RIP-based performance guarantee for low tubal rank tensor recovery

Feng Zhang[a], Wendong Wang[a], Jianwen Huang[a], Jianjun Wang[a,b,*]

[a]*School of Mathematics and Statistics, Southwest University, Chongqing 400715, China*
[b]*Research Center for Artificial Intelligence & Education Big Data, Southwest University, Chongqing, 400715, China*

Abstract

The essential task of multi-dimensional data analysis focuses on the tensor decomposition and the corresponding notion of rank. In this paper, by introducing the notion of tensor singular value decomposition (t-SVD), we establish a regularized tensor nuclear norm minimization (RTNNM) model for low tubal rank tensor recovery. On the other hand, many variants of the restricted isometry property (RIP) have proven to be crucial frameworks and analysis tools for recovery of sparse vectors and low-rank tensors. So, we initiatively define a novel tensor restrict isometry property (t-RIP) based on t-SVD. Besides, our theoretical results show that any third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ whose tubal rank is at most r can stably be recovered from its as few as measurements $\boldsymbol{y} = \boldsymbol{\mathfrak{M}}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$ with a bounded noise constraint $\|\boldsymbol{w}\|_2 \leq \epsilon$ via the RTNNM model, if the linear map $\boldsymbol{\mathfrak{M}}$ obeys t-RIP with

$$\delta_{tr}^{\mathfrak{M}} < \sqrt{\frac{t-1}{n_3^2 + t - 1}}$$

for certain fixed $t > 1$. Surprisingly, when $n_3 = 1$, our conditions coincide with Cai and Zhang's sharp work in 2013 for low-rank matrix recovery via the constrained nuclear norm minimization. We note that, as far as the authors are aware, such kind of result has not previously been reported in the literature.

*Corresponding author

Email addresses: `zhangf@email.swu.edu.cn` (Feng Zhang), `d.sylan@foxmail.com` (Wendong Wang), `hjw1303987297@126.com` (Jianwen Huang), `wjj@swu.edu.cn` (Jianjun Wang)

Keywords: Low-rank tensor recovery, tensor singular value decomposition, tensor restrict isometry property, regularized
2010 MSC: 90C25, 65J22, 53A45, 15A69

1. Introduction

Utilizing tensor model, possessed of the ability to make full use of multi-linear structure, instead of traditional matrix-based model to analyze multi-dimensional data (tensor data) has widely attracted attention. Low-rank tensor recovery as a representative problem is not only a mathematical natural generalization of the compressed sensing and low-rank matrix recovery problem, but also there exists lots of reconstruction applications of data that have intrinsically many dimensions in the context of low-rank tensor recovery including signal processing [1], machine learning [2], data mining [3], and many others.

The purpose of low-rank tensor recovery is to reconstruct a low-rank tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ (this article considers only the third-order tensor without loss of generality) from linear noise measurements $\boldsymbol{y} = \boldsymbol{\mathfrak{M}}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$, where $\boldsymbol{\mathfrak{M}}$: $\mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$ ($m \ll n_1 n_2 n_3$) is a random map with i.i.d. Gaussian entries and $\boldsymbol{w} \in \mathbb{R}^m$ is a vector of measurement errors. To be specific, we consider addressing the following rank minimization problem

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \operatorname{rank}(\boldsymbol{\mathcal{X}}), \quad s.t. \ \|\boldsymbol{y} - \boldsymbol{\mathfrak{M}}(\boldsymbol{\mathcal{X}})\|_2 \leq \epsilon, \tag{1}$$

where ϵ is a positive constant. The key to dealing with the low-rank tensor recovery problem is how to define the rank of the tensor. Unlike in the matrix case, there are different notions of tensor rank which are induced by different tensor decompositions. Two classical decomposition strategies can be regarded as higher-order extensions of the matrix SVD: CANDECOMP/PARAFAC (CP) decomposition [4] and Tucker decomposition [5]. Those induced tensor ranks are called the CP rank and Tucker rank, respectively. Tucker decomposition is the most widely used decomposition method at present. In particular, based on the Tucker decomposition, a convex surrogate optimization model [1] of the non-convex minimization problem (1) that is NP-hard regardless of the choice of the tensor decomposition has been studied as follows:

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \|\boldsymbol{\mathcal{X}}\|_{\text{SNN}}, \quad s.t. \quad \|\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{X}})\|_2 \leq \epsilon, \tag{2}$$

where $\|\boldsymbol{\mathcal{X}}\|_{\text{SNN}} := \sum_i^{n_3} \frac{1}{n_3} \|\boldsymbol{X}^{\{i\}}\|_*$ is referred as the Sum of Nuclear Norms (SNN) and $\boldsymbol{X}^{\{i\}}$ denotes the mode-i matricization of $\boldsymbol{\mathcal{X}}$, $\|\boldsymbol{X}^{\{i\}}\|_*$ is the trace norm of the matrix $\boldsymbol{X}^{\{i\}}$. This popular approach (2), however, has its limitations. Firstly, the Tucker decomposition is highly non-unique. Secondly, SNN is not the convex envelop of $\sum_i \text{rank}(\boldsymbol{X}^{\{i\}})$, which leads to a fact that the model (2) can be substantially suboptimal. Thirdly, the definition of SNN is inconsistent with the matrix case so that the existing analysis templates of low-rank matrix recovery can not be generalized to that for low-rank tensor recovery.

More recently, based on the definition of tensor-tensor product (t-product) and tensor singular value decomposition (t-SVD) [6, 7, 8] that enjoys many similar properties as the matrix case, Kilmer et al. proposed the tensor multi-rank definition and tubal rank definition [9] (see Definition 2.8) and Semerci et al. developed a new tensor nuclear norm (TNN) [10] . Continuing along this vein, Lu et al. given a new and rigorous way to define the average rank of tensor $\boldsymbol{\mathcal{X}}$ by $\text{rank}_a(\boldsymbol{\mathcal{X}})$ [11] (see Definition 2.9) and the nuclear norm of tensor $\boldsymbol{\mathcal{X}}$ by $\|\boldsymbol{\mathcal{X}}\|_*$ [11] (see Definition 2.10), and proved that the convex envelop of $\text{rank}_a(\boldsymbol{\mathcal{X}})$ is $\|\boldsymbol{\mathcal{X}}\|_*$ within the unit ball of the tensor spectral norm [11]. Furthermore, they pointed out that the low average rank assumption for tensor $\boldsymbol{\mathcal{X}}$ is weaker than the CP rank and Tucker rank assumptions and tensor $\boldsymbol{\mathcal{X}}$ always has low average rank if it has low tubal rank induced by t-SVD. Therefore, considering that this novel and computable tensor nuclear norm can address the shortcoming of SNN, a convex tensor nuclear norm minimization (TNNM) model based low tubal rank assumption for tensor recovery has been proposed in [11], which solves

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \|\boldsymbol{\mathcal{X}}\|_*, \quad s.t. \quad \|\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{X}})\|_2 \leq \epsilon, \tag{3}$$

where tensor nuclear norm $\|\boldsymbol{\mathcal{X}}\|_*$ is as the convex surrogate of tensor average rank $\text{rank}_a(\boldsymbol{\mathcal{X}})$. In order to facilitate the design of algorithms and the needs of practical applications, instead of considering the constrained-TNNM (3), in this paper, we present a theoretical analysis for regularized tensor nuclear norm minimization (RTNNM) model, which takes the form

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \|\boldsymbol{\mathcal{X}}\|_* + \frac{1}{2\lambda}\|\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{X}})\|_2^2, \tag{4}$$

where λ is a positive parameter. According to [12], there exists a $\lambda > 0$ such that the solution to the regularization problem (4) also solves the constrained problem (3) for any $\epsilon > 0$, and vice versa. However, model (4) is more commonly used than model (3) when the noise level is not given or cannot be accurately estimated. There exist many examples of solving RTNNM problem (4) based on the tensor nuclear norm heuristic. For instance, by exploiting the t-SVD, Semerci et al. [10] developed the tensor nuclear norm regularizer which can be solved by an alternating direction method of multipliers (ADMM) approach. Analogously, Lu et al. [11] and zhang et al. [13] used tensor nuclear norm to replace the tubal rank for low-rank tensor recovery from incomplete tensors (tensor completion) and tensor robust principal component analysis (TRPCA). Two kinds of problems can be regarded as special cases of (4). ADMM algorithm can also be applied to solve it. While the application and algorithm research of (4) is already well-developed, only few contributions on the theoretical results with regard to performance guarantee for low tubal rank tensor recovery are available so far. The restrict isometry property (RIP) introduced by Candès and Tao [14] is one of the most widely used frameworks in sparse vector/low-rank matrix recovery. In this paper, we generalize the RIP tool to tensor case based on t-SVD and hope to make up for the lack of research on low tubal tensor recovery.

Since different tensor decompositions induce different notions of tensor rank, and they also induce different notions of the tensor RIP. For example, in 2013, based on Tucker decomposition [5], Shi et al. defined tensor RIP [15] as follows:

Definition 1.1. *Let* $\mathfrak{S}_{(r_1,r_2,r_3)} : \{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3} : \operatorname{rank}(\boldsymbol{X}^{\{i\}}) \leq r_i, i = 1,2,3\}$*. The RIP constant* $\delta_{(r_1,r_2,r_3)}$ *of linear operator* $\mathfrak{F}$ *is the smallest value* *such that*

$$(1-\delta_{(r_1,r_2,r_3)})\|\boldsymbol{\mathcal{X}}\|_F^2 \leq \|\mathfrak{F}(\boldsymbol{\mathcal{X}})\|_2^2 \leq (1+\delta_{(r_1,r_2,r_3)})\|\boldsymbol{\mathcal{X}}\|_F^2$$

holds for all tensors $\boldsymbol{\mathcal{X}} \in \mathfrak{S}_{(r_1,r_2,r_3)}$.

