Robust Tensor Completion: Equivalent Surrogates, Error Bounds and Algorithms*

Xueying Zhao†, Minru Bai‡, Defeng Sun§, and Libin Zheng¶

Abstract. Robust Low-Rank Tensor Completion (RTC) problems have received considerable attention in recent years such as signal processing and computer vision. In this paper, we focus on the bound constrained RTC problem for third-order tensors which recovers a low-rank tensor from partial observations corrupted by impulse noise. A widely used convex relaxation of this problem is to minimize the tensor nuclear norm for low rank and the ℓ_1-norm for sparsity. However, it may result in biased solutions. To handle this issue, we propose a nonconvex model with a novel nonconvex tensor rank surrogate function and a novel nonconvex sparsity measure for RTC problems under limited sample constraints and two bound constraints, where these two nonconvex terms have a difference of convex functions (DC) structure. Then, a proximal majorization-minimization (PMM) algorithm is developed to solve the proposed model and this algorithm consists of solving a series of convex subproblems with an initial estimator to generate a new estimator which is used for the next subproblem. Theoretically, for this new estimator, we establish a recovery error bound for its recoverability and give the theoretical guarantee that lower error bounds can be obtained when a reasonable initial estimator is available. Then, by using the Kurdyka-Łojasiewicz property exhibited in the resulting problem, we show that the sequence generated by the PMM algorithm globally converges to a critical point of the problem. Extensive numerical experiments including color images and multispectral images show the high efficiency of the proposed model.

Key words. robust low-rank tensor completion, DC equivalent surrogates, proximal majorization-minimization, error bounds, impulse noise

AMS subject classifications. 15A69, 68U10, 90C26

1. Introduction.

Multi-dimensional data is becoming prevalent in many areas such as computer vision [27, 44], data mining [32], signal processing [10], and machine learning [39]. Tensor-based modeling has the capability of capturing these underlying multi-dimensional structures. However, the tensor data observed may suffer from information loss and be perturbed by different kinds of noise originating from human errors or signal interference. The purpose of this paper is to study Robust Low-Rank Tensor Completion (RTC) problems for third-order tensors, in which few available entries are defiled by impulse noise.

The original model of RTC problems is to minimize an optimization problem which consists of the tensor rank function plus the ℓ_0-norm under limited sample constraints, which is a generalization of Robust Matrix Completion (RMC) [8, 22]. As the rank function is

*Submitted to the editors DATE.

Funding: The work of the second author was supported in part by the National Natural Science Foundation of China under grant 11971159, Hunan Provincial Key Laboratory of Intelligent information processing and Applied Mathematics, Changsha 410082, PR China. The work of the third author was supported in part by the Hong Kong Research Grant Council under grant 15304019 .

†School of Mathematics, Hunan University, Changsha, Hunan 410082, China (xueying_zhao@hnu.edu.cn).

‡School of Mathematics, Hunan University, Changsha, Hunan 410082, China (minru-bai@hnu.edu.cn).

§Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (defeng.sun@polyu.edu.hk).

¶School of Mathematics, Hunan University, Changsha, Hunan 410082, China (lszt301@163.com).

nonconvex, the nuclear norm is widely used to approximate the rank function. Candès et al. [8] studied the RMC problem by solving a convex optimization problem that minimizes a weighted combination of the nuclear norm and the ℓ_1-norm under limited sample constraints, and theoretical conditions to ensure the perfect recovery in the probabilistic sense have been analyzed. Although the nuclear norm is a convex relaxation of the rank function, this kind of surrogate may make the solution seriously deviate from the solution of rank minimization. To improve the recovery quality of the solution for matrix completion with fixed basis coefficient, Miao et al. [31] proposed a rank-corrected procedure to generate an estimator with a pre-estimator and established a non-asymptotic recovery error bound. Liu et al. [28] recently reformulated the rank regularized problem as a family of nonconvex equivalent surrogates by establishing its global exact penalty.

Compared with RMC, RTC is more difficult to solve due to the fact that the rank of a tensor is not unique. The two commonly used tensor ranks are the CANDECOMP/PARAFAC (CP) rank [9] and the Tucker rank [43]. However, computing the CP rank of a given tensor is known to be NP-hard [16]. Liu et al. [27] proposed the sum of nuclear norms of unfolding matrices (SNN) of a tensor to approximate the Tucker rank to solve the low-rank tensor completion problem, which has since appeared frequently in practical settings. Although the SNN is easy to compute, Romera-Paredes et al. [36] showed that it is not the tightest convex envelope of the sum of entries of the Tucker rank. Recently, Huang et al. [17] proposed a tensor ring (TR) decomposition that factorizes a high-order tensor into a sequence of three-order tensors and used a number of TR unfoldings for RTC problems. However, the matricization of a tensor may break the intrinsic structures and correlations in the tensor data, hence the rank defined by the unfolding matrices cannot accurately describe the low-rank property of the tensor. Different from the rank based matricization above, Kilmer et al. [19] proposed the tensor multi-rank and tubal rank definitions based on a tensor singular value decomposition (t-SVD) framework [20] and Semerci et al. [37] developed a new tubal nuclear norm (TNN), which is a convex surrogate of the multi-rank [57]. In recent years, the tubal rank and the TNN have been widely studied for tensor recovery problems [18, 29, 45, 55]. Jiang et al. [18] showed that one can recover a low tubal rank tensor exactly with overwhelming probability by solving a convex program, where the objective function is a weighted combination of the TNN and the ℓ_1-norm. However, as pointed out in [38], the low-rank property of most natural images is mainly affected by a few large singular values, which present a heavy-tailed distribution. It means that the larger singular values are expected to be penalized mildly while the smaller ones are penalized severely. Nevertheless, the TNN treats the singular values with the same penalty, which will over-penalize large singular values and hence get the suboptimal performance. To address this issue, Zhang et al. [55] proposed a corrected TNN (CTNN) model for third-order tensor recovery from partial observations corrupted by Gaussian noise based on the rank-corrected procedure [31] and provided a non-asymptotic error bound of the CTNN model. However, [55] is not able to address the observations with impulse noise and the outer loop convergence of the adaptive correction procedure is unknown.

On the other hand, it is challenging to solve the ℓ_0 regularization problem since it is NP-hard [33]. As a convex relaxation of the ℓ_0-norm, the ℓ_1-norm has been widely used for sparsity in statistics. The least absolute shrinkage and selection operator (lasso) problem is the ℓ_1-norm penalized least squares method, which was proposed in [42] and has been used extensively in high-dimensional statistics and machine learning. However, as indicated by [12], the ℓ_1-norm has long been known by statisticians to yield biased estimators and cannot achieve the best estimation performance, and might not be statistically optimal in more challenging scenarios. Hence, to solve the above mentioned problems, some nonconvex penalties have been proposed to substitute sparsity measures [13, 14, 41, 50, 51, 58]. In [41], a sparse semismooth Newton based proximal majorization-minimization (PMM) algorithm for nonconvex square-root-loss regression problems was introduced where the nonconvex regularizer has the difference of convex functions (DC) structure. Ahn et al. [1] gave a unified DC representation for a family of surrogate sparsity functions that are employed as approximations of the ℓ_0-norm in statistical learning and established some sparsity properties of the directional stationary points. Yang et al. [51] proposed nonconvex models for RTC by the regularizing redescending M-estimators as sparsity measures and developed the linearized and proximal block coordinate methods to solve the nonconvex problems. Zhao et al. [58] studied a nonconvex model, consisting of the data-fitting term combined with the TNN and the nonconvex data fidelity term, for RTC problems and presented a Gauss-Seidel DC algorithm (GS-DCA) to solve the resulting optimization. By numerical experiments, [51] and [58] all showed that these nonconvex penalties outperformed the ℓ_1-norm penalty. Actually, the TNN is the sum of nuclear norms of all frontal slides of the tensor in the Fourier domain, which is the ℓ_1-norm of all singular vectors. In other words, the TNN results in a biased estimator as well as the ℓ_1-norm does. Therefore, some works [26, 49, 50, 54] proposed nonconvex penalties to replace the ℓ_1-norm in TNN. For example, Li et al. [26] established a nonconvex ℓ_p-norm relaxation model for low Tucker rank tensor recovery problem, which can recover the data in lower sampling ratios compared to the convex nuclear norm relaxation model, and the alternating direction method of multipliers (ADMM) was used to solve the resulting model. Xu et al. [49] proposed a novel nonconvex surrogate for the tensor multi-rank based on the Laplace function, which can more tightly approximate to the ℓ_0-norm than the tensor nuclear norm. However, there are few works on the mechanism to produce equivalent surrogates for the rank and the zero-norm optimization problems, although much research has been considering the nonconvex surrogates. What's more, prior studies mentioned above only focused on the algorithm and its convergence analysis, but statistical error bounds of obtained solutions were rarely discussed.

With an eye toward statistical performance, some researchers have studied the error bound for various models. Wu [48] proposed a two-stage rank-sparsity-correction procedure to deal with the problem of noisy low-rank and sparse matrix decomposition by adding adaptive rank-correction terms designed in [31], and examined its recovery performance by developing an error bound. However, [48] did not establish any theoretical guarantee that the recovery error bound obtained by the corrected model is smaller than that of the model without correction terms. Furthermore, it is difficult to generalize the error bound to tensor cases directly. In the tensor algebra framework, Bai et al. [4] proposed an adaptive correction approach for higher-order tensor completion and showed that the correction term with a suitable estimator could reduce the error bound of the corrected model, while the corrected model mainly deals with data missing problems without noises. In order to derive solutions with higher accuracy, zhang et al. [55] presented the CTNN model for low-rank tensor recovery and provided a non-asymptotic error bound, but this model could not address the sparse outliers.

To address the above problems, in this paper, we not only pay attention to nonconvex surrogates of the rank function and the ℓ_0-norm to overcome biased estimators yielded by the ℓ_1-norm penalty and the TNN penalty, but also study the statistical performance analysis of our method by establishing the recovery error bounds. We propose a bound constrained Nonconvex Robust Tensor Completion (BCNRTC) model which aims to recover a third-order tensor corrupted by impulse noise with partial observations. The proposed model consists of two nonconvex regularization terms with the DC structure for low-rank and sparsity under limited sample constraints and two bound constraints. These two nonconvex penalties can be chosen as the minimax concave penalty (MCP) function, the smoothly clipped absolute deviation (SCAD) function since such functions are continuous, sparsity promoting, and nearly unbiased [12, 52]. In addition, we prove the equivalence of global solutions between the bound constrained RTC problems and our proposed nonconvex model in theory. Recently, some works [6, 15, 40, 46] have been proposed to solve nonconvex and nonsmooth problems. Unfortunately, these works could not be applied to solve our proposed model directly. For example, Bolte et al. [6] proposed a proximal alternating linearized minimization (PALM) algorithm to solve the nonconvex and nonsmooth problems, but no constraints were considered. Guo et al. [15] studied the convergence of ADMM for minimizing the sum of two nonconvex functions with linear constraints, however, one of the nonconvex functions was required to be differentiable. [46] analyzed the convergence of ADMM for minimizing a nonconvex problem with coupled linear equality constraints, but the objective functions also needed to be Lipschitz differentiable. Therefore, for the proposed nonconvex and nonsmooth model, we design a proximal majorization-minimization (PMM) algorithm similar to [24, 41, 53] to solve it. The key idea of the PMM algorithm is to solve a series of convex subproblems with an initial estimator to generate a new estimator which is used for the next subproblem. Specifically, each subproblem solves a convex program which is to minimize a weighted combination of the TNN and the ℓ_1-norm minus two linear terms, where the linear terms can be seen as the rank-correction term and sparsity-correction term constructed on the initial estimator. Meanwhile, we establish the recovery error bound between new estimators and initial estimators and also discuss the impact of the correction term on recovery error. Compared with the one obtained without these two linear terms, the error bound has a certain degree of reduction. Finally, the convergence of the PMM algorithm is established by using the Kurdyka-Łojasiewicz property and extensive numerical experiments are presented to demonstrate the efficiency of the proposed BCNRTC model. Therefore, our work not only improves the tensor rank surrogate function but also modifies the tensor sparsity measure.

The main contributions of this paper are four aspects.

- We produce and prove equivalent nonconvex surrogates with DC structures in the sense that they have the same global optimal solution set as RTC problems with the tensor average rank and the ℓ_0-norm do. We also show that these equivalent surrogates include the popular MCP function and SCAD function in statistics as special cases.
- A proximal majorization-minimization (PMM) algorithm with convergence analysis is presented to solve the BCNRTC model, which is a nonconvex optimization problem with linear constraints and bound constraints. Each subproblem of the PMM algorithm is to solve a convex program, where the two linear terms obtained by majorization can be seen as the tensor rank-correction term and the sparsity-correction

term constructed on the initial estimator.

- We establish a non-asymptotic recovery error bound for the subproblem of the PMM algorithm, which gives the theoretical guarantee that under the mild condition the subproblem of the PMM algorithm can reduce recovery error bounds. Our results of recovery error bounds also suggest a criterion for constructing a suitable rank-correction function and a sparsity-correction function. We show that rank-correction functions and sparsity-correction functions constructed by the MCP function and SCAD function satisfy the above criterion.
- Numerically, we confirm that the error bounds decrease as the number of outer iterations increases. Moreover, extensive numerical experiments on color images and multispectral images demonstrate the superiority of the proposed model over several existing methods.

The rest of this paper is organized as follows. Some notations used throughout this paper are introduced in Section 2. The bound constrained Nonconvex Robust Tensor Completion (BCNRTC) model is proposed in Section 3. The PMM algorithm is presented to solve the resulting model and its global convergence is also established in Section 4. In Section 5, we establish a recovery error bound for the estimator generated from the PMM algorithm. Finally, we report numerical results to validate the efficiency of our proposed model in Section 6 and draw conclusions in Section 7.

2. Preliminaries.

Throughout this paper, tensors are denoted by Euler script letters, e.g., $\mathcal{X}$. Matrices are denoted by boldface capital letters, e.g., $\boldsymbol{X}$. Vectors are denoted by bold lowercase letters, e.g., $\boldsymbol{x}$, and scalars are denoted by ordinary letters, e.g., x. The fields of real numbers and complex numbers are denoted as $\mathbb{R}$ and $\mathbb{C}$, respectively. For a third-order tensor $\mathcal{X} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, we denote its (i,j,k)-th entry as $\mathcal{X}_{ijk}$. A slice of a tensor $\mathcal{X}$ is a matrix defined by fixing all indices but two. We use the notation $\mathcal{X}(i,:,:), \mathcal{X}(:,i,:)$ and $\mathcal{X}(:,:,i)$ to denote the i-th horizontal, lateral and frontal slice, respectively. Specifically, the front slice $\mathcal{X}(:,:,i)$ is also denoted by $\boldsymbol{X}^{(i)}$. A fiber of a tensor $\mathcal{X}$ is a vector defined by fixing all indices but one. The fiber along the third dimension $\mathcal{X}(i,j,:)$ is also called as the (i,j)-th tube of $\mathcal{X}$. We denote $\lfloor t \rfloor$ as the nearest integer less than or equal to t and $\lceil t \rceil$ as the one greater than or equal to t.

For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\pi(\mathcal{X}) \in \mathbb{R}^{n_1 n_2 n_3}$ means the vector obtained by arranging the entries of $|\mathcal{X}|$ in a non-increasing order, where $|\mathcal{X}|$ means the tensor whose (i,j,k)-th component is $|\mathcal{X}_{ijk}|$; and $\pi_i(\cdot)$ denotes the i-th entry of $\pi(\cdot)$. For $\boldsymbol{X} \in \mathbb{C}^{n_1 \times n_2}$, $\sigma(\boldsymbol{X})$ means the singular value vector of $\boldsymbol{X}$ with entries arranged in a non-increasing order; and $\sigma_i(\cdot)$ denotes the i-th entry of $\sigma(\cdot)$. For any given vector $\boldsymbol{x}$, $\mathrm{Diag}(\boldsymbol{x})$ denotes a rectangular diagonal matrix of suitable size with the i-th diagonal entry being x_i. For any matrix $\boldsymbol{X}$, $\mathrm{diag}(\boldsymbol{X})$ denotes a vector of suitable size with the i-th diagonal entry being x_{ii}. Denote the function $\mathrm{sign} : \mathbb{R} \to \mathbb{R}$ by $\mathrm{sign}(t) = 1$ if $t > 0$, $\mathrm{sign}(t) = -1$ if $t < 0$, and $\mathrm{sign}(t) = 0$ if $t = 0$, for $t \in \mathbb{R}$. For any $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, let $\mathrm{sign}(\mathcal{X})$ be the sign tensor of $\mathcal{X}$ where $[\mathrm{sign}(\mathcal{X})]_{ijk} = \mathrm{sign}(\mathcal{X}_{ijk})$.

The inner product of two matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\mathbb{C}^{n_1 \times n_2}$ is defined as $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle := \mathrm{Tr}(\boldsymbol{X}^H \boldsymbol{Y})$, where $\boldsymbol{X}^H$ denotes the conjugate transpose of $\boldsymbol{X}$, and $\mathrm{Tr}(\cdot)$ denotes the matrix trace. The inner product of two tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle := \sum_{i=1}^{n_3} \langle \boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)} \rangle$. The Frobenius norm of a tensor $\mathcal{X}$ is defined as $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$. And the infinity norm and the

l_1-norm of a tensor are defined as $\|\mathcal{X}\|_\infty = \max_{ijk} |\mathcal{X}_{ijk}|$ and $\|\mathcal{X}\|_1 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3} |\mathcal{X}_{ijk}|$, respectively. For any $\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times n_3}$, the complex conjugate of $\mathcal{X}$ is denoted as $\mathrm{conj}(\mathcal{X})$ which takes the complex conjugate of each entry of $\mathcal{X}$.

For any tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, we denote $\widehat{\mathcal{X}} \in \mathbb{C}^{n_1\times n_2\times n_3}$ as the results of the Fast Fourier Transform (FFT) of all tubes along the third dimension. Using MATLAB command fft, $\widehat{\mathcal{X}} = \mathrm{fft}(\mathcal{X}, [\,], 3)$. One can also compute $\mathcal{X}$ from $\widehat{\mathcal{X}}$ by using the inverse FFT operation along the third-dimension, i.e., $\mathcal{X} = \mathrm{ifft}(\widehat{\mathcal{X}}, [\,], 3)$. Let $\overline{\boldsymbol{X}}$ denote the block diagonal matrix of the tensor $\widehat{\mathcal{X}}$, where the i-th diagonal block of $\overline{\boldsymbol{X}}$ is the i-th frontal slice $\widehat{\boldsymbol{X}}^{(i)}$ of $\widehat{\mathcal{X}}$, i.e.,

$$\overline{\boldsymbol{X}} := \mathrm{bdiag}(\widehat{\mathcal{X}}) = \begin{bmatrix} \widehat{\boldsymbol{X}}^{(1)} & & & \\ & \widehat{\boldsymbol{X}}^{(2)} & & \\ & & \ddots & \\ & & & \widehat{\boldsymbol{X}}^{(n_3)} \end{bmatrix}.$$

We define a block circular matrix from the frontal slices $\boldsymbol{X}^{(i)}$ of $\mathcal{X}$ as

$$\mathrm{bcirc}(\mathcal{X}) := \begin{bmatrix} \boldsymbol{X}^{(1)} & \boldsymbol{X}^{(n_3)} & \cdots & \boldsymbol{X}^{(2)} \\ \boldsymbol{X}^{(2)} & \boldsymbol{X}^{(1)} & \cdots & \boldsymbol{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}^{(n_3)} & \boldsymbol{X}^{(n_3-1)} & \cdots & \boldsymbol{X}^{(1)} \end{bmatrix}.$$

It can be block diagonalized by using the FFT, i.e., $(\boldsymbol{F}_{n_3} \otimes \boldsymbol{I}_{n_1}) \cdot \mathrm{bcirc}(\mathcal{X}) \cdot (\boldsymbol{F}_{n_3}^{-1} \otimes \boldsymbol{I}_{n_2}) = \overline{\boldsymbol{X}}$, where $\boldsymbol{F}_n$ is the $n \times n$ discrete Fourier matrix, $\boldsymbol{I}_n$ is the $n\times n$ identity matrix, $\otimes$ denotes the Kronecker product, and $(\boldsymbol{F}_{n_3} \otimes \boldsymbol{I}_{n_1})/\sqrt{n_3}$ is unitary. The command $\mathrm{unfold}(\mathcal{X})$ takes $\mathcal{X}$ into a block $n_1n_3 \times n_2$ matrix:

$$\mathrm{unfold}(\mathcal{X}) := \begin{bmatrix} \boldsymbol{X}^{(1)} \\ \boldsymbol{X}^{(2)} \\ \vdots \\ \boldsymbol{X}^{(n_3)} \end{bmatrix}.$$

The inverse operator fold takes $\mathrm{unfold}(\mathcal{X})$ into a tensor form: $\mathrm{fold}(\mathrm{unfold}(\mathcal{X})) = \mathcal{X}$. It is showed in [29] that

$$\mathrm{conj}(\widehat{\boldsymbol{X}}^{(i)}) = \widehat{\boldsymbol{X}}^{(n_3-i+2)} \quad \forall i = 2, \ldots, \left\lfloor \frac{n_3+1}{2} \right\rfloor.$$

The tensor spectral norm of $\mathcal{X}$ is defined as $\|\mathcal{X}\| := \|\overline{\boldsymbol{X}}\|$, i.e., the spectral norm of the block diagonal matrix $\overline{\boldsymbol{X}}$ in the Fourier domain. The following properties will be used frequently: $\langle \mathcal{X}, \mathcal{Y}\rangle = \frac{1}{n_3}\langle \overline{\boldsymbol{X}}, \overline{\boldsymbol{Y}}\rangle$, $\|\mathcal{X}\|_F = \frac{1}{\sqrt{n_3}}\|\overline{\boldsymbol{X}}\|_F$.

Now we give some basic definitions about tensors, which serve as the foundation for our further analysis.

Definition 2.1 (T-product [20]). *The t-product $\mathcal{X} * \mathcal{Y}$ of $\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times n_3}$ and $\mathcal{Y} \in \mathbb{C}^{n_2\times n_4\times n_3}$ is a tensor $\mathcal{Z} \in \mathbb{C}^{n_1\times n_4\times n_3}$ given by $\mathcal{Z} = \mathrm{fold}(\mathrm{bcirc}(\mathcal{X}) \cdot \mathrm{unfold}(\mathcal{Y}))$. Moreover, we have the following equivalence: $\mathcal{X} * \mathcal{Y} = \mathcal{Z} \Leftrightarrow \overline{\boldsymbol{X}}\ \overline{\boldsymbol{Y}} = \overline{\boldsymbol{Z}}$.*

Definition 2.2 (Tensor transpose [20]). *The conjugate transpose of a tensor* $\mathcal{X} \in \mathbb{C}^{n_1\times n_2\times n_3}$ *is the tensor* $\mathcal{X}^H \in \mathbb{C}^{n_2\times n_1\times n_3}$ *obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices* 2 *through* n_3*.*

Definition 2.3 (F-diagonal tensor [20]). *A tensor* $\mathcal{X}$ *is called* f*-diagonal if each frontal slice* $\boldsymbol{X}^{(i)}$ *is a diagonal matrix.*

Definition 2.4 (Tensor Singular Value Decomposition: t-SVD [20]). *For* $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$*, the t-SVD of* $\mathcal{X}$ *is given by* $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^H$*, where* $\mathcal{U} \in \mathbb{R}^{n_1\times n_1\times n_3}$ *and* $\mathcal{V} \in \mathbb{R}^{n_2\times n_2\times n_3}$ *are orthogonal tensors, and* $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ *is a f-diagonal tensor, respectively. The entries in* $\mathcal{S}$ *are called the singular fibers of* $\mathcal{X}$*.*

Definition 2.5 (Tubal multi-rank [19, 57]). *The multi-rank of a tensor* $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ *is a vector* $\boldsymbol{r} \in \mathbb{R}^{n_3}$ *with its* i*-th entry as the rank of the* i*-th frontal slice* $\widehat{\boldsymbol{X}}^{(i)}$ *of* $\widehat{\mathcal{X}}$*, i.e.,* $r_i = \operatorname{rank}(\widehat{\boldsymbol{X}}^{(i)})$*.*

Definition 2.6 (Tensor average rank [29]). *For* $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$*, the tensor average rank, denoted as* $\operatorname{rank}_a(\mathcal{X})$*, is defined as* $\operatorname{rank}_a(\mathcal{X}) = \frac{1}{n_3}\sum_{i=1}^{n_3} \operatorname{rank}(\widehat{\boldsymbol{X}}^{(i)})$*.*

Definition 2.7 (Tubal nuclear norm [29]). *The tubal nuclear norm of* $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$*, denoted as* $\|\mathcal{X}\|_{\mathrm{TNN}}$*, is the average of the nuclear norm of all the frontal slices of* $\widehat{\mathcal{X}}$*, i.e.,* $\|\mathcal{X}\|_{\mathrm{TNN}} = \frac{1}{n_3}\sum_{i=1}^{n_3}\|\widehat{\boldsymbol{X}}^{(i)}\|_*$*, where* $\|\cdot\|_*$ *denote the nuclear norm of matrix, i.e., the sum of all singular values of matrix.*

Definition 2.8 (Tensor basis [56]). *The column basis, denoted by* $\vec{e}_i$ *is a tensor of size* $n_1 \times 1 \times n_3$ *with the* $(i,1,1)$*-th entry equaling to 1 and the rest equaling to 0. The row basis is the transpose of* $\vec{e}_i$*, i.e.,* $\vec{e}_i^{\,T}$*. The tube basis, denoted by* $\mathring{e}_i$*, is a tensor of size* $1\times 1\times n_3$ *with the* $(1,1,k)$*-th entry equaling to 1 and the rest equaling to 0. Hence, one can obtain a unit tensor* $\Theta_{ijk} \in \mathbb{R}^{n_1\times n_2\times n_3}$ *with the* (i,j,k)*-th nonzero entry equaling 1 via* $\Theta_{ijk} = \vec{e}_i * \mathring{e}_k * \vec{e}_j^{\,T}$*. Now for any tensor* $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$*, its description based on the basis form can be given as follows:* $\mathcal{X} = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\sum_{k=1}^{n_3}\langle\Theta_{ijk},\mathcal{X}\rangle\Theta_{ijk}$*.*

Other notations will be defined in appropriate sections if necessary.

3. The Equivalent Surrogates for Robust Tensor Completion Model.

Since the tensor is bounded in many practical applications, such as an 8-byte image with elements ranging from 0 to 255, in this section, we introduce a nonconvex optimization model for bound constrained robust low-rank tensor completion problems.

