Noisy Tensor Completion via Low-Rank Tensor Ring

Yuning Qiu, Guoxu Zhou, Qibin Zhao, Senior Member, IEEE, and Shengli Xie, Fellow, IEEE

Abstract—Tensor completion is a fundamental tool for incomplete data analysis, where the goal is to predict missing entries from partial observations. However, existing methods often make the explicit or implicit assumption that the observed entries are noise-free to provide a theoretical guarantee of exact recovery of missing entries, which is quite restrictive in practice. To remedy such drawback, this article proposes a novel noisy tensor completion model, which complements the incompetence of existing works in handling the degeneration of high-order and noisy observations. Specifically, the tensor ring nuclear norm (TRNN) and least-squares estimator are adopted to regularize the underlying tensor and the observed entries, respectively. In addition, a nonasymptotic upper bound of estimation error is provided to depict the statistical performance of the proposed estimator. Two efficient algorithms are developed to solve the optimization problem with convergence guarantee, one of which is specially tailored to handle large-scale tensors by replacing the minimization of TRNN of the original tensor equivalently with that of a much smaller one in a heterogeneous tensor decomposition framework. Experimental results on both synthetic and real-world data demonstrate the effectiveness and efficiency of the proposed model in recovering noisy incomplete tensor data compared with state-of-the-art tensor completion models.

Index Terms—Image/video inpainting, low-rank tensor recovery, tensor completion, tensor ring (TR) decomposition.

I. INTRODUCTION

TENSOR is an array of numbers, giving a faithful and effective representation to maintain the intrinsic structure of multidimensional data [1]. Many data collected in real-world applications can be naturally expressed as high-order tensors. For example, a color video sequence can be viewed as a fourth-order tensor due to its spatial, color, and temporal variables; a light field image can be formulated as a fifth-order tensor indexed by one color, two spatial, and two angular variables. For this reason, numerous theoretical and numerical tools for tensor data analysis have been developed and applied in many fields, including computer vision [2], multiview clustering [3]–[5], blind source separation [6], bioinformatics [7], [8], and pattern recognition [9]. Among them, tensor completion is one of the most practical and significant problems, which aims at predicting the missing entries of a low-rank tensor with partial observations. Recent works have demonstrated that the incomplete low-rank tensor can be exactly or approximately recovered under some appropriate assumptions and have been successfully implemented to broad tensor completion applications [10]–[12], such as images and videos inpainting [13], [14] and multirelational link prediction [15], [16], and achieve promising performance.

Low-rank tensor completion (LRTC) can be viewed as a multidimensional extension of low-rank matrix completion (LRMC), which aims at recovering the intrinsic low-rank matrix with incomplete observations [17]–[21]. However, this extension is rather nontrivial since it is difficult to find a well-defined tensor rank for complex multilinear structure. Two most popular tensor rank definitions are CANDECOMP/ PARAFAC (CP) rank [22] and Tucker rank [23]. For a given...
tensor, CP decomposition factorizes a tensor into the sum of rank-1 components, and the minimum number of these rank-1 components is defined as the CP rank. However, CP decomposition often suffers from the issue of ill-posedness, i.e., the optimal low-rank approximation does not always exist, which may lead to a poor fit in practical applications [24], [25]. Compared with CP rank, Tucker rank is a more promising alternative, which is defined on the rank of unfolding matrices along each mode [1]. Such treatment is often convenient and reasonable since it can obtain multidimensional low-rank structure maintained in high-order tensors [26]–[28]. However, this unbalanced (i.e., one versus the others) unfolding scheme often yields fat matrices that have been blamed for their poor restoration performance in the LRMC problem. As a result, the performance of Tucker rank-based tensor completion methods often tends to degrade due to this inevitable issue [10], [29].

In recent years, several novel tensor rank definitions have been proposed to achieve the state-of-the-art tensor completion performance, e.g., tensor train (TT) rank [30], tensor tree rank [31], and tensor ring (TR) rank [32]. In particular, TR rank is one of the most promising methods. According to the circular structure of TR decomposition, given a high-order tensor, the TR rank minimization problem can be squared to multiple rank minimization subproblems of square unfolding matrix [33], which can conceptually solve the existing issues of Tucker and TT rank. Fig. 1 shows the intuitive representation of TR decomposition and differences between three unfolding schemes. Wang et al. [34] first introduced a low-rank TR completion model by alternatively minimizing the core tensors. To avoid manually adjusting TR rank, Yuan et al. [35] established a new TR rank minimization scheme by minimizing nuclear norm on all core tensors. Yu et al. [33] proposed a tensor ring nuclear norm (TRNN) minimization model for LRTC. To guarantee the completion performance of TRNN, and Huang et al. [12] further demonstrated that given a tensor satisfied strong TR incoherence condition, it can be exactly recovered with high probability. Yu et al. [36] proposed a parallel matrix factorization-based method for low-TR-rank tensor completion.

Most LRTC methods mentioned above assume that the observed entries are noiseless. However, in real-world applications, data with missing issue caused by sensor failures, storage errors, or other mistakes also usually suffer from noise corruptions. Thus, it is rather desirable to predict tensor data from both noisy and incomplete observations. Although noisy tensor completion problem is more practical and worthwhile, there is very limited literature and only a few related works on this aspect. Tensor decomposition or total variation regularization-based models have been proposed to alleviate the noisy completion problems [14], [37], [38], and however, the statistical performance of all these methods is still unclear. For this reason, Wang et al. [39]–[42] introduced several tensor tubal rank-based methods to solve the third-order noisy tensor completion problem with theoretical guarantee. Albeit interesting and promising, tensor tubal rank is very sensitive to the choice of the third mode and particularly difficult to capture complex intramode and intermode correlations for high-order tensors.

In this article, we propose to recover tensor data from incomplete noisy observations, which substantially generalizes the noisy matrix completion problem by not only deriving theoretical results and practical applications of noisy completion to high-order tensors but also providing more scalable and efficient algorithms. The main contributions of this article can be summarized as follows.

1) We propose a novel tensor completion model to predict the missing entries with noisy observations based on the low-rank TR model, which complements the incompetence of existing works in handling the degeneration of high-order and noisy observations. We analyze a nonasymptotic upper bound of the estimation error to reveal the statistical performance of the proposed model.

2) We derive two algorithms based on alternating direction method of multipliers (ADMM) to solve the optimization problem with convergence guarantee, namely, noisy tensor ring completion (NTRC), and Fast NTRC (FaNTRC). For FaNTRC, we minimize the TRNN equivalently on a much smaller tensor in a heterogeneous tensor decomposition framework, which has been shown to significantly improve the computational efficiency, especially dealing with large-scale low-TR-rank tensor data.

3) The proposed NTRC and FaNTRC are successfully implemented in various noisy tensor data completion problems, and the experimental results demonstrate their superiority compared with state-of-the-art LRTC methods.
The remainder of this article is organized as follows. Section II presents a brief review on some notations and preliminaries. In Section III, we introduce the noisy TR completion model and analyze its nonasymptotic upper bound on the Frobenius norm. In Section IV, we develop two efficient algorithms to solve the optimization problem with convergence guarantee. Section V gives experimental results of noisy tensor completion tasks on both synthetic and real-world tensor data.

II. NOTATIONS AND PRELIMINARIES

A. Notations

We list some notations and their corresponding abbreviations in the Nomenclature. Let $\mathbf{x} \in \mathbb{R}^N$ be independent and identically distributed (i.i.d.) Rademacher sequences. Let $c_0$ and $c_1$ be the generic absolute constants. Let $D$ be the number of total entries of the given tensor, i.e., $D = \prod_{k=1}^{K} d_k$. For any $a, b \in \mathbb{R}$, we let $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. For the circular mode-$(k, s)$ unfolding matrix $\mathbf{T}_{(k,s)}$ of size $d_1 \times d_2$, we let $\mathbf{d}_k = d_1 \times d_2$, $\mathbf{d}_k = d_1 \times d_2$, and $k^r = \arg\min_{k \in [K]}(d_1 \times d_2)$.

B. Tensor Preliminaries

Definition 1 (TR Decomposition [32]): The TR decomposition represents a $K$-th-order tensor $\mathcal{T} \in \mathbb{R}^{d_1 \times \cdots \times d_K}$ by the circular multilinear product over a sequence of third-order core tensors, i.e., $\mathcal{T} = \mathcal{T}(\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \ldots, \mathcal{G}^{(K)})$, where $\mathcal{G}^{(k)} \in \mathbb{R}^{d_k \times d_k \times d_{k+1}}$, $k \in [K]$, and $r_{K+1} = r_1$. Elementwisely, it can be represented as

$$\mathcal{T}(i_1, i_2, \ldots, i_K) = \prod_{k=1}^{K} \mathcal{G}^{(k)}(v_{ik}, v_{ik+1}).$$

The size of cores, $r_k$, $k = 1, \ldots, K$, denoted by a vector $[r_1, \ldots, r_K]$, is called TR rank.

