsymbol{\tau}(r,\kappa,\delta_1,\delta_2) \tag{A30}$$

Note that the bound on $\|\boldsymbol{\Delta}^s\|_\Pi$ is given in Lemma A4, and the values of λ_ι and λ_s can be set according to Lemmas A8 and A9, respectively. Then, by putting Equations (A18), (A19) and (A52) together, and using Lemmas A8 and A9 to bound associated norms of the stochastic quantities $\boldsymbol{\mathcal{E}}$, $\boldsymbol{\mathcal{W}}$, and $\boldsymbol{\mathcal{R}}_\Sigma$ in the error term, we can obtain the bound on $\|\boldsymbol{\Delta}^\iota\|_{\mathrm{F}}^2+\|\boldsymbol{\Delta}^s\|_{\mathrm{F}}^2$ and complete the proof. □

Appendix A.2.2. Lemmas for the Proof of Theorem 3

Lemma A1. *Letting $\lambda_s\ge 4(\|\boldsymbol{\mathcal{E}}\|_\infty+2a\|\boldsymbol{\mathcal{W}}\|_\infty)$, it holds that*

$$\|\boldsymbol{\Delta}^s_{\Theta_s^\perp}\|_1\le 3\|\boldsymbol{\Delta}^s_{\Omega_s}\|_1+\frac{1}{N\lambda_s}(4a^2N_s+\sum_{i\in\Omega_s}\xi_i^2) \tag{A31}$$

Proof of Lemma A1. By the standard condition for optimality over a convex set, it holds that for any feasible $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})$

$$\left\langle(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}}),\partial\mathscr{F}(\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{S}}})\right\rangle\ge 0 \tag{A32}$$

which further leads to

$$\begin{aligned}&-\frac{2}{N}\sum_{i=1}^N(y_i-\left\langle\boldsymbol{\mathcal{X}}_i,\hat{\boldsymbol{\mathcal{L}}}+\hat{\boldsymbol{\mathcal{S}}}\right\rangle)\left\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{L}}+\boldsymbol{\mathcal{S}}-\hat{\boldsymbol{\mathcal{L}}}-\hat{\boldsymbol{\mathcal{S}}}\right\rangle \\ &+\lambda_\iota\left\langle\partial\|\hat{\boldsymbol{\mathcal{L}}}\|_\star,\boldsymbol{\mathcal{L}}-\hat{\boldsymbol{\mathcal{L}}}\right\rangle+\lambda_s\left\langle\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1,\boldsymbol{\mathcal{S}}-\hat{\boldsymbol{\mathcal{S}}}\right\rangle\ge 0.\end{aligned} \tag{A33}$$

Letting $(\boldsymbol{\mathcal{L}},\boldsymbol{\mathcal{S}})\leftarrow(\hat{\boldsymbol{\mathcal{L}}},\boldsymbol{\mathcal{S}}^*)$, we have

$$-\frac{2}{N}\sum_{i=1}^N(y_i-\left\langle\boldsymbol{\mathcal{X}}_i,\hat{\boldsymbol{\mathcal{L}}}+\hat{\boldsymbol{\mathcal{S}}}\right\rangle)\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle+\lambda_s\left\langle\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1,\boldsymbol{\Delta}^s\right\rangle\ge 0. \tag{A34}$$

Note that

$$\begin{aligned}
&-\frac{2}{N}\sum_{i=1}^{N}(y_i-\left\langle \boldsymbol{\mathcal{X}}_i,\hat{\boldsymbol{\mathcal{L}}}+\hat{\boldsymbol{\mathcal{S}}}\right\rangle)\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&=-\frac{2}{N}\sum_{i=1}^{N}(\xi_i+\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle)\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&=-\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2-\frac{2}{N}\sum_{i=1}^{N}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&=-\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2-\left(\frac{2}{N}\sum_{i\in\Omega_s}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle+\frac{2}{N}\sum_{i\in\Omega_\iota}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle\right) \\
&\quad-\left(\frac{2}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle+\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle\right) \\
&=-\left(\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2+\frac{2}{N}\sum_{i\in\Omega_s}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle+\frac{2}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle\right) \\
&\quad-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&\le-\left(\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2-\frac{1}{N}\sum_{i\in\Omega_s}(\xi_i^2+\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2)-\frac{1}{N}\sum_{i\in\Omega_s}((\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle^2+\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2)\right) \\
&\quad-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&=\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{1}{N}\sum_{i\in\Omega_s}((\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle^2-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle
\end{aligned}\tag{A35}$$

Thus, we have

$$\begin{aligned}
&\lambda_s\left\langle\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1,\hat{\boldsymbol{\mathcal{S}}}-\boldsymbol{\mathcal{S}}^*\right\rangle \\
&\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle-\frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle \\
&\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}+2|\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s\rangle|+\frac{2}{N}\left|\sum_{i\in\Omega_\iota}\langle\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle\right| \\
&\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}+2\|\boldsymbol{\mathcal{E}}\|_\infty\|\boldsymbol{\Delta}^s\|_1+2\|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle \boldsymbol{\mathcal{X}}_i\|_\infty\|\boldsymbol{\Delta}^s\|_1 \\
&\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}+2\|\boldsymbol{\mathcal{E}}\|_\infty\|\boldsymbol{\Delta}^s\|_1+4a\|\boldsymbol{\mathcal{W}}\|_\infty\|\boldsymbol{\Delta}^s\|_1
\end{aligned}\tag{A36}$$

One the other hand, the definition of sub-differential indicates

$$\|\boldsymbol{\mathcal{S}}^*\|_1-\|\hat{\boldsymbol{\mathcal{S}}}\|_1\ge\left\langle\boldsymbol{\mathcal{S}}^*-\hat{\boldsymbol{\mathcal{S}}},\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1\right\rangle\tag{A37}$$

which implies

$$\lambda_s(\|\hat{\boldsymbol{\mathcal{S}}}\|_1-\|\boldsymbol{\mathcal{S}}^*\|_1)\le\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{4a^2|\Theta_s|}{N}+2\|\boldsymbol{\mathcal{E}}\|_\infty\|\boldsymbol{\Delta}^s\|_1+2\|\boldsymbol{\mathcal{W}}\|_\infty\|\boldsymbol{\Delta}^s\|_1$$

Also note that

$$\begin{aligned}\|\hat{\boldsymbol{\mathcal{S}}}\|_1 - \|\boldsymbol{\mathcal{S}}^*\|_1 &= \|\boldsymbol{\mathcal{S}}^* - \boldsymbol{\Delta}^s\|_1 - \|\boldsymbol{\mathcal{S}}^*\|_1 \\ &= \|\boldsymbol{\mathcal{S}}^*_{\Theta_s} - (\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp} + \boldsymbol{\Delta}^s_{\Theta_s})\|_1 - \|\boldsymbol{\mathcal{S}}^*_{\Theta_s}\|_1 \\ &\geq \|\boldsymbol{\mathcal{S}}^*_{\Theta_s} - \boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 - \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1 - \|\boldsymbol{\mathcal{S}}^*_{\Theta_s}\|_1 \\ &= \|\boldsymbol{\mathcal{S}}^*_{\Theta_s}\|_1 + \|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 - \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1 - \|\boldsymbol{\mathcal{S}}^*_{\Theta_s}\|_1 \\ &= \|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 - \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1\end{aligned} \tag{A38}$$

which implies

$$\lambda_s(\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 - \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1) \leq \frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2 + \frac{4a^2|\Theta_s|}{N} + 2\|\boldsymbol{\mathcal{E}}\|_\infty\|\boldsymbol{\Delta}^s\|_1 + 2\|\boldsymbol{\mathcal{W}}\|_\infty\|\boldsymbol{\Delta}^s\|_1$$

Since $\lambda_s \geq 4(\|\boldsymbol{\mathcal{E}}\|_\infty + 2a\|\boldsymbol{\mathcal{W}}\|_\infty)$, we have

$$\lambda_s(\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 - \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1) \leq \frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2 + \frac{4a^2|\Theta_s|}{N} + \frac{\lambda_s}{2}(\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 + \|\boldsymbol{\Delta}^s_{\Theta_s}\|_1) \tag{A39}$$