Their theoretical results show that a tensor $\boldsymbol{\mathcal{X}}$ with rank (r_1, r_2, r_3) can be exactly recovered in the noiseless case if $\mathfrak{F}$ satisfies the RIP with the constant $\delta_\Lambda < 0.4931$ for $\Lambda \in \{(2r_1, n_2, n_3), (n_1, 2r_2, n_3), (n_1, n_2, 2r_3)\}$. This is the first work to extend the RIP-based results from the sparse vector recovery to tensor case. In addition, in 2016, Rauhut et al. [16] also induced three notions of the tensor RIP by utilizing the higher order singular value decomposition (HOSVD), the tensor train format (TT), and the general hierarchical Tucker decomposition (HT). These decompositions can be considered as variants of Tucker decomposition whose uniqueness is not guaranteed such that all these induced definitions of tensor RIP depend on a rank tuple that differs greatly from the definition of matrix rank. In contrast, t-SVD is a higher-order tensor decomposition strategy with uniqueness and computability. So, based on t-SVD, we initiatively define a novel tensor restrict isometry property as follows:

Definition 1.2. *(t-RIP) A linear map* $\mathfrak{M} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^m$*, is said to satisfy the t-RIP with tensor restricted isometry constant (t-RIC)* $\delta_r^{\mathfrak{M}}$ *if* $\delta_r^{\mathfrak{M}}$ *is the smallest value* $\delta^{\mathfrak{M}} \in (0, 1)$ *such that*

$$(1-\delta^{\mathfrak{M}})\|\boldsymbol{\mathcal{X}}\|_F^2 \leq \|\mathfrak{M}(\boldsymbol{\mathcal{X}})\|_2^2 \leq (1+\delta^{\mathfrak{M}})\|\boldsymbol{\mathcal{X}}\|_F^2 \tag{5}$$

holds for all tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *whose tubal rank is at most* r.

Our definition of tensor RIP shows the same form with vector RIP [14] and matrix RIP [17]. In other words, vector RIP and matrix RIP are low-dimensional versions of our t-RIP when $n_2 = n_3 = 1$ and $n_3 = 1$, respectively, which will result in some existing analysis tools and techniques that can also be used for tensor cases. At the same time, the existing theoretical results will provide us with a great reference. For constrained sparse vector/low rank matrix recovery, different conditions on the restricted isometry constant (RIC) have been introduced and studied in the literature [18, 17, 19], etc. Among these sufficient conditions, especially, Cai and Zhang [20] showed that for any given $t \geq 4/3$, the sharp vector RIC $\delta_{tr}^{\boldsymbol{M}} < \sqrt{\frac{t-1}{t}}$ (matrix RIC $\delta_{tr}^{\boldsymbol{\mathcal{M}}} < \sqrt{\frac{t-1}{t}}$) ensures the exact recovery in the noiseless case and stable recovery in the noisy case for r-sparse signals and matrices with rank at most r. In addition, Zhang and Li [21] obtained another part of the sharp condition, that is $\delta_{tr}^{\boldsymbol{M}} < \frac{t}{4-t}(\delta_{tr}^{\boldsymbol{\mathcal{M}}} < \frac{t}{4-t})$ with $0 < t < 4/3$. The results mentioned above are currently the best in the field. In view of unconstrained sparse vector recovery, as far as we know that Zhu [22] first studied this kind of problem in 2008 and he pointed out that r-sparse signals can be recovered stably if $\delta_{4r}^{\boldsymbol{M}} + 2\delta_{5r}^{\boldsymbol{M}} < 1$. Next, in 2015, Shen et al. [23] got a sufficient condition $\delta_{2r}^{\boldsymbol{M}} < 0.2$ under redundant tight frames. Recently, Ge et al. [24] proved that if the noisy vector $\boldsymbol{w}$ satisfies the ℓ_∞ bounded noise constraint (i.e., $\|\boldsymbol{M}^*\boldsymbol{w}\|_\infty \leq \lambda/2$) and $\delta_{tr}^{\boldsymbol{M}} < \sqrt{\frac{t-1}{t+8}}$ with $t > 1$, then r-sparse signals can be stably recovered. Although there is no similar result for unconstrained low rank matrix recovery, the results presented in this paper also can depict the case of the matrix when $n_3 = 1$.

Equipped with the t-RIP, in this paper, we aim to construct sufficient conditions for stable low tubal rank tensor recovery and obtain an ideal upper bound of error via solving (4). The rest of the paper is organized as follows. In Section 2, we introduce some notations and definitions. In Section 3, we give some key lemmas. In Section 4, our main result is presented. In Section 5, some numerical experiments are conducted to support our analysis. The conclusion is addressed in Section 6. Finally, Appendix A and Appendix B provide the proof of Lemma 3.2 and Lemma 3.3, respectively.

2. Notations and Preliminaries

We use lowercase letters for the entries, e.g. x, boldface letters for vectors, e.g. $\boldsymbol{x}$, capitalized boldface letters for matrices, e.g. $\boldsymbol{X}$ and capitalized boldface calligraphic letters for tensors, e.g. $\boldsymbol{\mathcal{X}}$. For a third-order tensor $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{X}}(i,:,:)$, $\boldsymbol{\mathcal{X}}(:,i,:)$ and $\boldsymbol{\mathcal{X}}(:,:,i)$ are used to represent the ith horizontal, lateral, and frontal slice. The frontal slice $\boldsymbol{\mathcal{X}}(:,:,i)$ can also be denoted as $\boldsymbol{X}^{(i)}$. The tube is denoted as $\boldsymbol{\mathcal{X}}(i,j,:)$. We denote the Frobenius norm as $\|\boldsymbol{\mathcal{X}}\|_F = \sqrt{\sum_{ijk} |x_{ijk}|^2}$. Defining some norms of matrix is also necessary. We denote by $\|\boldsymbol{X}\|_F = \sqrt{\sum_{ij} |x_{ij}|^2} = \sqrt{\sum_i \sigma_i^2(\boldsymbol{X})}$ the Frobenius norm of $\boldsymbol{X}$ and denote by $\|\boldsymbol{X}\|_* = \sum_i \sigma_i(\boldsymbol{X})$ the nuclear norm of $\boldsymbol{X}$, where $\sigma_i(\boldsymbol{X})$'s are the singular values of $\boldsymbol{X}$ and $\sigma(\boldsymbol{X})$ represents the singular value vector of matrix $\boldsymbol{X}$. Given a positive integer κ, we denote $[\kappa] = \{1, 2, \cdots, \kappa\}$ and $\Gamma^c = [n_1] \setminus \Gamma$ for any $\Gamma \subset [n_1]$.

For a third-order tensor $\boldsymbol{\mathcal{X}}$, let $\bar{\boldsymbol{\mathcal{X}}}$ be the discrete Fourier transform (DFT) along the third dimension of $\boldsymbol{\mathcal{X}}$, i.e., $\bar{\boldsymbol{\mathcal{X}}} = \mathrm{fft}(\boldsymbol{\mathcal{X}}, [], 3)$. Similarly, $\boldsymbol{X}$ can be calculated from $\bar{\boldsymbol{\mathcal{X}}}$ by $\boldsymbol{\mathcal{X}} = \mathrm{fft}(\bar{\boldsymbol{\mathcal{X}}}, [], 3)$. Let $\bar{\boldsymbol{X}} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$ be the block diagonal matrix with each block on diagonal as the frontal slice $\bar{\boldsymbol{X}}^{(i)}$ of $\bar{\boldsymbol{\mathcal{X}}}$, i.e.,

$$\bar{\boldsymbol{X}} = \mathrm{bdiag}(\bar{\boldsymbol{\mathcal{X}}}) = \begin{pmatrix} \bar{\boldsymbol{X}}^{(1)} & & & \\ & \bar{\boldsymbol{X}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\boldsymbol{X}}^{(n_3)} \end{pmatrix},$$

and $\mathrm{bcirc}(\boldsymbol{\mathcal{X}}) \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$ be the block circular matrix, i.e.,

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

The unfold operator and its inverse operator fold are, respectively, defined as

$$\mathrm{unfold}(\boldsymbol{\mathcal{X}}) = \begin{pmatrix} \boldsymbol{X}^{(1)} & \boldsymbol{X}^{(2)} & \cdots & \boldsymbol{X}^{(n_3)} \end{pmatrix}^T, \quad \mathrm{fold}(\mathrm{unfold}(\boldsymbol{\mathcal{X}})) = \boldsymbol{\mathcal{X}}.$$

Then tensor-tensor product (t-product) between two third-order tensors can be defined as follows.