Remark 1 (Multiple States in TR Decomposition): TR decomposition can be divided into three states, i.e., supercritical ($d_k < r_k r_{k+1}$), critical ($d_k = r_k r_{k+1}$), and subcritical ($d_k > r_k r_{k+1}$) states [43]. Previous work [34] has also experimentally found this fact in some real-world tensor data in the sense that letting $r_k r_{k+1} > d_k$ to achieve more favorable recovery performance.

Lemma 1: Tensor data with TR supercritical or critical states are full-Tucker-rank.

Proof of Lemma 1 is a part of proof of Lemma 3, and it can be found in Appendix I of Supplementary Material. According to Lemma 1, incomplete tensor with TR supercritical or critical states may not be recovered by Tucker rank minimization-based methods since it is full-Tucker-rank approximation problem.

Next, we give three tensor-matrix unfolding schemes and show their relationship with tensor ranks.

Definition 2 (Multi-Index Operation [44]): The multi-index operation is given by

$$i_1, i_2, \ldots, i_K = i_1 + (i_2 - 1)d_1 + (i_3 - 1)d_1 d_2 + \cdots + (i_K - 1)d_1 \cdots d_{K-1}$$

where $i_k \in [d_k]$, $k \in [K]$.

Definition 3 (Canonical Mode- $k$ Unfolding [11]): Let $\mathcal{T}$ be a $K$-th-order tensor, and its canonical mode- $k$ unfolding is denoted by $\mathbf{T}_{(k)}$ of size $d_k \times \prod_{j \neq k} d_j$, whose elements are given by

$$\mathcal{T}_{(k)}(i_k, i_1, \ldots, i_{k-1}i_{k+1}, \ldots, i_K) = \mathcal{T}(i_1, i_2, \ldots, i_K).$$

Definition 4 (First $k$-Modes Unfolding [30]): Let $\mathcal{T}$ be a $K$-th-order tensor, and its first $k$-modes unfolding is denoted by $\mathbf{T}_{(k)}$ of size $\prod_{i=1}^{k-1} d_i \times \prod_{j=k+1}^{K} d_j$, whose elements are given by

$$\mathbf{T}_{(k)}(i_1i_2, \ldots, i_k, i_{k+1}, \ldots, i_K) = \mathcal{T}(i_1, i_2, \ldots, i_K).$$

Definition 5 (Circular Mode-$(k, s)$ Unfolding [33]): Let $\mathcal{T}$ be a $K$-th-order tensor, and its circular mode-$(k, s)$ unfolding is a matrix, denoted by $\mathbf{T}_{(k,s)}$ of size $d_1 \times d_2$, where $d_1 = \prod_{i=1}^{k-1} d_i$, $d_2 = \prod_{i=k}^{K} d_i$, and $s$ is the number of indexes included in $d_2$, and

$$l = \left\{ \begin{array}{ll} k + s - 1, & k + s \leq K \\ k + s - 1 - K, & \text{otherwise} \end{array} \right.$$

Alternatively, its elementwise form is given by

$$\mathbf{T}_{(k,s)}(i_1, i_2, \ldots, i_K) = \text{reshape} \left( \text{permute}(\mathcal{T}, [l+1, \ldots, k-1, \ldots, s, \ldots, l], \prod_{u=l+1}^{k-1} d_u, \prod_{v=s}^{K} d_v) \right).$$

Remark 2 (Relationship Between Different Tensor Ranks and Unfolding Schemes): Given an arbitrary $K$-th-order tensor with TR rank $[r_1, \ldots, r_K]$, the rank of each circular mode-$(k, s)$ unfolding matrix is bounded by $r_{k} r_{k+1}$ [33]. This relationship can also be found in Tucker (or TT) rank between canonical mode- $k$ (or first $k$-modes) unfolding [1], [30]. Therefore, the complex tensor rank minimization problem can be equivalently reduced to a series of matrix rank minimization subproblems.

According to Remark 2, to minimize TR rank, a natural option is to consider the sum of rank of unfolding matrices

$$\min_{\mathcal{T}} \sum_{k=1}^{K} a_k \text{rank}(\mathbf{T}_{(k,s)}).$$

However, problem (6) is computationally intractable in general. Motivated by the proxy of rank function, the sum of nuclear norm has been adopted as a convex surrogate of (6) and revealed in the following definition.

Definition 6 TRNN [33]): If $\mathcal{T}$ is a tensor with TR rank $[r_1, \ldots, r_K]$, then its TRNN is given by

$$\|\mathcal{T}\|_{\text{trnn}} = \sum_{k=1}^{K} a_k \|\mathbf{T}_{(k,s)}\|_*$$

where $a_k \in [0, 1]$ and $\sum_{k=1}^{K} a_k = 1$ corresponds to the weight of mode- $(k, s)$ unfolding.
TRNN is defined by the sum of nuclear norm on circular mode-$(k,s)$ unfolding matrices, which is similar to Tucker nuclear norm (TcNN) [26] and tensor train nuclear norm (TTNN) [29]. TcNN always minimizes the nuclear norm on multiple fat matrices due to one versus the others unfolding (see Definition 3), while the unfolding matrices of TTNN are usually fat or thin matrices when $k$ approaches to one or $K$ (see Definition 4). Instead, by simply setting $s = [K/2]$, TRNN is defined on a series of square matrices. Interestingly, minimizing nuclear norm on these square matrices often tends to carefully adjusting the optimal value as in tensor completion problem [29]. TcNN always minimizes the nuclear norm on these square matrices often tends to one or $K$ tensor basis drawn from uniform distribution, and $\xi$ and $\sigma$ are the random noise and standard deviation, respectively.

Note that (9) is an ill-posed problem without introducing any priors in $\mathcal{F}$. Such dilemma can be handled by the following TR rank minimization problem:

$$\min_{\mathcal{F}} \text{rank}_{tr}(\mathcal{F})$$

s.t. $\frac{1}{2}||y - \hat{X}(\mathcal{F})||^2_2 \leq \epsilon, \ ||\mathcal{F}||_\infty \leq \delta$ (10)

where $\epsilon$ controls the noise level and $\text{rank}_{tr}(-)$ denotes the TR rank. $\hat{X}(\cdot)$ is the tensor uniform sampling operator, given by $\hat{X}(\mathcal{F}) = ((\mathcal{X}_1, \mathcal{F}), (\mathcal{X}_2, \mathcal{F}), \ldots, (\mathcal{X}_N, \mathcal{F}))^T \in \mathbb{R}^N$. Since (10) is an NP-hard problem, according to (7), we relax it by minimizing the following convex TRNN function:

$$\min_{\mathcal{F}} ||\mathcal{F}||_{trnn}$$

s.t. $\frac{1}{2}||y - \hat{X}(\mathcal{F})||^2_2 \leq \epsilon, \ ||\mathcal{F}||_\infty \leq \delta$. (11)

Furthermore, we can equivalently formulate the inequality constraint to the following regularization problem:

$$\min_{\mathcal{F}} \frac{1}{2}||y - \hat{X}(\mathcal{F})||^2_2 + \lambda ||\mathcal{F}||_{trnn}$$

s.t. $||\mathcal{F}||_\infty \leq \delta$ (12)

where $\lambda$ is a positive scalar to achieve a tradeoff between the degree of fitting error of the observed entries and low-TR-rank penalty.

B. Nonasymptotic Upper Bound

Based on the above noisy tensor completion problem, we analyze a nonasymptotic upper bound of the estimation error on the Frobenius norm to reveal the statistical performance of the proposed estimator (12).

**Theorem 1:** Let $\hat{X}(\cdot)$ be a tensor uniform sampling operator, and the true low-TR-rank tensor $\mathcal{F}^*$ and random noise variables $\xi_n$ satisfy Assumptions 1 and 2, respectively. If $\lambda \geq c_0 \sigma (N\log(d_{tr})/d_{tr})^{1/2}$, then we have probability at least $1 - 3/d_{tr}$ such that

$$\frac{||\mathcal{F}^* - \hat{X}(\mathcal{F})||^2_F}{D} \leq c_1 \max\left\{\delta^2 \sqrt{\log d_{tr}}, \frac{K_d \cdot R_0 \log(d_{tr})}{N}, \delta^2 \sqrt{\log d_{tr}} \right\}$$

where $R_0 = (\sum_{k=1}^{F} a_k r_k r_{k+1})^{1/2}$, $c_0$ and $c_1$ denote positive constants, and $\mathcal{F}$ is the flexible solution of problem (12).
We leave the proof of Theorem 1 in Appendix C of Supplementary Material. The result shows that for incomplete low-TR-rank tensor, the per-entry estimation error can be upper bounded by the right-hand side of (13) with high probability. Note that when $K = 2$, TRNN reduces to matrix nuclear norm. Thus, the noisy matrix completion is a special case of the proposed method, and the nonasymptotic upper bound of noisy matrix completion in [21, Th. 7] is a special case of our Theorem 1. In addition, the tensor algebra issues are more complicated than those in noisy matrix cases, making the proofs different from [21]. For example, our proofs require establishing the upper bound of the spectral norm on random tensors (see Lemmas 6 and 7). It is more challenging than matrix completion since some important results in the matrix’s case, e.g., Bernstein inequality, do not have an equivalent form for high-order tensors. The proofs of Theorem 1 also require generalizing several important properties of TRNN, such as the decomposability of TRNN and inequality relationship between Frobenius norm and TRNN on tensors. Furthermore, as discussed in the following paragraph, our result in Theorem 1 is more promising than that of matrix and Tucker rank-based methods.