Thus, it holds that

$$\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 \leq 3\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1 + \frac{1}{N\lambda_s}(4a^2|\Theta_s| + \sum_{i\in\Omega_s}\xi_i^2) \tag{A40}$$

which complete the proof. □

Lemma A2. *It holds that*

$$\|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota\rangle\boldsymbol{\mathcal{X}}_i\|_\infty \leq 2a\|\boldsymbol{\mathcal{W}}\|_\infty \tag{A41}$$

Proof of Lemma A2. Note that

$$\begin{aligned}\|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota\rangle\boldsymbol{\mathcal{X}}_i\|_\infty &\overset{(i)}{\leq} \sup_{\|\boldsymbol{\mathcal{T}}\|_1\leq 1}\left\langle\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota\rangle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\right\rangle \\ &\leq 2a\sup_{\|\boldsymbol{\mathcal{T}}\|_1\leq 1}\left\langle\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\frac{\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota\rangle}{2a}\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\right\rangle \\ &\overset{(ii)}{\leq} 2a\sup_{\|\boldsymbol{\mathcal{T}}'\|_1\leq 1}\left\langle\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}'\right\rangle \\ &\leq 2a\|\boldsymbol{\mathcal{W}}\|_\infty\end{aligned} \tag{A42}$$

where (i) hold since $\|\cdot\|_\infty$ is the dual norm of $\|\cdot\|_1$; (ii) holds since $|\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota\rangle| \leq \|\boldsymbol{\Delta}^\iota\|_\infty \leq 2a$ and the tensor ℓ_1-norm $\|\cdot\|_1$ is invariant to changes in sign. □

Lemma A3. *By letting $\lambda_\iota \geq 4\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}}$ and $\lambda_s \geq \|\boldsymbol{\mathcal{E}}\|_\infty$, we have*

$$\|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star \leq 3\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star + \frac{2a\lambda_s}{\lambda_\iota}N_s + \frac{2}{N\lambda_\iota}\sum_{i\in\Omega_s}\xi_i^2 \tag{A43}$$

Proof of Lemma A3. In Equation (A33), letting $(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{S}}) \leftarrow (\boldsymbol{\mathcal{L}}^*, \boldsymbol{\mathcal{S}}^*)$, we obtain

$$-\frac{2}{N}\sum_{i=1}^{N}(y_i - \left\langle\boldsymbol{\mathcal{X}}_i, \hat{\boldsymbol{\mathcal{L}}} + \hat{\boldsymbol{\mathcal{S}}}\right\rangle)\langle\boldsymbol{\mathcal{X}}_i, \boldsymbol{\Delta}^\iota + \boldsymbol{\Delta}^s\rangle + \lambda_\iota\left\langle\partial\|\hat{\boldsymbol{\mathcal{L}}}\|_\star, \boldsymbol{\Delta}^\iota\right\rangle + \lambda_s\left\langle\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1, \boldsymbol{\Delta}^s\right\rangle \geq 0 \tag{A44}$$

First, note that

$$
\begin{aligned}
&-\frac{2}{N}\sum_{i=1}^{N}(y_i-\left\langle \boldsymbol{\mathcal{X}}_i,\hat{\boldsymbol{\mathcal{L}}}+\hat{\boldsymbol{\mathcal{S}}}\right\rangle)\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle\\
&=-\frac{2}{N}\sum_{i=1}^{N}(\xi_i+\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle)\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle\\
&=-\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle^2-\frac{2}{N}\sum_{i=1}^{N}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle\\
&=-\frac{2}{N}\sum_{i=1}^{N}\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle^2-\frac{2}{N}\sum_{i\in\Omega_s}\xi_i\langle \boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota+\boldsymbol{\Delta}^s\rangle-2\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^\iota\rangle-2\left\langle \boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\right\rangle
\end{aligned}
\tag{A45}
$$

Also, we have according to the convexity of $\|\cdot\|_\star$ and $\|\cdot\|_1$ that

$$
\|\boldsymbol{\mathcal{L}}^*\|_\star-\|\hat{\boldsymbol{\mathcal{L}}}\|_\star\geq\left\langle \boldsymbol{\mathcal{L}}^*-\hat{\boldsymbol{\mathcal{L}}},\partial\|\hat{\boldsymbol{\mathcal{L}}}\|_\star\right\rangle,\ \text{and}\quad \|\boldsymbol{\mathcal{S}}^*\|_1-\|\hat{\boldsymbol{\mathcal{S}}}\|_1\geq\left\langle \boldsymbol{\mathcal{S}}^*-\hat{\boldsymbol{\mathcal{S}}},\partial\|\hat{\boldsymbol{\mathcal{S}}}\|_1\right\rangle
\tag{A46}
$$

Thus, we have

$$
\lambda_\iota(\|\hat{\boldsymbol{\mathcal{L}}}\|_\star-\|\boldsymbol{\mathcal{L}}^*\|_\star)+\lambda_s(\|\hat{\boldsymbol{\mathcal{S}}}\|_1-\|\boldsymbol{\mathcal{S}}^*\|_1)\leq 2\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}}\|\boldsymbol{\Delta}^\iota\|_\star+2\|\boldsymbol{\mathcal{E}}\|_\infty\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1+\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2.
$$

Moreover, it is often used that

$$
\|\hat{\boldsymbol{\mathcal{L}}}\|_\star-\|\boldsymbol{\mathcal{L}}^*\|_\star\geq\|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star-\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star,
\tag{A47}
$$

Since we set $\lambda_\iota\geq 4\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}}$ and $\lambda_s\geq 4\|\boldsymbol{\mathcal{E}}\|_\infty$, we have

$$
\begin{aligned}
&\lambda_\iota(\|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star-\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star)+\lambda_s(\|\hat{\boldsymbol{\mathcal{S}}}\|_1-\|\boldsymbol{\mathcal{S}}^*\|_1)\\
&\leq\frac{\lambda_\iota}{2}(\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star+\|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star)+\frac{\lambda_s}{2}\|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1+\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2
\end{aligned}
\tag{A48}
$$

where we use $\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}=-\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}$. It implies

$$
\frac{\lambda_\iota}{2}\|\boldsymbol{\mathcal{P}}^\perp(\boldsymbol{\Delta}^\iota)\|_\star+\lambda_s\|\hat{\boldsymbol{\mathcal{S}}}_{\Theta_s}\|_1+\frac{\lambda_s}{2}\|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1\leq\frac{3\lambda_\iota}{2}\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star+\lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1+\frac{1}{N}\sum_{i\in\Omega_s}\xi_i^2
\tag{A49}
$$

Note that, $\|\boldsymbol{\mathcal{S}}^*\|_1\leq aN_s$. Thus, we have

$$
\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star\leq 3\lambda_\iota\|\boldsymbol{\mathcal{P}}^\star(\boldsymbol{\Delta}^\iota)\|_\star+\frac{2a\lambda_s}{\lambda_\iota}N_s+\frac{2}{N\lambda_\iota}\sum_{i\in\Omega_s}\xi_i^2
\tag{A50}
$$

□

Lemma A4. *If $N_\iota\geq\tilde{d}$ and $\lambda_s\geq 4(\|\boldsymbol{\mathcal{E}}\|_\infty+2a\|\boldsymbol{\mathcal{W}}\|_\infty)$, then on the event* $\mathbf{E}$ *defined in Equation (A9), we have*

$$
\|\boldsymbol{\Delta}^s\|_1\leq\frac{N_s}{N\lambda_s}(4a^2+8aN\lambda_s+C\sigma^2\log\tilde{d})
\tag{A51}
$$

and it holds with probability at least $1-2.5/\tilde{d}$ that

$$
\begin{aligned}
&\|\boldsymbol{\Delta}^s\|_\Pi\leq\Delta_1:=\\
&C(\varrho\frac{2N_s}{N_\iota}(4a^2+2aN\lambda_\iota+CN_s\sigma^2\log\tilde{d})+\frac{16aN_s}{N\lambda_s}(4a^2+8aN\lambda_s+C\sigma^2\log\tilde{d})\mathbb{E}[\|\boldsymbol{\mathcal{E}}\|_\infty])
\end{aligned}
\tag{A52}
$$

Proof of Lemma A4. We first prove Equation (A51), and then prove Equation (A52).