3.1. Robust Tensor Completion Model.

Given the noisy data tensor $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$, only partial entries of $\mathcal{X}$ are observed, and the noisy data tensor $\mathcal{X}$ is an unknown low-rank tensor $\mathcal{L}^\star \in \mathbb{R}^{n_1\times n_2\times n_3}$ corrupted by an unknown sparse noise $\mathcal{M}^\star \in \mathbb{R}^{n_1\times n_2\times n_3}$. Then, we can recover the low-rank tensor $\mathcal{L}^\star$ by solving the following bound constrained Robust Tensor Completion model:

$$
\begin{aligned}
&\min_{\mathcal{L},\mathcal{M}} \operatorname{rank}_a(\mathcal{L}) + \lambda\|\mathcal{M}\|_0 \\
&\text{s.t. } \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \quad \|\mathcal{M}\|_\infty \le b_m, \quad \|\mathcal{L}\| \le b_l,
\end{aligned}
\tag{3.1}
$$

whereb_l, $b_m > 0$ are given constants, $\lambda > 0$ is a regularization parameter, $\|\cdot\|_0$ denotes the number of non-zero elements, $\operatorname{rank}_a(\mathcal{L})$ is the tensor average rank, $\|\cdot\|_\infty$ denotes the infinity norm, $\|\cdot\|$ is the tensor spectral norm, Ω is an index set, and $\mathcal{P}_\Omega$ is the orthogonal projection operator on Ω, i.e.,

$$\mathcal{P}_\Omega(\mathcal{X}) := \begin{cases} \mathcal{X}_{ijk}, & (i,j,k) \in \Omega, \\ 0, & \text{otherwise}. \end{cases}$$

It is well known that the rank and zero-norm optimization problems are in general NP-hard. Next, in terms of the variational characterization of the rank function and the zero-norm, we give its equivalent surrogates of (3.1) and prove that they have the same global optimal solution set as (3.1).

3.2. Equivalent Surrogates . Let Φ denote the family of closed proper convex functions $\phi : \mathbb{R} \to (-\infty, +\infty]$ satisfying $[0,1] \subseteq \operatorname{int}(\operatorname{dom}\phi)$, $\phi(1) = 1$ and $\phi(t^*_\phi) = 0$ where t^*_ϕ is the unique minimizer of ϕ over $[0,1]$. Let $\boldsymbol{e}$ be the vector of all ones. Then

$$\|\boldsymbol{z}\|_0 = \min_{\boldsymbol{w}}\{\Sigma_{i=1}^{p}\phi(w_i) \ \text{ s.t.}\langle \boldsymbol{e} - \boldsymbol{w}, |\boldsymbol{z}|\rangle = 0, 0 \le \boldsymbol{w} \le \boldsymbol{e}\} \tag{3.2}$$

and

$$\operatorname{rank}(\boldsymbol{X}) = \min_{\boldsymbol{W}}\{\Sigma_{i=1}^{n}\phi(\sigma_i(\boldsymbol{W})) \ \text{ s.t.}\|\boldsymbol{X}\|_* - \langle \boldsymbol{W}, \boldsymbol{X}\rangle = 0, \|\boldsymbol{W}\| \le 1\}, \tag{3.3}$$

which are introduced in [28]. By the variational characterization of the zero-norm and the rank function in (3.2) and (3.3), the rank plus zero-norm minimization problem (3.1) is equivalent to the problem

$$\begin{aligned} \min_{\mathcal{L},\mathcal{M},\mathcal{B},\mathcal{S}} \ & \frac{1}{n_3}\Sigma_{i=1}^{n_3}\Sigma_{j=1}^{\widetilde{n}}\phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \lambda\Sigma_{i=1}^{n_1}\Sigma_{j=1}^{n_2}\Sigma_{k=1}^{n_3}\phi(\mathcal{B}_{ijk}) \\ \text{s.t.} \ & \frac{1}{n_3}\Sigma_{i=1}^{n_3}(\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)}\rangle) + \lambda\langle \mathcal{E} - \mathcal{B}, |\mathcal{M}|\rangle = 0, \quad 0 \le \mathcal{B} \le \mathcal{E}, \quad \|\widehat{\boldsymbol{S}}^{(i)}\| \le 1, \\ & \mathcal{P}_\Omega(\mathcal{L} + \mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \ \|\mathcal{M}\|_\infty \le b_m, \quad \|\mathcal{L}\| \le b_l, \end{aligned} \tag{3.4}$$

where $\widetilde{n} = \min\{n_1, n_2\}$ and $\mathcal{E}$ is the tensor of all ones. Notice that $\frac{1}{n_3}\Sigma_{i=1}^{n_3}(\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)}\rangle) + \lambda\langle \mathcal{E} - \mathcal{B}, |\mathcal{M}|\rangle = 0$, $0 \le \mathcal{B} \le \mathcal{E}$, and $\|\widehat{\boldsymbol{S}}^{(i)}\| \le 1$ if and only if $\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)}\rangle = 0$, $\langle \mathcal{E} - \mathcal{B}, |\mathcal{M}|\rangle = 0$, $0 \le \mathcal{B} \le \mathcal{E}$, and $\|\widehat{\boldsymbol{S}}^{(i)}\| \le 1$, which can be obtained by the definition of the dual norm.

For brevity, we denote $J := \{(i,j,k)\}$. Now we consider the following penalty problem:

$$\begin{aligned} \min_{\mathcal{L},\mathcal{M},\mathcal{B},\mathcal{S}} \ & \frac{1}{n_3}\Sigma_{i=1}^{n_3}\Sigma_{j=1}^{\widetilde{n}}\phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \lambda\Sigma_{J}^{(n_1,n_2,n_3)}\phi(\mathcal{B}_J) + \frac{\rho}{n_3}\Sigma_{i=1}^{n_3}(\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)}\rangle) \\ & + \rho\lambda\langle \mathcal{E} - \mathcal{B}, |\mathcal{M}|\rangle \\ \text{s.t.} \ & 0 \le \mathcal{B} \le \mathcal{E}, \quad \|\widehat{\boldsymbol{S}}^{(i)}\| \le 1, \quad \mathcal{P}_\Omega(\mathcal{L} + \mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \quad \|\mathcal{M}\|_\infty \le b_m, \quad \|\mathcal{L}\| \le b_l, \end{aligned} \tag{3.5}$$

where $\rho > 0$ is the penalty factor. Next, we show that the penalty problem (3.5) is a global exact penalty for (3.4) in the sense that it has the same global optimal solution set as (3.4) does. The proof follows the line of [28, Theorem 5.1] in the matrix case by proving that the problem (3.4) is partially calm in its optimal solution set. The partial calmness is defined in [28], which is also given in Appendix A.

Theorem 3.1. *Let $\phi \in \Phi$. The penalty problem (3.5) is a global exact penalty for (3.4).*

Proof. Let $(\mathcal{L}^*, \mathcal{M}^*, \mathcal{B}^*, \mathcal{S}^*)$ be an arbitrary global optimal solution of (3.4) and consequently $\mathcal{L}^* \neq 0$ and $\mathcal{M}^* \neq 0$. For all $i \in \{1, 2, \ldots, n_3\}$, we write $r_i^* = \operatorname{rank}(\widehat{\boldsymbol{L}^*}^{(i)})$ and $s^* = \|\mathcal{M}^*\|_0$. Then $\sigma_{r_i^*}(\widehat{\boldsymbol{L}^*}^{(i)}) > 0$ and $\pi_{s^*}(\mathcal{M}^*) > 0$. By the continuity of $\sigma_{r_i^*}(\cdot)$ and $\pi_{s^*}(\cdot)$, there exists $\varepsilon > 0$ such that for any $(\mathcal{L}, \mathcal{M}) \in \mathbb{B}((\mathcal{L}^*, \mathcal{M}^*), \varepsilon)$,

$$\sigma_{r_i^*}(\widehat{\boldsymbol{L}^*}^{(i)}) \geq \alpha \quad \text{and} \quad \pi_{s^*}(\mathcal{M}) \geq \alpha \;\text{ with }\; \alpha = \min(\sigma_{r_i^*}(\widehat{\boldsymbol{L}^*}^{(i)}), \pi_{s^*}(\mathcal{M}^*))/2 \quad \forall i \in \{1, 2, \ldots, n_3\}. \tag{3.6}$$

We consider the perturbed problem of (3.4) whose feasible set takes the following form:

$$\mathcal{F}_\epsilon := \Bigg\{ (\mathcal{L}, \mathcal{M}, \mathcal{B}, \mathcal{S}) \;\Bigg|\; \frac{1}{n_3} \Sigma_{i=1}^{n_3} (\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)} \rangle) + \lambda(\|\mathcal{M}\|_1 - \langle \mathcal{B}, |\mathcal{M}| \rangle) = \epsilon,$$
$$0 \leq \mathcal{B} \leq \mathcal{E}, \;\; \|\widehat{\boldsymbol{S}}^{(i)}\| \leq 1, \;\; \mathcal{P}_\Omega(\mathcal{L} + \mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \;\; \|\mathcal{M}\|_\infty \leq b_m, \|\mathcal{L}\| \leq b_l \Bigg\}.$$

Fix an arbitrary $\epsilon \in \mathbb{R}$. It suffices to consider the case $\epsilon \geq 0$. Let $(\mathcal{L}, \mathcal{M}, \mathcal{B}, \mathcal{S})$ be an arbitrary point from $\mathcal{F}_\epsilon \bigcap \mathbb{B}((\mathcal{L}^*, \mathcal{M}^*, \mathcal{B}^*, \mathcal{S}^*), \varepsilon)$. Then, with $\bar{\rho} = \phi_-^{'}(1)/\alpha$,

$$\begin{aligned}
&\frac{1}{n_3} \Sigma_{i=1}^{n_3} \Sigma_{j=1}^{\widetilde{n}} \phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \lambda \Sigma_J^{(n_1,n_2,n_3)} \phi(\mathcal{B}_J) + \frac{\bar{\rho}}{n_3} \Sigma_{i=1}^{n_3} (\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \langle \widehat{\boldsymbol{S}}^{(i)}, \widehat{\boldsymbol{L}}^{(i)} \rangle) \\
&+ \bar{\rho}\lambda(\|\mathcal{M}\|_1 - \langle \mathcal{B}, |\mathcal{M}| \rangle) \\
\geq& \frac{1}{n_3} \Sigma_{i=1}^{n_3} \Sigma_{j=1}^{\widetilde{n}} [\phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \bar{\rho}\sigma_j(\widehat{\boldsymbol{L}}^{(i)})(1 - \sigma_j(\widehat{\boldsymbol{S}}^{(i)}))] + \lambda \Sigma_{j=1}^{n_1 n_2 n_3} [\phi(\pi_j(\mathcal{B})) + \bar{\rho}\pi_j(\mathcal{M})(1 - \pi_j(\mathcal{B}))] \\
\geq& \frac{1}{n_3} \Sigma_{i=1}^{n_3} \Sigma_{j=1}^{r_i^*} [\phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \bar{\rho}\sigma_{r_i^*}(\widehat{\boldsymbol{L}}^{(i)})(1 - \sigma_j(\widehat{\boldsymbol{S}}^{(i)}))] + \lambda \Sigma_{j=1}^{s^*} [\phi(\pi_j(\mathcal{B})) + \bar{\rho}\pi_{s^*}(\mathcal{M})(1 - \pi_j(\mathcal{B}))] \\
\geq& \frac{1}{n_3} \Sigma_{i=1}^{n_3} \Sigma_{j=1}^{r_i^*} [\phi(\sigma_j(\widehat{\boldsymbol{S}}^{(i)})) + \phi_-^{'}(1)(1 - \sigma_j(\widehat{\boldsymbol{S}}^{(i)}))] + \lambda \Sigma_{j=1}^{s^*} [\phi(\pi_j(\mathcal{B})) + \phi_-^{'}(1)(1 - \pi_j(\mathcal{B}))] \\
\geq& (\frac{1}{n_3} \Sigma_{i=1}^{n_3} r_i^* + \lambda s^*)\phi(1) = \frac{1}{n_3} \Sigma_{i=1}^{n_3} \operatorname{rank}(\widehat{\boldsymbol{L}^*}^{(i)}) + \lambda \|\mathcal{M}^*\|_0,
\end{aligned} \tag{3.7}$$

where the first inequality is by the von Neumann's inequality and $\langle \mathcal{B}, |\mathcal{M}| \rangle \leq \langle \pi(\mathcal{B}), \pi(\mathcal{M}) \rangle$, the second one is by the nonnegativity of ϕ in $[0, 1]$, the third one is due to (3.6) and $\bar{\rho} = \phi_-^{'}(1)/\alpha$, and the last one is using $\phi(t) \geq \phi(1) + \phi_-^{'}(1)(t - 1)$ for $t \in [0, 1]$. Since $\frac{1}{n_3} \Sigma_{i=1}^{n_3} \operatorname{rank}(\widehat{\boldsymbol{L}^*}^{(i)}) + \lambda \|\mathcal{M}^*\|_0$ is exactly the optimal value of (3.4), by the arbitrariness of ϵ in $\mathbb{R}$ and that of $(\mathcal{L}, \mathcal{M}, \mathcal{B}, \mathcal{S})$ in $\mathcal{F}_\epsilon \bigcap \mathbb{B}((\mathcal{L}^*, \mathcal{M}^*, \mathcal{B}^*, \mathcal{S}^*), \varepsilon)$, (3.7) shows that (3.4) is partially calm at $(\mathcal{L}^*, \mathcal{M}^*, \mathcal{B}^*, \mathcal{S}^*)$, where the definition of partial calmness and its properties are introduced in [28]. By the arbitrariness of $(\mathcal{L}^*, \mathcal{M}^*, \mathcal{B}^*, \mathcal{S}^*)$ in the global optimal solution set, it is partially calm in its optimal solution set. Since the feasible set of problem (3.5) is compact, the penalty problem (3.5) is a global exact penalty for (3.4) follows from [28, Proposition 2.1(b)]. ∎

Then, by letting $\psi(t) := \begin{cases} \phi(t), & t \in [0,1], \\ +\infty, & \text{otherwise} \end{cases}$ and using the conjugate ψ^* of ψ, i.e., $\psi^*(s) := \sup_{t\in\mathbb{R}}\{st - \psi(t)\}$, we can obtain the following conclusion.

Corollary 3.2. *Let $\phi \in \Phi$. There exists $\rho^* > 0$ such that the problem (3.1) has the same global optimal solution set as the following problem with $\rho > \rho^*$ does:*

$$\text{(3.8)}\quad \begin{aligned} &\min_{\mathcal{L},\mathcal{M}} \frac{\rho}{n_3}\Sigma_{i=1}^{n_3}\|\widehat{\boldsymbol{L}}^{(i)}\|_* - \frac{1}{n_3}\Sigma_{i=1}^{n_3}\Sigma_{j=1}^{\widetilde{n}}\psi^*(\rho\sigma_j(\widehat{\boldsymbol{L}}^{(i)})) + \lambda(\rho\|\mathcal{M}\|_1 - \Sigma_J\psi^*(\rho|\mathcal{M}_J|)) \\ &s.t.\ \ \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}),\quad \|\mathcal{M}\|_\infty \le b_m,\quad \|\mathcal{L}\| \le b_l. \end{aligned}$$

Let $u > 0$. Denote

$$\text{(3.9)}\qquad \widetilde{\theta}(s) := u\theta(\rho s)$$

with $\theta(s) := |s| - \psi^*(|s|)$. Then the problem (3.8) is equivalent to the following problem:

$$\text{(3.10)}\quad \begin{aligned} &\min_{\mathcal{L},\mathcal{M}} \frac{1}{n_3}\Sigma_{i=1}^{n_3}\Sigma_{j=1}^{\widetilde{n}}\widetilde{\theta}(\sigma_j(\widehat{\boldsymbol{L}}^{(i)})) + \lambda\Sigma_J\widetilde{\theta}(|\mathcal{M}_J|) \\ &\text{s.t.}\ \ \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}),\quad \|\mathcal{M}\|_\infty \le b_m,\quad \|\mathcal{L}\| \le b_l. \end{aligned}$$

It is worth noting that ϕ can be chosen as different functions satisfying $\phi \in \Phi$. In particular, if ϕ is chosen as the one in Example 3.1, then $\widetilde{\theta}$ becomes the MCP function (3.14); if ϕ is chosen as the one in Example 3.2, then $\widetilde{\theta}$ becomes the SCAD function (3.16).

Example 3.1. *Let $\phi(t) := \frac{\varphi(t)}{\varphi(1)}$ with $\varphi(t) := \frac{a^2}{4}t^2 - \frac{a^2}{2}t + at + \frac{(a-2)_+^2}{4}$, where $a > 0$ is a constant. Clearly, $\phi \in \Phi$ with $t_\phi^* = \frac{(a-2)_+}{a}$. Simple calculations show that ψ^* takes the following form:*

$$\psi^*(s) = \begin{cases} -\frac{(a-2)_+^2}{4}, & \text{if } s \le \frac{a-a^2/2}{\varphi(1)}, \\ \frac{1}{a^2\varphi(1)}(\frac{a^2-2a}{2} + s\varphi(1))^2 - \frac{(a-2)_+^2}{4\varphi(1)}, & \text{if } \frac{a-a^2/2}{\varphi(1)} < s \le \frac{a}{\varphi(1)}, \\ s-1, & \text{if } s > \frac{a}{\varphi(1)}. \end{cases}$$

When $a \ge 2$, we have $\varphi(1) = 1$ and $\theta(s) = |s| - \psi^*(|s|) = \begin{cases} \frac{2|s|}{a} - \frac{s^2}{a^2}, & |s| \le a, \\ 1, & |s| > a. \end{cases}$ Set $s := \frac{as}{\gamma}$ for some constants $\gamma > 0$, we have $\frac{\gamma}{2}\theta(\frac{as}{\gamma}) = \frac{\gamma}{2}(\frac{a|s|}{\gamma} - \psi^*(\frac{a|s|}{\gamma})) = \begin{cases} |s| - \frac{s^2}{2\gamma}, & |s| \le \gamma, \\ \frac{\gamma}{2}, & |s| > \gamma. \end{cases}$ If $\rho = \frac{a}{\gamma}$, $u = \frac{\gamma}{2}$ and $a \ge 2$, then the function $\widetilde{\theta}(s)$ defined in (3.9) is the MCP function.

Example 3.2. *Let $\phi(t) := \frac{\varphi(t)}{\varphi(1)}$ with $\varphi(t) := \frac{a-1}{2}t^2 + t$, where $a > 1$ is a constant. Clearly, $\phi \in \Phi$. Then,*

$$\psi^*(s) = \begin{cases} 0, & s \le \frac{1}{\varphi(1)}, \\ s-1, & s > \frac{a}{\varphi(1)}, \\ \frac{1}{2(a-1)\varphi(1)}(s\varphi(1)-1)^2, & \frac{1}{\varphi(1)} < s \le \frac{a}{\varphi(1)}. \end{cases}$$

Then, $\theta(s) = |s| - \psi^*(|s|) = \begin{cases} |s|, & |s| \le \frac{1}{\varphi(1)}, \\ 1, & |s| > \frac{a}{\varphi(1)}, \\ |s| - \frac{1}{2(a-1)\varphi(1)}(|s|\varphi(1)-1)^2, & \frac{1}{\varphi(1)} < |s| \le \frac{a}{\varphi(1)}. \end{cases}$ *Set* $s := \frac{s}{\gamma\varphi(1)}$ *for some constants* $\gamma > 0$, *we have*

$$\theta\Big(\frac{s}{\gamma\varphi(1)}\Big) = \frac{|s|}{\gamma\varphi(1)} - \psi^*\Big(\frac{|s|}{\gamma\varphi(1)}\Big) = \begin{cases} \frac{|s|}{\gamma\varphi(1)}, & |s| \le \gamma, \\ 1, & |s| > a\gamma, \\ \frac{|s|}{\gamma\varphi(1)} - \frac{1}{2(a-1)\varphi(1)}(|s|/\gamma - 1)^2, & \gamma < |s| \le a\gamma, \end{cases}$$

and $\gamma^2\varphi(1)\theta(\frac{s}{\gamma\varphi(1)}) = \gamma^2\varphi(1)(\frac{|s|}{\gamma\varphi(1)} - \psi^*(\frac{|s|}{\gamma\varphi(1)})) = \begin{cases} \gamma|s|, & |s| \le \gamma, \\ \frac{\gamma^2(a+1)}{2}, & |s| > a\gamma, \\ \frac{-s^2+2a|s|\gamma-\gamma^2}{2(a-1)}, & \gamma < |s| \le a\gamma. \end{cases}$ *If* $\rho = \frac{1}{\gamma\varphi(1)}$, $u = \gamma^2\varphi(1)$ *and* $a > 1$, *then the function* $\widetilde{\theta}(s)$ *defined in (3.9) is the SCAD function.*

3.3. BCNRTC for RTC Problems.

From the above discussion, the equivalent surrogates problem (3.10) can be rewritten in a simplified bound constrained Nonconvex Robust Tensor Completion (BCNRTC for short) form as follows:

$$\begin{aligned} &\min_{\mathcal{L},\mathcal{M}} \|\mathcal{L}\|_{\text{TNN}} - H_1(\mathcal{L}) + \lambda(\|\mathcal{M}\|_1 - H_2(\mathcal{M})) \\ &\text{s.t. } \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \quad \|\mathcal{M}\|_\infty \le b_m, \quad \|\mathcal{L}\| \le b_l, \end{aligned} \tag{3.11}$$

where H_1 and H_2 are defined as

$$H_1(\mathcal{L}) = \frac{1}{n_3}\Sigma_{i=1}^{n_3} g(\sigma(\widehat{\boldsymbol{L}}^{(i)})), \qquad H_2(\mathcal{M}) = \Sigma_{i=1}^{n_1}\Sigma_{j=1}^{n_2}\Sigma_{k=1}^{n_3} h(\mathcal{M}_{ijk}), \tag{3.12}$$

where $g(\boldsymbol{x}) = \Sigma_{j=1}^{\dim(\boldsymbol{x})} h(\boldsymbol{x}_j)$, h is a convex and continuous differentiable function which can be defined as

$$h(x) := \begin{cases} \frac{x^2}{2\gamma}, & |x| \le \gamma, \\ |x| - \frac{\gamma}{2}, & |x| > \gamma, \end{cases} \tag{3.13}$$

which is related to the MCP function ϖ_M with $h(x) = |x| - \varpi_M(x)$, where

$$\varpi_M(x) = \begin{cases} |x| - \frac{x^2}{2\gamma}, & |x| \le \gamma, \\ \frac{\gamma}{2}, & |x| > \gamma. \end{cases} \tag{3.14}$$

The convex function h can also be defined as

$$h(x) := \begin{cases} 0, & |x| \le \gamma_1, \\ \frac{x^2 - 2\gamma_1|x| + \gamma_1^2}{2(\gamma_2-\gamma_1)}, & \gamma_1 < |x| \le \gamma_2, \\ |x| - \frac{\gamma_1+\gamma_2}{2}, & |x| > \gamma_2, \end{cases} \tag{3.15}$$

which is related to the SCAD function ϖ_S with $h(x) = |x| - \varpi_S(x)$, where

$$\varpi_S(x) = \begin{cases} |x|, & |x| \le \gamma_1, \\ \frac{2\gamma_2|x| - x^2 - \gamma_1^2}{2(\gamma_2-\gamma_1)}, & \gamma_1 < |x| \le \gamma_2, \\ \frac{\gamma_1+\gamma_2}{2}, & |x| > \gamma_2. \end{cases} \tag{3.16}$$

Remark 3.3. When $H_1 \equiv 0$ and $H_2 \equiv 0$, the BCNRTC model (3.11) reduces to a convex model (CRTC for short)

$$\begin{aligned}(3.17)\qquad &\min_{\mathcal{L},\mathcal{M}} \|\mathcal{L}\|_{\mathrm{TNN}} + \lambda\|\mathcal{M}\|_1 \\ &\text{s.t.}\ \ \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \quad \|\mathcal{M}\|_\infty \le b_m, \quad \|\mathcal{L}\| \le b_l,\end{aligned}$$

which is actually a reformulation of the Robust Tensor Completion (RTCℓ_1) [18] with two bound constraints. We use the symmetric Gauss-Seidel based alternating direction method of multipliers (sGS-ADMM) to solve the CRTC which will be illustrated in Subsection 6.2 for a warm start of BCNRTC.

Notice that the feasible set of the problem (3.11) is bounded and closed, and the objective function is continuous and proper, by Weierstrass Theorem, the solution set of (3.11) is nonempty and compact.

In the next section, we will propose an algorithm to solve the BCNRTC model (3.11).

4. The Proximal Majorization-Minimization Algorithm.

In this section, we will develop a proximal majorization-minimization (PMM) algorithm to solve the BCNRTC model (3.11).

By using the indicator function, we can rewrite the BCNRTC model (3.11) to an unconstrained optimization problem as follows:

$$(4.1)\qquad \min_{\mathcal{L},\mathcal{M}} \|\mathcal{L}\|_{\mathrm{TNN}} - H_1(\mathcal{L}) + \lambda(\|\mathcal{M}\|_1 - H_2(\mathcal{M})) + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}),$$

where $D_1 := \{\mathcal{M} \mid \|\mathcal{M}\|_\infty \le b_m\}$, $D_2 := \{\mathcal{L} \mid \|\mathcal{L}\| \le b_l\}$, $\Gamma_1 := \{(\mathcal{L},\mathcal{M}) \mid \mathcal{P}_\Omega(\mathcal{L}+\mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X})\}$, and $\delta_{D_1}(\mathcal{M})$ is the indicator function of the nonempty set D_1.

The proposed PMM algorithm is to linearize the concave terms $-H_1(\cdot)$ and $-H_2(\cdot)$ in the objective function of (4.1) at each iteration with respect to the current iterate, say $(\mathcal{L}^k, \mathcal{M}^k)$, and generate the next iterate $(\mathcal{L}^{k+1}, \mathcal{M}^{k+1})$ by solving a convex subproblem inexactly:

(4.2)
$$\begin{aligned}\min_{\mathcal{L},\mathcal{M}} \Big\{ F(\mathcal{L},\mathcal{M};\mathcal{L}^k,\mathcal{M}^k) :=& \|\mathcal{L}\|_{\mathrm{TNN}} - H_1(\mathcal{L}^k) - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}-\mathcal{L}^k\rangle + \lambda(\|\mathcal{M}\|_1 - H_2(\mathcal{M}^k) \\ &- \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}-\mathcal{M}^k\rangle) + \frac{\eta}{2}\|\mathcal{M}-\mathcal{M}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{L}-\mathcal{L}^k\|_F^2 \\ &+ \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}) \Big\}.\end{aligned}$$

Let $\mathcal{L}^k = \mathcal{U}^k * \Sigma^k * (\mathcal{V}^k)^H$ be the t-SVD, then it holds that $\nabla H_1(\mathcal{L}^k) = \mathcal{U}^k * \mathcal{R}^k * (\mathcal{V}^k)^H$, where $\mathcal{R}^k = \mathrm{ifft}(\widehat{\mathcal{R}^k}, [\,], 3)$ and $\widehat{\boldsymbol{R}^k}^{(i)} = \mathrm{Diag}(\nabla g(\mathrm{diag}\,(\widehat{\boldsymbol{\Sigma}^k}^{(i)}))) = \mathrm{Diag}(\nabla g(\sigma(\widehat{\boldsymbol{L}^k}^{(i)})))$. For brevity, the proximal parameter $\eta > 0$ is assumed to be a constant, although it is frequently varying in practice to accelerate convergence.