For simplicity, we assume $r_k = r, d_k = d$, and $a_k = 1/K$, $k \in [K]$, and Theorem 1 states that the estimation error satisfies

$$\frac{\|\mathcal{T}^* - \hat{\mathcal{T}}\|^2_F}{D} \leq \frac{r^2 K d \left[\frac{\epsilon}{d}\right] \log\left(d\left[\frac{\epsilon}{d}\right] + d\left[\frac{\epsilon}{d}\right]\right)}{N} \tag{14}$$

with high probability. Thus, the sample complexity of the proposed estimator is

$$O\left(r^2 K d \left[\frac{\epsilon}{d}\right] \log\left(d\left[\frac{\epsilon}{d}\right] + d\left[\frac{\epsilon}{d}\right]\right)\right). \tag{15}$$

Note that (15) is suboptimal when compared to TR decomposition with degree of freedom $O(r^2 d K)$. This is actually a common issue in sum of nuclear norm convex surrogates for low-rank tensor recovery problems [50]–[53]. Nevertheless, the result in Theorem 1 is still promising. The sample complexity is substantially much lower than the number of tensor entries $d^K$. In addition, even when $r$ approaches $d$, the sample complexity of our estimator is still significantly lower than $d^K$. This result is surprising since it is not aligned with matrix and Tucker rank-based methods that require even full observations as the rank approaches $d$.

C. Minimax Lower Bound

In this section, we present a minimax lower bound of Theorem 1. We let $\inf_{\mathcal{T}} \mathcal{F}$ be the infimum over all the flexible solution of our model and $\sup_{\mathcal{T}} \mathcal{F}$ be the supremum over all the “true” tensor $\mathcal{T}$. We come up with the following theorem.

**Theorem 2:** Suppose that the random variables $z_n$ are i.i.d. Gaussian $\mathcal{N}(0, \sigma^2)$, $n \in [N]$, and $\sigma > 0$. Then, there exist absolute constants $\vartheta \in (0, 1)$ and $c > 0$ such that

$$\inf_{\mathcal{T}} \sup_{\mathcal{T}} \mathbb{P}_{\mathcal{T}} \left(\frac{\|\mathcal{T}^* - \hat{\mathcal{T}}\|^2_F}{D} > c (\delta^2 + d^2) \frac{K d \epsilon R_e}{N}\right) \geq \vartheta \tag{16}$$

where $R_e = (\sum_{k=1}^K a_k (r_k r_{k+1})^{1/2})^2$.

The proof of Theorem 2 can be found in Appendix H of Supplementary Material. Comparing (13) and (16), we can observe that the rate is minimax optimal up to a logarithm factor.

D. Comparisons With Previous Work

Both TR with balanced unfolding (TRBU) [12] and the proposed estimator are able to give statistical or exact recovery performance for low-TR-rank tensor completion. Here, we analyze the superiority of the proposed estimator.

1) TRBU requires structural assumption on core tensors, which extends matrix strong incoherence conditions by letting canonical mode-2 unfolding of core tensors not be aligned with the standard basis, i.e.,

$$G^{(k)}(\mathcal{G}^{(k)})^T \frac{r_k r_{k+1}}{d_k} I_d \leq \mu_k \left(\frac{r_k r_{k+1}}{d_k}\right), k \in [K] \tag{17}$$

where $\mu_k > 0, k \in [K]$. Compared with TRBU, priors on our estimator are significantly mild. The only assumption on unknown tensor is its upper bound on the $\ell_\infty$ norm, which can be easily verified in real-world data. However, it is still unclear how to estimate the existence of strong TR incoherence conditions on incomplete tensor data. Furthermore, even if the incomplete tensor is coherent, the proposed model can still be reliable since it has not involved with the structure assumption on the intrinsic low-TR-rank tensor.

2) TRBU can only obtain exact recovery performance with high probability in noiseless tensor completion case. The proposed model achieves the statistical performance for noisy observations, which is more practical and feasible in real-world applications.

We show the superiority of the proposed estimator compared with tensor completion or recovery methods based on CP rank, Tucker rank, and TT rank, respectively.

1) Yuan and Zhang [11] claimed that recovering incomplete third-order tensor of size $d \times d \times d$ with CP rank $r_{cp}$ required sample size $r_{cp} (d \log d)3/2$. However, the optimal low-CP-rank approximation is NP-complete and is typically intractable.

2) Tomioka et al. [51] claimed that recovering $K$th-order tensor of size $d \times \cdots \times d$ with Tucker rank $[r_0, \ldots, r_K]$ requires $O(r_0 (d^{(K/2)}))$ Gaussian measurements for tensor compressive sensing problem, and Mu et al. [54] further reduced the complexity to $O(r_0 (d^{(K/2)})^3 (d^{(K/2)})^3)$. However, the Tucker rank-based method suffers from the similar flaw as LRMC, that is, as $r_0$ approaches $d$, nearly full observations are required to recover the incomplete tensor.

3) Raubut et al. [55] proposed to solve low-TT-rank tensor compressive sensing using iterative hard thresholding (IHT) algorithm and provided a convergence result based on tensor restricted isometry property (TRIP). The authors claimed that recovering a $K$-order tensor of size $d \times d \times \cdots \times d$ with TT rank $[r_0, \ldots, r_L]$ using sub-Gaussian linear maps satisfied TRIP with level $\delta_r$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
A. NTRC Optimization Algorithm

In problem (12), the variable \(\{T_{(k,t)}\}_{k=1}^K\) shares the same entries \(K\) times in computing TRNN, which makes it difficult to solve the optimization problem directly. Thus, we introduce \(K\) auxiliary variables \(\{M^k\}_{k=1}^K\) to relax such constraint

\[
\min_{\mathcal{M}^k, \mathcal{T}} \frac{1}{2} \|y - \mathcal{X}(\mathcal{T})\|_2^2 + \lambda \sum_{k=1}^K \alpha_k \|M^k_{(k,t)}\|_\infty + \kappa^\infty_{\delta}(\mathcal{T})
\]

s.t. \(\mathcal{M}^k = \mathcal{T}, \ k \in [K]\) \hspace{1cm} (18)

where \(\mathcal{M}^k \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}\) denotes the \(k\)th auxiliary variable. \(\kappa^\infty_{\delta}(\cdot)\) denotes the \(\ell_\infty\) norm indicator function

\[
\kappa^\infty_{\delta}(\mathcal{T}) = \begin{cases} 
0, & \text{if } \mathcal{T} \in \mathbb{S}_\delta \\
\infty, & \text{otherwise}
\end{cases}
\]

where \(\mathbb{S}_\delta := \{\mathcal{T} \in \mathbb{R}^{d_1 \times d_2 \times \cdots \times d_K}, \|\mathcal{T}\|_\infty \leq \delta\}\). In order to solve the equality constraints, we formulate the augmented Lagrangian function of (18) as

\[
\ell^1_{\mu} (\{M^k\}_{k=1}^K, \{Q^k\}_{k=1}^K, \mathcal{T}) = \frac{1}{2} \|y - \mathcal{X}(\mathcal{T})\|_2^2 + \kappa^\infty_{\delta}(\mathcal{T})
\]

\[
+ \sum_{k=1}^K \left(\lambda \alpha_k \|M^k_{(k,t)}\|_\infty + \|Q^k - \mathcal{M}^k - \mathcal{T}\|_F^2\right) \hspace{1cm} (20)
\]

where \(Q^k\) denotes the \(k\)th dual variable and \(\mu\) is a positive penalty scalar. Note that it is rather difficult to simultaneously optimize multiple sets of variables in this objective function. An alternative scheme is to solve each set of variables using the alternating direction method (ADM) [56].

Algorithm 1 NTRC Optimization Algorithm

Input: \(y, \{X_n\}_{n=1}^N\)

Initialization: \((\alpha_k)_{k=1}^K, \mu = 10^{-4}, v = 1.1, \mu_{max} = 10^{10}\), \(tol = 10^{-6}\), zero filled with \(M^k, Q^k, k \in [K], t=1\).