(I) The proof of Equation (A51): Recall that Lemma A1 implies

$$\|\mathbf{\Delta}^s_{\mathbf{\Theta}_s^\perp}\|_1 \le 3\|\mathbf{\Delta}^s_{\Omega_s}\|_1 + \frac{1}{N\lambda_s}\big(4a^2 N_s + \sum_{i\in\Omega_s}\xi_i^2\big)$$

Then, we have on event **E**:

$$\begin{aligned}\|\mathbf{\Delta}^s\|_1 = \|\mathbf{\Delta}^s_{\mathbf{\Theta}_s^\perp}\|_1 + \|\mathbf{\Delta}^s_{\mathbf{\Theta}_s}\|_1 &\overset{(i)}{\le} \|\mathbf{\Delta}^s_{\mathbf{\Theta}_s^\perp}\|_1 + \|\mathbf{\Delta}^s_{\Omega_s}\|_1 \\ &\overset{(ii)}{\le} 4\|\mathbf{\Delta}^s_{\Omega_s}\|_1 + \frac{1}{N\lambda_s}\big(4a^2 N_s + \sum_{i\in\Omega_s}\xi_i^2\big) \\ &\overset{(iii)}{\le} \frac{N_s}{N\lambda_s}(4a^2 + 8aN\lambda_s + C\sigma^2 \log \tilde{d})\end{aligned} \tag{A53}$$

where (i) holds due to Assumption 1.I; (ii) is a direct use of Lemma A1; (iii) stems from the facts that $\|\mathbf{\Delta}^s\|_\infty \le 2a$ and the definition of event **E** in Equation (A9).

Thus, Equation (A51) is proved.

(II) The proof of Equation (A52): According to the optimality of $(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}})$ to Problem (8), we have

$$\mathscr{F}(\hat{\boldsymbol{\mathcal{L}}}, \hat{\boldsymbol{\mathcal{S}}}) \le \mathscr{F}(\hat{\boldsymbol{\mathcal{L}}}, \boldsymbol{\mathcal{S}}^*) \tag{A54}$$

which implies

$$\frac{1}{N}\sum_{i=1}^{N_\iota}(\xi_i + \langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota + \mathbf{\Delta}^s\rangle)^2 + \lambda_s\|\hat{\boldsymbol{\mathcal{S}}}\|_1 \le \frac{1}{N}\sum_{i=1}^{N_\iota}(\xi_i + \langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota\rangle)^2 + \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1 \tag{A55}$$

which further leads to

$$\begin{aligned}&\frac{1}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle^2 + \frac{1}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle^2 + \frac{2}{N}\sum_{i\in\Omega_s}\xi_i\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle + \frac{2}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle \\ &\qquad + \frac{2}{N}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota\rangle\left\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s_{\mathbf{\Theta}_s^\perp}\right\rangle + 2\sum_{i\in\Omega_\iota}\left\langle \boldsymbol{\mathcal{E}}, \mathbf{\Delta}^s_{\mathbf{\Theta}_s^\perp}\right\rangle + \lambda_s\|\hat{\boldsymbol{\mathcal{S}}}\|_1 \le \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1\end{aligned} \tag{A56}$$

Note that by using $2ab > -(1/2a^2 + 2b^2)$, we have

$$\begin{aligned}&\frac{1}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle^2 + \frac{2}{N}\sum_{i\in\Omega_s}\xi_i\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle + \frac{2}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota\rangle\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^s\rangle \\ &\ge -\big(\frac{2}{N}\sum_{i\in\Omega_s}\xi_i^2 + \frac{2}{N}\sum_{i\in\Omega_s}\langle \boldsymbol{\mathcal{X}}_i, \mathbf{\Delta}^\iota\rangle^2\big)\end{aligned}$$

Thus on the event $\mathbf{E}$ defined in Equation (A9), we have

$$\begin{aligned}
\frac{1}{N}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2 \le{}& \left|\frac{2}{N}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle\left\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\right\rangle\right| + \left|2\sum_{i\in\Omega_\iota}\left\langle\boldsymbol{\mathcal{E}},\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\right\rangle\right| \\
&+\lambda_s(\|\boldsymbol{\mathcal{S}}^*\|_1-\|\hat{\boldsymbol{\mathcal{S}}}\|_1)+\frac{2}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{2}{N}\sum_{i\in\Omega_s}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle^2 \\
\overset{(i)}{\le}{}& (4a\|\boldsymbol{\mathcal{W}}\|_\infty+2\|\boldsymbol{\mathcal{E}}\|_\infty)\|\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp}\|_1+\lambda_s(\|\boldsymbol{\mathcal{S}}^*\|_1-\|\hat{\boldsymbol{\mathcal{S}}}\|_1) \\
&+\frac{2}{N}\sum_{i\in\Omega_s}\xi_i^2+\frac{2}{N}\sum_{i\in\Omega_s}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle^2 \qquad\qquad \text{(A57)}\\
\overset{(ii)}{\le}{}& \frac{\lambda_s}{2}\|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1+\lambda_s(\|\boldsymbol{\mathcal{S}}^*\|_1-\|\hat{\boldsymbol{\mathcal{S}}}\|_1)+\frac{2}{N}\sum_{i\in\Omega_s}(\xi_i^2+\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle^2) \\
\overset{(iii)}{\le}{}& \lambda_s\|\boldsymbol{\mathcal{S}}^*\|_1+\frac{2}{N}\sum_{i\in\Omega_s}(\xi_i^2+4a^2) \\
\overset{(iv)}{\le}{}& \frac{2N_s}{N}(4a^2+2aN\lambda_\iota+CN_s\sigma^2\log\tilde{d})
\end{aligned}$$

where (i) holds due to Lemma A2; (ii) holds because $\lambda_s \ge 4(2a\|\boldsymbol{\mathcal{W}}\|_\infty+\|\boldsymbol{\mathcal{E}}\|_\infty)$, and $\boldsymbol{\Delta}^s_{\boldsymbol{\Theta}_s^\perp} = -\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}$; (iii) holds because $\|\hat{\boldsymbol{\mathcal{S}}}\|_1 = \|\hat{\boldsymbol{\mathcal{S}}}_{\Theta_s}\|_1+\|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1 \ge \|\hat{\boldsymbol{\mathcal{S}}}_{\boldsymbol{\Theta}_s^\perp}\|_1$, and $|\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^\iota\rangle| \le \|\boldsymbol{\Delta}^\iota\|_\infty \le 2a$; (iv) holds as a consequence of $\|\boldsymbol{\mathcal{S}}^*\|_\infty \le a$, $|\Theta_s| \le N_s$ (due to Assumption 1.I), and the definition of event $\mathbf{E}$.

Now, we discuss the bound of $\|\boldsymbol{\Delta}^s\|_\Pi$ in two cases.

Case 1. If $\|\boldsymbol{\Delta}^s\|_\Pi^2 \le \beta = 4a^2\sqrt{\frac{128\log\tilde{d}}{N_\iota}}$, then Equation (A52) holds trivially.

Case 2. If $\|\boldsymbol{\Delta}^s\|_\Pi^2 \ge 4a^2\sqrt{\frac{128\log\tilde{d}}{N_\iota}}$, then we have

$$\frac{\boldsymbol{\Delta}^s}{2a} \in \mathbf{D}\left(\frac{N_s}{2aN\lambda_s}(4a^2+8aN\lambda_s+C\sigma^2\log\tilde{d}),\delta\right)$$

due to the fact $\|\boldsymbol{\Delta}^s/(2a)\|_\infty \le 1$ and Equation (A51). Then according to Lemma A5, it holds with probability at least 1 - $1/\tilde{d}^2$ that

$$\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\Delta}^s\rangle^2 \ge \frac{1}{2}\|\boldsymbol{\Delta}^s\|_\Pi^2-\frac{16aN_s}{N\lambda_s}(4a^2+8aN\lambda_s+C\sigma^2\log\tilde{d})\mathbb{E}[\|\boldsymbol{\mathcal{E}}\|_\infty]. \qquad \text{(A58)}$$

Combing Equations (A57) and (A58) yields the bound on $\|\boldsymbol{\Delta}^s\|_\Pi$.