Definition 2.1. *(t-product [6]) For tensors $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the t-product $\boldsymbol{\mathcal{A}} \star \boldsymbol{\mathcal{B}}$ is defined to be a tensor of size $n_1 \times n_4 \times n_3$,*

$$\boldsymbol{\mathcal{A}} \star \boldsymbol{\mathcal{B}} = \mathrm{fold}(\mathrm{bcirc}(\boldsymbol{\mathcal{A}}) \cdot \mathrm{unfold}(\boldsymbol{\mathcal{B}})).$$

Definition 2.2. *(Conjugate transpose [6]) The conjugate transpose of a tensor $\boldsymbol{\mathcal{X}}$ of size $n_1 \times n_2 \times n_3$ is the $n_2 \times n_1 \times n_3$ tensor $\boldsymbol{\mathcal{X}}^*$ 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. *(Identity tensor [6]) The identity tensor $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor whose first frontal slice is the $n \times n$ identity matrix, and other frontal slices are all zeros.*

Definition 2.4. *(Orthogonal tensor [6]) A tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies*

$$\boldsymbol{\mathcal{Q}}^* \star \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} \star \boldsymbol{\mathcal{Q}}^* = \boldsymbol{\mathcal{I}}.$$

Definition 2.5. *(F-diagonal tensor [6]) A tensor is called F-diagonal if each of its frontal slices is a diagonal matrix.*

Theorem 2.6. *(t-SVD [6]) Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the t-SVD factorization of tensor $\boldsymbol{\mathcal{X}}$ is*

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star \boldsymbol{\mathcal{S}} \star \boldsymbol{\mathcal{V}}^*,$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an F-diagonal tensor. Figure 1 illustrates the t-SVD factorization.

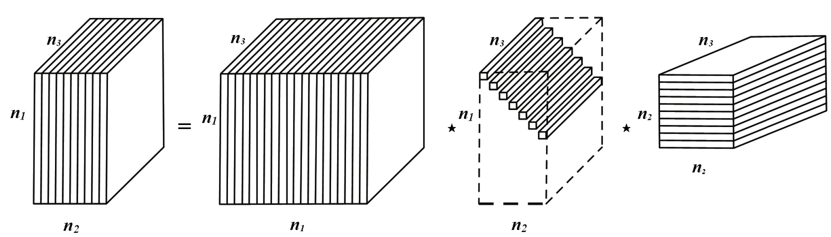


Figure 1: An illustration of the t-SVD of an $n_1 \times n_1 \times n_3$ tensor.

Remark 2.7. *For $\kappa = \min(n_1, n_2)$, the t-SVD of $\boldsymbol{\mathcal{X}}$ can be written*

$$\boldsymbol{\mathcal{X}} = \sum\nolimits_{i=1}^{\kappa} \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{X}}}(:, i, :) \star \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{X}}}(i, i, :) \star \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}(:, i, :)^*.$$

The diagonal vector of the first frontal slice of $\boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{X}}}$ *is denoted as* $\boldsymbol{s}_{\boldsymbol{\mathcal{X}}}$. *The best* r*-term approximation of* $\boldsymbol{\mathcal{H}}$ *with the tubal rank at most* r *is denoted by*

$$\boldsymbol{\mathcal{X}}_{\max(r)} = \arg \min_{\operatorname{rank}_t(\tilde{\boldsymbol{\mathcal{X}}}) \leq r} \|\boldsymbol{\mathcal{X}} - \tilde{\boldsymbol{\mathcal{X}}}\|_F = \sum\nolimits_{i=1}^{r} \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{X}}}(:,i,:) \star \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{X}}}(i,i,:) \star \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}(:,i,:)^*,$$

and $\boldsymbol{\mathcal{X}}_{-\max(r)} = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}_{\max(r)}$. *In addition, for index set* Γ, *we have*

$$\boldsymbol{\mathcal{X}}_{\Gamma} = \sum\nolimits_{i \in \Gamma} \boldsymbol{\mathcal{U}}_{\boldsymbol{\mathcal{X}}}(:,i,:) \star \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{X}}}(i,i,:) \star \boldsymbol{\mathcal{V}}_{\boldsymbol{\mathcal{X}}}(:,i,:)^*.$$

Definition 2.8. *(Tensor tubal rank [9]) For* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *the tensor tubal rank, denoted as* $\operatorname{rank_t}(\boldsymbol{\mathcal{X}})$, *is defined as the number of nonzero singular tubes of* $\boldsymbol{\mathcal{S}}$, *where* $\boldsymbol{\mathcal{S}}$ *is from the t-SVD of* $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star \boldsymbol{\mathcal{S}} \star \boldsymbol{\mathcal{V}}^*$. *We can write*

$$\operatorname{rank}_t(\boldsymbol{\mathcal{X}}) = \sharp\{i : \boldsymbol{\mathcal{S}}(i,i,:) \neq \mathbf{0}\} = \sharp\{i : \boldsymbol{\mathcal{S}}(i,i,1) \neq 0\}.$$

Definition 2.9. *(Tensor average rank [11]) For* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *the tensor average rank, denoted as* $\operatorname{rank}_a(\boldsymbol{\mathcal{X}})$, *is defined as*

$$\operatorname{rank}_a(\boldsymbol{\mathcal{X}}) = \frac{1}{n_3} \operatorname{rank}(\operatorname{bcirc}(\boldsymbol{\mathcal{X}})) = \frac{1}{n_3} \operatorname{rank}(\operatorname{bdiag}(\bar{\boldsymbol{X}})).$$

Definition 2.10. *(Tensor nuclear norm [11]) Let* $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{U}} \star \boldsymbol{\mathcal{S}} \star \boldsymbol{\mathcal{V}}^*$ *be the t-SVD of* $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. *The tensor nuclear norm of* $\boldsymbol{\mathcal{X}}$ *is defined as* $\|\boldsymbol{\mathcal{X}}\|_* := \sum_{i=1}^{r} \boldsymbol{\mathcal{S}}(i,i,1)$, *where* $r = \operatorname{rank}_t(\boldsymbol{\mathcal{X}})$.

Proposition 2.11. *For a third-order tensor* $\boldsymbol{\mathcal{X}}$, *we have the following properties*

$$\begin{aligned} \|\boldsymbol{\mathcal{X}}\|_F &= \frac{1}{\sqrt{n_3}} \|\bar{\boldsymbol{X}}\|_F, & (6) \\ \|\boldsymbol{\mathcal{X}}\|_* &= \frac{1}{n_3} \|\bar{\boldsymbol{X}}\|_*. & (7) \\ \operatorname{rank}(\bar{\boldsymbol{X}}) &\leq n_3 \operatorname{rank}_t(\boldsymbol{\mathcal{X}}). & (8) \end{aligned}$$

3. Some Key Lemmas

We present the following lemmas, which will play a key role in proving our sufficient conditions for low tubal rank tensor recovery.

Lemma 3.1. *[20] For a positive number ϕ and a positive integer s, define the polytope $T(\phi, s) \subset \mathbb{R}^n$ by*

$$T(\phi, s) = \{\boldsymbol{v} \in \mathbb{R}^n : \|\boldsymbol{v}\|_\infty \leq \phi, \|\boldsymbol{v}\|_1 \leq s\phi\}.$$

For any $\boldsymbol{v} \in \mathbb{R}^n$, define the set of sparse vectors $U(\phi, s, \boldsymbol{v}) \subset \mathbb{R}^n$ by

$$U(\phi, s, \boldsymbol{v}) = \{\boldsymbol{u} \in \mathbb{R}^n : \mathrm{supp}(\boldsymbol{u}) \subseteq \mathrm{supp}(\boldsymbol{v}), \|\boldsymbol{u}\|_0 \leq s, \|\boldsymbol{u}\|_1 = \|\boldsymbol{v}\|_1, \|\boldsymbol{u}\|_\infty \leq \phi\}.$$

Then $\boldsymbol{v} \in T(\phi, s)$ if and only if $\boldsymbol{v}$ is in the convex hull of $U(\phi, s, \boldsymbol{v})$. In particular, any $\boldsymbol{v} \in T(\phi, s)$ can be expressed as

$$\boldsymbol{v} = \sum_{i=1}^{N} \gamma_i \boldsymbol{u}_i$$

where $\boldsymbol{u}_i \in U(\phi, s, \boldsymbol{v})$ and $0 \leq \gamma_i \leq 1$, $\sum_{i=1}^{N} \gamma_i = 1$.

This elementary technique introduced by T. Cai and A. Zhang [20] shows that any point in a polytope can be represented as a convex combination of sparse vectors and makes the analysis surprisingly simple.

The following lemma shows that a suitable t-RIP condition implies the robust null space property [25] of the linear map $\mathfrak{M}$.

Lemma 3.2. *Let the linear map $\mathfrak{M} : \mathbb{R}^{n_1 \times n_2 \times n_3} \to \mathbb{R}^n$ satisfies the t-RIP of order $tr(t > 1)$ with t-RIC $\delta_{tr}^{\mathfrak{M}} \in (0, 1)$. Then for any tensor $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and any subset $\Gamma \subset [\kappa]$ with $|\Gamma| = r$ and $\kappa = \min(n_1, n_2)$, it holds that*

$$\|\boldsymbol{\mathcal{H}}_\Gamma\|_F \leq \eta_1 \|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 + \frac{\eta_2}{\sqrt{r}} \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*, \tag{9}$$

where

$$\eta_1 \triangleq \frac{2}{(1 - \delta_{tr}^{\mathfrak{M}})\sqrt{1 + \delta_{tr}^{\mathfrak{M}}}}, \quad \text{and} \quad \eta_2 \triangleq \frac{\sqrt{\mathrm{n}_3}\delta_{\mathrm{tr}}^{\mathfrak{M}}}{\sqrt{(1 - (\delta_{\mathrm{tr}}^{\mathfrak{M}})^2)(\mathrm{t} - 1)}}.$$

Proof. Please see Appendix A. □

In order to prove the main result we still need the following lemma.