1: while not converged do
2: for \(k = 1,\ldots, K\) do
3: Update \(M^{k,t+1}\) using (21).
4: end for
5: Update \(\mathcal{T}^{t+1}\) using (25).
6: for \(k = 1,\ldots, K\) do
7: Update \(Q^{k,t+1}\) using (26).
8: end for
9: \(\mu_{t+1} = \min(\mu_{max}, v \mu_t)\).
10: Check the convergence condition: \(\frac{\|\mathcal{T}^{t+1} - \mathcal{T}^t\|_F}{\|\mathcal{T}\|_F} \leq tol\).
11: \(t = t + 1\).
12: end while

1) \(
\{M^k\}_{k=1}^K\) Subproblems: For \(\{M^k_{(k,t)}\}_{k=1}^K\) subproblems, they can be solved by updating \(M^k_{(k,t)}\) while fixing the others

\[
M^{k,t+1}_{(k,t)} = \arg \min_{M_{(k,t)}} \mu \cdot \|M^k_{(k,t)}\|_\infty + \frac{1}{2} \left\|M^k_{(k,t)} - T^k_{(k,t)} + \frac{1}{\mu} Q^{k,t}_{(k,t)}\right\|_F^2
\]

\[
= \mathcal{P}_{\tau}^\infty (T^k_{(k,t)} - \frac{1}{\mu} Q^{k,t}_{(k,t)})
\]

where \(\tau = \frac{\lambda \alpha_k}{\mu}\) and \(\mathcal{P}_{\tau}^\infty(A)\) denotes the singular value thresholding (SVT) [57] method

\[
\mathcal{P}_{\tau}^\infty(A) = \text{Umax}(\Sigma - \tau, 0)\Sigma^T
\]

2) \(\mathcal{T}\)-Subproblem: The \(\mathcal{T}\)-subproblem can be equivalently formulated as the following vectorization form:

\[
\text{vec}(\mathcal{T}^{t+1}) = \arg \min_{\mathcal{T}} \kappa^\infty_{\delta}(\text{vec}(\mathcal{T}))
\]

\[
+ \left\|\text{vec}(\mathcal{T}) - (X^T X + \mu^t K I_d)^{-1} \left\|X^T Y\right\|_F \right\|_F^2
\]

\[
= \left(\mu^t \sum_{k=1}^K \text{vec}(Q^{k,t} + M^{k,t}) - XY\right)
\]

which can be solved by deriving the Karush–Kuhn–Tucker (KKT) condition

\[
\text{vec}(\mathcal{T}^{t+1}) = \mathcal{P}_{\tau}^\infty (f^{t+1}) = \text{sgn}(f^{t+1}) \min_{f^{t+1}} (\|f^{t+1}\|_1, \delta)
\]

where \(f^{t+1} = (X^T X + \mu^t K I_d)^{-1} (\mu^t \sum_{k=1}^K \text{vec}(Q^{k,t} + M^{k,t}) - XY)\) and \(X(:, m) = \text{vec}(X_m)\).

3) \(
\{Q^k\}_{k=1}^K\) Subproblems: The dual variables \(\{Q^k\}_{k=1}^K\) can be updated by the gradient ascent method

\[
Q^{k,t+1} = Q^{k,t} + \mu^t (M^{k,t+1} - \mathcal{T}^{t+1}) \hspace{1cm} (26)
\]

The details of the optimization procedure are summarized in Algorithm 1.
B. FaNTRC Optimization Algorithm

The main computational cost of NTRC is singular value decomposition (SVD) for nuclear norm minimization on unfolding matrices, which may prevent its applications in large-scale tensor completion problems. In order to alleviate such bottleneck, we introduce a fast NTRC (FaNTRC) algorithm, i.e., to minimize TRNN equivalently on a much smaller tensor. To establish the relationship between NTRC and FaNTRC, we further show that its global optimal solution is exactly the same as that of NTRC.

We first explore the relationship between TR and Tucker decomposition.

Lemma 3: Given an arbitrary tensor \( \mathcal{T} \in \mathbb{R}^{d_1 \times \ldots \times d_k} \), if the TR decomposition of \( \mathcal{T} \) is \( \mathcal{T} = \mathcal{T} \times_1 U_1, \ldots, \times_K U_K \), then \( \mathcal{T} \) can be equivalently represented by the Tucker decomposition format \( \mathcal{T} = \mathcal{F} \times_1 U_1, \ldots, \times_K U_K \), where \( U_k \in \mathbb{R}^{d_k \times R_k}, \ k \in [K] \), and \( R_k = (r_k r_{k+1} + d_k) \), denote the column orthogonal matrices.

We leave the proof of Lemma 3 in Appendix I of Supplementary Material. Lemma 3 depicts that given a tensor with TR rank \( [r_1, \ldots, r_K] \), it can also be represented by the Tucker decomposition format, which motivates us to come up with the following theorem.

Theorem 3: Given a tensor \( \mathcal{T} \in \mathbb{R}^{d_1 \times \ldots \times d_k} \) with TR rank \( [r_1, \ldots, r_K] \), \( U_k \in \text{St}(d_k, R_k) \), \( (r_k r_{k+1} + d_k) \leq R_k \leq d_k \), \( k \in [K] \), and \( \mathcal{F} \) satisfy \( \mathcal{F} = \mathcal{F} \times_1 U_1 \times_2 U_2, \ldots, \times_K U_K \), then we have

\[
\| \mathcal{T} \|_{\text{TTRNN}} = \| \mathcal{F} \|_{\text{TTRNN}} \tag{27}
\]

where \( \text{St}(d_k, R_k) := \{ U_k, \ U_k \in \mathbb{R}^{d_k \times R_k}, \ U_k^T U_k = I_{R_k} \} \) denotes the Stiefel manifold.

The proof of Theorem 3 can be found in Appendix J of Supplementary Material.

Equipped with Theorem 3, we can solve the high computational TRNN minimization problem in a more efficient manner than the existing methods, i.e., computing SVD on a much smaller circular unfolding matrix \( \mathcal{T} \times_1 U_1, \ldots, \times_K U_K \), where \( R_{1:k} = \prod_{j=1}^{k-1} r_j \) and \( R_{k:k} = \prod_{j=k}^{K} r_j \). Though this similar strategy has been used in tensor nuclear norm minimization problem [27], [58], [59], our method is very different from that of these works since our method is implemented in a heterogeneous framework in the sense that equivalently minimizing TRNN in a Tucker decomposition format. For example, we have to explore the connection between TR and Tucker decomposition to meet our analysis. Moreover, it is also crucial for us to establish the relationship of circular unfolding between the original tensor and its Tucker decomposition format. Therefore, these dilemmas make our analysis more complicated than those homogeneous frameworks in previous research works.

Consequently, according to Theorem 3, problem (12) can be equivalently reformulated as

\[
\min_{\mathcal{T}, \mathcal{F}, \{U_k\}_{k=1}^K} \frac{1}{2} \| y - \mathcal{X} (\mathcal{T}) \|^2 + \lambda \| \mathcal{F} \|_{\text{TTRNN}} + \kappa \| \mathcal{F} \|_{\text{F}} \tag{28}
\]

s.t. \( \mathcal{F} = \mathcal{F} \times_1 U_1, \ldots, \times_K U_K \)
\( U_k \in \text{St}(d_k, R_k), \ k \in [K]. \)

Now, we show that with appropriate choice of \( R_k, k \in [K] \), the global optimal solution of (28) is the same as that of (12).

Theorem 4: Suppose that \( \{ (U_k)_{k=1}^K, \mathcal{F} \} \) and \( \{ \mathcal{F}, \{ (U_k)_{k=1}^K \} \} \) are the global optimal solutions of problem (28) and problem (12) with \( (r_k r_{k+1} + d_k) \leq R_k \leq d_k, k \in [K] \), respectively, and then \( \mathcal{F} \times_1 U_1, \ldots, \times_K U_K \) is also the optimal solution of problem (12).

The proof of Theorem 4 can be found in Appendix K of Supplementary Material.

To tackle the constrained optimization problem (28), we formulate its augmented Lagrangian function as

\[
\ell^2_\eta (\mathcal{F}, \{ (U_k)_{k=1}^K \}) = \frac{1}{2} \| y - \mathcal{X} (\mathcal{T}) \|^2 + \| \mathcal{F} - \mathcal{T} \times_1 U_1, \ldots, \times_K U_K \|_F^2 + \sum_{k=1}^K \alpha_k \| L_{(k),s} \|_F^2 + \frac{\eta}{2} \| \mathcal{F} - \tilde{\mathcal{F}} \|_F^2 \tag{29}
\]

s.t. \( U_k \in \text{St}(d_k, R_k), \ k \in [K] \)

where \( \eta > 0 \) is penalty parameter. Based on the augmented Lagrangian function (29), we solve each subproblem alternatively by fixing the others.