□

Lemma A5. *Define the following set*

$$\mathbf{D}(\delta,\beta) := \left\{\boldsymbol{\mathcal{B}} \in \mathbb{R}^{d_1\times d_2\times d_3} \;\middle|\; \|\boldsymbol{\mathcal{B}}\|_\infty \le 1,\ \|\boldsymbol{\mathcal{B}}\|_\Pi^2 \ge \beta, \|\boldsymbol{\mathcal{B}}\|_1 \le \delta\right\} \qquad \text{(A59)}$$

Then, it holds with probability at least $1-2/\tilde{d}^3$ *that*

$$\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle\boldsymbol{\mathcal{X}}_i,\boldsymbol{\mathcal{B}}\rangle^2 \ge \frac{1}{2}\|\boldsymbol{\mathcal{B}}\|_\Pi^2-8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] \qquad \text{(A60)}$$

for any $\boldsymbol{\mathcal{B}} \in \mathbf{D}(\delta,\beta)$.

Proof of Lemma A5. We prove this lemma using a standard peeling argument. First, define the following

$$\mathbf{G} := \left\{ \exists \boldsymbol{\mathcal{B}} \in \mathbf{D}(\delta,\beta) \text{ such that } \Big|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{B}}\rangle - \|\boldsymbol{\mathcal{B}}\|_\Pi^2\Big| \geq \frac{1}{2}\|\boldsymbol{\mathcal{B}}\|_\Pi^2 + 8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] \right\}$$

We partition this set to simpler events with $l \in \mathbb{N}_+$:

$$\begin{aligned}\mathbf{G}_l := \Big\{ &\exists \boldsymbol{\mathcal{B}} \in \mathbf{D}(\delta,\beta) \cap \mathbf{C}'(t) \;\text{ with } t \in [\alpha^{l-1}\beta, \alpha^l\beta] \\ &\text{such that } \Big|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{B}}\rangle - \|\boldsymbol{\mathcal{B}}\|_\Pi^2\Big| \geq \frac{1}{2}\|\boldsymbol{\mathcal{B}}\|_\Pi^2 + 8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] \Big\}\end{aligned} \tag{A61}$$

Note that according to Lemma A7, we have with $t \in [\alpha^{l-1}\beta, \alpha^l\beta)$:

$$\mathbb{P}[\mathbf{G}_l] \leq \mathbb{P}\left[\sup_{\boldsymbol{\mathcal{B}}\in\mathbf{C}'(t)} \Big|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{B}}\rangle - \|\boldsymbol{\mathcal{B}}\|_\Pi^2\Big| \geq \frac{1}{2\alpha}\alpha^l\beta + 8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty]\right] \leq \exp(-\frac{n(\alpha^l\beta)^2}{32\alpha^2})$$

Thus, we have

$$\begin{aligned}\mathbb{P}[\mathbf{G}] \leq \mathbb{P}\Big[\bigcup_{l=1}^{\infty}\mathbf{G}_l\Big] \leq \sum_{l=1}^{\infty}\mathbb{P}[\mathbf{G}_l] &\leq \sum_{l=1}^{\infty}\exp(-\frac{N_\iota(\alpha^l\beta)^2}{32\alpha^2}) \\ &= \exp(-\frac{N_\iota\beta^2}{32}) + \sum_{l=2}^{\infty}\exp(-\frac{N_\iota\beta^2}{32}\alpha^{2(l-1)}) \\ &\overset{(i)}{\leq} \exp(-\frac{N_\iota\beta^2}{32}) + \sum_{l=2}^{\infty}\exp(-\frac{N_\iota\beta^2}{32}\cdot 2(l-1)\log\alpha) \\ &\leq \exp(-\frac{N_\iota\beta^2}{32}) + \frac{\exp(-\frac{N_\iota\beta^2}{16}\log\alpha)}{1-\exp(-\frac{N_\iota\beta^2}{16}\log\alpha)}\end{aligned} \tag{A62}$$

where (i) is due to $x \geq \log x$ for positive x. By setting $\alpha = e$ and recalling the value of $\beta = \sqrt{\frac{128\log\tilde{d}}{N_\iota}}$, the lemma is proved. □

Lemma A6. *For any* $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) \in \mathbf{C}(r,\kappa,\beta) \cap \mathbb{R}^{d_1\times d_2\times d_3} \times \mathbf{B}(\delta_1,\delta_2)$*, it holds with probability at least* $1 - 2/\tilde{d}$ *that*

$$\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\rangle^2 \geq \frac{1}{2}\|\boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\|_\Pi^2 - \boldsymbol{\tau}(r,\kappa,\delta_1,\delta_2) \tag{A63}$$

where

$$\boldsymbol{\tau}(r,\kappa,\delta_1,\delta_2) = 4(16\alpha+1)rd_3|\boldsymbol{\Theta}_s^\perp|\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}]^2 + 8\kappa\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}] + 8\delta_2\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] + 4\delta_1^2 \tag{A64}$$

Proof of Lemma A6. The proof is very similar to that of Lemma A5, and we simply omit it. □

Lemma A7. *Define the set*

$$\mathbf{C}'(t) := \{\boldsymbol{\mathcal{T}} \in \mathbb{R}^{d_1\times d_2\times d_3} \mid \|\boldsymbol{\mathcal{T}}\|_\Pi^2 \leq t, \|\boldsymbol{\mathcal{T}}\|_\infty \leq 1\} \tag{A65}$$

and

$$Z_t := \sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)}\left|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\rangle^2 - \|\boldsymbol{\mathcal{T}}\|_\Pi^2\right|. \tag{A66}$$

Then, it holds that

$$\mathbb{P}[Z_t \geq \mathbb{E}[Z_t] + \frac{t}{4\alpha}] \leq \exp(-\frac{N_\iota t^2}{128\alpha^2}), \text{ and } \quad \mathbb{E}[Z_t] \leq 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)} |\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{T}}\rangle|\Big] \tag{A67}$$

Proof of Lemma A7. First, we study the tail behavior of Z_t by directly using the Massart's inequality in Theorem 14.2 of [64]. According to the Massart's inequality, it holds for any $s > 0$

$$\mathbb{P}[Z_t \geq \mathbb{E}[Z_t] + s] \leq \exp(-\frac{N_\iota s^2}{8}) \tag{A68}$$

By letting $s = t/(4\alpha)$, the first inequality in Equation (A67) is proved.

Then, we will upper bound the expectation of Z_t. By standard symmetrization argument [65], we have

$$\begin{aligned}\mathbb{E}[Z_t] &= \mathbb{E}\left[\sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)}\left|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\rangle^2 - \mathbb{E}\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{T}}\rangle^2\right|\right] \\ &= 2\mathbb{E}\left[\sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)}\left|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\varepsilon_i\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\rangle^2\right|\right]\end{aligned} \tag{A69}$$

where ε_i's are i.i.d. Randemacher variables. Further, according to the contraction principle [66], it holds that

$$\mathbb{E}[Z_t] \leq 8\mathbb{E}\left[\sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)}\left|\frac{1}{N_\iota}\sum_{i\in\Omega_\iota}\varepsilon_i\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{T}}\rangle\right|\right] = 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{T}}\in\mathbf{C}'(t)} |\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{T}}\rangle|\Big] \tag{A70}$$

In the following, we consider the two cases:

Case 1. Consider $\boldsymbol{\mathcal{T}} \in \mathbf{D}(\delta,\beta) \cap \mathbf{C}'(t)$, we have

$$\mathbb{E}[Z_t] \leq 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{T}}}|\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{T}}\rangle|\Big] \leq 8\mathbb{E}\Big[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty\|\boldsymbol{\mathcal{T}}\|_1\Big] \leq 8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty]. \tag{A71}$$

By letting $s = t/(2\alpha)$ in Equation (A68), we obtain

$$\mathbb{P}\Big[Z_t \geq 8\delta\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] + \frac{t}{2\alpha}\Big] \leq \exp(-\frac{N_\iota t^2}{32\alpha^2}) \tag{A72}$$

when $\boldsymbol{\mathcal{T}} \in \mathbf{D}(\delta,\beta) \cap \mathbf{C}'(t)$.