Lemma 3.3. *If the noisy measurements $\boldsymbol{y} = \mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$ of tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are observed with noise level $\|\boldsymbol{w}\|_2 \leq \epsilon$, then for any subset $\Gamma \subset [\kappa]$ with $|\Gamma| = r$ and $\kappa = \min(n_1, n_2)$, the minimization solution $\hat{\boldsymbol{\mathcal{X}}}$ of (4) satisfies*

$$\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2\epsilon\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 \leq 2\lambda(\|\boldsymbol{\mathcal{H}}_\Gamma\|_* - \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*), \tag{10}$$

and

$$\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* \leq \|\boldsymbol{\mathcal{H}}_\Gamma\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* + \frac{\epsilon}{\lambda}\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2, \tag{11}$$

where $\boldsymbol{\mathcal{H}} \triangleq \hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}$.

Proof. Please see Appendix B. □

4. Main Results

With preparations above, now we present our main result.

Theorem 4.1. *For any observed vector $\boldsymbol{y} = \mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$ of tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ corrupted by an unknown noise $\boldsymbol{w}$, with bounded constrain $\|\boldsymbol{w}\|_2 \leq \epsilon$, if $\mathfrak{M}$ satisfies t-RIP with*

$$\delta_{tr}^{\mathfrak{M}} < \sqrt{\frac{t-1}{n_3^2 + t - 1}} \tag{12}$$

for certain $t > 1$, then we have

$$\|\mathfrak{M}(\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}})\|_2 \leq C_1\|\boldsymbol{\mathcal{X}}_{-\max(r)}\|_* + C_2, \tag{13}$$

and

$$\|\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}\|_F \leq C_3\|\boldsymbol{\mathcal{X}}_{-\max(r)}\|_* + C_4, \tag{14}$$

where $\hat{\boldsymbol{\mathcal{X}}}$ is the solution to (4), and $C_i, i = 1, 2, 3, 4$ are denoted as

$$\begin{aligned}
C_1 &= \frac{2}{\sqrt{r}\eta_1}, \quad C_2 = 2\sqrt{r}\eta_1\lambda + 2\epsilon, \\
C_3 &= \frac{2\sqrt{r}\eta_1(2\sqrt{n_3 r} + 1 + \eta_2)\lambda + 2(\sqrt{n_3 r} + \eta_2)\epsilon}{r\eta_1(1-\eta_2)\lambda}, \\
C_4 &= \frac{(\sqrt{n_3 r} + 1)\eta_1\lambda + (\sqrt{n_3 r} - \sqrt{n_3}\eta_2 + \sqrt{n_3} + 1)\epsilon}{(1-\eta_2)\lambda(2\sqrt{r}\eta_1\lambda + 2\epsilon)^{-1}}.
\end{aligned}$$

Proof. Please see Appendix C. □

Remark 4.2. *We note that the obtained t-RIC condition (12) is related to the length n_3 of the third dimension. This is due to the fact that the discrete Fourier transform (DFT) is performed along the third dimension of $\boldsymbol{\mathcal{X}}$. Further, we want to stress that this crucial quantity n_3 is rigorously deduced from the t-product and makes the result of the tensor consistent with the matrix case. For general problems, let n_3 be the smallest size of three modes of the third-order tensor, e.g. $n_3 = 3$ for the third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{h \times w \times 3}$ from a color image with size $h \times w$, where three frontal slices correspond to the R, G, B channels; $n_3 = 8$ for 3-D face detection using tensor data $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{h \times w \times 8}$ with column h, row w, and depth mode 8. Specially, when $n_3 = 1$, our model (4) returns to the case of low rank matrix recovery and the condition (12) degenerates to $\delta_{tr}^{\boldsymbol{\mathcal{M}}} < \sqrt{(t-1)/t}$ which has also been proved to be sharp by Cai, et al. [19] for stable recovery via the constrained nuclear norm minimization for $t > 4/3$. For unconstrained low rank matrix recovery, the degenerated sufficient condition $\delta_{tr}^{\boldsymbol{\mathcal{M}}} < \sqrt{(t-1)/t}$ for $t > 1$ and error upper bound estimation can be easily derived. Because these works are trivial, we omit these potential corollaries. We note that, to the best of our knowledge, results like our Theorem 4.1 has not previously been reported in the literature.*

Remark 4.3. *Theorem 4.1 not only offers a sufficient condition for stably recovering tensor $\boldsymbol{\mathcal{X}}$ based on solving (4), but also provides an error upper bound estimate for the recovery of tensor $\boldsymbol{\mathcal{X}}$ via RTNNM model. This result clearly depicts the relationship among reconstruction error, the best r-term approximation, noise level ϵ and λ. There exist some special cases of Theorem 4.1 which is worth studying. For examples, one can associate the ℓ_2-norm bounded noise level ϵ with the trade-off parameter λ (such as $\epsilon = \lambda/2$) as [24, 23, 17]. This case can be summarized by Corollary 4.4. Notice that we can take a λ which is close to zero such that $\tilde{C}_2\lambda$ and $\tilde{C}_4\lambda$ in (16),(17) is close to zero for the noise-free case $\boldsymbol{w} = \mathbf{0}$. Then Corollary 4.4 shows that tensor $\boldsymbol{\mathcal{X}}$ can be approximately recovery by solving (4) if $\|\boldsymbol{\mathcal{X}}_{-\max(r)}\|_*$ is small.*

Corollary 4.4. *Suppose that the noise measurements $\boldsymbol{y} = \mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$ of tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are observed with noise level $\|\boldsymbol{w}\|_2 \leq \epsilon = \frac{\lambda}{2}$. If $\mathfrak{M}$ satisfies t-RIP with*

$$\delta_{tr}^{\mathfrak{M}} < \sqrt{\frac{t-1}{n_3^2 + t - 1}} \tag{15}$$

for certain $t > 1$, then we have

$$\|\mathfrak{M}(\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}})\|_2 \leq \tilde{C}_1 \|\boldsymbol{\mathcal{X}}_{-\max(r)}\|_* + \tilde{C}_2 \lambda, \tag{16}$$

and

$$\|\hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}\|_F \leq \tilde{C}_3 \|\boldsymbol{\mathcal{X}}_{-\max(r)}\|_* + \tilde{C}_4 \lambda, \tag{17}$$

where $\hat{\boldsymbol{\mathcal{X}}}$ is the solution to (4), and $\tilde{C}_i, i = 1, 2, 3, 4$ are denoted as

$$\begin{aligned}
\tilde{C}_1 &= \frac{2}{\sqrt{r}\eta_1}, \quad \tilde{C}_2 = 2\sqrt{r}\eta_1 + 1, \\
\tilde{C}_3 &= \frac{2\sqrt{r}\eta_1(2\sqrt{n_3 r} + 1 + \eta_2) + \sqrt{n_3 r} + \eta_2}{r\eta_1(1-\eta_2)}, \\
\tilde{C}_4 &= \frac{2(\sqrt{n_3 r} + 1)\eta_1 + \sqrt{n_3 r} - \sqrt{n_3}\eta_2 + \sqrt{n_3} + 1}{2(1-\eta_2)(2\sqrt{r}\eta_1 + 1)^{-1}}.
\end{aligned}$$

5. Numerical experiments

In this section, we present several numerical experiments to corroborate our analysis. We use the tensor Singular Value Thresholding (t-SVT) algorithm in [11] to compute the RTNNM model (4). All numerical experiments are tested on a PC with 4 GB of RAM and Intel core i5-4200M (2.5GHz). In order to avoid the randomness, we perform 50 times against each test and report the average result.

We perform $\boldsymbol{y} = \boldsymbol{M}\mathrm{vec}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$ to get the linear noise measurements instead of $\boldsymbol{y} = \mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$. Then the RTNNM model can be reformulated as

$$\min_{\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \|\boldsymbol{\mathcal{X}}\|_* + \frac{1}{2\lambda}\|\boldsymbol{y} - \boldsymbol{M}\mathrm{vec}(\boldsymbol{\mathcal{X}})\|_2^2, \tag{18}$$

where $\boldsymbol{y}, \boldsymbol{w} \in \mathbb{R}^m$, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\boldsymbol{M} \in \mathbb{R}^{m \times (n_1 n_2 n_3)}$ is a Gaussian measurement ensemble and $\mathrm{vec}(\boldsymbol{\mathcal{X}})$ denotes the vectorization of $\boldsymbol{\mathcal{X}}$. According to

Theorem 4.2 in [11], the closed-form of the proximal operator of (18) can be presented by exploiting the t-SVT.