By casting some constants, the subproblem (4.2) can be rewritten as follows:

$$\begin{aligned}(4.3)\qquad &\min_{\mathcal{L},\mathcal{M}} \|\mathcal{L}\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}\rangle + \lambda(\|\mathcal{M}\|_1 - \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}\rangle) + \frac{\eta}{2}\|\mathcal{M}-\mathcal{M}^k\|_F^2 \\ &\quad + \frac{\eta}{2}\|\mathcal{L}-\mathcal{L}^k\|_F^2 + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}).\end{aligned}$$

For convenience, we define $\mathcal{W} := (\mathcal{L}, \mathcal{M})$. Note that $F(\mathcal{W}; \mathcal{W}^k)$ is strongly convex, by [35, Theorem 1.9, Theorem 2.6], we obtain that $F(\mathcal{W}; \mathcal{W}^k)$ has a unique minimizer.

Motivated by [3], we use an error criterion to describe the inexact solution in (4.3), i.e., we need to find $\mathcal{W}^{k+1}$ and $\mathcal{C}^{k+1} := (\mathcal{C}_{\mathcal{L}}^{k+1}, \mathcal{C}_{\mathcal{M}}^{k+1})$ such that

$$\mathcal{C}^{k+1} \in \partial F(\mathcal{L}^{k+1}, \mathcal{M}^{k+1}; \mathcal{L}^k, \mathcal{M}^k) \quad \text{and} \quad \|\mathcal{C}^{k+1}\|_F \leq \eta c \|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F, \tag{4.4}$$

where $0 \leq c < \frac{1}{2}$ is a given constant.

Now, we summarize the PMM algorithm for solving the BCNRTC (3.11) in Algorithm 4.1.

Algorithm 4.1 The PMM algorithm for solving the BCNRTC (3.11).

1: **Input**: $\mathcal{L}^0, \mathcal{M}^0, \mathcal{P}_\Omega(\mathcal{X}), \lambda$, γ and η. Set $k = 0$.
2: Find $\mathcal{W}^{k+1}$, $\mathcal{C}^{k+1}$ such that $\mathcal{C}^{k+1} \in \partial F(\mathcal{L}^{k+1}, \mathcal{M}^{k+1}; \mathcal{L}^k, \mathcal{M}^k)$ and $\|\mathcal{C}^{k+1}\|_F \leq \eta c \|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F$.
3: If a termination criterion is met, set $\mathcal{L}^* := \mathcal{L}^{k+1}$, $\mathcal{M}^* := \mathcal{M}^{k+1}$; else, set $k := k+1$, return to 2.

4.1. Convergence Analysis.

In this section, we establish the global convergence of the PMM algorithm when h is chosen as the one in (3.13) or (3.15). Recall that the notation $\mathcal{W} := (\mathcal{L}, \mathcal{M})$. Let

$$Q(\mathcal{W}) := \|\mathcal{L}\|_{\mathrm{TNN}} - H_1(\mathcal{L}) + \lambda(\|\mathcal{M}\|_1 - H_2(\mathcal{M})) + \delta_{\Gamma_1}(\mathcal{L}, \mathcal{M}) + \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}).$$

It is easy to see that $F(\mathcal{W}^k; \mathcal{W}^k) = Q(\mathcal{W}^k)$. Firstly, we show a descent lemma for $Q(\mathcal{W})$.

Lemma 4.1. *Let $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 4.1. Then, for any $\eta > 0$ and $0 \leq c < \frac{1}{2}$,*

$$Q(\mathcal{W}^{k+1}) + \frac{\eta}{2}(1 - 2c)\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F^2 \leq Q(\mathcal{W}^k) \quad \forall k \geq 0,$$

and furthermore, $\lim_{k\to\infty} \|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F = 0$, where $\|\mathcal{W}^k\|_F = \sqrt{\|\mathcal{L}^k\|_F^2 + \|\mathcal{M}^k\|_F^2}$.

Next, we show $Q(\mathcal{W})$ satisfies the relative error condition.

Lemma 4.2. *Let $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 4.1, $\mathcal{W}^*$ be a cluster point and $\mathcal{B}^{k+1} := (\mathcal{B}_{\mathcal{L}}^{k+1}, \mathcal{B}_{\mathcal{M}}^{k+1}) \in \partial Q(\mathcal{W}^{k+1})$. Then, there exist $\delta_0 > 0$ and $\widetilde{m} > 0$ such that*

$$\|\mathcal{B}^{k+1}\|_F \leq (\widetilde{m} + \lambda/\gamma + \eta + \eta c)\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F \quad \forall\ \mathcal{W}^k, \mathcal{W}^{k+1} \in B(\mathcal{W}^*, \delta_0).$$

Lemma 4.3. *The function $Q(\mathcal{W})$ is a KL function when h is chosen as the one in (3.13) or (3.15).*

The proofs of Lemma 4.1, Lemma 4.2 and Lemma 4.3 are given in Appendix C. Combining Lemmas 4.1 - 4.3, we obtain the following convergence result of the PMM algorithm.

Theorem 4.4. *Let h be chosen as the one in (3.13) or (3.15), $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ be the sequence generated by Algorithm 4.1 and $\mathcal{W}^*$ be a cluster point. Then, for any $\eta > 0$ and $0 \leq c < \frac{1}{2}$,* *the sequence $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ converges to $\mathcal{W}^*$ as k goes to infinity, and $\mathcal{W}^*$ is a critical point of BCNRTC model (3.11), i.e., $0 \in \partial Q(\mathcal{W}^*)$. Moreover, the sequence $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ has a finite length ,i.e., $\sum_{k=0}^{\infty} ||\mathcal{W}^{k+1} - \mathcal{W}^k||_F < \infty$.*

Proof. As mentioned in Lemma 4.2, the sequence $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ generated by Algorithm 4.1 is bounded which admits a converging subsequence, i.e., there exists a subsequence $\mathcal{W}^{k_j}$ such that $\mathcal{W}^{k_j} \to \mathcal{W}^*$, as $k_j \to \infty$. Moreover, $\mathcal{W}^k$ belongs to Γ_1, D_1 and D_2, which leads to $\delta_{\Gamma_1}(\mathcal{L}^{k_j}, \mathcal{M}^{k_j}) = 0$, $\delta_{D_1}(\mathcal{M}^{k_j}) = 0$ and $\delta_{D_2}(\mathcal{L}^{k_j}) = 0$. So we have

$$
\begin{aligned}
Q(\mathcal{W}^{k_j}) =& \|\mathcal{L}^{k_j}\|_{\mathrm{TNN}} - H_1(\mathcal{L}^{k_j}) + \lambda(\|\mathcal{M}^{k_j}\|_1 - H_2(\mathcal{M}^{k_j})) + \delta_{\Gamma_1}(\mathcal{L}^{k_j}, \mathcal{M}^{k_j}) \\
&+ \delta_{D_1}(\mathcal{M}^{k_j}) + \delta_{D_2}(\mathcal{L}^{k_j}) \\
=& \|\mathcal{L}^{k_j}\|_{\mathrm{TNN}} - H_1(\mathcal{L}^{k_j}) + \lambda(\|\mathcal{M}^{k_j}\|_1 - H_2(\mathcal{M}^{k_j})) \\
\to& \|\mathcal{L}^*\|_{\mathrm{TNN}} - H_1(\mathcal{L}^*) + \lambda(\|\mathcal{M}^*\|_1 - H_2(\mathcal{M}^*)), \text{ as } k_j \to \infty,
\end{aligned} \tag{4.5}
$$

where the last limit holds by the continuity of $\|\cdot\|_{\mathrm{TNN}} - H_1(\cdot) + \lambda(\|\cdot\|_1 - H_2(\cdot))$. Since the sets Γ_1, D_1 and D_2 are closed and $\mathcal{W}^k$ belongs to Γ_1, D_1 and D_2, we have $\mathcal{W}^*$ belongs to Γ_1, D_1 and D_2, and so $Q(\mathcal{W}^*) = \|\mathcal{L}^*\|_{\mathrm{TNN}} - H_1(\mathcal{L}^*) + \lambda(\|\mathcal{M}^*\|_1 - H_2(\mathcal{M}^*))$, which together with (4.5), implies that $Q(\mathcal{W}^{k_j}) \to Q(\mathcal{W}^*)$ as $k_j \to \infty$. Combining Lemma 4.1 - Lemma 4.3, the conclusion is obtained according to [3, Theorem 2.9]. This completes the proof. ∎

4.2. Solving the Subproblem. In this section, the symmetric Gauss-Seidel based alternating direction method of multipliers (sGS-ADMM)[25] is applied to solve the subproblem in the PMM algorithm. Each PMM iteration solves a strongly convex subproblem of the following form inexactly:

$$
\begin{aligned}
&\min_{\mathcal{L},\mathcal{M}} \|\mathcal{L}\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}\rangle + \lambda(\|\mathcal{M}\|_1 - \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}\rangle) + \frac{\eta}{2}\|\mathcal{M} - \mathcal{M}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{L} - \mathcal{L}^k\|_F^2 \\
&\text{s.t. } \mathcal{P}_\Omega(\mathcal{L} + \mathcal{M}) = \mathcal{P}_\Omega(\mathcal{X}), \ \|\mathcal{M}\|_\infty \le b_m, \ \ \|\mathcal{L}\| \le b_l.
\end{aligned} \tag{4.6}
$$

Let $\mathcal{L} + \mathcal{M} = \mathcal{Z}$ and add a proximal term. The problem (4.6) can be rewritten as

$$
\begin{aligned}
\min_{\mathcal{L},\mathcal{M},\mathcal{Z}} \ & \|\mathcal{L}\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}\rangle + \lambda(\|\mathcal{M}\|_1 - \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}\rangle) + \frac{\eta}{2}\|\mathcal{M} - \mathcal{M}^k\|_F^2 \\
& + \frac{\eta}{2}\|\mathcal{L} - \mathcal{L}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{Z} - \mathcal{Z}^k\|_F^2 + \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}) \\
\text{s.t.} \ \ & \mathcal{L} + \mathcal{M} = \mathcal{Z}, \ \ \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Z}).
\end{aligned} \tag{4.7}
$$

Let $\Gamma_2 := \{\mathcal{Z} | \mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Z})\}$. The augmented Lagrangian function associated with (4.7) is defined by

$$
\begin{aligned}
\mathscr{L}(\mathcal{L}, \mathcal{M}, \mathcal{Z}; \mathcal{Y}) :=& \|\mathcal{L}\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}\rangle + \lambda(\|\mathcal{M}\|_1 - \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}\rangle) + \langle \mathcal{Y}, \mathcal{Z} - \mathcal{L} - \mathcal{M}\rangle \\
&+ \frac{\eta}{2}\|\mathcal{M} - \mathcal{M}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{L} - \mathcal{L}^k\|_F^2 + \frac{\mu}{2}\|\mathcal{L} + \mathcal{M} - \mathcal{Z}\|_F^2 + \frac{\eta}{2}\|\mathcal{Z} - \mathcal{Z}^k\|_F^2 \\
&+ \delta_{D_1}(\mathcal{M}) + \delta_{D_2}(\mathcal{L}),
\end{aligned}
$$

where $\mu > 0$ is the penalty parameter and $\mathcal{Y}$ is a multiplier. The iterative scheme of sGS-ADMM is given explicitly by

$$\mathcal{Z}^{t+\frac{1}{2}} = \arg\min_{\mathcal{Z}\in\Gamma_2}\{\mathscr{L}(\mathcal{L}^t,\mathcal{M}^t,\mathcal{Z};\mathcal{Y}^t)\}, \tag{4.8}$$

$$\mathcal{L}^{t+1} = \arg\min_{\mathcal{L}}\{\mathscr{L}(\mathcal{L},\mathcal{M}^t,\mathcal{Z}^{t+\frac{1}{2}};\mathcal{Y}^t)\}, \tag{4.9}$$

$$\mathcal{Z}^{t+1} = \arg\min_{\mathcal{Z}\in\Gamma_2}\{\mathscr{L}(\mathcal{L}^{t+1},\mathcal{M}^t,\mathcal{Z};\mathcal{Y}^t)\}, \tag{4.10}$$

$$\mathcal{M}^{t+1} = \arg\min_{\mathcal{M}}\{\mathscr{L}(\mathcal{L}^{t+1},\mathcal{M},\mathcal{Z}^{t+1};\mathcal{Y}^t)\}, \tag{4.11}$$

$$\mathcal{Y}^{t+1} = \mathcal{Y}^t - \tau\mu(\mathcal{L}^{t+1}+\mathcal{M}^{t+1}-\mathcal{Z}^{t+1}), \tag{4.12}$$

where $\tau \in (0,(1+\sqrt{5})/2)$ is the step-length. Next, we turn to compute the concrete forms of solutions in each subproblem.

The optimal solution with respect to $\mathcal{Z}$ is given explicitly by

$$\mathcal{Z} = \mathcal{P}_\Omega(\mathcal{X}) + \frac{1}{\mu+\eta}\mathcal{P}_{\overline{\Omega}}(\mu(\mathcal{L}+\mathcal{M})+\eta\mathcal{Z}^k-\mathcal{Y}).$$

Before giving the solution of the problem (4.9), we need to present the following lemma.

Lemma 4.5. *For any $\mathcal{Y}\in\mathbb{R}^{n_1\times n_2\times n_3}$, $\tau>0$ and $\rho>0$. Let $\mathcal{Y}=\mathcal{U}*\Sigma*\mathcal{V}^H$ be the t-SVD. Then the optimal solution of the following problem*

$$\min_{\mathcal{X}\in\mathbb{R}^{n_1\times n_2\times n_3}}\left\{\tau\|\mathcal{X}\|_{TNN}+\frac{1}{2}\|\mathcal{X}-\mathcal{Y}\|_F^2 \mid \|\mathcal{X}\|\le\rho\right\}$$

is given by $\mathcal{X}^* = \mathcal{U}*\mathcal{D}_{\tau,\rho}*\mathcal{V}^H$, where $\mathcal{D}_{\tau,\rho} = \mathit{ifft}(\min\{\max\{\widehat{\Sigma}-\tau,0\},\rho\},[\,],3)$.

Lemma 4.5 can be proved easily. For brevity, we omit it here. It follows from Lemma 4.5 that the optimal solution with respect to $\mathcal{L}$ in (4.9) can be given by

$$\begin{aligned}\mathcal{L}^{t+1} &= \arg\min_{\|\mathcal{L}\|\le b_l}\left\{\|\mathcal{L}\|_{\mathrm{TNN}} - \langle\nabla H_1(\mathcal{L}^k)-\mathcal{Y}_1^t,\mathcal{L}\rangle + \frac{\mu}{2}\|\mathcal{L}+\mathcal{M}^t-\mathcal{Z}^{t+\frac{1}{2}}\|_F^2 + \frac{\eta}{2}\|\mathcal{L}-\mathcal{L}^k\|_F^2\right\}\\ &= \arg\min_{\|\mathcal{L}\|\le b_l}\left\{\|\mathcal{L}\|_{\mathrm{TNN}} + \frac{\eta+\mu}{2}\|\mathcal{L}-\mathcal{A}\|_F^2\right\} = \mathcal{U}^t*\mathcal{D}_{\tau,\rho}^t*(\mathcal{V}^t)^H,\end{aligned} \tag{4.13}$$

where $\mathcal{A} = (-\mu\mathcal{M}^t+\mu\mathcal{Z}^{t+\frac{1}{2}}+\eta\mathcal{L}^k+\mathcal{Y}_1^t+\nabla H_1(\mathcal{L}^k))/(\eta+\mu) = \mathcal{U}^t*\Sigma^t*(\mathcal{V}^t)^H$ and $\mathcal{D}_{\tau,\rho}^t = \mathrm{ifft}(\min\{\max\{\widehat{\Sigma^t}-1/(\eta+\mu),0\},b_l\},[\,],3)$.

On the other hand, the optimal solution with respect to (4.11) is given by

$$\begin{aligned}\mathcal{M}^{t+1} = &\arg\min_{\|\mathcal{M}\|_\infty\le b_m}\Big\{\lambda(\|\mathcal{M}\|_1-\langle\nabla H_2(\mathcal{M}^k),\mathcal{M}\rangle)-\langle\mathcal{Y}_1^t,\mathcal{M}\rangle+\frac{\eta}{2}\|\mathcal{M}-\mathcal{M}^k\|_F^2\\ &+\frac{\mu}{2}\|\mathcal{M}+\mathcal{L}^{t+1}-\mathcal{Z}^{t+1}\|_F^2\Big\}\\ = &\arg\min_{\|\mathcal{M}\|_\infty\le b_m}\left\{\|\mathcal{M}\|_1+\frac{\eta+\mu}{2\lambda}\|\mathcal{M}-\mathcal{G}\|_F^2\right\},\end{aligned}$$

where $\mathcal{G} = (\lambda \nabla H_2(\mathcal{M}^k) + \mu \mathcal{Z}^{t+1} - \mu \mathcal{L}^{t+1} + \eta \mathcal{M}^k + \mathcal{Y}_1^t)/(\eta + \mu)$. Simple calculations show that the closed form solution with respect to $\mathcal{M}^{t+1}$ can be given by

$$\mathcal{M}_{ijk}^{t+1} = \begin{cases} \operatorname{sign}(\mathcal{G}_{ijk}) \max\{|\mathcal{G}_{ijk}| - \lambda/(\mu+\eta), 0\}, & |\mathcal{G}_{ijk}| \le b_m + \lambda/(\mu+\eta), \\ \operatorname{sign}(\mathcal{G}_{ijk}) b_m, & |\mathcal{G}_{ijk}| > b_m + \lambda/(\mu+\eta). \end{cases} \tag{4.14}$$

Now we are ready to state the sGS-ADMM for solving (4.7) in Algorithm 4.2.

Algorithm 4.2 A symmetric Gauss-Seidel ADMM for solving (4.7).

1: **Input**: τ, Ω, λ, γ, μ, η, $\mathcal{P}_\Omega(\mathcal{X})$, $\mathcal{L}^0$, $\mathcal{M}^0$, $\mathcal{Y}^0$, $\mathcal{M}^k$, $\mathcal{L}^k$ and $\mathcal{Z}^k$. Set $t = 0$.
2: Compute $\mathcal{Z}^{t+\frac{1}{2}}$ by $\mathcal{Z}^{t+\frac{1}{2}} = \mathcal{P}_\Omega(\mathcal{X}) + \frac{1}{\mu+\eta} \mathcal{P}_{\overline{\Omega}}(\mu(\mathcal{L}^t + \mathcal{M}^t) + \eta \mathcal{Z}^k - \mathcal{Y}^t)$.
3: Compute $\mathcal{L}^{t+1}$ via (4.13).
4: Compute $\mathcal{Z}^{t+1}$ by $\mathcal{Z}^{t+1} = \mathcal{P}_\Omega(\mathcal{X}) + \frac{1}{\mu+\eta} \mathcal{P}_{\overline{\Omega}}(\mu(\mathcal{L}^{t+1} + \mathcal{M}^t) + \eta \mathcal{Z}^k - \mathcal{Y}^t)$.
5: Compute $\mathcal{M}^{t+1}$ via (4.14).
6: Compute $\mathcal{Y}^{t+1}$ by (4.12).
7: If a termination criterion is not met, set $t := t + 1$ and return to 2.

Note that the objective function of (4.7) is nonsmooth with respect to $\mathcal{L}$, $\mathcal{M}$ and quadratic with respect to $\mathcal{Z}$. By [25, Theorem 3], we can show the convergence of Algorithm 4.2, which is summarized in the following theorem.

Theorem 4.6. *Let $\{(\mathcal{L}^t, \mathcal{M}^t, \mathcal{Z}^t, \mathcal{Y}^t)\}_{t\in\mathbb{N}}$ be generated by Algorithm 4.2. Choose $\mu > 0$ and $\gamma \in (0, (\sqrt{5}+1)/2)$, then the sequence $\{(\mathcal{L}^t, \mathcal{M}^t, \mathcal{Z}^t)\}_{t\in\mathbb{N}}$ converges to an optimal solution of the problem (4.7) and $\{\mathcal{Y}^t\}_{t\in\mathbb{N}}$ converges to an optimal solution of the dual problem of (4.7).*

Proof. Notice that the problem (4.7) has a unique minimizer and the following constraint qualification is satisfied:

$$\text{There exists } (\mathcal{L}^*, \mathcal{M}^*, \mathcal{Z}^*) \in \operatorname{ri}(D_2 \times D_1 \times \Gamma_2) \cap \mathfrak{C},$$

where $\mathfrak{C} := \{(\mathcal{L}, \mathcal{M}, \mathcal{Z}) | \mathcal{L} + \mathcal{M} = \mathcal{Z}\}$. By [25, Theorem 3], we can easily obtain the conclusion of this theorem. ■

Remark 4.7. Actually, Algorithm 4.2 shows the process of solving the CRTC model if η, $\mathcal{M}^k$, $\mathcal{L}^k$ and $\mathcal{Z}^k$ are all equal to zero. For simplicity, we don't give the specific algorithm frame here.

Next we give the computational cost of algorithms. At each iteration of solving the subproblem of PMM algorithm, we need to calculate (4.8)-(4.12). The main cost of (4.9) is tensor SVD. The number of the floating point operations of fft is $\mathcal{O}(n_3 \log_2(n_3))$, and we need to calculate $n_1 n_2$ times, so the total cost of tensor fft is $\mathcal{O}(n_3 \log_2(n_3) n_1 n_2)$. Meanwhile the cost of SVDs for n_3 n_1-by-n_2 matrix is $\mathcal{O}(\widetilde{n}\widetilde{m}^2 n_3)$, where $\widetilde{n} = \min\{n_1, n_2\}$ and $\widetilde{m} = \max\{n_1, n_2\}$. Therefore, the total cost of tensor SVD is $\mathcal{O}(n_3 \log_2(n_3) n_1 n_2 + \widetilde{n}\widetilde{m}^2 n_3)$ operations. The complexities of computing $\mathcal{Z}^{t+1}$, $\mathcal{M}^{t+1}$ and $\mathcal{Y}^{t+1}$ are all $\mathcal{O}(n_1 n_2 n_3)$ operations for the independency that operation on each entry of the tensor. Then the total cost of the subproblem of PMM algorithm at each iteration is $\mathcal{O}(n_3 \log_2(n_3) n_1 n_2 + \widetilde{n}\widetilde{m}^2 n_3)$. During the algorithm execution, the largest data we storage is the $n_1 \times n_2 \times n_3$ tensor, so the memory complexity is $\mathcal{O}(n_1 n_2 n_3)$.

5. Error Bounds.

In this section, we establish the error bound between the optimal solution $(\mathcal{L}^c, \mathcal{M}^c)$ of (4.3) and the ground-truth $(\mathcal{L}^\star, \mathcal{M}^\star)$ in Frobenius norm. Meanwhile, we give the analysis that the error bound of BCNRTC can be reduced compared with that of CRTC as long as the given initial estimator is not far from the ground truth.

We assume that $\|\mathcal{M}^\star\|_0 = \widetilde{s}$ and the tubal multi-rank of $\mathcal{L}^\star$ is $\boldsymbol{r} = (r_1, r_2, \dots, r_{n_3})$. Denote $\widetilde{\Delta}_{\mathcal{L}} := \mathcal{L}^c - \mathcal{L}^\star$ and $\widetilde{\Delta}_{\mathcal{M}} := \mathcal{M}^c - \mathcal{M}^\star$. Firstly, we provide the connection among $\|\widetilde{\Delta}_{\mathcal{L}}\|_{\mathrm{TNN}}$, $\|\widetilde{\Delta}_{\mathcal{M}}\|_1$ and the Frobenius norms of $\widetilde{\Delta}_{\mathcal{L}}$ and $\widetilde{\Delta}_{\mathcal{M}}$. Similar results have been studied in [55], which established the relationship between the TNN and the Frobenius norm of the tensor by using the tubal rank. We show a structure constructed by the average rank, which may provide a more clear result of the error bound.

In order to display the structure, we study the subgradient of the TNN at first. Considering the $\overline{\boldsymbol{L}^\star}$ with the structure $\overline{\boldsymbol{L}^\star} = \mathrm{Diag}(\widehat{\boldsymbol{L}^\star}^{(1)}, \widehat{\boldsymbol{L}^\star}^{(2)}, \dots, \widehat{\boldsymbol{L}^\star}^{(n_3)})$, where $\widehat{\boldsymbol{L}^\star}^{(i)} \in \mathbb{C}^{n_1 \times n_2}$ with the SVD $\widehat{\boldsymbol{L}^\star}^{(i)} = \boldsymbol{U}^{(i)} \boldsymbol{S}^{(i)} (\boldsymbol{V}^{(i)})^H$. Notice that $\mathrm{rank}(\widehat{\boldsymbol{L}^\star}^{(i)}) = r_i$, by dividing the first r_i columns and the last $n_1 - r_i$ columns, we have the $\boldsymbol{U}^{(i)} = [\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}]$, where $\boldsymbol{U}_1^{(i)} \in \mathbb{C}^{n_1 \times r_i}$ and $\boldsymbol{U}_2^{(i)} \in \mathbb{C}^{n_1 \times (n_1 - r_i)}$. Similarly, $\boldsymbol{V}^{(i)} = [\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}]$, where $\boldsymbol{V}_1^{(i)} \in \mathbb{C}^{n_2 \times r_i}$ and $\boldsymbol{V}_2^{(i)} \in \mathbb{C}^{n_2 \times (n_2 - r_i)}$. From the subgradient of nuclear norm of the matrix, we have

$$\left\{ \boldsymbol{U}_1^{(i)} (\boldsymbol{V}_1^{(i)})^H + \boldsymbol{U}_2^{(i)} \boldsymbol{W}^{(i)} (\boldsymbol{V}_2^{(i)})^H | \boldsymbol{W}^{(i)} \in \mathbb{C}^{(n_1 - r_i) \times (n_2 - r_i)}, \|\boldsymbol{W}^{(i)}\| \leq 1 \right\} = \partial \|\widehat{\boldsymbol{L}^\star}^{(i)}\|_*.$$

We denote that $\widehat{\boldsymbol{U}_1}^{(i)} = [\boldsymbol{U}_1^{(i)}, 0] \in \mathbb{C}^{n_1 \times r_{\max}}$, $\widehat{\boldsymbol{V}_1}^{(i)} = [\boldsymbol{V}_1^{(i)}, 0] \in \mathbb{C}^{n_2 \times r_{\max}}$, $\widehat{\boldsymbol{U}_2}^{(i)} = [0, \boldsymbol{U}_2^{(i)}] \in \mathbb{C}^{n_1 \times (n_1 - r_{\min})}$, $\widehat{\boldsymbol{V}_2}^{(i)} = [0, \boldsymbol{V}_2^{(i)}] \in \mathbb{C}^{n_2 \times (n_2 - r_{\min})}$ and

$$\widehat{\boldsymbol{W}}^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{W}^{(i)} \end{bmatrix} \in \mathbb{C}^{(n_1 - r_{\min}) \times (n_2 - r_{\min})},$$

where $r_{\max} = \max\{r_1, r_2, \dots, r_{n_3}\}$, $r_{\min} = \min\{r_1, r_2, \dots, r_{n_3}\}$ and $\|\boldsymbol{W}^{(i)}\| \leq 1$. Then we have $\widehat{\boldsymbol{U}_1}^{(i)} (\widehat{\boldsymbol{V}_1}^{(i)})^H + \widehat{\boldsymbol{U}_2}^{(i)} \widehat{\boldsymbol{W}}^{(i)} (\widehat{\boldsymbol{V}_2}^{(i)})^H = \boldsymbol{U}_1^{(i)} (\boldsymbol{V}_1^{(i)})^H + \boldsymbol{U}_2^{(i)} \boldsymbol{W}^{(i)} (\boldsymbol{V}_2^{(i)})^H \in \partial \|\widehat{\boldsymbol{L}^\star}^{(i)}\|_*$.

