



HAL
open science

Provable Nonnegative Tensor Ring Approximation: Stability and Optimality

Zhen Long, Ce Zhu, Jiani Liu, Yipeng Liu, Pierre Comon

► **To cite this version:**

Zhen Long, Ce Zhu, Jiani Liu, Yipeng Liu, Pierre Comon. Provable Nonnegative Tensor Ring Approximation: Stability and Optimality. 2022. hal-03781546v2

HAL Id: hal-03781546

<https://hal.science/hal-03781546v2>

Preprint submitted on 13 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Provable Nonnegative Tensor Ring Approximation: Stability and Optimality

Zhen Long, Ce Zhu, Jiani Liu, Yipeng Liu, Pierre Comon

Abstract—Tensor Ring (TR) decomposition provides a flexible and powerful format to represent multi-way data. However, most current works based on TR point out that the performance of TR degrades with a choice of large ranks and/or with few available entries. To alleviate this issue, optimal TR ranks selection strategies or adding regularization on TR cores are proposed. In this paper, we explain that the above issue is caused by the fact that tensor representations in TR format with predefined bounds on TR-ranks do not form a closed set, which makes the computation of approximate TR unstable. Based on this finding, we theoretically show that an optimal approximate non-negative TR (NTR) decomposition always exists and can be obtained. Computer experiments show that TR is more prone to instability when there are fewer observations or when missing data did not clearly exhibit a low rank. On the contrary, NTR is stable and performs well on image recovery. Furthermore, we discuss factor degeneracy of TR-cores, which is one of the reasons for the instability of the TR decomposition, and build a new bridge to existing works, where adding a norm regularization on TR-cores is also a way to avoid factor degeneracy.

Index Terms—tensor ring decomposition, factor degeneracy, non-negative TR (NTR) decomposition, image completion

I. INTRODUCTION

TENSORS, which can be regarded as higher-order extensions of vectors and matrices, provide a natural way to represent multidimensional data. For example, a color image, including two space indices and one color index, is a third-order tensor. As a powerful tool, tensor decomposition has become increasingly popular in a series of applications in signal processing [1]–[3], machine learning [4]–[6], or Chemometrics [7], [8], etc.

Among the existing tensor decompositions, the Canonical Polyadic (CP) decomposition [9], [10] and the Tucker decomposition [11]–[13] are the most popular. For a D th-order tensor, the aim of Tucker decomposition is to obtain a (preferably small) core tensor contracted with a factor matrix along each mode. In particular, when the core tensor is diagonal, the Tucker decomposition reduces to the CP decomposition. However, the storage complexity of Tucker decomposition grows exponentially with the order of tensor, which causes the curse of dimensionality when the order is increasing. To address this issue, tensor networks (TNs) are proposed in [14]–[17], which decompose higher-order tensors into sparsely interconnected lower-order core tensors. One instance of TNs is the Tensor Train (TT) decomposition [18],

which divides a D th-order tensor into $D-2$ third-order tensors and two matrices at the ends. Besides, in recent tensor based applications, TT-based methods perform well since they can better explain links between different modes. For example, in [19], authors used low-rank TT decomposition for tensor completion and showed that TT can better capture the global correlation among tensor entries from the perspective of von Neumann entropy in quantum information theory [20].

However, TT ranks are unbalanced, where ranks are large in the middle and small at the ends. Meanwhile, TT decomposition is sensitive to the permutation of tensor dimensions [21]–[23], which limits its representation flexibility. As an extension of TT, Tensor Ring (TR) decomposition [21], also referred to as *Matrix Product State* [24], [25], provides a more flexible and powerful representation, which factorizes a D th-order tensor into D third-order tensors, as shown in Fig. 1. Benefiting from the matrix trace operation, TR provides a more balanced structure due to its circular invariance, which shows an improvement in tensor completion [22], [26], [27], wide compression [28]–[30], supervised learning [31]–[33], or hyperspectral image processing [34]–[36]. All these applications are mainly based on a low-rank tensor ring approximation model, where TR ranks are given in advance. In addition, a series of low-rank TR based works [23], [37]–[41] point out that the performance of low-rank TR decomposition becomes worse when the TR ranks are imposed too large and/or when observed entries are few.

To address this drawback, one way is to find the optimal TR ranks. For example, Long et al [37] proposed a Bayesian tensor ring completion model to automatically infer TR ranks and adjust the trade-off between TR ranks and fitting error. Moreover, Liu et al [38] proposed a compact tensor ring regression model, which adds the group sparsity norm minimization on the TR-cores to enforce every two adjacent TR-cores to be group-sparse. These two works pruned the zero components in the TR-cores and reduced TR ranks in each iteration, which permitted to find the optimal TR ranks. Instead, Sedighin [39] proposed a new rank selection model, which increases TR ranks gradually in each iteration according to the sensitivity of the fitting error to each TR-core, achieving the optimal selection of TR ranks until the model achieves a desired approximation accuracy. Another way is to add a regularization term to the TR-core fitting error. For instance, Yuan et al [40] proposed a TR low-rank factor (TRLRF) method, which introduces the nuclear norm minimization on the TR cores to improve the performance of image completion. Furthermore, Yu et al [23] proposed a low-rank sparse tensor-ring completion (LRSTR) model, which is robust to rank

Z. Long, C. Zhu, J. Liu and Y. Liu are with School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China.

P. Comon is with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France.

selection by imposing the Frobenius norm regulation on the TR cores.