Case 2. Consider $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}$, where $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}) \in \mathbf{C}(r,\kappa,\beta)$, $\boldsymbol{\mathcal{B}} \in \mathbf{B}(\delta_1,\delta_2)$, and $\|\boldsymbol{\mathcal{T}}\|_\Pi^2 \leq t$. The goal in this case is to upper bound

$$\begin{aligned}\mathbb{E}[Z_t] &\leq 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}}|\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\rangle|\Big] \\ &\overset{(i)}{\leq} 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{A}}}|\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{A}}\rangle|\Big] + 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{B}}}|\langle \boldsymbol{\mathcal{R}}_\Sigma, \boldsymbol{\mathcal{B}}\rangle|\Big] \\ &\overset{(ii)}{\leq} 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{A}}}\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\text{sp}}\|\boldsymbol{\mathcal{A}}\|_\star\Big] + 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{B}}}\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty\|\boldsymbol{\mathcal{B}}\|_1\Big] \\ &\overset{(iii)}{\leq} 8\mathbb{E}\Big[\sup_{\boldsymbol{\mathcal{A}}}\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\text{sp}}\|\boldsymbol{\mathcal{A}}\|_\star\Big] + 8\delta_2\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty]\end{aligned} \tag{A73}$$

where (i) holds as a property of the sup operation; (ii) holds due to the definition of dual norm; (iii) stems from the condition $\boldsymbol{\mathcal{B}} \in \mathbf{B}(\delta_1,\delta_2)$.

It remains to upper bound $\|\boldsymbol{\mathcal{A}}\|_\star$ in Equation (A72). First, according to the definition of $\mathbf{C}(r,\kappa,\beta)$, we have

$$\|\boldsymbol{\mathcal{A}}\|_\star \le \sqrt{rd_3}\|\boldsymbol{\mathcal{A}}_{\boldsymbol{\Theta}_s^\perp}\|_{\mathrm{F}} + \kappa \tag{A74}$$

We also have

$$\|\boldsymbol{\mathcal{A}}\|_\Pi \overset{(i)}{\le} \|\boldsymbol{\mathcal{A}}+\boldsymbol{\mathcal{B}}\|_\Pi + \|\boldsymbol{\mathcal{B}}\|_\Pi \overset{(ii)}{\le} \sqrt{t} + \delta_1. \tag{A75}$$

where (i) holds due to the triangular inequality, and (ii) is a result of conditions $\|\boldsymbol{\mathcal{T}}\|_\Pi^2 \le t$ and $\boldsymbol{\mathcal{B}} \in \mathbf{B}(\delta_1,\delta_2)$. Since Assumption 1.II indicates $\|\boldsymbol{\mathcal{A}}\|_\Pi^2 = |\boldsymbol{\Theta}_s^\perp|^{-1}\|\boldsymbol{\mathcal{A}}_{\boldsymbol{\Theta}_s^\perp}\|_{\mathrm{F}}^2$, combing Equations (A74) and (A75) further yields an upper bound on $\|\boldsymbol{\mathcal{A}}\|_\star$ as follows:

$$\|\boldsymbol{\mathcal{A}}\|_\star \le \sqrt{rd_3|\boldsymbol{\Theta}_s^\perp|}(\sqrt{t}+\delta_1) + \kappa$$

which further leads to

$$8\mathbb{E}\left[\sup_{\boldsymbol{\mathcal{A}}} \|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}\|\boldsymbol{\mathcal{A}}\|_\star\right] \le 8\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}](\sqrt{rd_3|\boldsymbol{\Theta}_s^\perp|}(\sqrt{t}+\delta_1)+\kappa) \tag{A76}$$

The application of $2\sqrt{ab} \le a/c + bc$ is also used to further relax the above inequality:

$$\begin{aligned} 8\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}](\sqrt{rd_3|\boldsymbol{\Theta}_s^\perp|}\sqrt{t} &\le \frac{t}{4\alpha} + 64\alpha rd_3|\boldsymbol{\Theta}_s^\perp|\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}]^2 \\ 8\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}](\sqrt{rd_3|\boldsymbol{\Theta}_s^\perp|}\delta_1 &\le 4\delta_1^2 + 4rd_3|\boldsymbol{\Theta}_s^\perp|\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}]^2 \end{aligned} \tag{A77}$$

Thus, putting things in Equations (A72), (A76) and (A77) together, we obtain

$$\mathbb{E}[Z_t] \le \frac{t}{4\alpha} + \underbrace{4(16\alpha+1)rd_3|\boldsymbol{\Theta}_s^\perp|\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}]^2 + 8\kappa\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_{\mathrm{sp}}] + 8\delta_2\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] + 4\delta_1^2}_{=:\tau}$$

which further gives

$$\mathbb{P}\left[Z_t \ge \frac{t}{2\alpha} + \tau\right] \le \exp(-\frac{N_\iota t^2}{128}) \tag{A78}$$

for $\boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{A}} + \boldsymbol{\mathcal{B}}$, where $(\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{B}}) \in \mathbf{C}(r,\kappa,\beta)$, $\boldsymbol{\mathcal{B}} \in \mathbf{B}(\delta_1,\delta_2)$, and $\|\boldsymbol{\mathcal{T}}\|_\Pi^2 \le t$. □

Lemma A8. *Under Assumption 1, there exists an absolute constant* $C > 0$ *such that the following bounds on the tensor spectral norm of stochastic tensors* $\boldsymbol{\mathcal{E}}$ *and* $\boldsymbol{\mathcal{R}}_\Sigma$ *hold:*

(I) For tensor $\boldsymbol{\mathcal{E}}$*, we have*

$$\mathbb{P}\left[\|\boldsymbol{\mathcal{E}}\|_{\mathrm{sp}} \le C\sigma \max\left\{\sqrt{\frac{t+\log\tilde{d}}{\varrho N(d_1\wedge d_2)}} + \frac{\log(d_1\wedge d_2)(t+\log\tilde{d})}{N}\right\}\right] \le 1 - e^{-t} \tag{A79}$$

(II) For tensor $\boldsymbol{\mathcal{R}}_\Sigma$*, we have*

$$\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] \le \left(\sqrt{\frac{\log\tilde{d}}{N_\iota(d_1\wedge d_2)}} + \frac{\log^2\tilde{d}}{N_\iota}\right) \tag{A80}$$

Proof of Lemma A8. Equation (A79) can be seen as a special case of Lemma 8 in [18] when $k = 1$. Equation (A80) can be proved very similarly to Equation (A79) followed by tricks used in proof of Lemma 6 in [51]. We omit the details due to the high similarity. □

Lemma A9. *Under Assumption 1, there exists an absolute constant* $C > 0$ *such that the following bounds on the* l_∞*-norm of stochastic tensors* $\boldsymbol{\mathcal{E}}$*,* $\boldsymbol{\mathcal{W}}$*, and* $\boldsymbol{\mathcal{R}}_\Sigma$ *hold:*

(I) *For tensor* $\boldsymbol{\mathcal{E}}$, *we have*

$$\mathbb{P}\left[\|\boldsymbol{\mathcal{E}}\|_\infty \le C\sigma\left(\sqrt{\frac{t+\log\tilde{d}}{\varrho N d_1 d_2 d_3}} + \frac{t+\log d}{N}\right)\right] \le 1-e^{-t} \tag{A81}$$

$$\mathbb{E}[\|\boldsymbol{\mathcal{E}}\|_\infty] \le C\sigma\left(\sqrt{\frac{\log\tilde{d}}{\varrho N d_1 d_2 d_3}} + \frac{\log d}{N}\right) \tag{A82}$$

(II) *For tensor* $\boldsymbol{\mathcal{W}}$, *we have*

$$\mathbb{P}\left[\|\boldsymbol{\mathcal{W}}\|_\infty \le C\left(\frac{1}{\varrho d_1 d_2 d_3} + \sqrt{\frac{t+\log\tilde{d}}{\varrho N d_1 d_2 d_3}} + \frac{t+\log\tilde{d}}{N}\right)\right] \le 1-e^{-t} \tag{A83}$$

$$\mathbb{E}[\|\boldsymbol{\mathcal{W}}\|_\infty] \le C\left(\frac{1}{\varrho d_1 d_2 d_3} + \sqrt{\frac{\log\tilde{d}}{\varrho N d_1 d_2 d_3}} + \frac{\log\tilde{d}}{N}\right) \tag{A84}$$