Since $\widehat{\boldsymbol{U}_1}^{(i)} \in \mathbb{C}^{n_1 \times r_{\max}}$ have the same size for $i = 1, 2, \dots, n_3$, we can stack the matrices to form a tensor $\widehat{\mathcal{U}_1} \in \mathbb{C}^{n_1 \times r_{\max} \times n_3}$. Let $\widehat{\mathcal{U}_2}$, $\widehat{\mathcal{V}_1}$, $\widehat{\mathcal{V}_2}$ and $\widehat{\mathcal{W}}$ are constructed likewise, we can see the following proposition holds.

Proposition 5.1. *Let $\widehat{\mathcal{U}_1}$, $\widehat{\mathcal{U}_2}$, $\widehat{\mathcal{V}_1}$, $\widehat{\mathcal{V}_2}$ and $\widehat{\mathcal{W}}$ are defined as above, and $\mathcal{U}_1 = ifft(\widehat{\mathcal{U}_1}, [\,], 3)$, $\mathcal{U}_2 = ifft(\widehat{\mathcal{U}_2}, [\,], 3)$, $\mathcal{V}_1 = ifft(\widehat{\mathcal{V}_1}, [\,], 3)$, $\mathcal{V}_2 = ifft(\widehat{\mathcal{V}_2}, [\,], 3)$, $\mathcal{W} = ifft(\widehat{\mathcal{W}}, [\,], 3)$. Then we have*

$$S(\mathcal{L}^\star) := \left\{ \mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H | \mathcal{W} \in \mathbb{C}^{(n_1 - r_{\min}) \times (n_2 - r_{\min}) \times n_3}, \|\mathcal{W}\| \leq 1 \right\} = \partial \|\mathcal{L}^\star\|_{TNN}. \tag{5.1}$$

The proof of the Proposition 5.1 is given in Appendix D.1. Obviously, $\mathcal{U}_1 \in \mathbb{R}^{n_1 \times r_{\max} \times n_3}$ and $\mathcal{V}_1 \in \mathbb{R}^{n_2 \times r_{\max} \times n_3}$ have the same tubal multi-rank with $\mathcal{L}^\star$.

Remark 5.2. A similar work is given in [29]:

$$G(\mathcal{L}^\star) := \left\{ \mathcal{U}_s * \mathcal{V}_s^H + \mathcal{R} \mid \mathcal{U}_s^H * \mathcal{R} = \mathbf{0}, \mathcal{R} * \mathcal{V}_s = \mathbf{0}, \|\mathcal{R}\| \leq 1 \right\} = \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}},$$

where $\mathcal{L}^\star = \mathcal{U}_s * \mathcal{S}_s * \mathcal{V}_s^H$ is the skinny t-SVD of $\mathcal{L}^\star$. However, its proof is not given, and it is not shown how to construct $\mathcal{U}_s$ and $\mathcal{V}_s$. If $\mathcal{U}_s$ and $\mathcal{V}_s$ are constructed as same as those in [55] similarly to the skinny SVD of matrix, then $S(\mathcal{L}^\star) \supseteq G(\mathcal{L}^\star)$, and the "equality" relationship holds when $r_i = r_{\max}$ for $i = 1, 2, ..., n_3$. If $\mathcal{U}_s$ and $\mathcal{V}_s$ are constructed as same as ours, i.e., $\mathcal{U}_s = \mathcal{U}_1$ and $\mathcal{V}_s = \mathcal{V}_1$, then $S(\mathcal{L}^\star) = G(\mathcal{L}^\star)$.

Denote the set $\mathcal{T}$ by

$$\mathcal{T} := \{\mathcal{U}_1 * \mathcal{Y}^H + \mathcal{W} * \mathcal{V}_1^H \,|\, \mathcal{Y} \in \mathbb{R}^{n_2 \times r_{\max} \times n_3}, \mathcal{W} \in \mathbb{R}^{n_1 \times r_{\max} \times n_3}\},$$

and its orthogonal complement by $\mathcal{T}^\perp$. The set $\mathcal{T}$ is the tangent space with respect to the rank-constraint tensors $\{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3} | \mathrm{rank}_a(\mathcal{X}) \leq r_{\max}\}$ at $\mathcal{L}^\star$.

Proposition 5.3. *For any tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the orthogonal projection of $\mathcal{X}$ onto $\mathcal{T}$ and $\mathcal{T}^\perp$ are given by*

$$\mathcal{P}_{\mathcal{T}}(\mathcal{X}) = \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} + \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H - \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H,$$

$$\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{X}) = \mathcal{U}_2 * \mathcal{U}_2^H * \mathcal{X} * \mathcal{V}_2 * \mathcal{V}_2^H.$$

The proof of the Proposition 5.3 is given in Appendix D.2. For simplicity of subsequently analysis, we denote

$$d_{\mathcal{L}} := \frac{1}{\sqrt{r}} \|\mathcal{U}_1 * \mathcal{V}_1^H - \nabla H_1(\mathcal{L}^k)\|_F, \quad d_{\mathcal{M}} := \frac{1}{\sqrt{\widetilde{s}}} \|\mathrm{sign}(\mathcal{M}^\star) - \nabla H_2(\mathcal{M}^k)\|_F, \tag{5.2}$$

$r := \frac{\sum_{i=1}^{n_3} r_i}{n_3}$, $|\Omega| := m$, and $\widetilde{\Delta} := \widetilde{\Delta}_{\mathcal{L}} + \widetilde{\Delta}_{\mathcal{M}}$.

Denote Θ_{ijk} as a unit tensor with the (i, j, k)-th nonzero entry equaling 1. Let the set of the standard orthogonal basis of $\mathbb{R}^{n_1 \times n_2 \times n_3}$ be denoted by $\Theta := \{\Theta_{ijk} | 1 \leq i \leq n_1, 1 \leq j \leq n_2, 1 \leq k \leq n_3\}$. For each unit tensor Θ_{ijk} , there exists a unique index $\omega_l = j + (i-1)n_2 + (k-1)n_1 n_2$ such that $\Theta_{\omega_l} = \Theta_{ijk}$, $\omega_l \in \{1, 2, \ldots, n_1 n_2 n_3\}$, which is a bijective mapping from $\{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\} \times \{1, 2, \ldots, n_3\}$ to $\{1, 2, \ldots, n_1 n_2 n_3\}$. Then Ω be the multiset of all sampled i.i.d. indices $\omega_1, \ldots, \omega_m$ mapping to the subset of $\{1, 2, \ldots, n_1\} \times \{1, 2, \ldots, n_2\} \times \{1, 2, \ldots, n_3\}$.

Lemma 5.4. *For any $\eta > 0$ and $\lambda > 0$, we have*

$$\|\widetilde{\Delta}_{\mathcal{L}}\|_{TNN} \leq p_1 \|\widetilde{\Delta}_{\mathcal{L}}\|_F + p_2 \|\widetilde{\Delta}_{\mathcal{M}}\|_F, \quad \|\widetilde{\Delta}_{\mathcal{M}}\|_1 \leq q_1 \|\widetilde{\Delta}_{\mathcal{L}}\|_F + q_2 \|\widetilde{\Delta}_{\mathcal{M}}\|_F, \tag{5.3}$$

where $p_1 := \sqrt{2r} + d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F$, $p_2 := \lambda d_{\mathcal{M}}\sqrt{\widetilde{s}} + \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F$, $q_1 := (d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)/\lambda$ and $q_2 := \sqrt{\widetilde{s}} + d_{\mathcal{M}}\sqrt{\widetilde{s}} + \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F/\lambda$.

The proof of the Lemma 5.4 is given in Appendix D.3. Let p_{ijk} denote the probability to observe the (i, j, k)-th entry of $\mathcal{X}$, we suppose that each element is sampled with positive probability.

Assumption 5.1. *There exists a positive constant $\mu_1 \geq 1$ such that $p_{ijk} \geq (\mu_1 n_1 n_2 n_3)^{-1}$.*

Note that Assumption 5.1 implies that

$$\mathbb{E}[\langle \Theta, \mathcal{X} \rangle^2] = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} p_{ijk} \mathcal{X}_{ijk}^2 \geq (\mu_1 n_1 n_2 n_3)^{-1} \|\mathcal{X}\|_F^2. \tag{5.4}$$

Define the operator $\mathfrak{D}_\Omega : \mathbb{R}^{n_1\times n_2\times n_3} \to \mathbb{R}^m$ by $\mathfrak{D}_\Omega(\mathcal{X}) := (\langle\Theta_{\omega_1}, \mathcal{X}\rangle, \ldots, \langle\Theta_{\omega_m}, \mathcal{X}\rangle)^T$. The adjoint $\mathfrak{D}_\Omega^* : \mathbb{R}^m \to \mathbb{R}^{n_1\times n_2\times n_3}$ by $\mathfrak{D}_\Omega^*(\mathfrak{D}_\Omega(\mathcal{X})) = \sum_{l=1}^m \langle\Theta_{\omega_l}, \mathcal{X}\rangle\Theta_{\omega_l}$. Let $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^T$ be independent and identically distributed (i.i.d.) Rademacher sequence, i.e., i.i.d. sequence of Bernoulli random variables taking the values 1 and -1 with probability $\frac{1}{2}$. Define

$$\beta_{\mathcal{L}} := \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|, \quad \beta_{\mathcal{M}} := \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty . \tag{5.5}$$

The following Lemma shows that the sampling operator $\mathcal{P}_\Omega$ satisfies some property specified in a certain set with high probability. Similar results can also be found in [21].

Lemma 5.5. *Suppose that Assumption 5.1 holds. Given any positive numbers p_1, p_2, q_1, q_2 and t, define*

$$\begin{aligned} K(p, q, t) := \{&\Delta = \Delta_{\mathcal{L}} + \Delta_{\mathcal{M}} |\ \|\Delta_{\mathcal{L}}\|_{TNN} \le p_1\|\Delta_{\mathcal{L}}\|_F + p_2\|\Delta_{\mathcal{M}}\|_F, \\ &\|\Delta_{\mathcal{M}}\|_1 \le q_1\|\Delta_{\mathcal{L}}\|_F + q_2\|\Delta_{\mathcal{M}}\|_F, \|\Delta\|_\infty = 1, \|\Delta_{\mathcal{L}}\|_F^2 + \|\Delta_{\mathcal{M}}\|_F^2 \ge t\mu_1 n_1 n_2 n_3\}, \end{aligned} \tag{5.6}$$

where $p := (p_1, p_2)$ and $q := (q_1, q_2)$. Denote $\beta_{\mathcal{S}} := (\beta_{\mathcal{L}}^2 p_1^2 + \beta_{\mathcal{L}}^2 p_2^2 + \beta_{\mathcal{M}}^2 q_1^2 + \beta_{\mathcal{M}}^2 q_2^2)^{\frac{1}{2}}$. Then, it holds that for all $\Delta \in K(p, q, t)$,

$$\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2 \ge \mathbb{E}[\langle\Theta, \Delta\rangle^2] - \frac{\|\Delta_{\mathcal{L}}\|_F^2 + \|\Delta_{\mathcal{M}}\|_F^2}{2\mu_1 n_1 n_2 n_3} - 256\mu_1 n_1 n_2 n_3 \beta_{\mathcal{S}}^2 \tag{5.7}$$

with probability at least $1 - \frac{\exp[-mt^2\log(2)/64]}{1-\exp[-mt^2\log(2)/64]}$. In particular, the inequality (5.7) holds with probability at least $1 - \frac{1}{n_1+n_2+n_3}$ if $t = 8\sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$.

The proof of the Lemma 5.5 is given in Appendix D.4.

Proposition 5.6. *Suppose that Assumption 5.1 holds. Then, there exists $C_2 > 0$, such that, it holds that either*

$$\frac{\|\widetilde{\Delta}_{\mathcal{L}}\|_F^2 + \|\widetilde{\Delta}_{\mathcal{M}}\|_F^2}{n_1 n_2 n_3} \le 32(b_m + b_l)^2\mu_1\sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$$

or

$$\begin{aligned} \frac{\|\widetilde{\Delta}_{\mathcal{L}}\|_F^2 + \|\widetilde{\Delta}_{\mathcal{M}}\|_F^2}{n_1 n_2 n_3} \le &\frac{64 b_l^2}{n_1 n_2 n_3}\left[\frac{(d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)^2}{\lambda^2} + \left(\sqrt{s} + d_{\mathcal{M}}\sqrt{s} + \frac{\eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F}{\lambda}\right)^2\right] \\ &+ C_2\Big[\beta_{\mathcal{L}}^2(\sqrt{2r} + d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)^2 \\ &+ \beta_{\mathcal{L}}^2(\lambda d_{\mathcal{M}}\sqrt{s} + \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F)^2 + \frac{\beta_{\mathcal{M}}^2(d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)^2}{\lambda^2} \\ &+ \beta_{\mathcal{M}}^2\left(\sqrt{s} + d_{\mathcal{M}}\sqrt{s} + \frac{\eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F}{\lambda}\right)^2\Big] \end{aligned}$$

with probability at least $1 - \frac{1}{n_1+n_2+n_3}$.

Proof. Let $\widetilde{b} := \|\widetilde{\Delta}\|_\infty$. Since $(\mathcal{L}^c, \mathcal{M}^c)$ is the optimal and $(\mathcal{L}^\star, \mathcal{M}^\star)$ is feasible to the problem (4.3), we have $\|\widetilde{\Delta}_\mathcal{M}\|_\infty \le 2b_m$ and $\|\widetilde{\Delta}_\mathcal{L}\|_\infty \le \|\mathcal{L}^c\| + \|\mathcal{L}^\star\| \le 2b_l$. Hence, $\widetilde{b} \le \|\widetilde{\Delta}_\mathcal{L}\|_\infty + \|\widetilde{\Delta}_\mathcal{M}\|_\infty \le 2(b_m + b_l)$. We consider the following two cases:

Case 1: Suppose that $\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2 \le 8\widetilde{b}^2\mu_1 n_1 n_1 n_2 n_3 \sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$. Then we immediately obtain that

$$\frac{\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2}{n_1 n_2 n_3} \le 32(b_m + b_l)^2 \mu_1 \sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}.$$

Case 2: Suppose that $\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2 \ge 8\widetilde{b}^2\mu_1 n_1 n_1 n_2 n_3 \sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$. It follows from the definition of $\widetilde{b}$ that $\widetilde{\Delta}/\widetilde{b} \in K(p,q,t)$, where $t = 8\sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$, and $p = (p_1, p_2)$ and $q = (q_1, q_2)$ are given in Lemma 5.4. Due to (5.4) and Lemma 5.5, we obtain that with probability at least $1 - \frac{1}{n_1+n_2+n_3}$,

$$\frac{\|\widetilde{\Delta}\|_F^2}{n_1 n_2 n_3} \le \frac{\mu_1}{m}\|\mathcal{P}_\Omega(\widetilde{\Delta})\|_F^2 + \frac{\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2}{2n_1 n_2 n_3} + 256\mu_1^2 n_1 n_2 n_3 \beta_\mathcal{S}^2 \widetilde{b}^2. \tag{5.8}$$

Since $(\mathcal{L}^c, \mathcal{M}^c)$ is the optimal solution of (4.3) and $(\mathcal{L}^\star, \mathcal{M}^\star)$ is the true tensor, we obtain $\mathcal{P}_\Omega(\widetilde{\Delta}) = 0$. In addition, due to $\|\widetilde{\Delta}_\mathcal{L}\|_\infty \le 2b_l$, we then derive from (5.3) that

$$\begin{aligned}
\|\widetilde{\Delta}\|_F^2 \ge& \|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2 - 2\|\widetilde{\Delta}_\mathcal{L}\|_\infty \|\widetilde{\Delta}_\mathcal{M}\|_1 \\
\ge& \|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2 - 4b_l(q_1\|\widetilde{\Delta}_\mathcal{L}\|_F + q_2\|\widetilde{\Delta}_\mathcal{M}\|_F) \\
\ge& \|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2 - 16b_l^2(q_1^2 + q_2^2) - \frac{\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2}{4} \\
=& \frac{3}{4}(\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2) - 16b_l^2(q_1^2 + q_2^2).
\end{aligned} \tag{5.9}$$

By combining (5.8) with (5.9), we obtain that

$$\frac{\|\widetilde{\Delta}_\mathcal{L}\|_F^2 + \|\widetilde{\Delta}_\mathcal{M}\|_F^2}{n_1 n_2 n_3} \le \frac{64b_l^2(q_1^2 + q_2^2)}{n_1 n_2 n_3} + 1024\mu_1^2 n_1 n_2 n_3 \beta_\mathcal{S}^2 \widetilde{b}^2. \tag{5.10}$$

Recall that $\beta_\mathcal{S} := (\beta_\mathcal{L}^2 p_1^2 + \beta_\mathcal{L}^2 p_2^2 + \beta_\mathcal{M}^2 q_1^2 + \beta_\mathcal{M}^2 q_2^2)^{\frac{1}{2}}$. By plugging this together with Lemma 5.4 into (5.10) and taking $C_2 := 4096\mu_1^2 n_1 n_2 n_3 (b_m + b_l)^2$, we complete the proof. ∎

For the third-order tensor, we need to avoid the case that each fiber is sampled with very high probability. Let $R_{:jk} := \Sigma_{i=1}^{n_1} p_{ijk}$, $C_{i:k} := \Sigma_{j=1}^{n_2} p_{ijk}$, $T_{ij:} := \Sigma_{k=1}^{n_3} p_{ijk}$, the following assumption is used to avoid this situation.

Assumption 5.2. *There exists a positive constant $\mu_2 \ge 1$ such that* $\max_{\{i,j,k\}}\{R_{:jk}, C_{i:k}, T_{ij:}\} \le \frac{\mu_2}{\min\{n_1,n_2,n_3\}}$.

We now estimate an upper bound of $\mathbb{E}\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\|$. First, we give a brief introduction about Orlicz ψ_s-norm. Given any $s \ge 1$, the Orlicz ψ_s-norm of a random variable z is defined by $\|z\|_{\psi_s} := \inf\{t > 0 | \mathbb{E}\exp(|z|^s/t^s) \le 2\}$. The proofs of the followings two lemmas are given in Appendix D.5 and Appendix D.6, respectively.

Lemma 5.7. *Under Assumption 5.2, for $m \geq \widetilde{n}\log((n_1+n_2)n_3)(\log(\widetilde{n}))^2/\mu_2$, there exists a positive constant C_1 such that*

$$\beta_{\mathcal{L}} = \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}^*_\Omega(\epsilon)\right\| \leq C_1\sqrt{\frac{3e\mu_2\log((n_1+n_2)n_3)}{\widetilde{n}m}},$$

where $\widetilde{n} := \min\{n_1, n_2\}$.

Lemma 5.8. *There exist $C > 0$ and $M > 0$ that depend on the Orlicz ψ_1-norm of ϵ_l such that*

$$\beta_{\mathcal{M}} = \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}^*_\Omega(\epsilon)\right\|_\infty \leq \frac{M(\log(2m)+1)}{Cm}.$$

We first define two fundamental terms

$$\begin{cases}\Upsilon_1 := (\frac{d_{\mathcal{L}}\sqrt{r}+\eta\|\mathcal{L}^\star-\mathcal{L}^k\|_F}{\lambda})^2 + (\sqrt{\bar{s}} + d_{\mathcal{M}}\sqrt{\bar{s}} + \frac{\eta\|\mathcal{M}^\star-\mathcal{M}^k\|_F}{\lambda})^2,\\ \Upsilon_2 := (\sqrt{2r} + d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)^2 + (d_{\mathcal{M}}\sqrt{\bar{s}}\lambda + \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F)^2.\end{cases}$$

By combining Proposition 5.6 with Lemma 5.7 and Lemma 5.8, we can easily establish the following error bound results.

Theorem 5.9. *Suppose that Assumption 5.1 and Assumption 5.2 hold. Then, for $m \geq \widetilde{n}\log((n_1+n_2)n_3)(\log(\widetilde{n}))^2/\mu_2$, there exist constants $C > 0$, $C_1 > 0$ and $C_2 > 0$ such that*

$$\begin{aligned}&\frac{\|\widetilde{\Delta}_{\mathcal{L}}\|_F^2 + \|\widetilde{\Delta}_{\mathcal{M}}\|_F^2}{n_1n_2n_3}\\ \leq&\frac{64b_l^2}{n_1n_2n_3}\Upsilon_1 + C_2\left[\frac{C_1^2 3e\mu_2\log((n_1+n_2)n_3)}{\widetilde{n}m}\Upsilon_2 + \left(\frac{M(\log(2m)+1)}{Cm}\right)^2\Upsilon_1\right]\end{aligned} \tag{5.11}$$

with probability at least $1 - \frac{1}{n_1+n_2+n_3}$.

When $H_1 \equiv 0$, $H_2 \equiv 0$ and $\eta \equiv 0$, the error bound in Theorem 5.9 is just the error bound of the CRTC problem (3.17). From Theorem 5.9, we can see that the second term in the maximum of (5.11) dominates the first term. Thus, the error bound is dominated by the second term. Now, we denote the second term as $\mathfrak{L}_m$. In fact, when $H_1 \equiv 0$ and $H_2 \equiv 0$, we obtain that $d_{\mathcal{L}} = 1$ and $d_{\mathcal{M}} = 1$ according to (5.2). In this case, we denote the second term as $\mathfrak{L}_m'$. Note that $\mathfrak{L}_m < \mathfrak{L}_m'$ when $d_{\mathcal{L}} < 1$ and $d_{\mathcal{M}} < 1$.

Let $\widehat{\boldsymbol{U}_1^k}^{(i)}$ and $\widehat{\boldsymbol{V}_1^k}^{(i)}$ denote the first r_i columns of $\widehat{\boldsymbol{U}^k}^{(i)}$ and $\widehat{\boldsymbol{V}^k}^{(i)}$. Next, we show that the error bound of (4.3) is lower than that of (3.17), i.e., $d_{\mathcal{L}} < 1$ and $d_{\mathcal{M}} < 1$.

Theorem 5.10. *Let $\varepsilon_{\nabla H_1}(\widehat{\boldsymbol{L}^k}^{(i)}) := \frac{1}{\sqrt{r_i}}\left\|\widehat{\nabla H_1(\mathcal{L}^k)}^{(i)} - \widehat{\boldsymbol{U}_1^k}^{(i)}(\widehat{\boldsymbol{V}_1^k}^{(i)})^H\right\|_F$ for $i = 1, \cdots, n_3$, and assume that*

$$\frac{\|\widehat{\boldsymbol{L}^k}^{(i)} - \widehat{\boldsymbol{L}^\star}^{(i)}\|_F}{\sigma_{r_i}(\widehat{\boldsymbol{L}^\star}^{(i)})} < \min\left\{\frac{1}{\sqrt{2}}\left(1 - \exp\left(-\sqrt{2r_i}\left(1 - \varepsilon_{\nabla H_1}(\widehat{\boldsymbol{L}^k}^{(i)})\right)\right)\right), \frac{1}{2}\right\}, \tag{5.12}$$

then $d_{\mathcal{L}} < 1$.

Proof. Let $\widehat{\boldsymbol{L}^\star}^{(i)} = \boldsymbol{U}^{(i)}\boldsymbol{S}^{(i)}(\boldsymbol{V}^{(i)})^H$ with $\boldsymbol{U}^{(i)} = [\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}]$ and $\boldsymbol{V}_i = [\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}]$, $\boldsymbol{U}_1^{(i)} \in \mathbb{C}^{n_1\times r_i}$, $\boldsymbol{V}_1^{(i)} \in \mathbb{C}^{n_2\times r_i}$, for $i = 1, \cdots, n_3$. Note that

$$\|\widehat{\boldsymbol{U}_1^k}^{(i)}(\widehat{\boldsymbol{V}_1^k}^{(i)})^H - \boldsymbol{U}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H\|_F \le -\frac{1}{\sqrt{2}}\log\left(1-\sqrt{2}\frac{\|\widehat{\boldsymbol{L}^k}^{(i)} - \widehat{\boldsymbol{L}^\star}^{(i)}\|_F}{\sigma_{r_i}(\widehat{\boldsymbol{L}^\star}^{(i)})}\right) < \sqrt{r_i}(1-\varepsilon_{\nabla H_1}(\widehat{\boldsymbol{L}^k}^{(i)})),$$

where the first inequality follows from the proof of [31, Theorem 3] and the second inequality is due to the inequality (5.12). So we obtain

$$\begin{aligned}\|\widehat{\nabla H_1(\mathcal{L}^k)}^{(i)} - \boldsymbol{U}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H\|_F \le&\|\widehat{\nabla H_1(\mathcal{L}^k)}^{(i)} - \widehat{\boldsymbol{U}_1^k}^{(i)}(\widehat{\boldsymbol{V}_1^k}^{(i)})^H\|_F + \|\widehat{\boldsymbol{U}_1^k}^{(i)}(\widehat{\boldsymbol{V}_1^k}^{(i)})^H - \boldsymbol{U}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H\|_F\\ <&\sqrt{r_i}\varepsilon_{\nabla H_1}(\widehat{\boldsymbol{L}^k}^{(i)}) + \sqrt{r_i}(1-\varepsilon_{\nabla H_1}(\widehat{\boldsymbol{L}^k}^{(i)})) = \sqrt{r_i}.\end{aligned}$$

On the other hand, it follows from $\widehat{\boldsymbol{U}_1}^{(i)} = [\boldsymbol{U}_1^{(i)}, 0] \in \mathbb{C}^{n_1\times r_{\max}}$ and $\widehat{\boldsymbol{V}_1}^{(i)} = [\boldsymbol{V}_1^{(i)}, 0] \in \mathbb{C}^{n_2\times r_{\max}}$ that

$$d_{\mathcal{L}}^2 = \frac{1}{r}\|\mathcal{U}_1 * \mathcal{V}_1^H - \nabla H_1(\mathcal{L}^k)\|_F^2 = \frac{1}{rn_3}\sum_{i=1}^{n_3}\|\widehat{\nabla H_1(\mathcal{L}^k)}^{(i)} - \widehat{\boldsymbol{U}_1}^{(i)}(\widehat{\boldsymbol{V}_1}^{(i)})^H\|_F^2 < \frac{1}{rn_3}\sum_{i=1}^{n_3} r_i = 1.$$

This completes the proof. ∎

Theorem 5.10 guarantees that $d_{\mathcal{L}} < 1$ if the estimator $\mathcal{L}^k$ does not deviate too much from $\mathcal{L}^\star$.