In this paper, we provide a new understanding of why fitting a tensor to a TR format performs worse with large TR ranks or with few observations. Firstly, we show that the problem is mainly caused by the tensor ring cyclic structure, which makes TR approximation with predefined TR-ranks an ill-posed problem [42]–[45]. In this case, the best low-rank TR approximation does not exist, similar to the CP decomposition [46]–[48]. This is often referred to as *instability* [24, ch.9]. In the nonnegative reals, this problem does not exist. We give the proof in Section III that the optimal nonnegative tensor ring (NTR) approximation always exists. Lee-Seung multiplicative update formulate [49] is used to compute the best low-rank NTR approximation model. Moreover, we employ NTR for image completion. Finally, experimental results on simulated data show that TR decomposition often proves unstable when there are fewer entries in the observed data, e.g. $SR < 0.2$ or when missing data do not have an obvious low rank, e.g. $R^2 > I$, R being the TR ranks and I the data size. In addition, experiments on image completion show that NTR is stable with few observations and large TR ranks. The recovery performance of NTR is generally superior to others in terms of Relative Square Error, Peak Signal to Noise Ratio, and CPU time. Note that some works [50], [51] also consider NTR for multiway representation learning, where their focus is on the application and non-negative constraints on TR cores are to provide physically meaningful features for better interpreting the boost performance.

Compared to previous works, the contributions of our work are three-fold:

- We provide the basic reason why TR decomposition is sensitive to rank selection: it is due to the fact that TR format is not topologically closed, which often leads to TR factor divergence. Furthermore, we prove that the nonnegative tensor ring (NTR) format is closed, which avoids TR factor divergence.
- We verify that TR decomposition is more prone to instability when there are fewer observations, or when the missing data do not have a low rank, but in these cases, the NTR decomposition remains stable.
- We also build a bridge with previous works, such as LRTRF and LRSTR, which demonstrates that imposing norm-constrained on TR-cores is a way to alleviate the TR-core factor divergence.

The rest of this article is organized as follows. Section II introduces some notations and preliminaries for TR decomposition. Section III provides a proof showing that the optimal nonnegative TR approximation can be obtained and describes how to apply it on tensor completion. In section IV, some experiments on simulated data and color image completion are provided. The conclusion is drawn in section V.

II. NOTATIONS AND PRELIMINARIES

A scalar, a vector, a matrix, and a tensor are written as x , \mathbf{x} , \mathbf{X} , and \mathcal{X} , respectively. The i -th entry of a vector \mathbf{x} is denoted as x_i or $\mathbf{x}(i)$, the (i, j) -th element of a matrix \mathbf{X} is denoted as

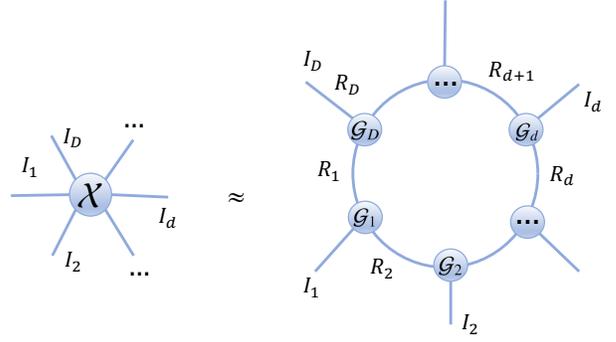


Fig. 1: TR decomposition.

$x_{i,j}$ or $\mathbf{X}(i, j)$ and (i, j, k) -th element of a third-order tensor \mathcal{X} is denoted as $x_{i,j,k}$ or $\mathcal{X}(i, j, k)$. Indices typically range from 1 to their capital version, e.g., $i = 1, \dots, I$. A nonnegative set in \mathbb{R}^I is defined by $\mathbb{R}_+^I = \{\mathbf{x} \in \mathbb{R}^I | \mathbf{x} \geq 0\}$ and the unit sphere in \mathbb{R}^I is defined by $\mathbb{S}^{I-1} = \{\mathbf{x} \in \mathbb{R}^I | \|\mathbf{x}\|_2 = 1\}$.

The definitions given in this section have been already introduced in the literature by several authors, for instance, [24], [25], [52]. We recommend [45], which is the most general and particularly elegant.