(III) *For tensor* $\boldsymbol{\mathcal{R}}_\Sigma$, *we have*

$$\mathbb{P}\left[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty \le C\left(\sqrt{\frac{t+\log\tilde{d}}{N_\iota d_1 d_2 d_3}} + \frac{t+\log d}{N_\iota}\right)\right] \le 1-e^{-t} \tag{A85}$$

$$\mathbb{E}[\|\boldsymbol{\mathcal{R}}_\Sigma\|_\infty] \le C\sigma\left(\sqrt{\frac{\log\tilde{d}}{N_\iota d_1 d_2 d_3}} + \frac{\log\tilde{d}}{N_\iota}\right) \tag{A86}$$

Proof of Lemma A9. Since this lemma can be straightforwardly proved in the same way as Lemma 10 in [36], we omit the proof. □

Appendix A.3. Proof of Theorem 4

The proof of Theorem 4 follows those of Theorems 2 and 3 in [36] for robust matrix completion. Given $\boldsymbol{\mathcal{L}}^*$ and $\boldsymbol{\mathcal{S}}^*$, let $\mathbb{P}_{(\boldsymbol{\mathcal{L}}^*,\boldsymbol{\mathcal{S}}^*)}[\cdot]$ be the probability with respect to the random design tensors $\{\boldsymbol{\mathcal{X}}_i\}$ and random noises $\{\xi_i\}$ according to the observation model in Equation (6). Without loss of generality, we assume $d_1 \ge d_2$.

Proof of Theorem 4. For element-wisely sparse $\boldsymbol{\mathcal{S}}^*$, we first construct a set $\mathsf{L} \subset \mathsf{L}(r,a)$ which satisfies the following conditions:

(i) For any tensor $\boldsymbol{\mathcal{T}}$ in L, we have $r_{\text{tb}}(\boldsymbol{\mathcal{T}}) \le r$;
(ii) For any two tensors $\boldsymbol{\mathcal{T}}_1, \boldsymbol{\mathcal{T}}_2$ in L, we have $r_{\text{tb}}(\boldsymbol{\mathcal{T}}_1 - \boldsymbol{\mathcal{T}}_2) \le r$;
(iii) For any tensor $\boldsymbol{\mathcal{T}} = (\boldsymbol{\mathcal{T}}_{ijk})$ in L, any of its entries $\boldsymbol{\mathcal{T}}_{ijk}$ are in $\{0,\alpha\}$, where $\alpha \le a$,

Motivated by the proof of Theorem 2 in [36], L is constructed as follows

$$\mathsf{L} := \left\{\boldsymbol{\mathcal{L}} \in \mathbb{R}^{d_1\times d_2\times d_3} : \forall k \in [d_3],\ \boldsymbol{\mathcal{L}}^{(k)} = (\mathbf{M}_k|\cdots|\mathbf{M}_k|\mathbf{0}) \in \mathbb{R}^{d_1\times d_2}, \text{where } \mathbf{M}_k \in \{0,\alpha\}^{d_1\times r}\right\}, \tag{A87}$$

where $\mathbf{0} \in \mathbb{R}^{d_1\times(d_2 - r\lfloor\frac{d_2}{2r}\rfloor)}$ is the zero matrix, and $\alpha = \gamma(a\wedge\sigma)\sqrt{rd_1d_3/N_\iota}$ with $\gamma \le 1$ being a small enough constant such that $\alpha \le a$.

Next, according to the Varshamov-Gilbert lemma (see Lemma 2.9 in [67]), there exists a set $\mathsf{L}_0 \subset \mathsf{L}$ containing the zero tensor $\mathbf{0} \in \mathbb{R}^{d_1\times d_2\times d_3}$, such that

(i) its cardinality $|\mathsf{L}_0| \ge 2^{rd_1d_3/8} + 1$, and

(ii) for any two distinct tensors $\boldsymbol{\mathcal{T}}_1$ and $\boldsymbol{\mathcal{T}}_2$ in $\mathbf{L}_0$,

$$\|\boldsymbol{\mathcal{T}}_1 - \boldsymbol{\mathcal{T}}_2\|_{\mathrm{F}}^2 \geq \frac{d_1 d_3 r}{8} \cdot \gamma^2 (a \wedge \sigma)^2 \frac{r d_1 d_3}{N_\iota} \cdot \lfloor \frac{d_2}{2r} \rfloor \geq \frac{\gamma^2 d_1 d_2 d_3}{16} \cdot \frac{(\sigma \wedge a)^2 r d_1 d_3}{N_\iota}. \tag{A88}$$

Let $\mathbb{P}_{(\boldsymbol{\mathcal{L}},\mathbf{0})}$ denote the probabilistic distribution of random variables $\{y_i\}$ observed when the underlying tensor is $(\boldsymbol{\mathcal{L}},\mathbf{0})$ in the observation model (6). Note that, the distribution of the random noise $\xi_i \overset{i.i.d.}{\sim} \mathcal{N}(0,\sigma^2)$. Thus, for any $\boldsymbol{\mathcal{L}} \in \mathbf{L}_0$, the KL divergence $\mathscr{K}(\mathbb{P}_{\mathbf{0},\mathbf{0}}, \mathbb{P}_{(\boldsymbol{\mathcal{L}},\mathbf{0})})$ between $\mathbb{P}_{(\mathbf{0},\mathbf{0})}$ and $\mathbb{P}_{(\boldsymbol{\mathcal{L}},\mathbf{0})}$ satisfies

$$\mathscr{K}(\mathbb{P}_{(\mathbf{0},\mathbf{0})}, \mathbb{P}_{(\boldsymbol{\mathcal{L}},\mathbf{0})}) = \frac{|\Omega_\iota|}{2\sigma^2} \|\boldsymbol{\mathcal{L}}\|_{\Pi}^2 \leq \frac{|\Omega_\iota|}{2\sigma^2} \gamma^2 (a \wedge \sigma)^2 \frac{r d_1 d_3}{N_\iota} \leq \frac{\gamma^2 r d_1 d_3}{2}. \tag{A89}$$

Hence, if we choose $\gamma \in (0, \sqrt{b \log 2}/2]$, then it holds that

$$\frac{1}{|\mathbf{T}_0| - 1} \sum_{\boldsymbol{\mathcal{L}} \in \mathbf{L}_0} \mathscr{K}(\mathbb{P}_{\mathbf{0}}, \mathbb{P}_{\boldsymbol{\mathcal{L}}}) \leq b \log(|\mathbf{L}_0| - 1), \tag{A90}$$

for any $b \in (0, 1/8)$.

According to Theorem 2.5 in [67], using Equations (A88) and (A90), there exists a constant $c > 0$, such that

$$\inf_{(\hat{\boldsymbol{\mathcal{L}}},\hat{\boldsymbol{\mathcal{S}}})} \sup_{\boldsymbol{\mathcal{L}}^* \in \mathbf{L}\{r,a\}} \mathbb{P}_{(\boldsymbol{\mathcal{L}}^*,\mathbf{0})} \left[\frac{\|\hat{\boldsymbol{\mathcal{L}}} - \boldsymbol{\mathcal{L}}^*\|_{\mathrm{F}}^2}{d_1 d_2 d_3} > c \frac{(\sigma \wedge a)^2 r d_1 d_3}{N_\iota} \right] \geq \beta(b, r, d_1, d_2, d_3), \tag{A91}$$

where

$$\beta(b, r, d_1, d_2, d_3) = \frac{1}{1 + 2^{-r d_1 d_3/16}} \left(1 - 2b - 4\sqrt{\frac{b}{r d_1 d_3 \log 2}} \right) > 0. \tag{A92}$$

Note that b can be chosen to be arbitrarily small, then low-rank part of Theorem 4 is proved.