Remark 5.11. Theorem 5.10 removes the rank constraint condition $r_1 < \frac{6}{4n_3-7}(r_2 + \cdots + r_{n_3})$ in [54, Lemma 4.2].

Theorem 5.12. *Let $\boldsymbol{M}^\star := Diag(vec(\mathcal{M}^\star))$, $\boldsymbol{M}^k := Diag(vec(\mathcal{M}^k))$, and $\varepsilon_{\nabla H_2}(\mathcal{M}^k) := \frac{1}{\sqrt{\widetilde{s}}}\|\nabla H_2(\mathcal{M}^k) - sign(\mathcal{M}^k)\|_F$. Assume that*

$$\frac{\|\boldsymbol{M}^k - \boldsymbol{M}^\star\|_F}{\sigma_{\widetilde{s}}(\boldsymbol{M}^\star)} < \min\left\{\frac{1}{\sqrt{2}}(1-\exp(-\sqrt{2\widetilde{s}}(1-\varepsilon_{\nabla H_2}(\mathcal{M}^k)))), \frac{1}{2}\right\},$$

where $\sigma_{\widetilde{s}}(\boldsymbol{M}^\star) := \min\{|\mathcal{M}_{ijk}^\star| |\mathcal{M}_{ijk}^\star \neq 0\}$. Then, we have $d_{\mathcal{M}} < 1$.

Proof. We can obtain the following decomposition

$$\begin{aligned}\boldsymbol{M}^\star =&\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}(\mathcal{M}^\star)))\mathrm{Diag}(\mathrm{vec}(|\mathcal{M}^\star|))\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}^2(\mathcal{M}^\star)))\\ =&\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}(\mathcal{M}^\star)))\boldsymbol{P}_1\boldsymbol{P}_2...\boldsymbol{P}_{\widetilde{s}}\mathrm{Diag}(\pi(\mathrm{vec}(|\mathcal{M}^\star|)))\\ &\boldsymbol{P}_{\widetilde{s}}^H\boldsymbol{P}_{\widetilde{s}-1}^H...\boldsymbol{P}_1^H\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}^2(\mathcal{M}^\star))),\end{aligned}$$

where $\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_{\widetilde{s}}$ are elementary transformation matrices. Let $\boldsymbol{M}^\star = \boldsymbol{U}^\star\boldsymbol{\Sigma}^\star(\boldsymbol{V}^\star)^H$ be the SVD, where $\boldsymbol{U}^\star = [\boldsymbol{U}_1^\star\ \boldsymbol{U}_2^\star]$, $\boldsymbol{V}^\star = [\boldsymbol{V}_1^\star\ \boldsymbol{V}_2^\star]$, $\boldsymbol{U}_1^\star \in \mathbb{R}^{n_1n_2n_3\times\widetilde{s}}$ and $\boldsymbol{V}_1^\star \in \mathbb{R}^{n_1n_2n_3\times\widetilde{s}}$. This implies that

$$\begin{aligned}\boldsymbol{U}_1^\star(\boldsymbol{V}_1^\star)^H =&[\boldsymbol{U}_1^\star\ 0]\begin{bmatrix}(\boldsymbol{V}_1^\star)^H\\ 0\end{bmatrix} = \boldsymbol{U}^\star(\boldsymbol{V}^\star)^H\\ =&\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}(\mathcal{M}^\star)))\boldsymbol{P}_1\boldsymbol{P}_2...\boldsymbol{P}_{\widetilde{s}}\boldsymbol{P}_{\widetilde{s}}^H\boldsymbol{P}_{\widetilde{s}-1}^H...\boldsymbol{P}_1^H\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}^2(\mathcal{M}^\star)))\\ =&\mathrm{Diag}(\mathrm{vec}(\mathrm{sign}(\mathcal{M}^\star))).\end{aligned} \tag{5.13}$$

Notice that $\sigma_{\widetilde{s}}(\boldsymbol{M}^\star) = \min\{|\mathcal{M}^\star_{ijk}||\mathcal{M}^\star_{ijk} \neq 0\}$, we have

$$\begin{aligned} d_{\mathcal{M}} =& \frac{1}{\sqrt{\widetilde{s}}}\|\nabla H_2(\mathcal{M}^k) - \mathrm{sign}(\mathcal{M}^\star)\|_F = \frac{1}{\sqrt{\widetilde{s}}}\|\mathrm{Diag}(\mathrm{vec}(\nabla H_2(\mathcal{M}^k))) - \mathrm{Diag}(\mathrm{vec}(\mathrm{sign}(\mathcal{M}^\star)))\|_F \\ =& \frac{1}{\sqrt{\widetilde{s}}}\|\mathrm{Diag}(\mathrm{vec}(\nabla H_2(\mathcal{M}^k))) - \boldsymbol{U}_1^\star(\boldsymbol{V}_1^\star)^H\|_F \\ \leq& -\frac{1}{\sqrt{2\widetilde{s}}}\log\left(1 - \sqrt{2}\frac{\|\boldsymbol{M}^k - \boldsymbol{M}^\star\|_F}{\sigma_{\widetilde{s}}(\boldsymbol{M}^\star)}\right) + \varepsilon_{\nabla H_2}(\mathcal{M}^k) < 1, \end{aligned}$$

where the third equation follows from (5.13), and the first inequality follows from [31, Theorem 3]. ■

The above theorem demonstrates that $d_{\mathcal{M}} < 1$ if $\mathcal{M}^k$ does not deviate too much from $\mathcal{M}^\star$.

Now, we analyze the constructions of ∇H_1 and ∇H_2. In order to get a small error bound, according to Theorem 5.9, we desire $d_{\mathcal{L}}$ and $d_{\mathcal{M}}$ as small as possible, i.e., $\nabla H_1(\mathcal{L}^k)$ is close to $\mathcal{U}_1 * \mathcal{V}_1^H$ and $\nabla H_2(\mathcal{M}^k)$ is close to $\mathrm{sign}(\mathcal{M}^\star)$. Firstly, let $\nabla H_1(\mathcal{L}^k) = \mathcal{U}^k * \mathcal{R}^k * (\mathcal{V}^k)^H$, where $\mathcal{U}^k = [\mathcal{U}_1^k\ \mathcal{U}_2^k]$ and $\mathcal{V}^k = [\mathcal{V}_1^k\ \mathcal{V}_2^k]$ with $\mathcal{U}_1^k \in \mathbb{R}^{n_1\times r_{\max}\times n_3}$ and $\mathcal{V}_1^k \in \mathbb{R}^{n_2\times r_{\max}\times n_3}$. If $\mathcal{L}^k$ is close to $\mathcal{L}^\star$, we desire $\nabla H_1(\mathcal{L}^k)$ is close to $\mathcal{U}_1^k * (\mathcal{V}_1^k)^H$. Notice from (3.13) that

$$h'(x) := \begin{cases} \frac{x}{\gamma}, & |x| \leq \gamma, \\ \mathrm{sign}(x), & |x| > \gamma. \end{cases} \tag{5.14}$$

It is observed from (5.14) that the function h' is S-shaped with two inflection points at $\pm\gamma$ and the parameter γ mainly controls the shape of h', the steepness of h' increase when γ decrease. So, there exist some $\gamma \in (0, b_l]$ such that the following property holds:

$$(\nabla g(\sigma(\widehat{\boldsymbol{L}^k}^{(i)})))_j = h'(\sigma_j(\widehat{\boldsymbol{L}^k}^{(i)})) \approx \begin{cases} 1, & 1 \leq j \leq r_i, \\ 0, & \text{otherwise}, \end{cases} \quad \forall i = 1, \ldots, n_3. \tag{5.15}$$

Similarly, the SVD of $\boldsymbol{M}^k$ is given by $\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{V}})^H$. Let $\widetilde{\boldsymbol{U}}_1$ and $\widetilde{\boldsymbol{V}}_1$ denote the first $\widetilde{s}$ columns of $\widetilde{\boldsymbol{U}}$ and $\widetilde{\boldsymbol{V}}$. If $\mathcal{M}^k$ is close to $\mathcal{M}^\star$, we desire $\mathrm{Diag}(\mathrm{vec}(\nabla H_2(\mathcal{M}^k)))$ is close to $\widetilde{\boldsymbol{U}}_1\widetilde{\boldsymbol{V}}_1^H$. So, there also exist some $\gamma \in (0, b_m]$ such that the following property holds:

$$h'(\boldsymbol{M}^k_{jj}) \approx \begin{cases} 1, & \boldsymbol{M}^k_{jj} > 0, \\ -1, & \boldsymbol{M}^k_{jj} < 0, \\ 0, & \text{otherwise}. \end{cases} \tag{5.16}$$

Remark 5.13. Notice that if ∇H_1 and ∇H_2 are obtained from the derivative of (3.15), i.e.,

$$h'(x) := \begin{cases} 0, & |x| \leq \gamma_1, \\ \frac{x - \gamma_1 \mathrm{sign}(x)}{\gamma_2 - \gamma_1}, & \gamma_1 < |x| \leq \gamma_2, \\ \mathrm{sign}(x), & |x| > \gamma_2, \end{cases} \tag{5.17}$$

then, the properties (5.15) and (5.16) hold. And the results can also be established if ∇H_1 and ∇H_2 are chosen as the correction function in [31].

Remark 5.14. By numerical experiments, we verify that $d_{\mathcal{L}} < 1$ and $d_{\mathcal{M}} < 1$ when h is chosen as the one in (3.13). The relevant results can be found in Table 1.

6. Numerical Experiments.

In this section, we present numerical experiments to show the effectiveness of our BCNRTC method in recovering color images and multispectral images, and compare it with the Robust Tensor Ring Completion (RTRC) [17], the Robust Tensor Completion (RTCℓ_1) [18] and the Nonconvex Robust Tensor Completion (NCRTC) [58]. The RTCℓ_1 model is a convex model and the NCRTC model is nonconvex, which gives the nonconvex approximation of the sparse term compared to the RTCℓ_1. The superior performance of NCRTC compared to the RTCℓ_1 in terms of recovery quality has been demonstrated in [58] via extensive numerical results. To show the effectiveness of the BCNRTC more clearly, we also present results of RTCℓ_1. For fair comparisons, the parameters in each method are tuned to give optimal performance. All experiments are performed on an Intel i7-2600 CPU desktop computer with 8 GB of RAM and MATLAB R2020a.

We define the sample ratio (SR) as SR:$=\frac{|\Omega|}{n_1n_2n_3}$ for an $n_1 \times n_2 \times n_3$ tensor, where Ω is generated uniformly at random and $|\Omega|$ represents the cardinality of Ω. Meanwhile, we use α to represent the impulse noise level. For each tensor, we randomly add the salt-and-pepper impulse noise with ratio α, and the observed tensor $\mathcal{P}_\Omega(\mathcal{X})$ is generated by the given SR.

To evaluate the performance of different methods, the peak signal-to-noise ratio (PSNR) is used to measure the quality of the recovered tensors, which is defined as follows:

$$\mathrm{PSNR}(\mathcal{L}) := 10\log_{10}\frac{n_1n_2n_3(\max_{i,j,k}\mathcal{L}^\star - \min_{i,j,k}\mathcal{L}^\star)^2}{\|\mathcal{L}^\star - \mathcal{L}\|_F^2},$$

where $\mathcal{L}$ and $\mathcal{L}^\star$ are the recovered tensor and the ground-truth tensor, respectively. The relative error (RE) between the recovered and the true tensor is defined by RE $:= \frac{\|\mathcal{L}-\mathcal{L}^\star\|_F}{\|\mathcal{L}^\star\|_F}$.

6.1. Stopping Criteria.

6.1.1. The stopping criterion for the PMM algorithm.

For the nonconvex BCNRTC model (3.11), we adopt the relative KKT residual

$$\eta_{\mathrm{kkt}} := \max\{\eta_{\mathcal{L}}, \eta_{\mathcal{M}}, \eta_P\} \leq 3\times 10^{-3} \tag{6.1}$$

to measure the accuracy of an approximate optimal solution obtained by the PMM algorithm, where

$$\begin{aligned}
&\eta_P := \frac{\|\mathcal{L}+\mathcal{M}-\mathcal{Z}\|_F}{1+\|\mathcal{Z}\|_F+\|\mathcal{L}\|_F+\|\mathcal{M}\|_F},\ \eta_{\mathcal{L}} := \frac{\|\mathcal{L}-\mathrm{Prox}_{\|\cdot\|_{\mathrm{TNN}}+\delta_{D_2}(\cdot)}(\mathcal{Y}+\mathcal{L}+\nabla H_1(\mathcal{L}))\|_F}{1+\|\mathcal{Y}\|_F+\|\mathcal{L}\|_F+\|\nabla H_1(\mathcal{L})\|_F},\\
&\eta_{\mathcal{M}} := \frac{\|\mathcal{M}-\mathrm{Prox}_{\lambda\|\cdot\|_1+\delta_{D_1}(\cdot)}(\mathcal{Y}+\mathcal{M}+\lambda\nabla H_2(\mathcal{M}))\|_F}{1+\|\mathcal{Y}\|_F+\|\mathcal{M}\|_F+\|\lambda\nabla H_2(\mathcal{M})\|_F}
\end{aligned} \tag{6.2}$$

with

$$\mathrm{Prox}_{\lambda f}(\mathbf{x}) := \arg\min_{\mathbf{w}\in\mathbb{R}^p} f(\mathbf{w}) + \frac{1}{2\lambda}\|\mathbf{w}-\mathbf{x}\|_F^2$$

denoting the proximal mapping of f with parameter λ [35].

6.1.2. The stopping criterion for the sGS-ADMM algorithm. In order to evaluate the performance of sGS-ADMM for solving convex subproblem (4.7), we use the primal infeasibility η_P and relative duality gap defined by

$$\eta_{\text{gap}} := \frac{|\text{pobj} - \text{dobj}|}{1 + |\text{pobj}| + |\text{dobj}|},$$

where

$$\begin{aligned}\text{pobj} :=& \|\mathcal{L}\|_{\text{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \mathcal{L}\rangle + \lambda(\|\mathcal{M}\|_1 - \langle \nabla H_2(\mathcal{M}^k), \mathcal{M}\rangle) + \frac{\eta}{2}\|\mathcal{M} - \mathcal{M}^k\|_F^2 \\ &+ \frac{\eta}{2}\|\mathcal{L} - \mathcal{L}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{Z} - \mathcal{Z}^k\|_F^2,\end{aligned}$$

and

$$\begin{aligned}\text{dobj} :=& \lambda \min_{\|\mathcal{M}\|_\infty \le b_m} \left[\|\mathcal{M}\|_1 + \frac{\eta}{2\lambda}\left\| \mathcal{M} - \left(\mathcal{M}^k + \frac{\lambda \nabla H_2(\mathcal{M}^k) + \mathcal{Y}}{\eta}\right)\right\|_F^2 \right] - \frac{\eta}{2}\left\| \mathcal{L}^k + \frac{\mathcal{Y} + \nabla H_1(\mathcal{L}^k)}{\eta}\right\|_F^2 \\ &+ \min_{\|\mathcal{L}\| \le b_l} \left[\|\mathcal{L}\|_{\text{TNN}} + \frac{\eta}{2}\left\| \mathcal{L} - \left(\mathcal{L}^k + \frac{\mathcal{Y} + \nabla H_1(\mathcal{L}^k)}{\eta}\right)\right\|_F^2 \right] - \frac{\eta}{2}\left\| \mathcal{M}^k + \frac{\lambda \nabla H_2(\mathcal{M}^k) + \mathcal{Y}}{\eta}\right\|_F^2 \\ &+ \min_{\mathcal{P}_\Omega(\mathcal{X}) = \mathcal{P}_\Omega(\mathcal{Z})} \left[\frac{\eta}{2}\left\| \mathcal{Z} - \left(\mathcal{Z}^k - \frac{\mathcal{Y}}{\eta}\right)\right\|_F^2 \right] + \langle \mathcal{Y}, \mathcal{Z}^k\rangle + \frac{\eta}{2}\|\mathcal{L}^k\|_F^2 + \frac{\eta}{2}\|\mathcal{M}^k\|_F^2 - \frac{1}{2\eta}\|\mathcal{Y}\|_F^2\end{aligned}$$

are the primal and dual objective function values, respectively. For given tolerance Tol_S, we will terminate the sGS-ADMM when $\max\{\eta_{\text{gap}}, \eta_P\} \le \text{Tol}_\text{S}$ or the number of iterations reaches the maximum of 200. We initialize Tol_S^0 to be 3×10^{-2} and decrease it by a ratio, i.e., $\text{Tol}_\text{S}^{k+1} = \text{Tol}_\text{S}^k / 1.1$.

6.2. The Setting of Parameters. In order to improve the convergence speed of Algorithm 4.2, based on the KKT optimality conditions of problem (4.7), we adopt the following relative residuals of $\mathcal{L}$ and $\mathcal{M}$ to update the penalty parameter μ in the augmented Lagrangian function:

$$\begin{aligned}\eta_{D_1} &= \frac{\left\| \mathcal{L} - \text{Prox}_{\frac{1}{\eta}(\|\cdot\|_{\text{TNN}} + \delta_{D_2}(\cdot))}\left(\mathcal{L}^k + \frac{\mathcal{Y} + \nabla H_1(\mathcal{L}^k)}{\eta}\right)\right\|_F}{1 + \frac{1}{\eta}\|\mathcal{Y}\|_F + \|\mathcal{L}^k\|_F + \frac{1}{\eta}\|\nabla H_1(\mathcal{L}^k)\|_F}, \\ \eta_{D_2} &= \frac{\left\| \mathcal{M} - \text{Prox}_{\frac{1}{\eta}(\lambda\|\cdot\|_1 + \delta_{D_1}(\cdot))}\left(\mathcal{M}^k + \frac{\mathcal{Y} + \lambda\nabla H_2(\mathcal{M}^k)}{\eta}\right)\right\|_F}{1 + \frac{1}{\eta}\|\mathcal{Y}\|_F + \|\mathcal{M}^k\|_F + \frac{\lambda}{\eta}\|\nabla H_2(\mathcal{M}^k)\|_F},\end{aligned}$$

which is a similar strategy as [23]. Let $\eta_D := \max\{\eta_{D_1}, \eta_{D_2}\}$. Specifically, set $\mu^0 = 0.1$. At the t-th iteration, compute $\chi^{t+1} = \frac{\eta_P^{t+1}}{\eta_D^{t+1}}$ and then set

$$\mu^{t+1} = \begin{cases} \xi\mu^t, & \chi^{t+1} > 7, \\ \xi^{-1}\mu^t, & \frac{1}{\chi^{t+1}} > 7, \\ \mu^t, & \text{otherwise} \end{cases} \quad \text{with} \quad \xi = \begin{cases} 1.1, & \max\left\{\chi^{t+1}, \frac{1}{\chi^{t+1}}\right\} \le 50, \\ 2, & \max\left\{\chi^{t+1}, \frac{1}{\chi^{t+1}}\right\} > 500, \\ 1.5, & \text{otherwise.} \end{cases}$$

For the proximal term in the PMM algorithm, the parameter η^0 is initialized as 10^{-4} and gradually decreased by some factors $\varsigma \in (0,1)$, i.e., $\eta^{k+1} = \varsigma\eta^k$, where η^k denotes the penalty parameter value at the k-th PMM iteration.

In our following experiments, the function h in (3.13) which is related to the MCP function is used in both H_1 and H_2 for simplicity. Meanwhile, we use γ_1 and γ_2 to denote the parameters in H_1 and H_2, respectively. The parameters λ, γ_1 and γ_2 are sensitive to the recovery performance. For different sample ratios and different noise levels, we use the grid search method to get the best values of λ, γ_1 and γ_2 in terms of PSNR values of recovered images. These best values show that the value of λ depends on the sample ratio, noise level, γ_2 and the size of tensors. By using the data fitting method, we obtain the fitting function of λ, i.e., $\lambda = \frac{\tilde{c}}{\sqrt{SR\gamma_2\alpha n_3 \widetilde{m}}}$, where $\tilde{c}$ is chosen from $\{0.4, 0.5, 0.6, 0.7\}$ to get the best recovery performance. The parameter γ_1 is chosen as $10(1.2 - \text{SR})$ and γ_2 is chosen from $\{0.3, 0.4\}$, respectively. For practical problems, we adjust the above parameters slightly to obtain the best possible results. The step length τ in (4.12) can vary in the range$(0, (\sqrt{5}+1)/2)$ [25]. In our numerical test, we find that the larger the step length, the faster the convergence speed. Hence, we set $\tau = 1.618$ in all the experiments. In experiments, all testing images are normalized to [0, 1]. Therefore, we set $b_m = 1$ and $\|\mathcal{L}\|_\infty \leq 1$. According to the equivalence between norms, we have $\|\mathcal{L}\| \leq \sqrt{n_1 n_2} n_3 \|\mathcal{L}\|_\infty$. So we set $b_l = \sqrt{n_1 n_2} n_3$ in our numerical experiments.

As mentioned in Theorem 5.10 and Theorem 5.12, a lower recovery error bound can be obtained if the estimator $(\mathcal{L}^k, \mathcal{M}^k)$ in the PMM algorithm does not deviate from the ground-truth $(\mathcal{L}^\star, \mathcal{M}^\star)$ too much. Therefore, we use the solution obtained from solving the CRTC problem (3.17) as the initial estimator to warm-start our PMM algorithm. The sGS-ADMM is implemented to solve the CRTC method and will be terminated if (6.1) is satisfied or the number of iterations reaches the maximum of 200, where $\nabla H_1(\cdot)$ and $\nabla H_2(\cdot)$ in (6.2) vanish. We use the grid search method to get the best choice of λ, i.e., a value that gives nearly the highest possible PSNR value. And we use a similar strategy as [23] to update the penalty parameter μ.

6.3. Error Bounds and the Performance of the PMM Algorithm. In this subsection, we test error bounds and the performance of the PMM algorithm in different outer iterations. The test image is Pepper, and the test results are given in Table 1 which reports $d_{\mathcal{L}}$, $d_{\mathcal{M}}$, relative error and PSNR values of the CRTC and the first three outer iterations. In all experiments in Table 1, the stopping criterion of the PMM algorithm is achieved in the third outer iteration.

We can see from Table 1 that $d_{\mathcal{L}} = 1$ and $d_{\mathcal{M}} = 1$ in CRTC, and $d_{\mathcal{L}} < 1$ and $d_{\mathcal{M}} < 1$ in each outer iteration of PMM algorithm, which verifies the results of Theorem 5.10 and Theorem 5.12. The PMM algorithm substantially reduces $d_{\mathcal{L}}$ and $d_{\mathcal{M}}$ in the first iteration. The first outer iteration improves the recovery quality at least 33% in terms of the relative error with respect to the CRTC model.

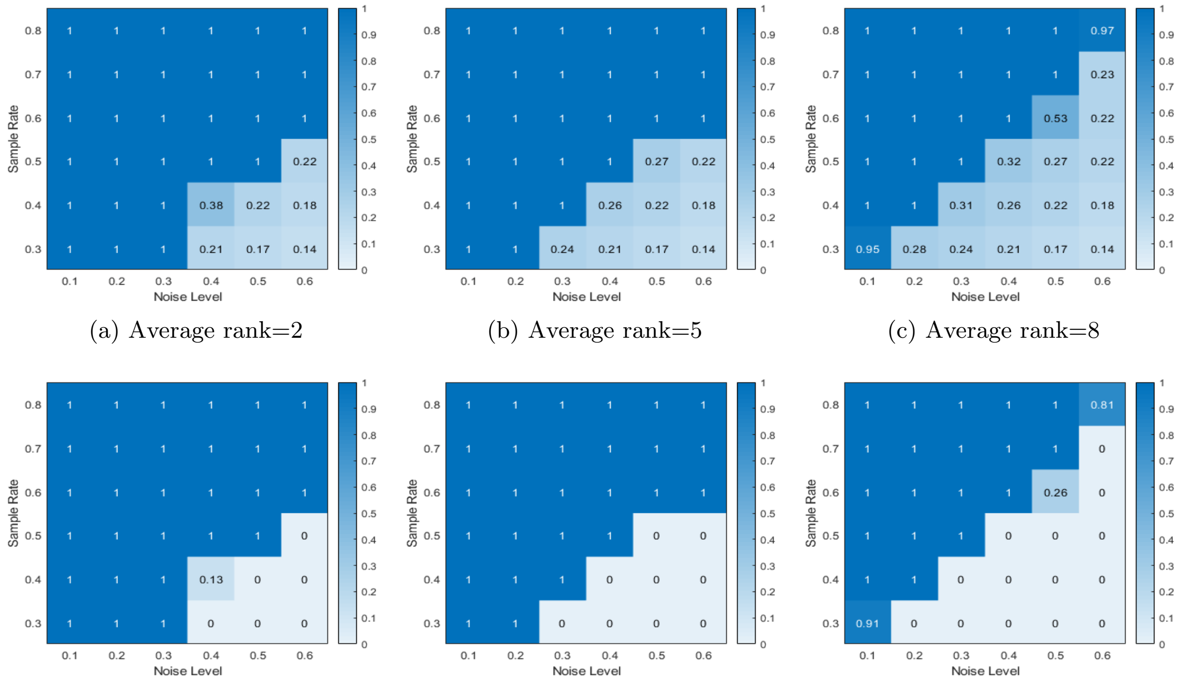
Table 1 also shows that $d_{\mathcal{L}}$ and $d_{\mathcal{M}}$ continue to decrease as the number of outer iterations increases, which implies that the upper error bounds in (5.11) in Theorem 5.9 continue to decrease. The PMM algorithm significantly improves the recovery quality in terms of both the relative error and the PSNR values.

Table 1

The values of $d_{\mathcal{L}}$, $d_{\mathcal{M}}$ and the performance of the PMM algorithm for Pepper image in different outer iterations with different sample ratios and noise levels.