First, we generate a tubal rank r tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n\times n\times n_3}$ as a product $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_1 \star \boldsymbol{\mathcal{X}}_2$ where $\boldsymbol{\mathcal{X}}_1 \in \mathbb{R}^{n\times r\times n_3}$ and $\boldsymbol{\mathcal{X}}_2 \in \mathbb{R}^{r\times n\times n_3}$ are two tensors with entries independently sampled from a standard Gaussian distribution. Next, we generate a measurement matrix $\boldsymbol{M} \in \mathbb{R}^{m\times(n^2 n_3)}$ with entries drawn independently from a $\mathcal{N}(0, 1/m)$ distribution. Using $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{M}$, the measurements $\boldsymbol{y}$ are produced by $\boldsymbol{y} = \boldsymbol{M}\text{vec}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$, where $\boldsymbol{w}$ is the Gaussian white noise with mean 0 and variance σ^2. We uniformly evaluate the recovery performance of the model by signal-to-noise ratio (SNR) defined as $20\log(\|\boldsymbol{\mathcal{X}}\|_F/\|\boldsymbol{\mathcal{X}} - \hat{\boldsymbol{\mathcal{X}}}\|_F)$ in decibels (dB) (the greater the SNR, the better the reconstruction). The key to studying the RTNNM model (4) is to explain the relationship among reconstruction error, noise level ϵ and λ. Therefor, we design two sets of experiments to explain it. Case 1: $n = 10$, $n_3 = 5$, $r = 0.2n$; Case 2: $n = 30$, $n_3 = 5$, $r = 0.3n$. The number of samples m in all experiments is set to $3r(2n - r)n_3 + 1$ as [26]. The results are given in Table 1, Table 2 and Figure 2. It can be seen that there exist two accordant conclusions for low tubal tensor recovery at different scales. For a fixed regularization parameter λ, as the standard deviation σ increases (the greater σ, the greater the noise level ϵ), the SNR gradually decreases. In addition, for each fixed noise level, the smaller the regularization parameter, the larger the SNR, which means the low tubal rank tensor can be better recovered. Thus, these experiments clearly demonstrate the quantitative correlation among reconstruction error, noise level ϵ and λ.

6. Conclusion

In this paper, a heuristic notion of tensor restricted isometry property (t-RIP) has been introduced based on tensor singular value decomposition (t-SVD). Comparing with other definitions [15, 16], it is more representative as a higher-order generalization of the traditional RIP for vector and matrix recovery. since the forms and properties of t-RIP and t-SVD are consistent with

Table 1: SNR for different noise levels and regularization parameters in Case 1.

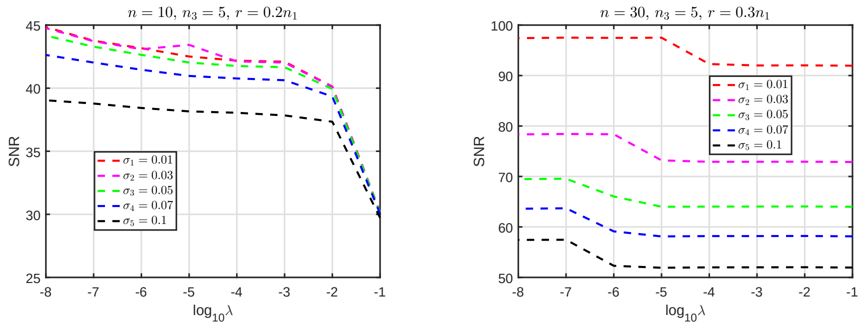
$n = 10$, $n_3 = 5$, $r = 0.2n$

SNR	$\sigma_1 = 0.01$	$\sigma_2 = 0.03$	$\sigma_3 = 0.05$	$\sigma_4 = 0.07$	$\sigma_5 = 0.1$
$\lambda_1 = 1e-1$	**30.15**	30.14	30.12	30.04	29.73
$\lambda_2 = 1e-2$	**40.13**	40.11	39.94	39.34	37.34
$\lambda_3 = 1e-3$	**42.10**	42.03	41.66	40.63	37.84
$\lambda_4 = 1e-4$	**42.18**	42.12	41.76	40.77	38.05
$\lambda_5 = 1e-5$	42.51	**43.43**	42.03	40.97	38.16
$\lambda_6 = 1e-6$	**43.17**	43.09	42.64	41.46	38.43
$\lambda_7 = 1e-7$	**43.76**	43.72	43.29	42.03	38.78
$\lambda_8 = 1e-8$	**44.84**	**44.76**	**44.16**	**42.63**	**39.04**

Table 2: SNR for different noise levels and regularization parameters in Case 2.

$n = 30$, $n_3 = 5$, $r = 0.3n$

SNR	$\sigma_1 = 0.01$	$\sigma_2 = 0.03$	$\sigma_3 = 0.05$	$\sigma_4 = 0.07$	$\sigma_5 = 0.1$
$\lambda_1 = 1e-1$	**91.95**	72.86	63.99	58.14	51.95
$\lambda_2 = 1e-2$	**92.02**	72.93	64.06	58.21	52.02
$\lambda_3 = 1e-3$	**92.00**	72.91	64.04	58.19	52.00
$\lambda_4 = 1e-4$	**92.28**	72.91	64.03	58.19	51.99
$\lambda_5 = 1e-5$	**97.50**	73.19	64.00	58.12	51.91
$\lambda_6 = 1e-6$	**97.46**	78.37	66.04	59.12	52.30
$\lambda_7 = 1e-7$	**97.51**	**78.42**	**69.55**	**63.70**	**57.46**
$\lambda_8 = 1e-8$	97.41	78.34	69.47	63.62	57.43



(a) (b)

Figure 2: SNR for different noise levels and regularization parameters. (a) λ versus SNR with $n = 10$, $n_3 = 5$ and $r = 0.2n$. (b) λ versus SNR with $n = 30$, $n_3 = 5$ and $r = 0.3n$.

the vector/matrix case. This point is crucial because this guarantees that our theoretical investigation can be done in a similar way as sparse vector/low rank matrix recovery. A sufficient condition was presented, based on the RTNNM model, for stably recovering a given low tubal rank tensor that are corrupted with an ℓ_2-norm bounded noise. However, this condition only considers the $\delta_{tr}^{\mathfrak{M}}$ of the map $\mathfrak{M}$ when t is limited to $t > 1$. In the future, we hope to provide a complete answer for $\delta_{tr}^{\mathfrak{M}}$ when $0 < t \leq 1$. Another important topic is to establish the guarantee for stable recovery based on (4) in the context of the required number of measurements.

Acknowledgment

This work was supported by National Natural Science Foundation of China (Grant No. 61273020, 61673015).

Appendix A. Proof of Lemma 3.2

Proof. ***Step 1: Sparse Representation of a Polytope.***

Without loss of generality, assume that tr is an integer for a given $t > 1$. Next we divide the index set Γ^c into two disjoint subsets, that is,

$$\Gamma_1 = \{i \in \Gamma^c : \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{H}}}(i,i,1) > \phi\}, \quad \Gamma_2 = \{i \in \Gamma^c : \boldsymbol{\mathcal{S}}_{\boldsymbol{\mathcal{H}}}(i,i,1) \leq \phi\},$$

where $\phi \triangleq \|\mathcal{H}_{\Gamma^c}\|_*/((t-1)r)$. Clearly,

$$\Gamma_1 \cup \Gamma_2 = \Gamma^c \text{ and } \Gamma_1 \cap \Gamma = \emptyset,$$

which implies that $\mathcal{H} = \mathcal{H}_{\Gamma} + \mathcal{H}_{\Gamma^c} = \mathcal{H}_{\Gamma} + \mathcal{H}_{\Gamma_1} + \mathcal{H}_{\Gamma_2}$ and $\|\mathcal{H}_{\Gamma}\|_F \le \|\mathcal{H}_{\Gamma\cup\Gamma_1}\|_F$, respectively. In order to prove (9), we only need to check

$$\|\mathcal{H}_{\Gamma\cup\Gamma_1}\|_F \le \eta_1 \|\mathfrak{M}(\mathcal{H})\|_2 + \frac{\eta_2}{\sqrt{r}} \|\mathcal{H}_{\Gamma^c}\|_*. \tag{A.1}$$

Let $\|\boldsymbol{s}_{\mathcal{H}_{\Gamma_1}}\|_1 \triangleq \sum_{i\in\Gamma_1} \mathcal{S}_{\mathcal{H}}(i,i,1) = \|\mathcal{H}_{\Gamma_1}\|_*$, where $\boldsymbol{s}_{\mathcal{H}_{\Gamma_1}}$ is denoted as the diagonal vector of first frontal slice of $\mathcal{S}_{\mathcal{H}}$ whose element $\mathcal{S}_{\mathcal{H}_{\Gamma_1}}(i,i,1) = \mathcal{S}_{\mathcal{H}}(i,i,1)$ for $i \in \Gamma_1$ and $\mathcal{S}_{\mathcal{H}_{\Gamma_1}}(i,i,1) = 0$ otherwise. Since all non-zero entries of vector $\boldsymbol{s}_{\mathcal{H}_{\Gamma_1}}$ have magnitude larger than ϕ, we have,

$$\|\boldsymbol{s}_{\mathcal{H}_{\Gamma_1}}\|_1 = \|\mathcal{H}_{\Gamma_1}\|_* > |\Gamma_1| \frac{\|\mathcal{H}_{\Gamma^c}\|_*}{(t-1)r} \ge |\Gamma_1| \frac{\|\mathcal{H}_{\Gamma_1}\|_*}{(t-1)r} = \frac{|\Gamma_1|}{(t-1)r} \|\boldsymbol{s}_{\mathcal{H}_{\Gamma_1}}\|_1.$$

Namely $|\Gamma_1| < (t-1)r$. Besides, we also have

$$\|\boldsymbol{s}_{\mathcal{H}_{\Gamma_2}}\|_1 = \|\mathcal{H}_{\Gamma_2}\|_* = \|\mathcal{H}_{\Gamma^c}\|_* - \|\mathcal{H}_{\Gamma_1}\|_* \le ((t-1)r - |\Gamma_1|)\phi$$

and

$$\|\boldsymbol{s}_{\mathcal{H}_{\Gamma_2}}\|_\infty \triangleq \max_{i\in\Gamma_2} \mathcal{S}_{\mathcal{H}}(i,i,1) \le \phi.$$

Now, since $\boldsymbol{s}_{\mathcal{H}_{\Gamma_2}} \in T(\phi, (t-1)r - |\Gamma_1|)$, applying Lemma 3.1, $\boldsymbol{s}_{\mathcal{H}_{\Gamma_2}}$ can be rewritten as:

$$\boldsymbol{s}_{\mathcal{H}_{\Gamma_2}} = \sum_{i=1}^{N} \gamma_i \boldsymbol{g}_i,$$

where $\boldsymbol{g}_i \in U(\phi, (t-1)r - |\Gamma_1|, \boldsymbol{s}_{\mathcal{H}_{\Gamma_2}})$ and $0 \le \gamma_i \le 1$, $\sum_{i=1}^{N} \gamma_i = 1$.