Then, we consider the sparse part of Theorem 4. Given a set $\Omega_{\mathrm{e}} \subset [d_1] \times [d_2 - \lfloor d_2/2 \rfloor] \times [d_3]$ with cardinality $s = |\Omega_{\mathrm{e}}| \leq (d_1 d_2 d_3)/2$, we also define $\mathbf{S}$ as follows

$$\mathbf{S} := \left\{ \boldsymbol{\mathcal{S}} = \left[\mathbf{0} | \boldsymbol{\mathcal{M}} \right], \text{where } \boldsymbol{\mathcal{M}}_{ijk} \in \left\{ \begin{array}{l} \{0, \alpha'\}, \text{if}(i,j,k) \in \Omega_{\mathrm{e}} \\ \{0\}, \quad \text{if}(i,j,k) \notin \Omega_{\mathrm{e}} \end{array} \right\} \right\}.$$

where $\mathbf{0} \in \mathbb{R}^{d_1 \times \lfloor \frac{d_2}{2} \rfloor \times d_3}$ is the zero tensor, and $\alpha' = \gamma'(\sigma \wedge a)$. Then, according to the Varshamov-Gilbert lemma (see Lemma 2.9 in [67]), there exists a set $\mathbf{S}_0 \subset \mathbf{S}$ containing the zero tensor $\mathbf{0} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$, such that: (i) its cardinality $|\mathbf{S}_0| \geq 2^{s/8} + 1$, and (ii) for any two distinct tensors $\boldsymbol{\mathcal{S}}_1$ and $\boldsymbol{\mathcal{S}}_2$ in $\mathbf{S}_0$,

$$\|\boldsymbol{\mathcal{S}}_1 - \boldsymbol{\mathcal{S}}_2\|_{\mathrm{F}}^2 \geq \frac{s \gamma'^2 (\sigma \wedge a)^2}{8}. \tag{A93}$$

Let $\mathbb{P}_{(\mathbf{0},\boldsymbol{\mathcal{S}})}$ denote the probabilistic distribution of random variables $\boldsymbol{\mathcal{Y}}$ observed when the underlying tensor is $(\mathbf{0}, \boldsymbol{\mathcal{S}})$ in the observation model (6). Thus, for any $\boldsymbol{\mathcal{S}} \in \mathbf{S}_0$, the KL divergence $\mathscr{K}(\mathbb{P}_{(\mathbf{0},\mathbf{0})}, \mathbb{P}_{(\mathbf{0},\boldsymbol{\mathcal{S}})})$ between $\mathbb{P}_{(\mathbf{0},\mathbf{0})}$ and $\mathbb{P}_{(\mathbf{0},\boldsymbol{\mathcal{S}})}$ satisfies

$$\mathscr{K}(\mathbb{P}_{(\mathbf{0},\mathbf{0})}, \mathbb{P}_{(\mathbf{0},\boldsymbol{\mathcal{S}})}) = N_s \frac{\boldsymbol{\mathcal{S}}_{ijk}^2}{2\sigma^2} \leq \frac{s \gamma'^2}{2}. \tag{A94}$$

Hence, if we choose $\gamma' \in (0, \sqrt{b' \log 2}/2]$, then it holds that

$$\frac{1}{|\mathbf{S}_0| - 1} \sum_{\boldsymbol{\mathcal{S}} \in \mathbf{S}_0} \mathscr{K}(\mathbb{P}_{(\mathbf{0},\mathbf{0})}, \mathbb{P}_{(\mathbf{0},\boldsymbol{\mathcal{S}})}) \leq b' \log(|\mathbf{S}_0| - 1), \tag{A95}$$

for any $b' \in (0, 1/8)$. According to Theorem 2.5 in [67], using Equations (A93) and (A95), there exists a constant $c' > 0$, such that

$$\inf_{(\hat{\mathcal{L}},\hat{\mathcal{S}})} \sup_{\mathcal{S}^*\in\mathbf{S}} \mathbb{P}_{(\mathbf{0},\mathcal{S}^*)}\left[\frac{\|\hat{\mathcal{S}}-\mathcal{S}^*\|_{\mathrm{F}}^2}{d_1d_2d_3} > c'\frac{(\sigma\wedge a)^2 s}{d_1d_2d_3}\right] \geq \beta'(b', s), \tag{A96}$$

where

$$\beta'(b', s) = \frac{1}{1+2^{-s/8}}\left(1 - 2b' - 4\sqrt{\frac{b'}{s\log 2}}\right) > 0. \tag{A97}$$

Note that b' can be chosen to be arbitrarily small, then sparse part of Theorem 4 is proved.

Thus, according to Equations (A91) and (A96), by setting

$$c_1' = \frac{c}{2},\ c_1'' = \frac{c'}{2},\ \text{and}\ \ \beta_1 = \min\left\{\beta(b, r, d_1, d_2, d_3),\ \beta'(b', s)\right\}, \tag{A98}$$

the following relationship holds

$$\inf_{(\hat{\mathcal{L}},\hat{\mathcal{S}})} \sup_{\substack{(\mathcal{L}^*,\mathcal{S}^*)\\ \in\mathbf{A}(r,s,a)}} \mathbb{P}_{(\mathcal{L}^*,\mathcal{S}^*)}\left[\frac{\|\mathbf{\Delta}^{\iota}\|_{\mathrm{F}}^2 + \|\mathbf{\Delta}^{s}\|_{\mathrm{F}}^2}{d_1d_2d_3} \geq \phi_{\mathrm{e}}\right] \geq \beta_1, \tag{A99}$$

where $\phi := (\sigma\wedge a)^2\big(c_1' r d_1 d_3/N_\iota + c_1'' s/(d_1d_2d_3)\big)$. Then according to Markov inequality, we obtain

$$\mathscr{M}(\mathbf{A}(r, s, a)) \geq \beta_1\phi. \tag{A100}$$

□

References

1. He, W.; Yokoya, N.; Yuan, L.; Zhao, Q. Remote sensing image reconstruction using tensor ring completion and total variation. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8998–9009. [CrossRef]
2. He, W.; Yao, Q.; Li, C.; Yokoya, N.; Zhao, Q.; Zhang, H.; Zhang, L. Non-local meets global: An integrated paradigm for hyperspectral image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef] [PubMed]
3. Davis, J.W.; Sharma, V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182. [CrossRef]
4. Bello, S.A.; Yu, S.; Wang, C.; Adam, J.M.; Li, J. Deep learning on 3D point clouds. *Remote Sens.* **2020**, *12*, 1729. [CrossRef]
5. Zheng, Y.B.; Huang, T.Z.; Zhao, X.L.; Chen, Y.; He, W. Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8450–8464. [CrossRef]
6. Liu, H.K.; Zhang, L.; Huang, H. Small target detection in infrared videos based on spatio-temporal tensor model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8689–8700. [CrossRef]
7. Zhou, A.; Xie, W.; Pei, J. Background modeling combined with multiple features in the Fourier domain for maritime infrared target detection. *IEEE Trans. Geosci. Remote. Sens.* **2021**. [CrossRef]
8. Jiang, Q.; Ng, M. Robust low-tubal-rank tensor completion via convex optimization. In Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; AAAI Press: Palo Alto, CA, USA, 2019; pp. 2649–2655.
9. Zhao, Q.; Zhou, G.; Zhang, L.; Cichocki, A.; Amari, S.I. Bayesian robust tensor factorization for incomplete multiway data. *IEEE Trans. Neural Networks Learn. Syst.* **2016**, *27*, 736–748. [CrossRef]
10. Liu, H.; Li, H.; Wu, Z.; Wei, Z. Hyperspectral image recovery using non-convex low-rank tensor approximation. *Remote Sens.* **2020**, *12*, 2264. [CrossRef]
11. Ma, T.H.; Xu, Z.; Meng, D. Remote sensing image denoising via low-rank tensor approximation and robust noise modeling. *Remote Sens.* **2020**, *12*, 1278. [CrossRef]
12. Fazel, M. Matrix Rank Minimization with Applications. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
13. Liu, J.; Musialski, P.; Wonka, P.; Ye, J. Tensor completion for estimating missing values in visual data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 208–220. [CrossRef]
14. Carroll, J.D.; Chang, J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Youn" decomposition. *Psychometrika* **1970**, *35*, 283–319. [CrossRef]
15. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311. [CrossRef] [PubMed]
16. Oseledets, I. Tensor-train decomposition. *SIAM J. Sci. Comput.* **2011**, *33*, 2295–2317. [CrossRef]
17. Zhao, Q.; Zhou, G.; Xie, S.; Zhang, L.; Cichocki, A. Tensor ring decomposition. *arXiv* **2016**, arXiv:1606.05535.