SR	α		CRTC	1	2	3
0.8	0.2	$d_{\mathcal{L}}$	1	0.9432	0.923	0.9131
		$d_{\mathcal{M}}$	1	0.5317	0.5153	0.5104
		RE	0.0681	0.0393	0.0294	0.0257
		PSNR	29.27	34.04	36.56	37.72
	0.3	$d_{\mathcal{L}}$	1	0.963	0.9379	0.9262
		$d_{\mathcal{M}}$	1	0.5339	0.5195	0.5146
		RE	0.094	0.0584	0.0447	0.039
		PSNR	26.47	30.6	32.93	34.12
	0.4	$d_{\mathcal{L}}$	1	0.9817	0.9559	0.9451
		$d_{\mathcal{M}}$	1	0.5364	0.5241	0.5195
		RE	0.1279	0.0866	0.0692	0.0611
		PSNR	23.8	27.18	29.13	30.21
0.7	0.2	$d_{\mathcal{L}}$	1	0.952	0.935	0.926
		$d_{\mathcal{M}}$	1	0.6143	0.6011	0.5968
		RE	0.0773	0.0478	0.0377	0.0334
		PSNR	28.17	32.34	34.4	35.46
	0.3	$d_{\mathcal{L}}$	1	0.9672	0.9474	0.9386
		$d_{\mathcal{M}}$	1	0.6262	0.6201	0.619
		RE	0.1054	0.0668	0.0535	0.0491
		PSNR	25.47	29.43	31.37	32.11
	0.4	$d_{\mathcal{L}}$	1	0.9802	0.963	0.9552
		$d_{\mathcal{M}}$	1	0.6253	0.6213	0.6209
		RE	0.1415	0.0961	0.079	0.0727
		PSNR	22.91	26.28	27.98	28.7

6.4. Random data. In this section, we present the results to analyze the success ratio on random data. We present the colormap of 3-order random tensors $\mathcal{L}$ with size $100 \times 100 \times 30$ and all entries $\mathcal{L}_{ijk} \in [0, 1]$. The tensor average ranks are 2, 5 and 8, respectively. The sample ratio SR increases from 0.3 to 0.8 with increment 0.1 and the noise level α increases from 0.1 to 0.6 with increment 0.1. For each pair (SR, α), we simulate 100 test instances. We consider two kinds of success ratios. One is defined by the percentage of successful entries ($|\mathcal{L}_{ijk} - \mathcal{L}^{\star}_{ijk}| < 10^{-2}$) from total entries. The another is defined by the relative error. If the relative error is smaller than 10^{-2}, then the tensor recovery is regarded as successful and the success ratio is denoted by $1 (= 100\%)$. Figure 1 reports the fraction of successful recovery for each pair. The first row reports the success ratio defined by the percentage of successful entries from total entries, and the second row reports the success ratio defined by relative error. The success ratio in the second row is defined by 1 if the recovered tensor $\mathcal{L}$ satisfies $\|\mathcal{L} - \mathcal{L}^{\star}\|_F / \|\mathcal{L}^{\star}\|_F < 10^{-2}$, and defined by 0 for others. Figure 1 shows: (1) the recovery success ratio is higher when the average rank is smaller; (2) the tensor data is more difficult to recover when the sample rate is lower and the noise level is higher; (3) in some cases, the entire tensor is judged to be failed to recover, but there are still some entries that can be successfully recovered. Numerical results in Figure 1 also show that the rank and noise level of tensors greatly affect the recovery of tensors. For example, under the setting that the average rank is 8 and the noise level is 0.6, it's hard to recover the data with sample rates from 0.3 to 0.7.



(a) Average rank=2 (b) Average rank=5 (c) Average rank=8

(d) Average rank=2 (e) Average rank=5 (f) Average rank=8

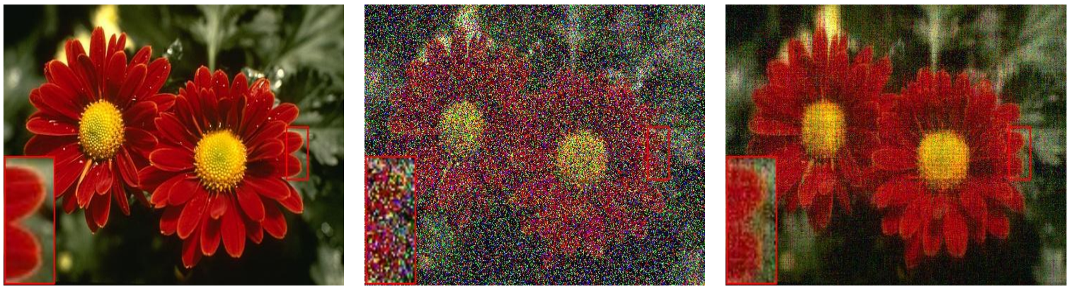
Figure 1. *The success ratio for varying sample ratio and noise level under different average ranks, where the success ratio in the first row is defined by the percentage of successful entries from total entries, and the success ratio in the second row is defined by relative error.*

6.5. Experiments on Color Images. In this subsection, we test color images including Pepper ($512 \times 512 \times 3$), Lena ($512 \times 512 \times 3$) [1] and Flower ($321 \times 481 \times 3$)[2]. Although the color images are not low-rank exactly, most information on each frontal slice of the color images is dominated by a few top singular values. In our experiments, these testing images are normalized on $[0, 1]$ and are all corrupted by removing arbitrary voxels and adding salt-and-pepper noise.

Figure 2 and Figure 3 show the recovered results and corresponding zoomed regions of RTRC, RTCℓ_1, NCRTC and BCNRTC. It can be observed that the BCNRTC performs better than others in terms of PSNR values and visual quality, where the BCNRTC preserves more details for Pepper image and many more sharp edges for Flower image than others.

In Table 2, we report the PSNR values of RTRC, RTCℓ_1, NCRTC and BCNRTC for three color images. We set SR = 0.6, 0.7 and 0.8 to illustrate the performance of methods and noise levels are considered as $\alpha \in \{0.2, 0.3, 0.4, 0.5\}$ simultaneously. It can be observed that the PSNR values obtained by our proposed BCNRTC model are much higher than those obtained by RTRC, RTCℓ_1 and NCRTC, especially for low noise levels. The PSNR values of the restored image by the BCNRTC increase at least 3dB relative to those of the RTCℓ_1 model.

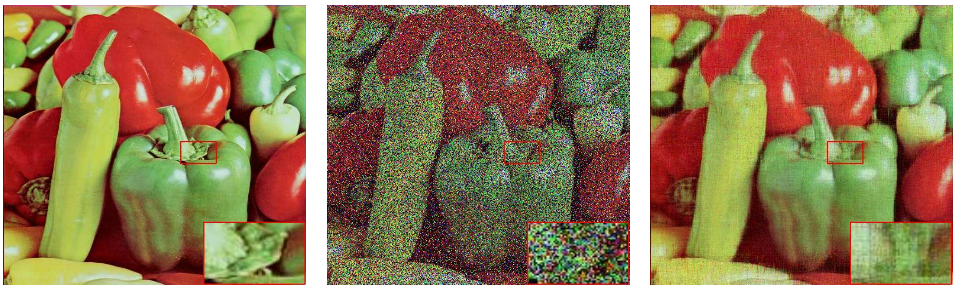
[1] http://sipi.usc.edu/database/
[2] https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/



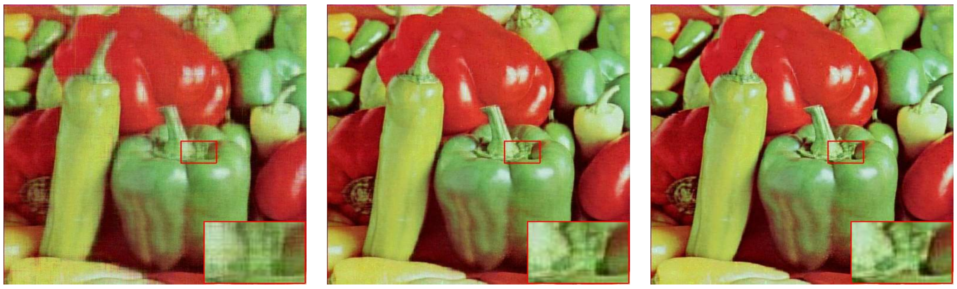
(a) Original (b) Observation (c) RTRC:17.07

(d) RTCℓ_1: 22.53 (e) NCRTC: 24.31 (f) BCNRTC: 26.53

Figure 2. *Recovered images (with PSNR(dB)) and zoomed regions of four different methods for the Flower image, where SR*$= 0.8$ *and* $\alpha = 0.4$.



(a) Original (b) Observation (c) RTRC:22.95



(d) RTCℓ_1: 25.16 (e) NCRTC: 29.74 (f) BCNRTC: 32.12

Figure 3. *Recovered images (with PSNR(dB)) and zoomed regions of four different methods for the Pepper image, where SR*$= 0.7$ *and* $\alpha = 0.3$.

Table 2

PSNR(dB) values for restoring results of different methods for color images corrupted by sample losing and salt-and-pepper noise. The boldface numbers are the best performance.

sample ratios	noise level	Pepper				Lena				Flower			
		RTRC	RTCℓ_1	NCRTC	BCNRTC	RTRC	RTCℓ_1	NCRTC	BCNRTC	RTRC	RTCℓ_1	NCRTC	BCNRTC
0.8	0.2	27.98	29.08	34.99	**37.72**	28.12	29.5	34.36	**36.31**	25.92	26.97	29.85	**32.54**
	0.3	24.15	26.09	31.24	**34.12**	24.78	26.98	31.48	**33.84**	23.68	24.64	26.85	**29.48**
	0.4	17.07	23.56	27.39	**30.21**	17.41	24.96	28.44	**30.6**	19.62	22.5	24.25	**26.37**
	0.5	11.66	21.25	23.87	**26.86**	11.79	23.07	25.26	**27.33**	14.9	20.36	21.72	**23.31**
0.7	0.2	27.01	27.85	32.82	**35.46**	27.25	28.43	32.58	**35.02**	25.17	26.02	28.55	**30.77**
	0.3	22.95	25.12	29.74	**32.11**	23.73	26.17	30.16	**31.98**	22.84	23.84	25.84	**28.03**
	0.4	16.11	22.71	25.98	**28.7**	16.44	24.3	27.29	**29.29**	18.88	21.75	23.37	**25.33**
	0.5	11.48	20.51	22.94	**25.1**	11.67	22.51	24.62	**26.48**	14.55	19.61	20.76	**22.11**
0.6	0.2	25.86	26.56	30.69	**33.31**	26.3	27.34	30.98	**32.92**	24.3	25.01	27.15	**29.07**
	0.3	21.6	24.09	27.98	**30.27**	22.52	25.32	28.72	**30.31**	21.86	22.94	24.8	**26.67**
	0.4	15.17	21.82	24.77	**27.05**	15.52	23.57	26.18	**27.96**	18.1	20.9	22.43	**24.17**
	0.5	11.32	19.75	21.94	**23.57**	11.54	21.81	23.86	**25.38**	14.19	18.82	19.69	**21.06**

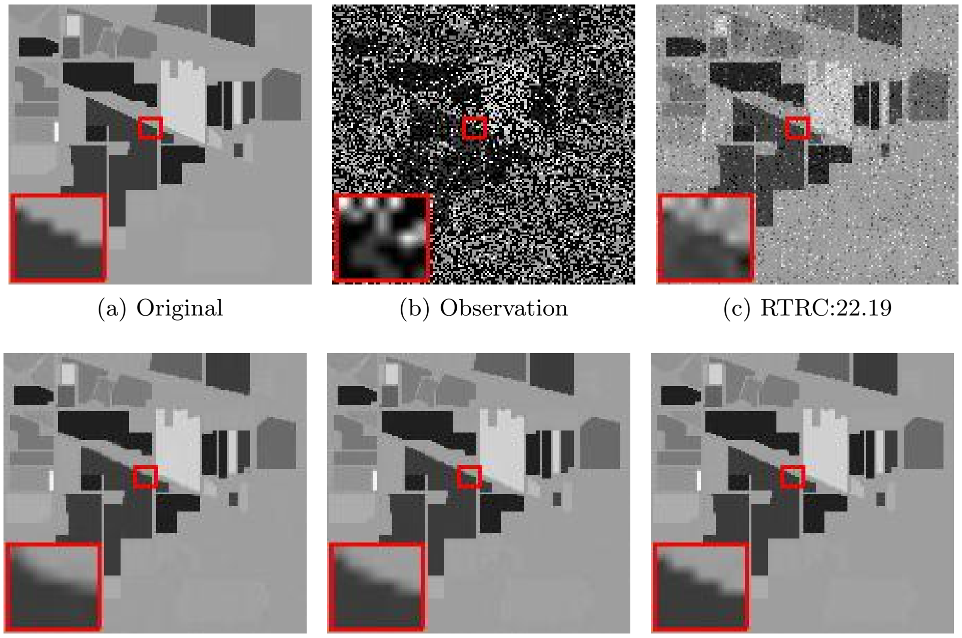
The performance of the nonconvex BCNRTC model can be improved greatly compared with that of the convex RTCℓ_1 model. The PSNR values of the restored image by the BCNRTC is at least 2dB higher than that of the nonconvex NCRTC model, which shows that both low-rank and sparse terms are nonconvex better than only sparse term is nonconvex.

6.6. Experiments on Multispectral Images. In this subsection, we test the multispectral images datasets including Cloth ($521 \times 521 \times 31$)[3] and the Indian Pines dataset ($145 \times 145 \times 224$)[4], which is a synthetic data. Since the Cloth dataset is too large, we resize the Cloth dataset to 128×128 in each image, and the size of the resulting tensor is $128 \times 128 \times 31$. This testing image is normalized on $[0, 1]$. For Multispectral Images, we compute the PSNR values between each ground-truth band and the recovered band, and then averaged them. This metric is denoted as mean PSNR (MPSNR).

In Figure 4, we show the 20-th band of the recovered images and corresponding zoomed regions of different methods for the Indian dataset, where SR$= 0.5$ and $\alpha = 0.2$. It is obvious that the details of the zoomed region obtained by BCNRTC are more clear than those obtained by RTRC and RTCℓ_1. The performance of NCRTC and BCNRTC is almost the same for the testing images in terms of visual quality. But PSNR values also show the BCNRTC is quite effective than NCRTC.

Table 3 presents detailed comparison results of four different methods for the two multispectral images with different sample ratios and noise levels, where the MPSNR values, the relative error (RE), the number of iterations (Iter) and the CPU time (in seconds) are given. Note that for the columns "Iter" and "Time" in the BCNRTC, we list the total inner sGS-ADMM iterations and CPU times outside brackets. Meanwhile, the values in brackets in this table mean the number of iterations and CPU times of CRTC for a warm start. In addition, the outer PMM iterations in Indian are four when SR$= 0.8$, 0.7, and the rest of cases are three. Table 3 shows the advantage of BCNRTC over other three methods no matter in terms of MPSNR values (largest) or relative errors (smallest). Meanwhile, the BCNRTC takes less

[3]https://www.cs.columbia.edu/CAVE/databases/multispectral/stuff/
[4]https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html



(a) Original (b) Observation (c) RTRC:22.19

(d) RTCℓ_1: 38.06 (e) NCRTC: 40.84 (f) BCNRTC: 45.19

Figure 4. *The 20-th band of recovered images (with PSNR(dB)) and zoomed regions of four different methods for the Indian dataset, where SR= 0.5 and $\alpha = 0.2$.*

Table 3
Numerical results of different methods for the multispectral images dataset with different SRs and α.

Images	α	SR	RTRC				RTCℓ_1				NCRTC				BCNRTC			
			MPSNR	RE	Iter	Time	MPSNR	RE	Iter	Time	MPSNR	RE	Iter	Time	MPSNR	RE	Iter	Time
Indian	0.2	0.8	23.19	1.17e-1	100	308	38.06	3.24e-2	68	349	42.7	2.54e-2	55	211	50.47	1.3e-2	26(26)	87(78)
		0.7	22.74	1.23e-1	100	291	37.87	3.26e-2	69	345	41.11	2.72e-2	57	219	48.66	1.46e-2	27(34)	89(100)
		0.6	22.25	1.3e-1	100	292	36.33	3.67e-2	69	339	39.61	2.97e-2	59	225	45.98	1.79e-2	24(35)	78(101)
		0.5	21.67	1.39e-1	100	295	35.39	3.92e-2	69	332	37.59	3.38e-2	59	225	43.74	2.03e-2	28(42)	89(119)
Cloth	0.4	0.8	18.34	5.53e-1	100	28	32.53	1.29e-1	58	20	37.18	7.39e-2	41	17	39.68	5.81e-2	33(15)	12(8)
		0.7	17.69	5.98e-1	100	27	31.25	1.42e-1	57	19	35.84	8.51e-2	42	17	38.14	6.67e-2	36(17)	13(6)
		0.6	17.45	6.14e-1	100	27	30.24	1.64e-1	58	19	34.1	1.02e-1	45	18	36.59	7.67e-2	40(17)	15(5)
		0.5	17.24	6.28e-1	100	27	28.96	1.88e-1	58	19	31.89	1.31e-1	50	20	34.85	1.25e-1	46(17)	17(5)

CPU time and iteration numbers than the others when a suitable initial point is given. Specifically, BCNRTC is able to outperform others by a factor of about 2-4 in terms of computation times for the Indian dataset.

7. Conclusions.

In this paper, we propose a BCNRTC model for the RTC problem which aims to recover a third-order low-rank tensor from partial observations corrupted by impulse noise. Then, we prove the equivalence of global solutions between RTC problems and our proposed nonconvex model, which gives the theoretical guarantee that the nonconvex penalties are superior to convex penalties. Due to the nonconvexity, the resulting model is difficult to solve. To tackle this problem, we devise the PMM algorithm to solve the nonconvex model and show that the sequence generated by the PMM algorithm globally converges to a critical point of the problem. Next, we establish a recovery error bound and give the theoretical guarantee that the proposed model can get lower error bounds when the initial estimator is close to the ground truth. Extensive numerical experiments including color images and multispectral images demonstrate that the proposed BCNRTC method outperforms several state-of-the-art methods.

In the future, it would be of great interest to extend the BCNRTC to higher-order tensors since some real datasets are higher-order tensors, such as color videos or traffic data.

Appendix A. Partial Calmless. The partial calmness is defined in detail in [28], which is used in the proof of Theorem 3.1. Let $\theta : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper lsc function, $h : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and Δ be a nonempty closed set of $\mathbb{R}^n$. Consider the following problem:

$$\text{(MP)} \quad \min_{\boldsymbol{z}} \{\theta(z) : h(\boldsymbol{z}) = 0, \boldsymbol{z} \in \Delta\}.$$

Let $\mathcal{F}$ and $\mathcal{F}^*$ denote the feasible set and the global optimal solution set of (MP), respectively, and $v^*(\text{MP})$ is the optimal value of (MP). Assume that $\mathcal{F}^* \neq \emptyset$. Consider the perturbed problem of (MP):

$$(\text{MP}_\epsilon) \quad \min_{\boldsymbol{z}} \{\theta(\boldsymbol{z}) : h(\boldsymbol{z}) = \epsilon, \boldsymbol{z} \in \Delta\},$$

where $\epsilon \in \mathbb{R}$, $\mathcal{F}_\epsilon$ denotes the feasible set of (MP_ϵ) associated to ϵ.

Definition A.1. *The problem (MP) is said to be partially calm at a solution point $\boldsymbol{z}^*$ if there exist $\varepsilon > 0$ and $\mu > 0$ such that for all $\epsilon \in [-\varepsilon, \varepsilon]$ and all $\boldsymbol{z} \in (\boldsymbol{z}^* + \varepsilon\mathbb{B}) \cap \mathcal{F}_\epsilon$, one has $\theta(\boldsymbol{z}) - \theta(\boldsymbol{z}^*) + \mu|h(\boldsymbol{z})| \geq 0$.*

The partial calmness plays a critical role in the proof of Theorem 3.1. [28, Proposition 2.1] shows that under the compactness of feasible set of problem (3.5), the partial calmness of (3.4) over its global optimal solution set implies the global exact penalization of (3.5).

Appendix B. The Kurdyka-Łojasiewicz property. The Kurdyka-Łojasiewicz property is defined in detailed in [3], which is used in the proof of Lemma 4.3.

Definition B.1. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper and lower semicontinuous function.*

(i) The function f is said to have the KL property at $\mathbf{x} \in \operatorname{dom}(\partial f)$ if there exist $\eta \in (0, +\infty]$, a neighborhood $\mathfrak{U}$ of $\mathbf{x}$ and a continuous concave function $\varphi : [0, \eta) \to [0, +\infty)$ such that: (a) $\varphi(0) = 0$; (b) φ is continuously differentiable on $(0, \eta)$, and continuous at 0; (c) $\varphi'(s) > 0$ for all $s \in (0, \eta)$; (d) for all $\mathbf{y} \in \mathfrak{U} \cap [\mathbf{y} \in \mathbb{R}^n : f(\mathbf{x}) < f(\mathbf{y}) < f(\mathbf{x}) + \eta]$, the following KL inequality holds:

$$\varphi'(f(\mathbf{y}) - f(\mathbf{x})) \operatorname{dist}(0, \partial f(\mathbf{y})) \geqslant 1.$$

(ii) If f satisfies the KL property at each point of $\operatorname{dom}(\partial f)$, then f is called a KL function.

Appendix C. Proofs of the results in Section 4. This part includes the proofs of part of results in Section 4.

C.1. Proof of Lemma 4.1. From the definition of Q, we have

$$Q(\mathcal{W}) - F(\mathcal{W};\mathcal{W}^k) = H_1(\mathcal{L}^k) - H_1(\mathcal{L}) + \langle \nabla H_1(\mathcal{L}^k), \mathcal{L} - \mathcal{L}^k\rangle + \lambda(H_2(\mathcal{M}^k) - H_2(\mathcal{M}) + \langle \nabla H_2(\mathcal{M}^k), \mathcal{M} - \mathcal{M}^k\rangle) - \frac{\eta}{2}\|\mathcal{W} - \mathcal{W}^k\|_F^2. \tag{C.1}$$

On the other hand, the convexity of H_1 and H_2 implies that

$$H_1(\mathcal{L}) \geq H_1(\mathcal{L}^k) + \langle \nabla H_1(\mathcal{L}^k), \mathcal{L} - \mathcal{L}^k\rangle, \quad H_2(\mathcal{M}) \geq H_2(\mathcal{M}^k) + \langle \nabla H_2(\mathcal{M}^k), \mathcal{M} - \mathcal{M}^k\rangle. \tag{C.2}$$

Combining (C.1) with (C.2), we obtain that $Q(\mathcal{W}) - F(\mathcal{W};\mathcal{W}^k) \leq -\frac{\eta}{2}\|\mathcal{W} - \mathcal{W}^k\|_F^2$. Thus, we obtain

$$Q(\mathcal{W}^{k+1}) + \frac{\eta}{2}\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F^2 \leq F(\mathcal{W}^{k+1};\mathcal{W}^k). \tag{C.3}$$

Since $\mathcal{C}^{k+1} \in \partial F(\mathcal{W}^{k+1};\mathcal{W}^k)$, we have

$$\begin{aligned} Q(\mathcal{W}^k) = F(\mathcal{W}^k;\mathcal{W}^k) &\geq F(\mathcal{W}^{k+1};\mathcal{W}^k) + \langle \mathcal{C}^{k+1}, \mathcal{W}^k - \mathcal{W}^{k+1}\rangle \\ &\geq F(\mathcal{W}^{k+1};\mathcal{W}^k) - \|\mathcal{C}^{k+1}\|_F\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F \\ &\geq F(\mathcal{W}^{k+1};\mathcal{W}^k) - \eta c\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F^2, \end{aligned} \tag{C.4}$$

where the last inequality follows from (4.4). Combining (C.3) with (C.4), we have

$$Q(\mathcal{W}^{k+1}) + \frac{\eta}{2}(1-2c)||\mathcal{W}^{k+1} - \mathcal{W}^k||_F^2 \leq Q(\mathcal{W}^k), \tag{C.5}$$

which completes the first part of the proof. Let N be a positive integer. Summing (C.5) from $k = 0$ to $N-1$, we get

$$\sum_{k=0}^{N-1}(\|\mathcal{L}^{k+1} - \mathcal{L}^k\|_F^2 + \|\mathcal{M}^{k+1} - \mathcal{M}^k\|_F^2) = \sum_{k=0}^{N-1}\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F^2 \leq \frac{2}{\eta(1-2c)}(Q(\mathcal{W}^0) - Q(\mathcal{W}^N)),$$

where the inequality is valid since the condition $\eta(1-2c) > 0$ holds. By the inequality (C.5), we can get the sequence $\{Q(\mathcal{W}^k)\}_{k\in\mathbb{N}}$ is non-increasing. Since $Q(\mathcal{W})$ is bounded below, the sequence $\{Q(\mathcal{W}^k)\}_{k\in\mathbb{N}}$ converges. Taking the limit as $N \to \infty$, we obtain that $\sum_{k=0}^{\infty}\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F^2 < \infty$ and the sequence $\{\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F\}_{k\in\mathbb{N}}$ converges to zero. Therefore, the conclusion is obtained.

C.2. Proof of Lemma 4.2. By [2, Proposition 2.1], [35, Exercise 8.8(c)] and $\mathcal{C}^{k+1} \in \partial F(\mathcal{W}^{k+1};\mathcal{W}^k)$, we have

$$\mathcal{C}_{\mathcal{L}}^{k+1} = \widetilde{\mathcal{Y}}^{k+1} - \nabla H_1(\mathcal{L}^k) + \eta(\mathcal{L}^{k+1} - \mathcal{L}^k), \quad \mathcal{C}_{\mathcal{M}}^{k+1} = \widetilde{\mathcal{Z}}^{k+1} - \nabla H_2(\mathcal{M}^k) + \eta(\mathcal{M}^{k+1} - \mathcal{M}^k) \tag{C.6}$$

for some $\widetilde{\mathcal{Y}}^{k+1} \in \partial_{\mathcal{L}}[\|\mathcal{L}\|_{\mathrm{TNN}} + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_2}(\mathcal{L})]_{\mathcal{W}=\mathcal{W}^{k+1}}$, $\widetilde{\mathcal{Z}}^{k+1} \in \partial_{\mathcal{M}}[\lambda\|\mathcal{M}\|_1 + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_1}(\mathcal{M})]_{\mathcal{W}=\mathcal{W}^{k+1}}$. From the definition of Q, we get

$$\partial_{\mathcal{L}}Q(\mathcal{W}) = \partial_{\mathcal{L}}[\|\mathcal{L}\|_{\mathrm{TNN}} + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_2}(\mathcal{L})] - \nabla H_1(\mathcal{L}),$$

$$\partial_{\mathcal{M}} Q(\mathcal{W}) = \partial_{\mathcal{M}}[\lambda\|\mathcal{M}\|_1 + \delta_{\Gamma_1}(\mathcal{L},\mathcal{M}) + \delta_{D_1}(\mathcal{M})] - \nabla H_2(\mathcal{M}).$$

By the definitions of $\widetilde{Y}^{k+1}$ and $\widetilde{Z}^{k+1}$, we obtain that

$$\mathcal{B}_{\mathcal{L}}^{k+1} := \widetilde{Y}^{k+1} - \nabla H_1(\mathcal{L}^{k+1}) \in \partial_{\mathcal{L}} Q(\mathcal{W}^{k+1}),\ \ \mathcal{B}_{\mathcal{M}}^{k+1} := \widetilde{Z}^{k+1} - \nabla H_2(\mathcal{M}^{k+1}) \in \partial_{\mathcal{M}} Q(\mathcal{W}^{k+1}).$$

Then, we have $\mathcal{B}^{k+1} \in \partial Q(\mathcal{W}^{k+1})$. Define

$$\mathcal{H}_{\mathcal{L}}^{k+1} := \widetilde{Y}^{k+1} - \nabla H_1(\mathcal{L}^k),\ \ \mathcal{H}_{\mathcal{M}}^{k+1} := \widetilde{Z}^{k+1} - \lambda\nabla H_2(\mathcal{M}^k). \tag{C.7}$$

We now have to estimate the norm of $\mathcal{B}^{k+1}$. By the definitions of $\mathcal{B}^{k+1}$ and $\mathcal{H}^{k+1}$, we have

$$\|\mathcal{B}^{k+1} - \mathcal{H}^{k+1}\|_F = \|(\nabla H_1(\mathcal{L}^k) - \nabla H_1(\mathcal{L}^{k+1}), \lambda(\nabla H_2(\mathcal{M}^k) - \nabla H_2(\mathcal{M}^{k+1})))\|_F. \tag{C.8}$$

Since $\mathcal{W}^k$ is an approximate solution of $F(\mathcal{W};\mathcal{W}^{k-1})$, by the definition of the indicator function, we get that $\mathcal{W}^k$ belongs to Γ_1, D_1 and D_2. Thus, $\{\mathcal{W}^k\}_{k\in\mathbb{N}}$ is bounded and $\mathcal{W}^*$ is a cluster point. Then, it follows from [11, Theorem 3.10] that there exist constants $\delta_0 > 0$ and $\widetilde{m} > 0$ such that for any $\mathcal{W}^k, \mathcal{W}^{k+1} \in B(\mathcal{W}^*, \delta_0)$,

$$\|\nabla H_1(\mathcal{L}^k) - \nabla H_1(\mathcal{L}^{k+1})\|_F \le \widetilde{m}\|\mathcal{L}^{k+1} - \mathcal{L}^k\|_F. \tag{C.9}$$

It follows from ∇H_2 is Lipschitz continuous with constant $\frac{1}{\gamma}$ that

$$\lambda\|\nabla H_2(\mathcal{M}^k) - \nabla H_2(\mathcal{M}^{k+1})\|_F \le \frac{\lambda}{\gamma}\|\mathcal{M}^{k+1} - \mathcal{M}^k\|_F. \tag{C.10}$$

By combining (C.6) with (C.7), we have that $\mathcal{H}^{k+1} = \mathcal{C}^{k+1} - \eta(\mathcal{W}^{k+1} - \mathcal{W}^k)$. Moreover, by $\|\mathcal{B}^{k+1} - \mathcal{H}^{k+1}\|_F \ge \|\mathcal{B}^{k+1}\|_F - \|\mathcal{H}^{k+1}\|_F$, we obtain that

$$\begin{aligned}
\|\mathcal{B}^{k+1}\|_F &\le \|\mathcal{B}^{k+1} - \mathcal{H}^{k+1}\|_F + \|\mathcal{H}^{k+1}\|_F \\
&\le \widetilde{m}\|\mathcal{L}^{k+1} - \mathcal{L}^k\|_F + \frac{\lambda}{\gamma}\|\mathcal{M}^{k+1} - \mathcal{M}^k\|_F + \|\mathcal{C}^{k+1}\|_F + \eta\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F \\
&\le (\widetilde{m} + \lambda/\gamma + \eta + \eta c)\|\mathcal{W}^{k+1} - \mathcal{W}^k\|_F,
\end{aligned}$$

where the second inequality holds by (C.8) and the last inequality holds by (4.4), (C.9) and (C.10). The desired result is proven.