Step 2: Consequence of t-RIP.

Furthermore, define

$$\begin{aligned}
\boldsymbol{s}_{\mathcal{B}_i} &= (1+\delta_{tr}^{\mathfrak{M}})\boldsymbol{s}_{\mathcal{H}_{\Gamma\cup\Gamma_1}} + \delta_{tr}^{\mathfrak{M}} \boldsymbol{s}_{\mathcal{G}_i}, \quad \boldsymbol{s}_{\mathcal{P}_i} = (1-\delta_{tr}^{\mathfrak{M}})\boldsymbol{s}_{\mathcal{H}_{\Gamma\cup\Gamma_1}} - \delta_{tr}^{\mathfrak{M}} \boldsymbol{s}_{\mathcal{G}_i},\\
\mathcal{G}_i &= \sum\nolimits_{j=1}^{\kappa} \mathcal{U}_{\mathcal{H}}(:,j,:) \star \mathcal{S}_{\mathcal{G}_i}(j,j,:) \star \mathcal{V}_{\mathcal{H}}(:,j,:)^*,\\
\mathcal{B}_i &= \sum\nolimits_{j=1}^{\kappa} \mathcal{U}_{\mathcal{H}}(:,j,:) \star \mathcal{S}_{\mathcal{B}_i}(j,j,:) \star \mathcal{V}_{\mathcal{H}}(:,j,:)^*,\\
\mathcal{P}_i &= \sum\nolimits_{j=1}^{\kappa} \mathcal{U}_{\mathcal{H}}(:,j,:) \star \mathcal{S}_{\mathcal{P}_i}(j,j,:) \star \mathcal{V}_{\mathcal{H}}(:,j,:)^*.
\end{aligned}$$

Then it is not hard to see that both $\boldsymbol{B}_i$ and $\boldsymbol{P}_i$ are all tensors with tubal rank at most tr for $i = 1, 2, \cdots, N$, and

$$\boldsymbol{\mathcal{H}}_{\Gamma_2} = \sum\nolimits_{i=1}^{N} \gamma_i \boldsymbol{\mathcal{G}}_i,\ \ \boldsymbol{\mathcal{B}}_i = (1+\delta_{tr}^{\mathfrak{M}})\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1} + \delta_{tr}^{\mathfrak{M}}\boldsymbol{\mathcal{G}}_i,\ \ \boldsymbol{\mathcal{P}}_i = (1-\delta_{tr}^{\mathfrak{M}})\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1} - \delta_{tr}^{\mathfrak{M}}\boldsymbol{\mathcal{G}}_i.$$

Now we estimate the upper bounds of

$$\xi \triangleq \sum\nolimits_{i=1}^{N} \gamma_i \left(\|\mathfrak{M}(\boldsymbol{\mathcal{B}}_i)\|_2^2 - \|\mathfrak{M}(\boldsymbol{\mathcal{P}}_i)\|_2^2 \right).$$

Applying Definition 1.2, we have

$$\begin{aligned} \xi &= 4\delta_{tr}^{\mathfrak{M}} \sum\nolimits_{i=1}^{N} \gamma_i \left\langle \mathfrak{M}(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}), \mathfrak{M}(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1} + \boldsymbol{\mathcal{G}}_i) \right\rangle \\ &\overset{(a)}{=} 4\delta_{tr}^{\mathfrak{M}} \left\langle \mathfrak{M}(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}), \mathfrak{M}\left(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1} + \sum\nolimits_{i=1}^{N} \gamma_i \boldsymbol{\mathcal{G}}_i \right) \right\rangle \\ &\overset{(b)}{=} 4\delta_{tr}^{\mathfrak{M}} \langle \mathfrak{M}(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}), \mathfrak{M}(\boldsymbol{\mathcal{H}}) \rangle \\ &\overset{(c)}{\leq} 4\delta_{tr}^{\mathfrak{M}} \|\mathfrak{M}(\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1})\|_2 \|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 \\ &\overset{(d)}{\leq} 4\delta_{tr}^{\mathfrak{M}} \sqrt{1+\delta_{tr}^{\mathfrak{M}}} \|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F \|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2, \end{aligned} \tag{A.2}$$

where (a) is due to $\sum_{i=1}^{N} \gamma_i = 1$, (b) is founded on the fact that $\boldsymbol{\mathcal{H}}_{\Gamma_2} = \sum_{i=1}^{N} \gamma_i \boldsymbol{\mathcal{G}}_i$ and $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{H}}_{\Gamma} + \boldsymbol{\mathcal{H}}_{\Gamma_1} + \boldsymbol{\mathcal{H}}_{\Gamma_2}$, (c) holds because of the Cauchy-Schwarz inequality, and (d) follows from (5), $|\Gamma_1| < (t-1)r$ and the monotonicity of t-RIC.

Next, we use the block diagonal matrix to estimate the lower bound of ξ. Let $\bar{\phi} \triangleq \|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma^c})\|_*/(t-1)r$. Repeat step 1 for the matrix $\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}})$ as we did for tensor $\boldsymbol{\mathcal{H}}$ and we have

$$\sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma_2})) \in T(\bar{\phi}, (t-1)r - |E_1|),\ \ \bar{\boldsymbol{g}}_i \in U(\bar{\phi}, (t-1)r - |E_1|, \sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma_2}))),$$

here, E_1 is an index set as the counterpart of Γ_1. By further defining

$$\begin{aligned} \bar{\boldsymbol{b}}_i &= (1+\delta_{tr}^{\mathfrak{M}})\sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1})) + \delta_{tr}^{\mathfrak{M}} \bar{\boldsymbol{g}}_i, \\ \bar{\boldsymbol{p}}_i &= (1-\delta_{tr}^{\mathfrak{M}})\sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1})) - \delta_{tr}^{\mathfrak{M}} \bar{\boldsymbol{g}}_i, \\ \bar{\boldsymbol{G}}_i &= \sum\nolimits_{j=1}^{\kappa} (\boldsymbol{u}_{\bar{\boldsymbol{H}}})_j \cdot (\bar{\boldsymbol{g}}_i)_j \cdot (\boldsymbol{v}_{\bar{\boldsymbol{H}}})_j^*, \\ \bar{\boldsymbol{B}}_i &= \sum\nolimits_{j=1}^{\kappa} (\boldsymbol{u}_{\bar{\boldsymbol{H}}})_j \cdot (\bar{\boldsymbol{b}}_i)_j \cdot (\boldsymbol{v}_{\bar{\boldsymbol{H}}})_j^*, \\ \bar{\boldsymbol{P}}_i &= \sum\nolimits_{j=1}^{\kappa} (\boldsymbol{u}_{\bar{\boldsymbol{H}}})_j \cdot (\bar{\boldsymbol{p}}_i)_j \cdot (\boldsymbol{v}_{\bar{\boldsymbol{H}}})_j^*. \end{aligned}$$

Then we can easily induce that

$$
\begin{aligned}
\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma_2}) &= \sum\nolimits_{i=1}^{N} \gamma_i \bar{\boldsymbol{G}}_i, \\
\bar{\boldsymbol{B}}_i &= (1+\delta_{tr}^{\mathfrak{M}}) \operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1}) + \delta_{tr}^{\mathfrak{M}} \bar{\boldsymbol{G}}_i, \\
\bar{\boldsymbol{P}}_i &= (1-\delta_{tr}^{\mathfrak{M}}) \operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1}) - \delta_{tr}^{\mathfrak{M}} \bar{\boldsymbol{G}}_i.
\end{aligned}
$$