1) \( \{ U_k \}_{k=1}^K \)-subproblems: According to (29), the \( U_k \)-subproblem can be formulated as the following maximization problem over factor matrix \( U_k \):

\[
\max_{U_k} \frac{1}{2} \| \mathcal{F} \times_1 U_k - \frac{1}{\eta} \| P + \mathcal{F} \|_F \times_1 U_k^T \)
\]

s.t. \( U_k \in \text{St}(d_k, R_k) \tag{30} \)

where \( \mathcal{F} ^\times \times K U_k \) denotes \( \mathcal{F} \times_1 U_1 \times_2 U_2, \ldots, \times_K U_K \), and it can be further reformulated as

\[
U_k \leftarrow U_k \left( \frac{1}{\eta} P_k + \mathcal{F} \right)^T B_k \left( \frac{1}{\eta} P_k + \mathcal{F} \right) \tag{31}
\]

s.t. \( U_k \in \text{St}(d_k, R_k) \)

where \( B_k \) denotes the canonical mode-\( k \) unfolding of \( \mathcal{F} \times_1 U_k \). Note that (31) is actually a well-known orthogonal procrustes problem [60], whose global optimal solution is given by SVD, i.e., \( U_k \leftarrow U_k \times_1 \tilde{U}_k \times_2 \tilde{V}_k \), where \( \tilde{U}_k \) and \( \tilde{V}_k \) denote the left and right singular vectors of \( (1/\eta) P_k + \mathcal{F} \), respectively. By alternatively solving the maximization problem (31), we can obtain the optimal solution of \( \{ U_k \}_{k=1}^K \).

2) \( \mathcal{F} \)-subproblem: By fixing all the other variables, the \( \mathcal{F} \)-subproblem is given by

\[
\min_{\mathcal{F}} \left\| \left( \frac{1}{\eta} P + \mathcal{F} \right) \times_1 U_1, \ldots, \times_K U_K \right\|_F^2 + \sum_{k=1}^K \| \mathcal{F} \times_1 U_k - \mathcal{F} \|_F^2 \tag{32}
\]
whose closed-form solution can be obtained directly

\[
\mathcal{T}^{t+1} = \frac{1}{K+1} \left( \frac{1}{\eta^t} \mathcal{P}^t + \mathcal{F} \right) \times_1 U_{1:t+1}^{1,\top}, \ldots, \times_K U_{K:t+1}^{1,\top} \\
+ \frac{1}{K+1} \sum_{k=1}^K \frac{1}{\eta^t} \mathcal{R}_k^{t+1} + \mathcal{L}^{t+1}. \tag{33}
\]

3) \{L_i^{K_{k=1}}\} Subproblems: \( \mathcal{L}_i^k \) can be solved by minimizing the following subproblem:

\[
\mathcal{L}^{k,t+1}(k,i) = \arg \min \frac{\lambda^k d_k}{\eta^t} \| \mathcal{L}^{k,t+1} \|_1 \\
+ \frac{1}{2} \| \mathcal{L}^{k,t+1} - \mathcal{T}^{t+1} \|_F^2. \tag{34}
\]

Similar to (s), it can be easy to obtain its closed-form solution by the SVT method.

4) \( \mathcal{F} \)-Subproblem: We update the variable \( \mathcal{F} \) by fixing the others

\[
\text{vec}(\mathcal{F}^{t+1}) = \arg \min_{\mathcal{F}} \frac{\lambda^0}{\eta^t} \| \text{vec}(\mathcal{F}^{t+1}) - (X^TX + \eta^t I_D)^{-1} \\
\times (Xy - \text{vec}((\mathcal{P}^t - \eta^t \mathcal{F}^{t+1} \times_1 U_{1:t+1}^{1,\top}, \ldots, \times_K U_{K:t+1}^{1,\top})) \|_2^2. \tag{35}
\]

whose closed-form solution can be obtained as (23).

5) \{P_i^{K_{k=1}}\} (K_k=1)-Subproblems: The dual variables \( \mathcal{P} \) and \( \mathcal{R}_k^{K_{k=1}} \) can be updated by the gradient ascent method

\[
P^{t+1} = \mathcal{P}^t + \eta^t \left( \mathcal{F}^{t+1} - \mathcal{T}^{t+1} \times_1 U_{1:t+1}^{1,\top}, \ldots, \times_K U_{K:t+1}^{1,\top} \right) \tag{36}
\]

\[
\mathcal{R}_k^{t+1} = \mathcal{R}_k^{t} + \eta^t \left( \mathcal{L}_k^{t+1} - \mathcal{T}^{t+1} \right), \quad k \in [K]. \tag{37}
\]

The details of the optimization procedure are summarized in Algorithm 2.

\begin{algorithm}
\caption{FaNTRC Optimization Algorithm}
\begin{algorithmic}
\State Input: \( y, [x_i]_{i=1}^N \).
\State Initialization: \( [a_i]_{i=1}^N, \eta = 10^{-4}, \nu = 1.1, \eta_{\text{max}} = 10^{10}, \)
\State \( \text{tol} = 10^{-6}, \) zero filled with \( \mathcal{M}^{K} \), \( \mathcal{Q}^{k}, k \in [K], t = 1. \)
\State while not converged do
\For {k = 1, \ldots, K}
\State Update \( U_{k:t+1} \) via solving (31).
\EndFor
\State Update \( \mathcal{T}^{t+1} \) using (33).
\State for \( k = 1, \ldots, K \) do
\State Update \( \mathcal{L}^{k,t+1} \) using (34).
\EndFor
\State Update \( \mathcal{T}^{t+1} \) using (35).
\State Update \( \mathcal{P}^{t+1} \) and \( \mathcal{R}_k^{t+1}, \) \( k \in [K] \), using (36).
\State \( \eta = \min(\eta_{\text{max}}, \nu \eta) \).
\State Check the convergence condition:
\State \( \frac{\| \mathcal{T}^{t+1} - \mathcal{T}^{t} \|_F}{\| \mathcal{T}^{t} \|_F} \leq \text{tol} \).
\State \( t = t + 1 \).
\EndWhile
\end{algorithmic}
\end{algorithm}

\subsection{C. Computational Complexity Analysis}

For simplicity, we assume that the observed tensor is of size \( d_1 = \cdots = d_K = d \), the given rank satisfies \( R_1 = \cdots = R_K = R \), and \( s = [K/2] \). For Algorithm 1, the main per-iteration cost lies in (21), which requires computing SVD of \( d^{2K} \times d^{2K/2} \) matrices. Therefore, the per-iteration complexity is \( \mathcal{O}(Kd^{3K/2}) \). For Algorithm 2, the main per-iteration cost lies in computing multiplication operation in (30) and (35) and calculating SVD in (31) and (34). The multiplication operations in (30) and (35) involve complexity \( \mathcal{O}(K Rd^k) \) and \( \mathcal{O}(\sum_{k=1}^K R^{k-1} d^k) \), respectively. The computational costs of SVD in (31) and (34) are \( \mathcal{O}(K R^2d^{K-1}) \) and \( \mathcal{O}(K R^{3K/2}) \), respectively. Hence, the total computational cost in one iteration of Algorithm 2 is given by

\[
\mathcal{O} \left( K Rd^K + K R^2d^{K-1} + \sum_{k=1}^K R^{K-1} d^k + K R^{3K/2} \right) \tag{38}
\]

Note that when \( R \ll d \), the computational cost of Algorithm 2 is significantly lower than that of Algorithm 1 and the other TRNN-based methods.

\subsection{D. Convergence Analysis}

The proposed NTRC adopts the classical two-block convex ADMM optimization framework, and the convergence analysis or discussion can be found in previous works [61], [62]. FaNTRC is a nonconvex ADMM algorithm for which the theoretical convergence guarantee is difficult to derive. Instead, we provide weak convergence results for FaNTRC under mild condition.

\textbf{Theorem 5:} Let \( (\mathcal{T}_i^t, \mathcal{F}^t, \{U_{k}^{t} \}_{k=1}^{K}, \{\mathcal{L}_{k}^{t} \}_{k=1}^{K}, \mathcal{P}^t, \{\mathcal{R}_{k}^{t} \}_{k=1}^{K}) \) be the sequences generated by Algorithm 2 (FaNTRC). Suppose that the sequence \( \{\mathcal{P}^t\} \) is bounded, and then, we have the following conclusions.

1) \( (\mathcal{T}_i^t, \mathcal{F}^t, \{U_{k}^{t} \}_{k=1}^{K}, \{\mathcal{L}_{k}^{t} \}_{k=1}^{K}) \) are Cauchy sequences.
2) Any accumulation point of sequences \( (\mathcal{T}_i^t, \mathcal{F}^t, \{U_{k}^{t} \}_{k=1}^{K}, \{\mathcal{L}_{k}^{t} \}_{k=1}^{K}) \) satisfies the KKT conditions for problem (28).