C.3. Proof of Lemma 4.3. It is easy to see that δ_{Γ_1}, δ_{D_1} and δ_{D_2} are semialgebraic [6]. On the other hand, the MCP function and the SCAD function are shown to be semialgebraic in [50], and $\|\mathcal{L}\|_{\mathrm{TNN}}$ is also shown to be semi-algebraic in [58]. Hence, the function $Q(\mathcal{W})$ is semi-algebraic since it is the finite sum of semialgebraic functions. Since $Q(\mathcal{W})$ is also proper lower semicontinuous, and it follows from [6, Theorem 3] that the function Q is a KL function, which completes the proof.

Appendix D. Proofs of the results in Section 5. This part includes the proofs of part of results in Section 5.

D.1. Proof of Proposition 5.1. Recall that

$$S(\mathcal{L}^\star) := \left\{\mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H \mid \mathcal{W} \in \mathbb{C}^{(n_1 - r_{\min}) \times (n_2 - r_{\min}) \times n_3}, \|\mathcal{W}\| \leq 1\right\}.$$

First we are going to show that $S(\mathcal{L}^\star) \subseteq \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}$. For any $\mathcal{Z} \in S(\mathcal{L}^\star)$, we have

$$\begin{aligned}
\langle \mathcal{Z}, \mathcal{L}^\star \rangle &= \langle \mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H, \mathcal{U} * \mathcal{S} * \mathcal{V}^H \rangle \\
&= \frac{1}{n_3} \sum_{i=1}^{n_3} \left\langle \widehat{\boldsymbol{U}_1}^{(i)} (\widehat{\boldsymbol{V}_1}^{(i)})^H + \widehat{\boldsymbol{U}_2}^{(i)} \widehat{\boldsymbol{W}}^{(i)} (\widehat{\boldsymbol{V}_2}^{(i)})^H, \widehat{\boldsymbol{U}}^{(i)} \widehat{\boldsymbol{S}}^{(i)} (\widehat{\boldsymbol{V}}^{(i)})^H \right\rangle \\
&= \frac{1}{n_3} \sum_{i=1}^{n_3} \left\langle \boldsymbol{U}_1^{(i)} (\boldsymbol{V}_1^{(i)})^H + \boldsymbol{U}_2^{(i)} \boldsymbol{W}^{(i)} (\boldsymbol{V}_2^{(i)})^H, \boldsymbol{U}^{(i)} \boldsymbol{S}^{(i)} (\boldsymbol{V}^{(i)})^H \right\rangle \\
&= \frac{1}{n_3} \sum_{i=1}^{n_3} \left\langle \boldsymbol{U}^{(i)} \begin{pmatrix} \boldsymbol{I}_{r_i} & 0 \\ 0 & \boldsymbol{W}^{(i)} \end{pmatrix} (\boldsymbol{V}^{(i)})^H, \boldsymbol{U}^{(i)} \begin{pmatrix} \mathrm{Diag}(\sigma(\widehat{\boldsymbol{L}^\star}^{(i)})) & 0 \\ 0 & 0 \end{pmatrix} (\boldsymbol{V}^{(i)})^H \right\rangle \\
&= \frac{1}{n_3} \sum_{i=1}^{n_3} \|\widehat{\boldsymbol{L}^\star}^{(i)}\|_* \\
&= \|\mathcal{L}^\star\|_{\mathrm{TNN}}.
\end{aligned}$$

It is easy to verify that $\|\mathcal{Z}\| \leq 1$. Then, by [47], we have $\mathcal{Z} \in \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}$. So we have $S(\mathcal{L}^\star) \subseteq \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}$.

Next, we are going to prove that $\partial \|\mathcal{L}^\star\|_{\mathrm{TNN}} \subseteq S(\mathcal{L}^\star)$. We argue it by contradiction. Assume that exist $\mathcal{G}' \in \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}$ but $\mathcal{G}' \notin S(\mathcal{L}^\star)$. It can be verified that $S(\mathcal{L}^\star)$ is convex and closed. Then, by Strict Separation Theorem [5], there exists $\mathcal{R} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ satisfying $\langle \mathcal{G}', \mathcal{R} \rangle > \langle \mathcal{H}, \mathcal{R} \rangle$ for any $\mathcal{H} \in S(\mathcal{L}^\star)$. So that

$$\max_{\mathcal{G} \in \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}} \langle \mathcal{G}, \mathcal{R} \rangle > \max_{\mathcal{H} \in S(\mathcal{L}^\star)} \langle \mathcal{H}, \mathcal{R} \rangle.$$

Let $f(\mathcal{L}^\star) := \|\mathcal{L}^\star\|_{\mathrm{TNN}}$. We use $f'(\mathcal{L}^\star; \mathcal{R})$ to denote the directional derivative of f at $\mathcal{L}^\star$ with the direction $\mathcal{R}$. It follows from [34, Theorem 23.4] that $f'(\mathcal{L}^\star; \mathcal{R}) = \max_{\mathcal{G} \in \partial \|\mathcal{L}^\star\|_{\mathrm{TNN}}} \langle \mathcal{G}, \mathcal{R} \rangle$.

Moreover,

$$
\begin{aligned}
f'(\mathcal{L}^\star;\mathcal{R}) &= \lim_{\gamma\to0^+}\frac{\|\mathcal{L}^\star+\gamma\mathcal{R}\|_{\mathrm{TNN}}-\|\mathcal{L}^\star\|_{\mathrm{TNN}}}{\gamma}\\
&= \lim_{\gamma\to0^+}\frac{1}{n_3}\sum_{i=1}^{n_3}\frac{\|\widehat{\boldsymbol{L}^\star+\gamma\boldsymbol{R}}^{(i)}\|_*-\|\widehat{\boldsymbol{L}^\star}^{(i)}\|_*}{\gamma}\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\lim_{\gamma\to0^+}\frac{\|\widehat{\boldsymbol{L}^\star}^{(i)}+\gamma\widehat{\boldsymbol{R}}^{(i)}\|_*-\|\widehat{\boldsymbol{L}^\star}^{(i)}\|_*}{\gamma}\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\sum_{j=1}^{n_1}\boldsymbol{d}_j^{(i)}(\boldsymbol{u}_j^{(i)})^H\widehat{\boldsymbol{R}}^{(i)}\boldsymbol{v}_j^{(i)}\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\langle\sum_{j=1}^{n_1}\boldsymbol{d}_j^{(i)}\boldsymbol{u}_j^{(i)}(\boldsymbol{v}_j^{(i)})^H,\widehat{\boldsymbol{R}}^{(i)}\rangle\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\langle\boldsymbol{U}^{(i)}\mathrm{Diag}(\boldsymbol{d}^{(i)})\boldsymbol{V}^{(i)H},\widehat{\boldsymbol{R}}^{(i)}\rangle\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\left\langle\begin{bmatrix}\boldsymbol{U}_1^{(i)} & \boldsymbol{U}_2^{(i)}\end{bmatrix}\begin{bmatrix}\mathrm{Diag}(\boldsymbol{d}_{\le r_i}^{(i)}) & 0\\ 0 & \mathrm{Diag}(\boldsymbol{d}_{>r_i}^{(i)})\end{bmatrix}\begin{bmatrix}(\boldsymbol{V}_1^{(i)})^H\\(\boldsymbol{V}_2^{(i)})^H\end{bmatrix},\widehat{\boldsymbol{R}}^{(i)}\right\rangle\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\left\langle\boldsymbol{U}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H+\boldsymbol{U}_2^{(i)}\mathrm{Diag}(\boldsymbol{d}_{>r_i}^{(i)})(\boldsymbol{V}_2^{(i)})^H,\widehat{\boldsymbol{R}}^{(i)}\right\rangle\\
&= \frac{1}{n_3}\sum_{i=1}^{n_3}\max_{\boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1}\left\langle\widehat{\boldsymbol{U}_1}^{(i)}(\widehat{\boldsymbol{V}_1}^{(i)})^H+\widehat{\boldsymbol{U}_2}^{(i)}\begin{bmatrix}0 & 0\\ 0 & \mathrm{Diag}(\boldsymbol{d}_{>r_i}^{(i)})\end{bmatrix}(\widehat{\boldsymbol{V}_2}^{(i)})^H,\widehat{\boldsymbol{R}}^{(i)}\right\rangle,
\end{aligned}
$$

where $\boldsymbol{u}_j^{(i)}$ is the j-th column of the $\boldsymbol{U}^{(i)}$ (also the j-th column of $\widehat{\boldsymbol{U}}^{(i)}$ when $j\le r_i$) and the fourth equality is due to [47, Theorem 1]. Notice that $|\boldsymbol{d}_j^{(i)}|\le 1$ when $j>r_i$. Denote

$$
\widehat{\boldsymbol{D}}^{(i)} := \begin{bmatrix}0 & 0\\ 0 & \mathrm{Diag}(\boldsymbol{d}_{>r_i}^{(i)})\end{bmatrix}\in\mathbb{C}^{(n_1-r_{\min})\times(n_2-r_{\min})}.
$$

Then we have $\widehat{\boldsymbol{D}}^{(i)}\in\{\widehat{\boldsymbol{W}}^{(i)}|\|\widehat{\boldsymbol{W}}^{(i)}\|\le 1\}$, which means that

$$
\{\widehat{\boldsymbol{D}}^{(i)}|\mathrm{diag}(\widehat{\boldsymbol{D}}^{(i)})=(0,\boldsymbol{d}_{>r_i}^{(i)})^H,\ \boldsymbol{d}^{(i)}\in\partial\|\boldsymbol{\sigma}^{(i)}\|_1\}\subseteq\{\widehat{\boldsymbol{W}}^{(i)}|\|\widehat{\boldsymbol{W}}^{(i)}\|\le 1\}.
$$

Let $\Lambda^{(i)} := \{\widehat{\boldsymbol{D}}^{(i)} | \mathrm{diag}(\widehat{\boldsymbol{D}}^{(i)}) = (0, \boldsymbol{d}^{(i)}_{>r_i})^H,\ \boldsymbol{d}^{(i)} \in \partial\|\boldsymbol{\sigma}^{(i)}\|_1\}$. Then we have

$$
\begin{aligned}
&\max_{\mathcal{H}\in S(\mathcal{L}^\star)} \langle \mathcal{H}, \mathcal{R}\rangle \\
&= \max_{\|\mathcal{W}\|\le 1} \left\langle \mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H, \mathcal{R}\right\rangle \\
&= \frac{1}{n_3}\sum_{i=1}^{n_3} \max_{\|\widehat{\boldsymbol{W}}^{(i)}\|\le 1} \left\langle \widehat{\boldsymbol{U}_1}^{(i)}\widehat{\boldsymbol{V}_1}^{(i)H} + \widehat{\boldsymbol{U}_2}^{(i)}\widehat{\boldsymbol{W}}^{(i)}(\widehat{\boldsymbol{V}_2}^{(i)})^H, \widehat{\boldsymbol{R}}^{(i)}\right\rangle \\
&\ge \frac{1}{n_3}\sum_{i=1}^{n_3} \max_{\widehat{\boldsymbol{W}}^{(i)}\in\Lambda^{(i)}} \left\langle \widehat{\boldsymbol{U}_1}^{(i)}\widehat{\boldsymbol{V}_1}^{(i)H} + \widehat{\boldsymbol{U}_2}^{(i)}\widehat{\boldsymbol{W}}^{(i)}(\widehat{\boldsymbol{V}_2}^{(i)})^H, \widehat{\boldsymbol{R}}^{(i)}\right\rangle \\
&= f'(\mathcal{L}^\star; \mathcal{R}),
\end{aligned}
$$

which implies $\max_{\mathcal{H}\in S(\mathcal{L}^\star)} \langle \mathcal{H}, \mathcal{R}\rangle \ge \max_{\mathcal{G}\in\partial\|\mathcal{L}^\star\|_{\mathrm{TNN}}} \langle \mathcal{G}, \mathcal{R}\rangle$. This contradicts the assumption. Therefore, we have $\partial\|\mathcal{L}^\star\|_{\mathrm{TNN}} \subseteq S(\mathcal{L}^\star)$. This completes the proof.

D.2. Proof of Proposition 5.3. Considering $\overline{\boldsymbol{X}} = \mathrm{Diag}(\widehat{\boldsymbol{X}}^{(1)}, \widehat{\boldsymbol{X}}^{(2)}, \ldots, \widehat{\boldsymbol{X}}^{(n_3)})$, $\forall i = 1, 2, \ldots, n_3$, we have

$$
\begin{aligned}
\widehat{\boldsymbol{X}}^{(i)} =& [\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}][\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}]^H \widehat{\boldsymbol{X}}^{(i)} [\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}][\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}]^H \\
=& [\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}] \begin{bmatrix} (\boldsymbol{U}_1^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_1^{(i)} & (\boldsymbol{U}_1^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_2^{(i)} \\ (\boldsymbol{U}_2^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_1^{(i)} & 0 \end{bmatrix} [\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}]^H + \\
& [\boldsymbol{U}_1^{(i)}, \boldsymbol{U}_2^{(i)}] \begin{bmatrix} 0 & 0 \\ 0 & (\boldsymbol{U}_2^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_2^{(i)} \end{bmatrix} [\boldsymbol{V}_1^{(i)}, \boldsymbol{V}_2^{(i)}]^H \\
=& \boldsymbol{U}_1^{(i)}(\boldsymbol{U}_1^{(i)})^H\widehat{\boldsymbol{X}}^{(i)} + \widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H - \boldsymbol{U}_1^{(i)}(\boldsymbol{U}_1^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_1^{(i)}(\boldsymbol{V}_1^{(i)})^H \\
& + \boldsymbol{U}_2^{(i)}(\boldsymbol{U}_2^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\boldsymbol{V}_2^{(i)}(\boldsymbol{V}_2^{(i)})^H \\
=& \widehat{\boldsymbol{U}_1}^{(i)}(\widehat{\boldsymbol{U}_1}^{(i)})^H\widehat{\boldsymbol{X}}^{(i)} + \widehat{\boldsymbol{X}}^{(i)}\widehat{\boldsymbol{V}_1}^{(i)}(\widehat{\boldsymbol{V}_1}^{(i)})^H - \widehat{\boldsymbol{U}_1}^{(i)}(\widehat{\boldsymbol{U}_1}^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\widehat{\boldsymbol{V}_1}^{(i)}(\widehat{\boldsymbol{V}_1}^{(i)})^H \\
& + \widehat{\boldsymbol{U}_2}^{(i)}(\widehat{\boldsymbol{U}_2}^{(i)})^H\widehat{\boldsymbol{X}}^{(i)}\widehat{\boldsymbol{V}_2}^{(i)}(\widehat{\boldsymbol{V}_2}^{(i)})^H,
\end{aligned}
$$

which means that

$$\overline{\boldsymbol{X}} = \overline{\boldsymbol{U}_1}\,\overline{\boldsymbol{U}_1}^H\overline{\boldsymbol{X}} + \overline{\boldsymbol{X}}\,\overline{\boldsymbol{V}_1}\,\overline{\boldsymbol{V}_1}^H - \overline{\boldsymbol{U}_1}\,\overline{\boldsymbol{U}_1}^H\,\overline{\boldsymbol{X}}\,\overline{\boldsymbol{V}_1}\,\overline{\boldsymbol{V}_1}^H + \overline{\boldsymbol{U}_2}\,\overline{\boldsymbol{U}_2}^H\overline{\boldsymbol{X}}\,\overline{\boldsymbol{V}_2}\,\overline{\boldsymbol{V}_2}^H.$$

So we have

$$\mathcal{X} = \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} + \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H - \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{U}_2^H * \mathcal{X} * \mathcal{V}_2 * \mathcal{V}_2^H.$$

By the definition of $\mathcal{T}$, we can see that

$$\mathcal{P}_{\mathcal{T}}(\mathcal{X}) = \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} + \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H - \mathcal{U}_1 * \mathcal{U}_1^H * \mathcal{X} * \mathcal{V}_1 * \mathcal{V}_1^H.$$

Therefore, it follows from $\mathcal{X} = \mathcal{P}_{\mathcal{T}}(\mathcal{X}) + \mathcal{P}_{\mathcal{T}^\perp}(\mathcal{X})$ that

$$\mathcal{P}_{\mathcal{T}^\perp}(\mathcal{X}) = \mathcal{U}_2 * \mathcal{U}_2^H * \mathcal{X} * \mathcal{V}_2 * \mathcal{V}_2^H.$$

This completes the proof.

D.3. Proof of Lemma 5.4. Since $(\mathcal{L}^c, \mathcal{M}^c)$ is optimal and $(\mathcal{L}^\star, \mathcal{M}^\star)$ is feasible to the problem (4.3), we have

$$\begin{aligned} 0 \geq & (\|\mathcal{L}^c\|_{\mathrm{TNN}} - \|\mathcal{L}^\star\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \widetilde{\Delta}_{\mathcal{L}}\rangle) + \lambda(\|\mathcal{M}^c\|_1 - \langle \nabla H_2(\mathcal{M}^k), \widetilde{\Delta}_{\mathcal{M}}\rangle - \|\mathcal{M}^\star\|_1) \\ & + \frac{\eta}{2}(\|\mathcal{L}^c - \mathcal{L}^k\|_F^2 - \|\mathcal{L}^\star - \mathcal{L}^k\|_F^2) + \frac{\eta}{2}(\|\mathcal{M}^c - \mathcal{M}^k\|_F^2 - \|\mathcal{M}^\star - \mathcal{M}^k\|_F^2). \end{aligned} \tag{D.1}$$

By (5.1), we know that $\{\mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H | \ \|\mathcal{W}\| \leq 1\} = \partial\|\mathcal{L}^\star\|_{\mathrm{TNN}}$. Thus, by the convexity of $\|\cdot\|_{\mathrm{TNN}}$, we have

$$\begin{aligned} & \|\mathcal{L}^c\|_{\mathrm{TNN}} - \|\mathcal{L}^\star\|_{\mathrm{TNN}} - \langle \nabla H_1(\mathcal{L}^k), \widetilde{\Delta}_{\mathcal{L}}\rangle \\ \geq & \langle \mathcal{U}_1 * \mathcal{V}_1^H + \mathcal{U}_2 * \mathcal{W} * \mathcal{V}_2^H, \widetilde{\Delta}_{\mathcal{L}}\rangle - \langle \nabla H_1(\mathcal{L}^k), \widetilde{\Delta}_{\mathcal{L}}\rangle \\ = & \frac{1}{n_3}\langle \overline{\boldsymbol{U}_1}\ \overline{\boldsymbol{V}_1}^H - \overline{\nabla H_1(\mathcal{L}^k)}, \overline{\widetilde{\Delta}_{\mathcal{L}}}\rangle + \frac{1}{n_3}\langle \overline{\boldsymbol{U}_2}\ \overline{\boldsymbol{W}}\ \overline{\boldsymbol{V}_2}^H, \overline{\widetilde{\Delta}_{\mathcal{L}}}\rangle \\ \geq & \frac{1}{n_3}\sup_{\|\overline{\boldsymbol{W}}\|\leq 1} \langle \overline{\boldsymbol{W}}, \overline{\boldsymbol{U}_2}^H \overline{\widetilde{\Delta}_{\mathcal{L}}}\overline{\boldsymbol{V}_2}\rangle - \frac{1}{n_3}\|\overline{\boldsymbol{U}_1}\ \overline{\boldsymbol{V}_1}^H - \overline{\nabla H_1(\mathcal{L}^k)}\|_F \|\overline{\widetilde{\Delta}_{\mathcal{L}}}\|_F \\ = & \frac{1}{n_3}\|\overline{\boldsymbol{U}_2}^H \overline{\widetilde{\Delta}_{\mathcal{L}}}\overline{\boldsymbol{V}_2}\|_* - \frac{1}{n_3}\|\overline{\boldsymbol{U}_1}\ \overline{\boldsymbol{V}_1}^H - \overline{\nabla H_1(\mathcal{L}^k)}\|_F \|\overline{\widetilde{\Delta}_{\mathcal{L}}}\|_F \\ = & \|\mathcal{U}_2 * \widetilde{\Delta}_{\mathcal{L}} * \mathcal{V}_2^H\|_{\mathrm{TNN}} - \|\mathcal{U}_1 * \mathcal{V}_1^H - \nabla H_1(\mathcal{L}^k)\|_F \|\widetilde{\Delta}_{\mathcal{L}}\|_F \\ = & \|\mathcal{P}_{\mathcal{T}^\perp}(\widetilde{\Delta}_{\mathcal{L}})\|_{\mathrm{TNN}} - d_{\mathcal{L}}\sqrt{r}\|\widetilde{\Delta}_{\mathcal{L}}\|_F, \end{aligned} \tag{D.2}$$

where the second equality directly from the definition of dual norm.

Similarly, we know that $\{\mathrm{sign}(\mathcal{M}^\star) + \mathcal{F} | \mathcal{P}\mathrm{supp}_{\mathcal{M}^\star}(\mathcal{F}) = 0, \|\mathcal{F}\|_\infty \leq 1\} \subseteq \partial\|\mathcal{M}^\star\|_1$, where $\mathrm{supp}_{\mathcal{X}} := \{(i,j,k) | \langle \Theta_{ijk}, \mathcal{X}\rangle \neq 0\}$. Thus, by the convexity of $\|\cdot\|_1$, we have

$$\begin{aligned} & \|\mathcal{M}^c\|_1 - \|\mathcal{M}^\star\|_1 - \langle \nabla H_2(\mathcal{M}^k), \widetilde{\Delta}_{\mathcal{M}}\rangle \\ \geq & \langle \mathrm{sign}(\mathcal{M}^\star) + \mathcal{P}\mathrm{supp}^c_{\mathcal{M}^\star}(\mathcal{F}), \widetilde{\Delta}_{\mathcal{M}}\rangle - \langle \nabla H_2(\mathcal{M}^k), \widetilde{\Delta}_{\mathcal{M}}\rangle \\ \geq & \sup_{\|\mathcal{F}\|_\infty \leq 1} \langle \mathcal{F}, \mathcal{P}\mathrm{supp}^c_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\rangle - \|\mathrm{sign}(\mathcal{M}^\star) - \nabla H_2(\mathcal{M}^k)\|_F \|\widetilde{\Delta}_{\mathcal{M}}\|_F \\ = & \|\mathcal{P}\mathrm{supp}^c_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1 - d_{\mathcal{M}}\sqrt{s}\|\widetilde{\Delta}_{\mathcal{M}}\|_F. \end{aligned} \tag{D.3}$$

By the convexity of $\|\cdot\|_F^2$, we also have

$$\begin{aligned} & \frac{\eta}{2}(\|\mathcal{L}^c - \mathcal{L}^k\|_F^2 - \|\mathcal{L}^\star - \mathcal{L}^k\|_F^2) + \frac{\eta}{2}(\|\mathcal{M}^c - \mathcal{M}^k\|_F^2 - \|\mathcal{M}^\star - \mathcal{M}^k\|_F^2) \\ \geq & \eta(\langle \mathcal{L}^\star - \mathcal{L}^k, \mathcal{L}^c - \mathcal{L}^\star\rangle + \langle \mathcal{M}^\star - \mathcal{M}^k, \mathcal{M}^c - \mathcal{M}^\star\rangle) \\ \geq & -\eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F \|\widetilde{\Delta}_{\mathcal{L}}\|_F - \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F \|\widetilde{\Delta}_{\mathcal{M}}\|_F. \end{aligned} \tag{D.4}$$

By substituting (D.2), (D.3) and (D.4) into (D.1), we get that

$$\begin{aligned} & \|\mathcal{P}_{\mathcal{T}^\perp}(\widetilde{\Delta}_{\mathcal{L}})\|_{\mathrm{TNN}} + \lambda\|\mathcal{P}\mathrm{supp}^c_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1 \\ \leq & (d_{\mathcal{L}}\sqrt{r} + \eta\|\mathcal{L}^\star - \mathcal{L}^k\|_F)\|\widetilde{\Delta}_{\mathcal{L}}\|_F + (\lambda d_{\mathcal{M}}\sqrt{s} + \eta\|\mathcal{M}^\star - \mathcal{M}^k\|_F)\|\widetilde{\Delta}_{\mathcal{M}}\|_F. \end{aligned}$$

Thus,

$$\text{(D.5)}\qquad \begin{aligned}&\max\{\|\mathcal{P}_{\mathcal{T}^\perp}(\widetilde{\Delta}_{\mathcal{L}})\|_{\text{TNN}}, \lambda\|\mathcal{P}\text{supp}^c_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1\}\\ \le&(d_{\mathcal{L}}\sqrt{r}+\eta\|\mathcal{L}^\star-\mathcal{L}^k\|_F)\|\widetilde{\Delta}_{\mathcal{L}}\|_F+(\lambda d_{\mathcal{M}}\sqrt{\widetilde{s}}+\eta\|\mathcal{M}^\star-\mathcal{M}^k\|_F)\|\widetilde{\Delta}_{\mathcal{M}}\|_F.\end{aligned}$$

It follows from Proposition 5.3 that $\text{rank}_a(\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}}))\le 2r$, which together with $\|\mathcal{P}\text{supp}_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_0\le\widetilde{s}$, we have

(D.6)

$$\begin{aligned}\|\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}})\|_{\text{TNN}}=\frac{1}{n_3}\|\overline{\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}})}\|_*\le&\frac{\sqrt{2rn_3}}{n_3}\|\overline{\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}})}\|_F=\sqrt{2r}\|\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}})\|_F\le\sqrt{2r}\|\widetilde{\Delta}_{\mathcal{L}}\|_F,\\ \|\mathcal{P}\text{supp}_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1\le&\sqrt{\widetilde{s}}\|\mathcal{P}\text{supp}_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_F\le\sqrt{\widetilde{s}}\|\widetilde{\Delta}_{\mathcal{M}}\|_F.\end{aligned}$$

Note that $\|\widetilde{\Delta}_{\mathcal{L}}\|_{\text{TNN}}\le\|\mathcal{P}_{\mathcal{T}}(\widetilde{\Delta}_{\mathcal{L}})\|_{\text{TNN}}+\|\mathcal{P}_{\mathcal{T}^\perp}(\widetilde{\Delta}_{\mathcal{L}})\|_{\text{TNN}}$ and $\|\widetilde{\Delta}_{\mathcal{M}}\|_1\le\|\mathcal{P}\text{supp}_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1+\|\mathcal{P}\text{supp}^c_{\mathcal{M}^\star}(\widetilde{\Delta}_{\mathcal{M}})\|_1$. By combining (D.5) and (D.6) together with the above inequalities, we complete the proof.