Thus, on the other hand, we also have

$$
\begin{aligned}
\xi &\overset{(a)}{\geq} \sum\nolimits_{i=1}^{N} \gamma_i \left((1-\delta_{tr}^{\mathfrak{M}})\|\boldsymbol{\mathcal{B}}_i\|_F^2 - (1+\delta_{tr}^{\mathfrak{M}})\|\boldsymbol{\mathcal{P}}_i\|_F^2\right) \\
&\overset{(b)}{=} \frac{1}{n_3}\sum\nolimits_{i=1}^{N} \gamma_i \left((1-\delta_{tr}^{\mathfrak{M}})\|\bar{\boldsymbol{B}}_i\|_2^2 - (1+\delta_{tr}^{\mathfrak{M}})\|\bar{\boldsymbol{P}}_i\|_2^2\right) \\
&\overset{(c)}{=} \frac{2}{n_3}\delta_{tr}^{\mathfrak{M}}(1-(\delta_{tr}^{\mathfrak{M}})^2)\|\sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1}))\|_2^2 - \frac{2}{n_3}(\delta_{tr}^{\mathfrak{M}})^3 \sum\nolimits_{i=1}^{N} \gamma_i \|\bar{\boldsymbol{g}}_i\|_2^2 \\
&\overset{(d)}{\geq} \frac{2}{n_3}\delta_{tr}^{\mathfrak{M}}(1-(\delta_{tr}^{\mathfrak{M}})^2)\|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1})\|_F^2 - \frac{2(\delta_{tr}^{\mathfrak{M}})^3}{n_3(t-1)r}\|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma^c})\|_*^2 \\
&\overset{(e)}{=} 2\delta_{tr}^{\mathfrak{M}}(1-(\delta_{tr}^{\mathfrak{M}})^2)\|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F^2 - \frac{2n_3(\delta_{tr}^{\mathfrak{M}})^3}{(t-1)r}\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*^2,
\end{aligned}
\tag{A.3}
$$

where (a) follows from t-RIP, (b) holds because of (6), (c) is due to $\langle \sigma(\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma\cup\Gamma_1})), \bar{\boldsymbol{g}}_i\rangle = 0$ for all $i = 1, 2, \cdots, N$, (d) is based on the fact that $\|\boldsymbol{X}\|_F = \|\sigma(\boldsymbol{X})\|_2$ for any matrix $\boldsymbol{X}$ and

$$
\|\bar{\boldsymbol{g}}_i\|_2^2 \leq \|\bar{\boldsymbol{g}}_i\|_0(\|\bar{\boldsymbol{g}}_i\|_\infty)^2 \leq ((t-1)r - |E_1|)\bar{\phi}^2 \leq \frac{\|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma^c})\|_*^2}{(t-1)r},
$$

and (e) follows from (7).

Combining (A.2) and (A.3), we get

$$
(1-(\delta_{tr}^{\mathfrak{M}})^2)\|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F^2 - \frac{n_3(\delta_{tr}^{\mathfrak{M}})^2}{(t-1)r}\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*^2 \leq 2\sqrt{1+\delta_{tr}^{\mathfrak{M}}}\|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2.
\tag{A.4}
$$

Obviously, (A.4) is a quadratic inequality in terms of $\|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F$. Using extract roots formula, we obtain

$$
\begin{aligned}
&\|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1}\|_F \\
\leq\ & \frac{2\sqrt{1+\delta_{tr}^{\mathfrak{M}}}\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 + \sqrt{(2\sqrt{1+\delta_{tr}^{\mathfrak{M}}}\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2)^2 + 4(1-(\delta_{tr}^{\mathfrak{M}})^2)\frac{n_3(\delta_{tr}^{\mathfrak{M}})^2}{(t-1)r}\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*^2}}{2(1-(\delta_{tr}^{\mathfrak{M}})^2)} \\
\leq\ & \frac{2}{(1-\delta_{tr}^{\mathfrak{M}})\sqrt{1+\delta_{tr}^{\mathfrak{M}}}}\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 + \frac{\sqrt{n_3}\delta_{tr}^{\mathfrak{M}}}{\sqrt{(1-(\delta_{tr}^{\mathfrak{M}})^2)(t-1)}}\frac{\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*}{\sqrt{r}},
\end{aligned}
$$

where the last inequality is based on the fact that $\sqrt{x^2+y^2} \leq |x|+|y|$. Therefor we prove (A.1). Since we also have $\|\boldsymbol{\mathcal{H}}\|_F \leq \|\boldsymbol{\mathcal{H}}_{\Gamma\cup\Gamma_1\|_F}$, it is easy to induce (9), which completes the proof. □

Appendix B. Proof of Lemma 3.3

Proof. Since $\hat{\boldsymbol{\mathcal{X}}}$ is the minimizer of (4), we have

$$\|\hat{\boldsymbol{\mathcal{X}}}\|_* + \frac{1}{2\lambda}\|\boldsymbol{y} - \mathfrak{M}(\hat{\boldsymbol{\mathcal{X}}})\|_2^2 \leq \|\boldsymbol{\mathcal{X}}\|_* + \frac{1}{2\lambda}\|\boldsymbol{y} - \mathfrak{M}(\boldsymbol{\mathcal{X}})\|_2^2.$$

Also because $\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{H}} + \boldsymbol{\mathcal{X}}$ and $\boldsymbol{y} = \mathfrak{M}(\boldsymbol{\mathcal{X}}) + \boldsymbol{w}$, so the above inequality is equivalent to

$$\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2\langle \boldsymbol{w}, \mathfrak{M}(\boldsymbol{\mathcal{H}})\rangle \leq 2\lambda(\|\boldsymbol{\mathcal{X}}\|_* - \|\hat{\boldsymbol{\mathcal{X}}}\|_*).$$

It follows from the Cauchy-Schwarz inequality and assumption $\|\boldsymbol{w}\|_2 \leq \epsilon$ that

$$\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2\langle \boldsymbol{w}, \mathfrak{M}(\boldsymbol{\mathcal{H}})\rangle \geq \|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2\epsilon\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2. \tag{B.1}$$

On the other hand, we have

$$\begin{aligned}
\|\hat{\boldsymbol{\mathcal{X}}}\|_* - \|\boldsymbol{\mathcal{X}}\|_* &= \|(\boldsymbol{\mathcal{H}}+\boldsymbol{\mathcal{X}})_\Gamma\|_* + \|(\boldsymbol{\mathcal{H}}+\boldsymbol{\mathcal{X}})_{\Gamma^c}\|_* - (\|\boldsymbol{\mathcal{X}}_\Gamma\|_* + \|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*) \\
&\geq (\|\boldsymbol{\mathcal{X}}_\Gamma\|_* - \|\boldsymbol{\mathcal{H}}_\Gamma\|_*) + (\|\boldsymbol{\mathcal{H}}_{\Gamma^c}|_* - \|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*) - (\|\boldsymbol{\mathcal{X}}_\Gamma\|_* + \|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*) \\
&\geq \|\boldsymbol{\mathcal{H}}_{\Gamma^c}|_* - \|\boldsymbol{\mathcal{H}}_\Gamma|_* - 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}|_*. \qquad \text{(B.2)}
\end{aligned}$$

Combining (B.1) and (B.2) and by a simple calculation, we get (10). As to (11), it is obtained by subtracting the term $\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2$ from the left-hand side of (10). □

Appendix C. Proof of Theorem 4.1

Proof. For convenience, let

$$T = \mathrm{supp}(\boldsymbol{s}_{\boldsymbol{\mathcal{X}}_{\max(r)}})$$

be an index set with cardinality $|T| \leq r$. In addition, if we set $\boldsymbol{\mathcal{H}} = \hat{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}$ and $\mathrm{rank}(\mathrm{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_\Gamma)) = \bar{r}$, then by inequality (9) and (10), we would get

$$
\begin{aligned}
\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2\epsilon\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 &\leq 2\lambda(\|\boldsymbol{\mathcal{H}}_\Gamma\|_* - \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*) \\
&\stackrel{(a)}{=} 2\lambda\left(\frac{1}{n_3}\|\,\mathrm{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_\Gamma)\|_* - \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*\right) \\
&\leq 2\lambda\left(\frac{\sqrt{\bar{r}}}{n_3}\|\,\mathrm{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_\Gamma)\|_F - \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*\right) \\
&\stackrel{(b)}{\leq} 2\lambda\left(\sqrt{r}\|\boldsymbol{\mathcal{H}}_\Gamma\|_F - \|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*\right) \\
&\leq 2\sqrt{r}\lambda\left(\eta_1\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 + \frac{\eta_2}{\sqrt{r}}\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*\right) - 2\lambda\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 4\lambda\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* \\
&= 2\sqrt{r}\eta_1\lambda\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 - 2(1-\eta_2)\lambda\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* + 4\lambda\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*,
\end{aligned}
$$

where (a) follows from (7) and (b) is due to (6), (8). The assumption (12) implies that

$$
1-\eta_2 = 1-\frac{\sqrt{n_3}\delta_{tr}^{\mathfrak{M}}}{\sqrt{(1-(\delta_{tr}^{\mathfrak{M}})^2)(t-1)}} > 1-\frac{\sqrt{n_3}\sqrt{(t-1)/(n_3^2+t-1)}}{\sqrt{(1-(t-1)/(n_3^2+t-1))\,(t-1)}} = 0,
$$

and hence

$$
\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2^2 - 2(\sqrt{r}\eta_1\lambda+\epsilon)\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 - 4\lambda\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* \leq 0,
$$

which implies that

$$
\begin{aligned}
\left(\|\mathfrak{M}(\boldsymbol{\mathcal{H}})\|_2 - (\sqrt{r}\eta_1\lambda+\epsilon)\right)^2 &\leq (\sqrt{r}\eta_1\lambda+\epsilon)^2 + 4\lambda\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* \\
&\leq \left(\sqrt{r}\eta_1\lambda+\epsilon+\frac{2\lambda\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*}{\sqrt{r}\eta_1\lambda+\epsilon}\right)^2 \\
&\leq \left(\sqrt{r}\eta_1\lambda+\epsilon+\frac{2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*}{\sqrt{r}\eta_1}\right)^2.
\end{aligned}
$$