The proof sketches of Theorem 5 can be found in Appendix L of Supplementary Material. In addition, we also experimentally investigate the relative error (RE) and peak signal-to-noise ratio (PSNR) values versus the iteration number for the proposed NTRC and FaNTRC in Fig. 12. We can clearly observe that the proposed methods usually converge after only approximately 50 iterations, indicating the effective convergence behavior of the proposed algorithms.

\section{V. EXPERIMENTAL RESULTS}

In this section, we conduct four experiments to evaluate noisy tensor completion performance of the proposed NTRC and FaNTRC on synthetic and real-world datasets, including color images, light field images, and video sequences. All experiments are run in MATLAB 9.4 on a Linux personal computer equipped with dual Intel E5 2640v4 and 128GB of RAM. We then present numerical results to compare with several state-of-the-art tensor completion methods.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Then, we construct a synthetic tensor by using tensor circular four TR cores \(G\). For the parameter \(\lambda\), we normalize \(N\) by \(d^r\) for the noisy condition, we produce an additive noise tensor from corrupted noise and is given by

\[
\hat{\mathcal{F}} = \mathcal{F} + \sigma(\epsilon) \cdot \mathcal{N}(0, \mathcal{N})
\]

where \(\sigma(\epsilon)\) is the maximum possible power of a signal and the power of corrupted noise is given by

\[
\text{PSNR} = 10 \log_{10} \left( \frac{\sqrt{D} \Vert \hat{\mathcal{F}} \Vert_F^2}{\Vert \hat{\mathcal{F}} - \mathcal{F} \Vert_F} \right).
\]

A. Application to Synthetic Tensor Completion

We conduct two experiments on synthetic data to evaluate the recovery performance of the proposed NTRC and FaNTRC for recovering incomplete and noisy low-rank tensors. We generate a fourth-order tensor \(\mathcal{F} \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}\) with TR rank \([r_1, r_2, r_3, r_4]\) as follows. First, we produce four TR cores \(\mathcal{G}^{(k)} \in \mathbb{R}^{d_1 \times d_2 \times d_3 \times d_4}, k \in [4]\), using MATLAB command \(\mathcal{G}^{(k)} = \text{rand}(d_1, d_2, d_3, d_4)\). For simplicity, we let \(d_k = d\) and \(r_k = r\), \(k \in [4]\), in all experiments next. Then, we construct a synthetic tensor by using tensor circular product \(\mathcal{T}^* = \mathcal{T}^{(1)} \odot \mathcal{T}^{(2)} \odot \mathcal{T}^{(3)} \odot \mathcal{T}^{(4)}\). In order to meet noisy condition, we produce an additive noise tensor from the \(\mathcal{N}(0, \sigma^2)\) distribution, where we set the standard deviation \(\sigma = c \Vert \mathcal{T}^* \Vert_F / \sqrt{D}\) to keep a constant signal-to-noise ratio (SNR). Finally, we sample \(N = pd^4\) entries uniformly to form the partially observed noisy tensor, where \(p\) denotes the SR.

1) Sharpness of the Proposed Upper Bound: In this experiment, we set the tensor size \(d_k = d, d \in \{10, 20, 30\}\), \(r_k \in [1, 2, 3, 4, 5, 6]\), \(k \in [4]\) following [39]. For the parameter \(\lambda\), we let \(\lambda_0 = \sigma (N \log(d_1) / d_1)^{1/2}\) and chose the optimal one by scaling it \(\lambda = a \lambda_0\), where \(a\) is experimentally selected in the candidate set \(\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}\). To fix the SNR regardless of tensor size [47], we normalize \(\mathcal{T}^*\) to have unit Frobenius norm and set the noise level \(c = 0.01\). We repeat 25 trials using the proposed NTRC algorithm and compute the mean of estimation error \(\Vert \hat{\mathcal{T}}^* - \mathcal{T}^* \Vert_F^2\) over all the trials. Fig. 2(a) shows the estimator error obtained by the NTRC algorithm versus the observation numbers. The curves of Fig. 2(a) decrease as the observation number \(N\) increases. In addition, synthetic tensor with larger size required a larger number of observations.

According to the constant SNR settings, we substitute \(\sigma = 0.01 \Vert \mathcal{T}^* \Vert_F / \sqrt{D}\) to the right-hand side of (13) and obtain

\[
\Vert \hat{\mathcal{T}} - \mathcal{T}^* \Vert_F^2 \leq c (10^{-4} \sqrt{D})^2 \sigma^2
\]

holds with high probability. Following [47], the proposed upper bound is sharp if the estimation error should scale like:

\[
\Vert \hat{\mathcal{T}} - \mathcal{T}^* \Vert_F^2 \leq \mathcal{O} \left( \frac{r^2 K d^{1/2} \log(d^{1/2} + d^{1/2})}{N} \right)
\]

By setting the rescaled observation number \(N_0\) defined by

\[
N_0 := \frac{N}{r^2 K d^{1/2} \log(d^{1/2} + d^{1/2})}
\]

the curves should be relatively aligned regardless of tensor sizes. We replot the estimation error versus the rescaled observation number in Fig. 2(b) and observe that all the three curves are well-aligned, which demonstrates the sharpness of the proposed upper bound.

2) Effects of Varying TR Ranks: In this experiment, we investigate the influence of TR rank on the estimation error as well as the noise distribution on RE. The experiments are conducted on synthetic tensor with size \(d_k = d, d \in \{10, 20, 30\}\), \(k \in [4]\), and TR rank \(r_k = r, r = [1, 2, 3, 4, 5, 6]\), \(k \in [4]\). We set noise level \(c = 0.01\) and SR = 40%. We repeat 25 trials using the proposed NTRC algorithm and compute the mean of estimation error over all the trials. We plot the estimation error versus the square of TR rank in Fig. 3. It can be seen that the estimation error scales linearly against the square of TR rank, which agrees with the result in (41).

3) Effects of Different Subexponential Noises: In this experiment, we investigate the effectiveness of the proposed estimator for dealing with different subexponential noises. We generate the synthetic tensor of size \(d_k = d, d \in \{10, 20\}\), \(k \in [4]\), and TR rank \(r_k = [\log^{1/2}(d_k)]\), \(k \in [4]\). We generate three types of noises include Gaussian noise, sub-Gaussian noise, and Poisson noise. For Gaussian distribution noise, we produce the noise tensor from \(\mathcal{N}(0, \sigma^2)\), where \(\sigma = 0.01 \Vert \mathcal{T}^* \Vert_F / \sqrt{D}\). For sub-Gaussian distribution noise, we

1https://tu-yangyang.github.io/TMac/
2https://www.cs.rochester.edu/u/jliu/publications.html
3https://sites.google.com/site/jbengua/home
4https://github.com/yuanlonghao/TLRF
5The code was provided by Dr. Jinshi Liu.
6The code was provided by Dr. Andong Wang.
NTRC achieves the lowest RE in all the cases, especially approximate the incomplete tensor. However, when different noise levels versus varying TR rank the mean of RE. Fig. 5 shows the RE of different methods on and the SR is set to 40%. We repeat 25 trials and compute the noise is generated with noise levels c = 0.01. (b) RE versus TR rank on synthetic tensor of size $\mathbb{R}^{30 \times 30 \times 30}$ with noise level $c = 0.05$.

generate the noise using the uniform distribution $U([-0.5, 0.5])$. For Poisson distribution noise, we generate the noise tensor from $\text{Pois}(0.01)$. The incomplete tensors are generated by uniformly selecting entries at random with SR from 10% to 90%. We repeat 25 trials using the NTRC algorithm and compute the mean of RE. Fig. 4 shows the RE versus varying SR. We can see that the RE on different noise distributions on two distinct tensor sizes is relatively low and decreases as the SR increases, which indicates the effectiveness of the proposed estimator for recovering incomplete tensors with different subexponential noises.

4) Effects of Multiple TR States: To verify the effectiveness of the proposed methods in multiple TR states, we investigate the RE of different methods versus varying TR ranks. We consider the synthetic tensor of size $\mathbb{R}^{30 \times 30 \times 30}$ and TR rank $r_k = r, r \in \{2, 3, 4, 5, 6, 7\}, k \in \{4\}$. The zero-mean Gaussian noise is generated with noise levels $c = 0.01$ and $c = 0.05$, and the SR is set to 40%. We repeat 25 trials and compute the mean of RE. Fig. 5 shows the RE of different methods on different noise levels versus varying TR rank $r$. The proposed NTRC achieves the lowest RE in all the cases, especially when the noise level is large. When $r \leq 5$, HaLRTC can well approximate the incomplete tensor. However, when $r \geq 6$, the synthetic tensor is supercritical ($r^2 > d$), i.e., full-Tucker-rank, the RE of HaLRTC grows dramatically as $r$ increases, which verifies the results revealed in Lemma 1.