D.4. Proof of Lemma 5.5. First, we will show that the following event holds with small probability:

$$E:=\left\{\exists\Delta\in K(p,q,t)\text{such that }\left|\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2-\mathbb{E}[\langle\Theta,\Delta\rangle^2]\right|\ge\frac{\|\Delta_{\mathcal{L}}\|_F^2+\|\Delta_{\mathcal{M}}\|_F^2}{2\mu_1n_1n_2n_3}\right.$$
$$\left.+256\mu_1n_1n_2n_3\beta_{\mathcal{S}}^2\right\}.$$

It is clear that the complement of the interested event is included in E. Now we estimate the probability of the event E. We decompose the set $K(p,q,t)$ into

$$K(p,q,t)=\bigcup_{j=1}^\infty\left\{\Delta\in K(p,q,t)\;\Big|\;2^{j-1}t\le\frac{\|\Delta_{\mathcal{L}}\|_F^2+\|\Delta_{\mathcal{M}}\|_F^2}{\mu_1n_1n_2n_3}\le 2^jt\right\}.$$

For any $s\ge t$, we define the set

$$K(p,q,t,s):=\left\{\Delta\in K(p,q,t)\;\Big|\;\frac{\|\Delta_{\mathcal{L}}\|_F^2+\|\Delta_{\mathcal{M}}\|_F^2}{\mu_1n_1n_2n_3}\le s\right\}.$$

Let

$$E_j:=\left\{\exists\Delta\in K(p,q,t,2^jt)\text{ s.t. }\left|\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2-\mathbb{E}[\langle\Theta,\Delta\rangle^2]\right|\ge 2^{j-2}t+256\mu_1n_1n_2n_3\beta_{\mathcal{S}}^2\right\}.$$

Note that $E\subseteq\bigcup_{j=1}^\infty E_j$. In the following, we estimate the probability of the event E_j. Let

$$Z_s:=\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2-\mathbb{E}[\langle\Theta,\Delta\rangle^2]\right|,$$

we have

$$\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2 - \mathbb{E}[\langle\Theta,\Delta\rangle^2] = \frac{1}{m}\sum_{l=1}^{m}(\langle\Theta_{\omega_l},\Delta\rangle^2 - \mathbb{E}[\langle\Theta,\Delta\rangle^2]). \tag{D.7}$$

Since $\|\Delta\|_\infty = 1$ for all $\Delta \in K(p,q,t)$, it follows that

$$|\langle\Theta_{\omega_l},\Delta\rangle^2 - \mathbb{E}[\langle\Theta,\Delta\rangle^2]| \le \max\{\langle\Theta_{\omega_l},\Delta\rangle^2, \mathbb{E}[\langle\Theta,\Delta\rangle^2]\} \le 1.$$

Thus, it follows from Massart's Hoeffding type concentration inequality [30, Theorem 1.4] that

$$\mathbb{P}(Z_s \ge \mathbb{E}[Z_s] + \varepsilon) \le \exp\left(-\frac{m\varepsilon^2}{2}\right), \quad \forall \varepsilon > 0. \tag{D.8}$$

In order to be able to apply the inequality (D.8), we need to estimate an upper bound of $\mathbb{E}[Z_s]$. By (D.7), we have

$$\begin{aligned}
\mathbb{E}[Z_s] =& \mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\|\mathcal{P}_\Omega(\Delta)\|_F^2 - \mathbb{E}[\langle\Theta,\Delta\rangle^2]\right|\right] \le 2\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\sum_{l=1}^{m}\epsilon_l\langle\Theta_{\omega_l},\Delta\rangle^2\right|\right] \\
\le& 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\sum_{l=1}^{m}\langle\epsilon_l\Theta_{\omega_l},\Delta\rangle\right|\right] = 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left|\frac{1}{m}\langle\mathfrak{D}_\Omega^*(\epsilon),\Delta\rangle\right|\right] \\
\le& 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left\|\frac{1}{m}\overline{\mathfrak{D}_\Omega^*(\epsilon)}\right\|\left\|\frac{1}{n_3}\overline{\Delta_\mathcal{L}}\right\|_* + \sup_{\Delta\in K(p,q,t,s)}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty\|\Delta_\mathcal{M}\|_1\right] \\
=& 8\mathbb{E}\left[\sup_{\Delta\in K(p,q,t,s)}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|\|\Delta_\mathcal{L}\|_{\mathrm{TNN}} + \sup_{\Delta\in K(p,q,t,s)}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty\|\Delta_\mathcal{M}\|_1\right] \\
\le& 8\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|\left(\sup_{\Delta\in K(p,q,t,s)}\|\Delta_\mathcal{L}\|_{\mathrm{TNN}}\right) + 8\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty\left(\sup_{\Delta\in K(p,q,t,s)}\|\Delta_\mathcal{M}\|_1\right),
\end{aligned}$$

where the first inequality is due to the symmetrization theorem [7, Theorem 14.3] and the second inequality follows from the contraction theorem [7, Theorem 14.4]. Notice that for any $u \ge 0$,$v \ge 0$ and $\Delta \in K(p,q,t,s)$,

$$u\|\Delta_\mathcal{L}\|_F + v\|\Delta_\mathcal{M}\|_F \le 32\mu_1 n_1 n_2 n_3(u^2+v^2) + \frac{\|\Delta_\mathcal{L}\|_F^2 + \|\Delta_\mathcal{M}\|_F^2}{128\mu_1 n_1 n_2 n_3} \le 32\mu_1 n_1 n_2 n_3(u^2+v^2) + \frac{s}{128},$$

where the first inequality follows from the fact $2ab \le a^2 + b^2$. Then, follows from (5.5), (5.6), the definition of $K(p,q,t)$ and the above inequality, we derive that

$$\begin{aligned}
\mathbb{E}[Z_s] \le& 8\left[\sup_{\Delta\in K(p,q,t,s)}\beta_\mathcal{L}(p_1\|\Delta_\mathcal{L}\|_F + p_2\|\Delta_\mathcal{M}\|_F) + \sup_{\Delta\in K(p,q,t,s)}\beta_\mathcal{M}(q_1\|\Delta_\mathcal{L}\|_F + q_2\|\Delta_\mathcal{M}\|_F)\right] \\
\le& 256\mu_1 n_1 n_2 n_3\beta_\mathcal{S}^2 + \frac{s}{8}.
\end{aligned} \tag{D.9}$$

Then it follows from (D.8) and (D.9) that

$$\mathbb{P}\left(Z_s \ge 256\mu_1 n_1 n_2 n_3\beta_\mathcal{S}^2 + \frac{s}{4}\right) \le \mathbb{P}\left(Z_s \ge \mathbb{E}[Z_s] + \frac{s}{8}\right) \le \exp\left(-\frac{ms^2}{128}\right).$$

This, together with the choice of $s = 2^j t$, implies that $\mathbb{P}(E_j) \le \exp\left(-\frac{4^j m t^2}{128}\right)$. Therefore, it follows from the simple fact $4^j > \log(4^j) = 2j\log(2)$ that

$$\mathbb{P}(E) \le \sum_{j=1}^{\infty} \mathbb{P}(E_j) \le \sum_{j=1}^{\infty} \exp\left(-\frac{4^j m t^2}{128}\right) \le \sum_{j=1}^{\infty} \exp\left(-\frac{j m t^2 \log(2)}{64}\right) \le \frac{\exp[-mt^2\log(2)/64]}{1-\exp[-mt^2\log(2)/64]}.$$

Then, taking $t = 8\sqrt{\frac{\log(n_1+n_2+n_3+1)}{m\log(2)}}$, we obtain that $\mathbb{P}(E) \le \frac{1}{n_1+n_2+n_3}$. The proof is completed.

D.5. Proof of Lemma 5.7. For $l = 1, \dots m$, define the random tensor $\mathcal{Z}_{\omega_l} := \epsilon_l \Theta_{\omega_l}$. Then $\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon) = \frac{1}{m}\sum_{l=1}^m \mathcal{Z}_{\omega_l}$. Since ϵ_l is an i.i.d. Rademacher sequence, we have that $|\epsilon_l| \le 1$, $\mathbb{E}[\epsilon_l]$=0 and $\mathbb{E}[\epsilon_l^2]$=1. Notice that ϵ_l and Θ_{ω_l} are independent, we get $\mathbb{E}[\mathcal{Z}_{\omega_l}] = \mathbb{E}[\epsilon_l]\mathbb{E}[\Theta_{\omega_l}] = 0$. Since $\|\Theta_{\omega_l}\|_F = 1$, we have

$$\|\mathcal{Z}_{\omega_l}\| \le \|\mathcal{Z}_{\omega_l}\|_F = |\epsilon_l|\|\Theta_{\omega_l}\|_F = |\epsilon_l|.$$

It is easy to obtain that there exists a constant $M > 0$ such that $\|\|\mathcal{Z}_{\omega_l}\|\|_{\psi_1} \le \|\epsilon_l\|_{\psi_1} \le M$ and $\mathbb{E}^{\frac{1}{2}}[\|\mathcal{Z}_{\omega_l}\|^2] \le \mathbb{E}^{\frac{1}{2}}[\epsilon_l^2] = 1$. Define

$$\sigma_{\mathcal{Z}} := \max\left\{\left\|\frac{1}{m}\sum_{l=1}^m \mathbb{E}[\mathcal{Z}_{\omega_l} * \mathcal{Z}_{\omega_l}^H]\right\|^{\frac{1}{2}}, \left\|\frac{1}{m}\sum_{l=1}^m \mathbb{E}[\mathcal{Z}_{\omega_l}^H * \mathcal{Z}_{\omega_l}]\right\|^{\frac{1}{2}}\right\}.$$

By direct calculations we can see that $\mathbb{E}[\mathcal{Z}_{\omega_l} * \mathcal{Z}_{\omega_l}^H] = \mathbb{E}[\epsilon_l^2 \Theta_{\omega_l} * \Theta_{\omega_l}^H] = \mathbb{E}[\Theta_{\omega_l} * \Theta_{\omega_l}^H]$. The calculation for $\mathbb{E}[\mathcal{Z}_{\omega_l}^H * \mathcal{Z}_{\omega_l}]$ is similar. We obtain from Assumption 5.2 that $\sigma_{\mathcal{Z}}^2 \le \frac{\mu_2}{\widetilde{n}}$. By applying [48, Lemma 2.6], we obtain

$$\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| \le C_1\left\{\sqrt{\frac{\mu_2(t+\log((n_1+n_2)n_3))}{\widetilde{n}m}}, \frac{(t+\log((n_1+n_2)n_3))\log(\widetilde{n})}{m}\right\}$$

with probability at least $1-\exp(-t)$. Set $\tau^* = \frac{\mu_2 C_1}{\widetilde{n}\log(\widetilde{n})}$. Then we can derive

$$\text{(D.10)} \qquad \mathbb{P}\left[\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| > \tau\right] \le \begin{cases} ((n_1+n_2)n_3)\exp\left(-\frac{\tau^2\widetilde{n}m}{C_1^2\mu_2}\right), & \tau \le \tau^*, \\ ((n_1+n_2)n_3)\exp\left(-\frac{\tau m}{C_1\log(\widetilde{n})}\right), & \tau > \tau^*. \end{cases}$$

We set $v_1 = \frac{\widetilde{n}m}{C_1^2\mu_2}$ and $v_2 = \frac{m}{C_1\log(\widetilde{n})}$. By Hölder's inequality, we get

$$\text{(D.11)} \qquad \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| \le \left[\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|^{2\log((n_1+n_2)n_3)}\right]^{\frac{1}{2\log((n_1+n_2)n_3)}}.$$

Combining (D.10) with (D.11), we obtain that

$$\text{(D.12)} \qquad \begin{aligned} \mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| \le& \left(\int_0^\infty \mathbb{P}\left(\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| > \tau^{\frac{1}{2\log((n_1+n_2)n_3)}}\right) d\tau\right)^{\frac{1}{2\log((n_1+n_2)n_3)}} \\ =&\sqrt{e}\left[\log((n_1+n_2)n_3)v_1^{-\log((n_1+n_2)n_3)}\Gamma(\log((n_1+n_2)n_3))\right. \\ &\left.+2\log((n_1+n_2)n_3)v_2^{-2\log((n_1+n_2)n_3)}\Gamma(2\log((n_1+n_2)n_3))\right]^{\frac{1}{2\log((n_1+n_2)n_3)}}. \end{aligned}$$

Since the Gamma-function satisfies the inequality $\Gamma(x) \le \left(\frac{x}{2}\right)^{x-1}$, $\forall x \ge 2$. Plugging this inequality into (D.12), we obtain that

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| \le & \sqrt{e}\Big[(\log((n_1+n_2)n_3))^{\log((n_1+n_2)n_3)} v_1^{-\log((n_1+n_2)n_3)} 2^{1-\log((n_1+n_2)n_3)} \\ & + 2(\log((n_1+n_2)n_3))^{2\log((n_1+n_2)n_3)} v_2^{-2\log((n_1+n_2)n_3)}\Big]^{\frac{1}{2\log((n_1+n_2)n_3)}}.\end{aligned}$$

Observe that $m \ge \widetilde{n}\log((n_1+n_2)n_3)(\log(\widetilde{n}))^2/\mu_2$ implies that $v_1\log((n_1+n_2)n_3)) \le v_2^2$. Thus, we have

$$\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\| \le \sqrt{\frac{3e\log((n_1+n_2)n_3)}{v_1}} = C_1\sqrt{\frac{3e\mu_2\log((n_1+n_2)n_3)}{\widetilde{n}m}}.$$

This completes the proof.

D.6. Proof of Lemma 5.8. For any index (i,j,k) such that $1 \le i \le n_1$, $1 \le j \le n_2$, $1 \le k \le n_3$ and $(\Theta_{\omega_l})_{ijk} \ne 0$ for some $\omega_l \in \Omega$, let $\omega^{ijk} := ((\Theta_{\omega_1})_{ijk}, \ldots, (\Theta_{\omega_l})_{ijk})^T$. Form [48, Lemma 2.4], we know that there exists a constant $C > 0$ such that for any $\tau > 0$,

$$\mathbb{P}\left[\left|\frac{1}{m}\sum_{l=1}^m \omega_l^{ijk}\epsilon_l\right| > \tau\right] \le 2\exp\left[-C\min\left(\frac{m^2\tau^2}{M^2\|\omega^{ijk}\|_2^2}, \frac{m\tau}{M\|\omega^{ijk}\|_\infty}\right)\right].$$

By taking a union bound, we get that

$$\mathbb{P}\left[\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty > \tau\right] \le 2m\exp\left[-C\min\left(\frac{m^2\tau^2}{M^2\max\|\omega^{ijk}\|_2^2}, \frac{m\tau}{M\max\|\omega^{ijk}\|_\infty}\right)\right],$$

where both of the maximums are taken over all such indices (i,j,k). Evidently, $\|\omega^{ijk}\|_2^2 \le 1$ and $\|\omega^{ijk}\|_\infty \le 1$. By letting

$$\begin{aligned}-t :&= -C\min\left(\frac{m^2\tau^2}{M^2}, \frac{m\tau}{M}\right) + \log(m) \\ &\ge -C\min\left(\frac{m^2\tau^2}{M^2\max\|\omega^{ijk}\|_2^2}, \frac{m\tau}{M\max\|\omega^{ijk}\|_\infty}\right) + \log(m),\end{aligned}$$

we obtain that with probability no greater than $2\exp(-t)$,

$$\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty > M\max\left\{\sqrt{\frac{\log(m)+t}{Cm^2}}, \frac{\log(m)+t}{Cm}\right\}.$$

Set $\tau^* = \max\left\{\frac{M}{m}, \frac{M(\log(2m))}{mC}\right\}$. Then we can derive that

$$\mathbb{P}\left[\left\|\frac{1}{m}\mathfrak{D}_\Omega^*(\epsilon)\right\|_\infty > \tau\right] \le \begin{cases} 1, & \tau \le \tau^*, \\ 2m\exp\left(-\frac{Cm}{M}\tau\right), & \tau > \tau^*. \end{cases}$$

Then it follows that

$$\mathbb{E}\left\|\frac{1}{m}\mathfrak{D}_{\Omega}^{*}(\epsilon)\right\|_{\infty} \leq \int_{0}^{\tau^{*}} 1\mathrm{d}\tau + \int_{\tau^{*}}^{+\infty} 2m\exp\left(-\frac{Cm}{M}\tau\right)\mathrm{d}\tau = \frac{M(\log(2m)+1)}{Cm},$$

which completes the proof.

REFERENCES

[1] M. Ahn, J.-S. Pang, and J. Xin, *Difference-of-convex learning: directional stationarity, optimality, and sparsity*, SIAM J. Optim., 27 (2017), pp. 1637–1665.

[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.

[3] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods*, Math. Program., 137 (2013), pp. 91–129.

[4] M. Bai, X. Zhang, G. Ni, and C. Cui, *An adaptive correction approach for tensor completion*, SIAM J. Imaging Sci., 9 (2016), pp. 1298–1323.

[5] D. Bertsekas, *Convex optimization theory*, Athena Sci., 2009.

[6] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.

[7] P. Bühlmann and S. Van De Geer, *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media, 2011.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, J. ACM, 58 (2011), pp. 1–37.

[9] J. D. Carroll and J.-J. Chang, *Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition*, Psychometrika, 35 (1970), pp. 283–319.

[10] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, *Tensor decompositions for signal processing applications: From two-way to multiway component analysis*, IEEE Signal Process. Mag., 32 (2015), pp. 145–163.

[11] C. Ding, *An introduction to a class of matrix optimization problems*, PhD thesis, Department of Mathematics, National University of Singapore, 2012.

[12] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc., 96 (2001), pp. 1348–1360.

[13] J. Fan, L. Xue, and H. Zou, *Strong oracle optimality of folded concave penalized estimation*, Ann. Statist., 42 (2014), pp. 819–849.

[14] G. Gu, S. Jiang, and J. Yang, *A tvscad approach for image deblurring with impulsive noise*, Inverse Problems, 33 (2017), p. 125008.

[15] K. Guo, D. Han, and T. Wu, *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints*, Int. J. Comput. Math., 94 (2017), pp. 1653–1669.

[16] C. J. Hillar and L.-H. Lim, *Most tensor problems are np-hard*, J. ACM, 60 (2013), pp. 1–39.

[17] H. Huang, Y. Liu, Z. Long, and C. Zhu, *Robust low-rank tensor ring completion*, Int. J. Comput. Math., 6 (2020), pp. 1117–1126.

[18] Q. Jiang and M. K. Ng, *Robust low-tubal-rank tensor completion via convex optimization.*, in Proceedings of IJCAI, 2019, pp. 2649–2655.

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

[20] M. E. Kilmer and C. D. Martin, *Factorization strategies for third-order tensors*, Linear Algebra Appl., 435 (2011), pp. 641–658.

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

[22] O. Klopp, K. Lounici, and A. B. Tsybakov, *Robust matrix completion*, Probab. Theory Relat. Fields., 169 (2017), pp. 523–564.
[23] X. Y. Lam, J. S. Marron, D. Sun, and K.-C. Toh, *Fast algorithms for large-scale generalized distance weighted discrimination*, J. Comput. Graph. Statist., 27 (2018), pp. 368–379.
[24] K. Lange, D. R. Hunter, and I. Yang, *Optimization transfer using surrogate objective functions*, J. Comput. Graph. Statist., 9 (2000), pp. 1–20.
[25] X. Li, D. Sun, and K.-C. Toh, *A schur complement based semi-proximal admm for convex quadratic conic programming and extensions*, Math. Program., 155 (2016), pp. 333–373.
[26] Y. Li, K. Shang, and Z. Huang, *Low tucker rank tensor recovery via admm based on exact and inexact iteratively reweighted algorithms*, J. Comput. Appl. Math., 331 (2018), pp. 64–81.
[27] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, IEEE Trans. Pattern Anal., 35 (2013), pp. 208–220.
[28] Y. Liu, S. Bi, and S. Pan, *Equivalent lipschitz surrogates for zero-norm and rank optimization problems*, J. Global Optim., 72 (2018), pp. 679–704.
[29] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, *Tensor robust principal component analysis with a new tensor nuclear norm*, IEEE Trans. Pattern Anal., 42 (2019), pp. 925–938.
[30] P. Massart, *Optimal constants for Hoeffding type inequalities*, Mathématique Université de Paris-Sud, 1998.
[31] W. Miao, S. Pan, and D. Sun, *A rank-corrected procedure for matrix completion with fixed basis coefficients*, Math. Program., 159 (2016), pp. 289–338.
[32] M. Mørup, *Applications of tensor (multiway array) factorizations and decompositions in data mining*, Data Min. Knowl. Discov., 1 (2011), pp. 24–40.
[33] B. K. Natarajan, *Sparse approximate solutions to linear systems*, SIAM J. Comput., 24 (1995), pp. 227–234.
[34] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, 1970.
[35] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Springer, Berlin, 2009.
[36] B. Romera-Paredes and M. Pontil, *A new convex relaxation for tensor completion*, in Proceedings of NIPS, 2013, pp. 2967–2975.
[37] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, *Tensor-based formulation and nuclear norm regularization for multienergy computed tomography*, IEEE Trans. Image Process., 23 (2014), pp. 1678–1693.
[38] F. Shang, J. Cheng, Y. Liu, Z.-Q. Luo, and Z. Lin, *Bilinear factor matrix norm minimization for robust pca: Algorithms and applications*, IEEE Trans. Pattern Anal., 40 (2017), pp. 2066–2080.
[39] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, *Tensor decomposition for signal processing and machine learning*, IEEE Trans. Signal Process., 65 (2017), pp. 3551–3582.
[40] T. Sun, P. Yin, L. Cheng, and H. Jiang, *Alternating direction method of multipliers with difference of convex functions*, Adv. Comput. Math., 44 (2018), pp. 723–744.
[41] P. Tang, C. Wang, D. Sun, and K.-C. Toh, *A sparse semismooth newton based proximal majorization-minimization algorithm for nonconvex square-root-loss regression problems*, J. Mach. Learn. Res., 21 (2020), pp. 1–38.
[42] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 58 (1996), pp. 267–288.
[43] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
[44] M. A. O. Vasilescu and D. Terzopoulos, *Multilinear subspace analysis of image ensembles*, in Proceedings of CVPR., 2003, pp. 93–99.
[45] A. Wang, D. Wei, B. Wang, and Z. Jin, *Noisy low-tubal-rank tensor completion through iterative singular tube thresholding*, IEEE Access, 6 (2018), pp. 35112–35128.
[46] Y. Wang, W. Yin, and J. Zeng, *Global convergence of admm in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 29–63.
[47] G. A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra Appl., 170 (1992), pp. 33–45.
[48] B. Wu, *High-dimensional analysis on matrix decomposition with application to correlation matrix esti-*

mation in factor models, PhD thesis, Department of Mathematics, National University of Singapore, 2014.

[49] W. H. Xu, X. L. Zhao, T. Y. Ji, J. Q. Miao, J. Q. Ma, S. Wang, and T. Z. Huang, *Laplace function based nonconvex surrogate for low-rank tensor completion*, Signal Process. Image, 73 (2019), pp. 62–69.

[50] L. Yang, T. K. Pong, and X. Chen, *Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction*, SIAM J. Imaging Sci., 10 (2017), pp. 74–110.

[51] Y. Yang, Y. Feng, and J. A. Suykens, *Robust low-rank tensor recovery with regularized redescending m-estimator*, IEEE Trans. Neural Netw. Learn. Syst., 27 (2015), pp. 1933–1946.

[52] C.-H. Zhang, *Nearly unbiased variable selection under minimax concave penalty*, Ann. Statist., 38 (2010), pp. 894–942.

[53] H. Zhang, P. Zhou, Y. Yang, and J. Feng, *Generalized majorization-minimization for non-convex optimization*, in Proceedings of IJCAI, 2019, pp. 4257–4263.

[54] X. Zhang, *A nonconvex relaxation approach to low-rank tensor completion*, IEEE Trans. Neural Netw. Learn. Syst., 30 (2018), pp. 1659–1671.

[55] X. Zhang and M. K. Ng, *A corrected tensor nuclear norm minimization method for noisy low-rank tensor completion*, SIAM J. Imaging Sci., 12 (2019), pp. 1231–1273.

[56] Z. Zhang and S. Aeron, *Exact tensor completion using t-svd*, IEEE Trans. Signal Process., 65 (2016), pp. 1511–1526.

[57] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, *Novel methods for multilinear data completion and de-noising based on tensor-svd*, in Proceedings of CVPR, 2014, pp. 3842–3849.

[58] X. Zhao, M. Bai, and M. K. Ng, *Nonconvex optimization for robust tensor completion from grossly sparse observations*, J. Sci. Comput, 85 (2020), pp. 1–32.