Therefore, we conclude that (13) holds. Plugging (13) into (11), by $\|\boldsymbol{\mathcal{H}}_\Gamma\|_* \leq \sqrt{r}\|\boldsymbol{\mathcal{H}}_\Gamma\|_F$, we get

$$
\begin{aligned}
\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_* &\leq \|\boldsymbol{\mathcal{H}}_\Gamma\|_* + 2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* + \frac{\epsilon}{\lambda}\left(\frac{2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*}{\sqrt{r}\eta_1} + 2\sqrt{r}\eta_1\lambda + 2\epsilon\right) \\
&\leq \sqrt{r}\|\boldsymbol{\mathcal{H}}_\Gamma\|_F + \frac{2(\sqrt{r}\eta_1\lambda+\epsilon)}{\sqrt{r}\eta_1\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_* + \frac{\epsilon}{\lambda}(2\sqrt{r}\eta_1\lambda+2\epsilon). \quad \text{(C.1)}
\end{aligned}
$$

Combining (9), (13) and (C.1) yields

$$
\begin{aligned}
\|\boldsymbol{\mathcal{H}}_\Gamma\|_F &\leq \eta_1\left(\frac{2\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*}{\sqrt{r}\eta_1}+2\sqrt{r}\eta_1\lambda+2\epsilon\right)\\
&\quad +\frac{\eta_2}{\sqrt{r}}\left(\sqrt{r}\|\boldsymbol{\mathcal{H}}_\Gamma\|_F+\frac{2(\sqrt{r}\eta_1\lambda+\epsilon)}{\sqrt{r}\eta_1\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+\frac{\epsilon}{\lambda}(2\sqrt{r}\eta_1\lambda+2\epsilon)\right)\\
&= \eta_2\|\boldsymbol{\mathcal{H}}_\Gamma\|_F+\frac{2\sqrt{r}\eta_1(1+\eta_2)\lambda+2\eta_2\epsilon}{r\eta_1\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+(\eta_1+\frac{\epsilon}{\lambda})(2\sqrt{r}\eta_1\lambda+2\epsilon).
\end{aligned}
$$

Note that $1-\eta_2>0$, so the above inequality leads to

$$
\|\boldsymbol{\mathcal{H}}_\Gamma\|_F\leq\frac{2\sqrt{r}\eta_1(1+\eta_2)\lambda+2\eta_2\epsilon}{r\eta_1(1-\eta_2)\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+\frac{(\eta_1\lambda+\epsilon)(2\sqrt{r}\eta_1\lambda+2\epsilon)}{(1-\eta_2)\lambda}. \quad \text{(C.2)}
$$

To prove (14), application of (C.1) and (C.2) yields

$$
\begin{aligned}
\|\boldsymbol{\mathcal{H}}\|_F &\leq \|\boldsymbol{\mathcal{H}}_\Gamma\|_F+\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_F\\
&\leq (\sqrt{n_3r}+1)\|\boldsymbol{\mathcal{H}}_\Gamma\|_F+\frac{2\sqrt{n_3}(\sqrt{r}\eta_1\lambda+\epsilon)}{\sqrt{r}\eta_1\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+\frac{\epsilon}{\lambda}\sqrt{n_3}(2\sqrt{r}\eta_1\lambda+2\epsilon)\\
&\leq (\sqrt{n_3r}+1)\left(\frac{2\sqrt{r}\eta_1(1+\eta_2)\lambda+2\eta_2\epsilon}{r\eta_1(1-\eta_2)\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+\frac{(\eta_1\lambda+\epsilon)(2\sqrt{r}\eta_1\lambda+2\epsilon)}{(1-\eta_2)\lambda}\right)\\
&\quad +\frac{2\sqrt{n_3}(\sqrt{r}\eta_1\lambda+\epsilon)}{\sqrt{r}\eta_1\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*+\frac{\epsilon}{\lambda}\sqrt{n_3}(2\sqrt{r}\eta_1\lambda+2\epsilon)\\
&\leq \frac{2\sqrt{r}\eta_1(2\sqrt{n_3r}+1+\eta_2)\lambda+2(\sqrt{n_3r}+\eta_2)\epsilon}{r\eta_1(1-\eta_2)\lambda}\|\boldsymbol{\mathcal{X}}_{\Gamma^c}\|_*\\
&+ \frac{(\sqrt{n_3r}+1)\eta_1\lambda+(\sqrt{n_3r}-\sqrt{n_3}\eta_2+\sqrt{n_3}+1)\epsilon}{(1-\eta_2)\lambda(2\sqrt{r}\eta_1\lambda+2\epsilon)^{-1}},
\end{aligned}
$$

where the second inequality is because of $\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_F=\frac{1}{\sqrt{n_3}}\|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma^c})\|_F\leq\frac{1}{\sqrt{n_3}}\|\operatorname{bdiag}(\bar{\boldsymbol{\mathcal{H}}}_{\Gamma^c})\|_*=\sqrt{n_3}\|\boldsymbol{\mathcal{H}}_{\Gamma^c}\|_*$. So far, we have completed the proof. □

References

[1] J. Liu, P. Musialski, P. Wonka, J. Ye, Tensor completion for estimating missing values in visual data, IEEE Trans. Pattern Anal. Mach. 35 (1) (2013) 208–220.

[2] B. Romera-Paredes, H. Aung, N. Bianchi-Berthouze, M. Pontil, Multilinear multitask learning, in: Int. Conf. on Machine Learning, Proceedings of Machine Learning Research, Atlanta, 2013, pp. 1444–1452.

[3] M. Mørup, Applications of tensor (multiway array) factorizations and decompositions in data mining, Wiley Interdiscip. Rev.-Data Mining Knowl. Discov. 1 (1) (2011) 24–40.

[4] H. A. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemometr. 14 (3) (2000) 105–122.

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

[6] M. E. Kilmer, C. D. Martin, Factorization strategies for third-order tensors, Linear Alg. Appl. 435 (3) (2011) 641–658.

[7] C. D. Martin, R. Shafer, B. LaRue, An order-p tensor factorization with applications in imaging, SIAM J. Sci. Comput. 35 (1) (2013) A474–A490.

[8] I. V. Oseledets, Tensor-train decomposition, SIAM J. Sci. Comput. 33 (5) (2011) 2295–2317.

[9] M. E. Kilmer, K. Braman, N. Hao, 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 (1) (2013) 148–172.

[10] O. Semerci, N. Hao, M. E. Kilmer, E. L. Miller, Tensor-based formulation and nuclear norm regularization for multienergy computed tomography, IEEE Trans. Image Process. 23 (4) (2014) 1678–1693.

[11] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm (2018). arXiv:1804.03728.

[12] H. Zhang, M. Yan, W. Yin, One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimizations, Adv. Comput. Math. 42 (6) (2016) 1381–1399.

[13] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-svd, in: Proc. CVPR, IEEE, Columbus, 2014, pp. 3842–3849.

[14] E. J. Candes, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215.

[15] Z. Shi, J. Han, T. Zheng, J. Li, Guarantees of augmented trace norm models in tensor recovery, in: IJCAI, AAAI, Beijing, 2013, pp. 1670–1676.

[16] H. Rauhut, R. Schneider, Ž. Stojanac, Low rank tensor recovery via iterative hard thresholding, Linear Alg. Appl. 523 (2017) 220–262.

[17] E. J. Candes, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, IEEE Trans. Inf. Theory 57 (4) (2011) 2342–2359.

[18] E. J. Candes, The restricted isometry property and its implications for compressed sensing, C. R. Math. 346 (9) (2008) 589–592.

[19] T. T. Cai, A. Zhang, Sharp rip bound for sparse signal and low-rank matrix recovery, Appl. Comput. Harmon. Anal. 35 (1) (2013) 74–93.

[20] T. T. Cai, A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, IEEE Trans. Inf. Theory 60 (1) (2014) 122–132.

[21] R. Zhang, S. Li, A proof of conjecture on restricted isometry property constants $\delta_{tk}(0 < t < 4/3)$, IEEE Trans. Inf. Theory 64 (3) (2018) 1699–1705.

[22] C. Zhu, Stable recovery of sparse signals via regularized minimization, IEEE Trans. Inf. Theory 54 (7) (2008) 3364–3367.

[23] Y. Shen, B. Han, E. Braverman, Stable recovery of analysis based approaches, Appl. Comput. Harmon. Anal. 39 (1) (2015) 161–172.

[24] H. Ge, J. Wen, W. Chen, J. Weng, M.-J. Lai, Stable sparse recovery with three unconstrained analysis based approaches, http://alpha.math.uga.edu/mjlai/papers/20180126.pdf (2018).

[25] S. Foucart, Stability and robustness of ℓ_1-minimizations with weibull matrices and redundant dictionaries, Linear Alg. Appl. 441 (2014) 4–21.

[26] C. Lu, J. Feng, Z. Lin, S. Yan, Exact low tubal rank tensor recovery from gaussian measurements (2018). arXiv:1806.02511.