5) Effects of the Given Rank for FaNTRC: We investigate the effects of the given rank for FaNTRC. We consider the synthetic tensor of size $d_k = d, d \in \{30, 50\}, k \in \{4\}$, and TR rank $r_k = r, r \in \{3, 4, 5, 6\}, k \in \{4\}$. We produce the noise tensor using the zero-mean Gaussian distribution with noise level $c = 0.01$ as in the above experiment and sample the incomplete tensor using uniform distribution with SR = 40%. Fig. 6 presents the RE of FaNTRC on different low-rank synthetic tensors versus the given rank $R_k$, where we set $R_k = R, k \in \{4\}$, and $R = \text{round}(c d^2), c \in \{0.7, 0.8, \ldots, 1.5\}$. We can observe that when $R < r^2$, the RE value is relatively large, and as long as $R \geq r^2$, the RE of FaNTRC tends to be stable with small RE value, which verifies the correctness of Theorem 3.

B. Application to Color Image Inpainting

In this section, we evaluate the proposed NTRC and FaNTRC against state-of-the-art LRTC methods on 20 benchmark images that are shown in Fig. 7. Each image can be

---

1http://decsai.ugr.es/cvg/dbimagenes/c512.php
treated as a third-order tensor of $512 \times 512 \times 3$ entries. For each image, 40% of pixels are sampled uniformly at random together with additive Gaussian noise with standard deviation $\sigma = 0.25 \times \|T\|_F/\sqrt{D}$. For TMac-inc, we set $a_k = 1/3$, $r_k^{\text{max}} = 60$, and maximum iteration to 500. For HaLRTC, SiLRTC-TT, and TRNNM, we employ their default parameter settings, which leads to good performance. For TRLRF, we set the rank to $[5, 5, 5]$ as suggested in [35]. Note that for SiLRTC-TT, TRNNM, and the proposed methods, we convert the images to fifth-order tensors of size $16 \times 16 \times 32 \times 32 \times 3$ by using visual data tensorization (VDT) technique [29], [63]. This simple and efficient trick has been shown to boost the tensor completion performance for TT-rank- and TR-rank-based methods [29], [63]. For the proposed NTRC and FaNTRC, we simply set the weights $a_k = 1/K, k \in [K]$, and the given rank of FaNTRC is set equal to $[10, 10, 18, 18, 3]$ empirically. For NoisyTNN, NTRC, and FaNTRC, the penalty parameter $\lambda$ is selected as in Section V-A.1. For TMac-inc, we simply set $R = 1/3$, $a_k = 1/4$ and $r_n^{\text{max}} = 40$ empirically. For HaLRTC, $\lambda$ is set to equal $[12, 20, 10, 15]$ since it achieves the final performance in most cases. For TRLRF, we let $r_1 = r_2 = r$ and $r_3 = r_4 = 5$ and select the optimal $r$ in the set $[11, 12, \ldots, 17]$ to achieve the best performance. For FaNTRC, we simply let $R_1 = R_2 = 110$ and $R_3 = R_4 = 5$ in all cases of these experiments. To use NoisyTNN, we concatenate the angular dimensions by reshaping the light field images into the third-order tensors of size $192 \times 192 \times 81$. The remaining parameters are set to the same as Section V-A.1.

Fig. 9 shows the results of applying the compared methods to the noisy and incomplete benchmark images. From Fig. 9(a) and (b), it can be observed that the proposed NTRC and FaNTRC perform the first and the second best in terms of RE and PSNR metrics, verifying their superiority in completing the real-world colorful images. Compared with NoisyTNN, our methods can estimate more complex low-rank structure residing along each order of images. Thus, our methods achieve more promising recovery results. Though both SiLRTC-TT and TRNNM utilize the fifth-order VDT images, TRNNM performs a little better than SiLRTC-TT. TMac-inc is a Tucker rank-based noisy tensor completion method and achieves better completion performance than HaLRTC in most cases. We also report the running time of each image in Fig. 9(c). The proposed FaNTRC is the fastest method, at least three times faster than the existing counterparts, which shows its scalability on such high spatial resolution images. We also observed that our methods are much faster than TRLRF since it requires time-consuming TR reconstruction operation at each iteration. Conversely, our methods merely approximate the low-TR-rank tensor on its unfolding matrices and avoid frequently estimating TR core tensors explicitly.

C. Application to Light Field Image Inpainting

Different from the conventional image that only records light intensity, light field image contains both light intensity and light direction in space, giving a high-dimensional representation of visual data for various computer vision applications. The tested four light-field images are shown in Fig. 8. In this experiment, we choose four light field images from the HCI light field image dataset.8 Each light field image can be represented as a fifth-order tensor of size $768 \times 768 \times 3 \times 9 \times 9$, where $768 \times 768$ and $9 \times 9$ denote spatial and angular resolutions, respectively. For the sake of computational efficiency, we downsample the spatial dimension to $192 \times 192$ and convert each image to the grayscale. Thus, we can form a fourth-order tensor of size $192 \times 192 \times 9 \times 9$. The input incomplete tensors are generated by uniformly selecting pixels at random with SRs from 10% to 70% with increment 5%, and the additive noises are produced by the i.i.d. Gaussian distribution with standard deviation $\sigma = 0.05 \times \|T\|_F/\sqrt{D}$. For TMac-inc, we let $a_k = 1/4$ and $r_n^{\text{max}} = 40$ empirically. For HaLRTC, $\lambda$ is set to equal $[2, 2, 10, 10]$ since it achieves the final performance in most cases. For TRLRF and FaNTRC, we simply let $R_1 = R_2 = 110$ and $R_3 = R_4 = 5$ in all cases of these experiments. To use NoisyTNN, we reshape the remaining parameters are set to the same as Section V-A.1.

D. Application to Video Inpainting

Finally, we evaluate our methods on publicly available YUV color video sequences,9 which have been widely used for tensor completion applications. In this section, we test our methods and the other compared methods on four of these sequences, namely, Akiyo, Carphone, Hall, and Suzie. For each video sequence, we adopt its first 50 frames, and each frame contains $144 \times 176 \times 3$ pixels. Therefore, the evaluated video sequences can be treated as the fourth-order tensors of size $144 \times 176 \times 3 \times 50$. We randomly select 40% of pixels and add i.i.d. Gaussian noise with standard deviation $\sigma = 0.1 \times \|T\|_F/\sqrt{D}$. For TMac-inc, we simply set $a_k = 1/4$ and $r_n^{\text{max}} = 40$. For HaLRTC, we empirically let $\lambda = [1, 1, 10^{-3}, 10^2]$. For TRLRF and FaNTRC, we set the rank to $[20, 20, 20, 20]$ and $[40, 40, 3, 20]$, respectively. To use NoisyTNN, we reshape the video sequences to $144 \times 176 \times 150$ tensors. All the other parameters are consistent with the above section.

8http://lightfieldgroup.iwr.uni-heidelberg.de

9http://trace.eas.asu.edu/yuv/
We present the inpainting result of the 21st frame in Fig. 11 of four video sequences. It can be observed that both NTRC and FaNTRC obtain more favorable performance in terms of RE and PSNR metrics than the other compared methods. More specifically, the proposed method can reconstruct more details of video frames. From the recovered frames of Hall and Suzie,
it can be observed that the proposed methods can present more clean background than the other LRTC methods. Moreover, for Akiyo, Carphone, and Suzie, the additive noise on the clothes and face can be efficiently removed by the proposed methods. Analogue to the experimental results in Sections V-B and V-C, TR-based methods consistently outperform the Tucker and TT methods. FaNTRC and TRNNM are the first and second fastest methods, respectively. Note that TRLRF is still the slowest method, where the reason has been analyzed above. We also present the RE and PSNR history over iterations of our methods on the Akiyo video in Fig. 12. It can be observed that both NTRC and FaNTRC in terms of two metrics tend to be stable within merely 50 iterations.

VI. CONCLUSION

We proposed a novel noisy tensor completion model to predict the missing entries with noisy observations based on the low-rank TR model. In order to illustrate the statistical performance of the proposed model, we theoretically analyze a nonasymptotic upper bound of the estimation error. Moreover, we propose two algorithms to solve the optimization problem with convergence guarantee, namely, NTRC and its fast version FaNTRC. The FaNTRC can significantly reduce the high computational cost of SVD on the large-scale circular unfolding matrix by equivalently minimizing TRNN on a much smaller one in a heterogeneous tensor decomposition framework and thus accelerates the algorithm. Extensive experiments on both synthetic and visual data evidence the merits of low-TR-rank model for modeling a tensor, the ability of noisy completion setting in recovering incomplete data with noise corruption, and the computational efficiency of heterogeneous tensor decomposition strategy for TRNN minimization.