18. Wang, A.; Zhou, G.; Jin, Z.; Zhao, Q. Tensor recovery via $*_L$-spectral k-support norm. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 522–534. [CrossRef]
19. Wang, A.; Li, C.; Jin, Z.; Zhao, Q. Robust tensor decomposition via orientation invariant tubal nuclear norms. In Proceedings of the The AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 6102–6109.
20. Zhang, Z.; Ely, G.; Aeron, S.; Hao, N.; Kilmer, M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 3842–3849.
21. Liu, X.; Aeron, S.; Aggarwal, V.; Wang, X. Low-tubal-rank tensor completion using alternating minimization. *IEEE Trans. Inf. Theory* **2020**, *66*, 1714–1737. [CrossRef]
22. Kilmer, M.E.; Braman, K.; Hao, N.; Hoover, R.C. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM J. Matrix Anal. A* **2013**, *34*, 148–172. [CrossRef]
23. Liu, X.Y.; Wang, X. Fourth-order tensors with multidimensional discrete transforms. *arXiv* **2017**, arXiv:1705.01576.
24. Kernfeld, E.; Kilmer, M.; Aeron, S. Tensor–tensor products with invertible linear transforms. *Linear Algebra Its Appl.* **2015**, *485*, 545–570. [CrossRef]
25. Zhang, X.; Ng, M.K.P. Low rank tensor completion with poisson observations. *IEEE Trans. Pattern Anal. Mach. Intell.*. **2021**. [CrossRef]
26. Lu, C.; Peng, X.; Wei, Y. Low-rank tensor completion with a new tensor nuclear norm induced by invertible linear transforms. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 5996–6004.
27. Song, G.; Ng, M.K.; Zhang, X. Robust tensor completion using transformed tensor singular value decomposition. *Numer. Linear Algebr.* **2020**, *27*, e2299. [CrossRef]
28. He, B.; Yuan, X. On the O(1/n) convergence rate of the Douglas–Rachford alternating direction method. *SIAM J. Numer. Anal.* **2012**, *50*, 700–709. [CrossRef]
29. Parikh, N.; Boyd, S. Proximal algorithms. *Found. Trends® Optim.* **2014**, *1*, 127–239. [CrossRef]
30. Kolda, T.G.; Bader, B.W. Tensor decompositions and applications. *SIAM Rev.* **2009**, *51*, 455–500. [CrossRef]
31. Kong, H.; Lu, C.; Lin, Z. Tensor Q-rank: New data dependent definition of tensor rank. *Mach. Learn.* **2021**, *110*, 1867–1900. [CrossRef]
32. Lu, C.; Zhou, P. Exact recovery of tensor robust principal component analysis under linear transforms. *arXiv* **2019**, arXiv:1907.08288.
33. Zhang, Z.; Aeron, S. Exact tensor completion using t-SVD. *IEEE Trans. Signal Process.* **2017**, *65*, 1511–1526. [CrossRef]
34. Wang, A.; Jin, Z.; Tang, G. Robust tensor decomposition via t-SVD: Near-optimal statistical guarantee and scalable algorithms. *Signal Process.* **2020**, *167*, 107319. [CrossRef]
35. Zhou, P.; Feng, J. Outlier-robust tensor PCA. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
36. Klopp, O.; Lounici, K.; Tsybakov, A.B. Robust matrix completion. *Probab. Theory Relat. Fields* **2017**, *169*, 523–564. [CrossRef]
37. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5249–5257.
38. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM (JACM)* **2011**, *58*, 11. [CrossRef]
39. Negahban, S.; Wainwright, M.J. Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Ann. Stat.* **2011**, *39*, 1069–1097. [CrossRef]
40. Wang, A.; Wei, D.; Wang, B.; Jin, Z. Noisy low-tubal-rank tensor completion through iterative singular tube thresholding. *IEEE Access* **2018**, *6*, 35112–35128. [CrossRef]
41. Boyd, S.; Parikh, N.; Chu, E. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
42. Wang, A.; Jin, Z. Near-optimal noisy low-tubal-rank tensor completion via singular tube thresholding. In Proceedings of the IEEE International Conference on Data Mining Workshop (ICDMW), New Orleans, LA, USA, 18–21 November 2017; pp. 553–560.
43. Wang, A.; Lai, Z.; Jin, Z. Noisy low-tubal-rank tensor completion. *Neurocomputing* **2019**, *330*, 267–279. [CrossRef]
44. Wang, A.; Song, X.; Wu, X.; Lai, Z.; Jin, Z. Generalized Dantzig selector for low-tubal-rank tensor recovery. In Proceedings of the The International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3427–3431.
45. Huang, B.; Mu, C.; Goldfarb, D.; Wright, J. Provable models for robust low-rank tensor completion. *Pac. J. Optim.* **2015**, *11*, 339–364.
46. Wang, A.; Song, X.; Wu, X.; Lai, Z.; Jin, Z. Robust low-tubal-rank tensor completion. In Proceedings of the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 3432–3436.
47. Fang, W.; Wei, D.; Zhang, R. Stable tensor principal component pursuit: Error bounds and efficient algorithms. *Sensors* **2019**, *19*, 5335. [CrossRef]
48. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

49. Chen, J.; Wang, C.; Ma, Z.; Chen, J.; He, D.; Ackland, S. Remote sensing scene classification based on convolutional neural networks pre-trained using attention-guided sparse filters. *Remote Sens.* **2018**, *10*, 290. [CrossRef]
50. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
51. Klopp, O. Noisy low-rank matrix completion with general sampling distribution. *Bernoulli* **2014**, *20*, 282–303. [CrossRef]
52. Li, N.; Zhou, D.; Shi, J.; Wu, T.; Gong, M. Spectral-locational-spatial manifold learning for hyperspectral images dimensionality reduction. *Remote Sens.* **2021**, *13*, 2752. [CrossRef]
53. Mayalu, A.; Kochersberger, K.; Jenkins, B.; Malassenet, F. Lidar data reduction for unmanned systems navigation in urban canyon. *Remote Sens.* **2020**, *12*, 1724. [CrossRef]
54. Hwang, Y.S.; Schlüter, S.; Park, S.I.; Um, J.S. Comparative evaluation of mapping accuracy between UAV video versus photo mosaic for the scattered urban photovoltaic panel. *Remote Sens.* **2021**, *13*, 2745. [CrossRef]
55. Lou, J.; Zhu, W.; Wang, H.; Ren, M. Small target detection combining regional stability and saliency in a color image. *Multimed. Tools Appl.* **2017**, *76*, 14781–14798. [CrossRef]
56. Hui, B.; Song, Z.; Fan, H. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background. *China Sci. Data* **2020**, *5*, 291–302.
57. Wang, Z.; Zeng, Q.; Jiao, J. An adaptive decomposition approach with dipole aggregation model for polarimetric SAR data. *Remote Sens.* **2021**, *13*, 2583. [CrossRef]
58. Wei, D.; Wang, A.; Feng, X.; Wang, B.; Wang, B. Tensor completion based on triple tubal nuclear norm. *Algorithms* **2018**, *11*, 94. [CrossRef]
59. Han, X.; Wu, B.; Shou, Z.; Liu, X.Y.; Zhang, Y.; Kong, L. Tensor FISTA-net for real-time snapshot compressive imaging. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 10933–10940.
60. Mu, C.; Zhang, Y.; Wright, J.; Goldfarb, D. Scalable robust matrix recovery: Frank–Wolfe meets proximal methods. *SIAM J. Sci. Comput.* **2016**, *38*, A3291–A3317. [CrossRef]
61. Wang, A.; Jin, Z.; Yang, J. A faster tensor robust PCA via tensor factorization. *Int. J. Mach. Learn. Cybern.* **2020**, *11*, 2771–2791. [CrossRef]
62. Lou, J.; Cheung, Y. Robust Low-Rank Tensor Minimization via a New Tensor Spectral k-Support Norm. *IEEE TIP* **2019**, *29*, 2314–2327. [CrossRef] [PubMed]
63. Negahban, S.; Yu, B.; Wainwright, M.J.; Ravikumar, P.K. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. In Proceedings of Advances in Neural Information Processing Systems, Vancouver, BC, USA, 7–10 December 2009; pp. 1348–1356.
64. Bühlmann, P.; Van De Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011.
65. Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*; Cambridge University Press: Cambridge, UK, 2018; Volume 47.
66. Talagrand, M. A new look at independence. *Ann. Probab.* **1996**, *24*, 1–34. [CrossRef]
67. Tsybakov, A.B. *Introduction to Nonparametric Estimation*; Springer: New York, NY, USA, 2011.