ACKNOWLEDGMENT

The authors would like to thank Dr. Andong Wang for his helpful discussions.

REFERENCES

[1] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” SIAM Rev., vol. 51, pp. 455–500, Aug. 2009.
[2] W. Hu, Y. Yang, W. Zhang, and Y. Xie, “Moving object detection using tensor-based low-rank and saliently fused-sparse decomposition,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 724–737, Feb. 2016.

[3] Y. Xie, D. Tao, W. Zhang, L. Zhang, and Y. Qu, “On unifying multi-view self-representations for clustering by tensor multi-rank minimization,” *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.

[4] C. Zhang et al., “Generalized latent multi-view subspace clustering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2018.

[5] Y. Qiu, G. Zhou, X. Chen, D. Zhang, X. Zhao, and Q. Zhao, “Semi-supervised non-negative Tucker decomposition for tensor data representation,” *Sci. China Technol. Sci.*, vol. 64, no. 9, pp. 1881–1892, Sep. 2021.

[6] A. Cichocki et al., “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[7] G. Zhou et al., “Linked component analysis from matrices to high-order tensors: Applications to biomedical data,” *Proc. IEEE*, vol. 104, no. 2, pp. 310–331, Feb. 2016.

[8] Y. Qiu, G. Zhou, Q. Zhao, and A. Cichocki, “Comparative study on the classification methods for breast cancer diagnosis,” *Bull. Polish Acad. Sci., Tech. Sci.*, vol. 66, no. 6, pp. 841–848, 2018.

[9] Y. Qu, X. Zhang, Y. Zhang, and S. Xie, “A generalized graph regularized non-negative tensor decomposition framework for tensor data representation,” *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 594–607, Jan. 2022.

[10] C. Mu, B. Huang, J. Wright, and D. Goldfarb, “Square deal: Lower bounds and improved relaxations for tensor recovery,” in *Proc. 31st Int. Conf. Mach. Learn. (ICML)*, vol. 2, 2014, pp. 1242–1250.

[11] M. Yuan and C.-H. Zhang, “On tensor completion via nuclear norm minimization,” *Found. Comput. Math.*, vol. 16, no. 4, pp. 1031–1068, 2016.

[12] H. Huang, Y. Liu, J. Liu, and C. Zhu, “Provable tensor ring completion,” *Signal Process.*, vol. 171, Jun. 2020, Art. no. 107486.

[13] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, “The twist tensor nuclear norm for video completion,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 2961–2973, Dec. 2021.

[14] Q. Zhao, L. Zhang, and A. Cichocki, “Bayesian CP factorization of incomplete tensors with automatic rank determination,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1751–1763, Sep. 2015.

[15] B. Ermis, E. Acar, and A. T. Cemgil, “Link prediction in heterogeneous data via generalized coupled tensor factorization,” *Data Mining Knowl. Discovery*, vol. 29, no. 1, pp. 203–236, 2015.

[16] M. Nickel, V. Tresp, K. Kersting, and F. Železný, “Tensor factorization for multi-relational learning,” in *Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer, 2013, pp. 617–621.

[17] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, p. 717, 2009.

[18] E. J. Candès and T. Tao, “The power of convex relaxation: Near-optimal matrix completion,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2053–2080, May 2010.

[19] B. Recht, M. Fazel, and P. A. Parrilo, “Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization,” *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.

[20] V. Kolchinskii, K. Lounici, and A. Tsybakov, “Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion,” *Ann. Statist.*, vol. 39, no. 5, pp. 2302–2329, Oct. 2011.

[21] O. Klopp, “Noisy low-rank matrix completion with general sampling distribution,” *Bernoulli*, vol. 20, no. 1, pp. 282–303, 2014.

[22] A. H. Kiers, “Towards a standardized notation and terminology in multiway analysis,” *J. Chemometrics, A J. Chemometrics Soc.*, vol. 14, no. 3, pp. 105–122, 2000.

[23] L. R. Tucker, “Some mathematical notes on three-factor mode analysis,” *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.

[24] J. Hästad, “Tensor rank is NP-complete,” *J. Algorithms*, vol. 11, no. 4, pp. 644–654, Nov. 1990.

[25] V. de Silva and L.-H. Lim, “Tensor rank and the ill-posedness of the best low-rank approximation problem,” *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1084–1127, 2008.

[26] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[27] Y. W. Fan, F. Shang, J. Cheng, and H. Cheng, “Generalized higher order orthogonal iteration for tensor learning and decomposition,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2551–2563, Dec. 2016.
[55] H. Rauhut, R. Schneider, and Z. Stojanac, “Low rank tensor recovery via iterative hard thresholding,” Linear Algebra Appl., vol. 523, pp. 220–262, Jun. 2017.

[56] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, Jan. 2011.

[57] J.-F. Cai, E. J. Candès, and Z. Shen, “A singular value thresholding algorithm for matrix completion,” SIAM J. Optim., vol. 20, no. 4, pp. 1956–1982, 2010.

[58] Y. Liu, F. Shang, W. Fan, J. Cheng, and H. Cheng, “Generalized higher-order orthogonal iteration for tensor decomposition and completion,” in Proc. Adv. Neural Inf. Process. Syst., vol. 2, 2014, pp. 1763–1771.

[59] Y. Qiu, G. Zhou, Z. Huang, Q. Zhao, and S. Xie, “Efficient tensor robust PCA under hybrid model of tucker and tensor train,” IEEE Signal Process. Lett., vol. 29, pp. 627–631, 2022.

[60] N. Higham and P. Papadimitriou, “Matrix Procrustes problems,” Rapport Technique, Univ. Manchester, Manchester, U.K., Tech. Rep., 1995.

[61] D. Gabay and B. Mercier, “A dual algorithm for the solution of nonlinear variational problems via finite element approximation,” Comput. Math. Appl., vol. 2, no. 1, pp. 17–40, 1976.

[62] Z. Lin, M. Chen, and Y. Ma, “The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices,” 2010, arXiv:1009.5055.

[63] L. Yuan, Q. Zhao, L. Gui, and J. Cao, “High-order tensor completion via gradient-based optimization under tensor train format,” Signal Process., Image Commun., vol. 73, pp. 53–61, Apr. 2019.

Yuning Qiu received the B.S. degree from the South China Agricultural University, Guangzhou, China, in 2016, and the Ph.D. degree from the Guangdong University of Technology, Guangzhou, in 2021. He was a Student Intern with the RIKEN Brain Science Institute, Tokyo, Japan, in 2016. He is currently a Post-Doctoral Researcher with the School of Automation, Guangdong University of Technology. His current research interests include machine learning, tensor factorization, tensor completion, and computer vision.

Guoxu Zhou received the Ph.D. degree in intelligent signal and information processing from the South China University of Technology, Guangzhou, China, in 2010. He was a Research Scientist with the RIKEN Brain Science Institute, Tokyo, Japan. He is currently a Full Professor with the School of Automation, Guangdong University of Technology, Guangzhou. His current research interests include statistical signal processing, tensor analysis, intelligent information processing, and machine learning.

Qibin Zhao (Senior Member, IEEE) received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2009. He was a Research Scientist with the RIKEN Brain Science Institute, Tokyo, Japan, from 2009 to 2017. He was with the Center for Advanced Intelligence Project, RIKEN, as a Unit Leader from 2017 to 2019, where he is currently a Team Leader of the Tensor Learning Team. He is also a Visiting Professor with the Tokyo University of Agriculture and Technology, Tokyo; the Saitama Institute of Technology, Fukaya, Japan; and the Guangdong University of Technology, Guangzhou, China. He has published more than 150 scientific articles and coauthored two monographs on tensor networks. His research interests include machine learning, tensor factorization and tensor networks, and brain signal processing.

Dr. Zhao serves as an Area Chair for top-tier ML conferences of International Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), International Conference on Artificial Intelligence and Statistics (AISTATS), the Association for the Advancement of Artificial Intelligence (AAAI), International Joint Conferences on Artificial Intelligence Organization (IJCAI), and Asian Conference on Machine Learning (ACML). He has (co)-organized several workshops on “tensor networks in machine learning” at NeurIPS 2020 and 2021 and IJCAI 2020.

Shengli Xie (Fellow, IEEE) received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992, and the Ph.D. degree in control theory and applications from the South China University of Technology, Guangzhou, China, in 1997. He is currently the Director of the Laboratory for Intelligent Information Processing (LIIP) and a Full Professor with the Guangdong University of Technology, Guangzhou. He has authored or coauthored two monographs and more than 100 scientific papers published in journals and conference proceedings. His current research interests include automatic control and signal processing, especially blind signal processing and image processing.