Definition 1. (mode- d unfolding) The mode- d unfolding of a D th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_D}$ is expressed as $\mathbf{X}_{(d)} = \text{unfold}_1(\mathcal{X}, d)$ or $\mathbf{X}_{[d]} = \text{unfold}_2(\mathcal{X}, d)$. Mathematically, the elements of $\mathbf{X}_{(d)}$ or $\mathbf{X}_{[d]}$ satisfy

$$\mathbf{X}_{(d)}(i_d, j_1) = \mathcal{X}(i_1, \dots, i_d, \dots, i_D)$$

or

$$\mathbf{X}_{[d]}(i_d, j_2) = \mathcal{X}(i_1, \dots, i_d, \dots, i_D)$$

where $j_1 = \overline{i_{d+1}, \dots, i_D, i_1, \dots, i_{d-1}}$ and $j_2 = \overline{i_1, \dots, i_{d-1}, i_{d+1}, \dots, i_D}$. The inverse operator is defined as $\mathcal{X} = \text{fold}_1(\mathbf{X}_{(d)}, \text{size}(\mathcal{X}), d)$ or $\mathcal{X} = \text{fold}_2(\mathbf{X}_{[d]}, \text{size}(\mathcal{X}), d)$.

Definition 2. (rank-one tensor) A rank one D th order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_D}$ can be written as the outer product of D vectors, i.e.,

$$\mathcal{X} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \dots \otimes \mathbf{u}_D$$

where \otimes represents the vector outer product. Then

$$\|\mathcal{X}\|_F = \|\mathbf{u}_1\|_F \|\mathbf{u}_2\|_F \dots \|\mathbf{u}_D\|_F. \quad (1)$$

Definition 3. (Tensor Ring decomposition) For a D th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_D}$, there exists a D -uplet, $[R_1, \dots, R_d, \dots, R_D]$, such that

$$\mathcal{X}(i_1, i_2, \dots, i_D) = \text{Trace}(\mathcal{G}_1(:, i_1, :) \mathcal{G}_2(:, i_2, :) \dots \mathcal{G}_D(:, i_D, :)). \quad (2)$$

This representation format is called a Tensor Ring (TR) decomposition, where $\mathcal{G}_d \in \mathbb{R}^{R_d \times I_d \times R_{d+1}}$, $d = 1, \dots, D$, $R_{D+1} = R_1$ are core factors. Meanwhile, it can be rewritten as the sum of rank one tensors, i.e.,

$$\mathcal{X} = \sum_{d=1}^D \sum_{r_d=1}^{R_d} \mathcal{G}_1(r_1, :, r_2) \otimes \mathcal{G}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{G}_D(r_D, :, r_1). \quad (3)$$

When we use $\mathcal{G}_d(r_d, :, r_{d+1}) = g_{r_d r_{d+1}} \mathcal{U}_d(r_d, :, r_{d+1})$, with $\|\mathcal{U}_d(r_d, :, r_{d+1})\|_F = 1$, the TR format can be rewritten as:

$$\mathcal{X} = \sum_{d=1}^D \sum_{r_d=1}^{R_d} z_{r_1, r_2, \dots, r_D} \mathcal{U}_1(r_1, :, r_2) \otimes \mathcal{U}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{U}_D(r_D, :, r_1). \quad (4)$$

where $z_{r_1, r_2, \dots, r_D} = g_{r_1 r_2} g_{r_2 r_3} \dots g_{r_D r_1}$, $r_d = 1, \dots, R_d$, $d = 1, \dots, D$.

For simplicity, we denote the TR decomposition of \mathcal{X} by $\mathcal{X} = \mathfrak{R}(\mathcal{G}_1, \dots, \mathcal{G}_D)$. The graphical illustration of TR decomposition is shown in Fig. 1.

Definition 4 (TR ranks). *Let $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_D}$ be a D th-order tensor. The TR ranks of \mathcal{X} is the D -uplet of minimal integers $[R_1, \dots, R_D]$, such that \mathcal{X} can be exactly written as in (2). The TR ranks of \mathcal{X} is then denoted:*

$$\text{rank}_{TR}(\mathcal{X}) = [R_1, \dots, R_D, \dots, R_D], \quad (5)$$

Definition 5 (TR closure). *Denote by*

$$TR(D; R_1, \dots, R_D; I_1, \dots, I_D)$$

the set of D th order tensors \mathcal{X} of size $I_1 \times \dots \times I_D$ whose TR ranks are bounded by (R_1, \dots, R_D) :

$$\text{rank}_{TR}(\mathcal{X}) \leq [R_1, \dots, R_D, \dots, R_D]$$

where the inequality is satisfied entry-wise. Also denote its Zariski closure as $\overline{TR}(D; R_1, \dots, R_D; I_1, \dots, I_D)$. In other words, every element of TR is the limit of a sequence in \overline{TR} .

Definition 6 (TR border rank). *Let $\mathcal{X} \in TR(D; R_1, \dots, R_D; I_1, \dots, I_D)$. Then it is clear that $\mathcal{X} \in \overline{TR}(D; R'_1, \dots, R'_D; I_1, \dots, I_D)$, with $R'_d \leq R_d$. If (R'_1, \dots, R'_D) is minimal, it is called the TR border rank of \mathcal{X} and denoted as*

$$\overline{\text{rank}}_{TR}(\mathcal{X}) = [R'_1, \dots, R'_D], \quad (6)$$

Similarly to the usual tensor rank (defined via the CP decomposition), the set $TR(D; R_1, \dots, R_D; I_1, \dots, I_D)$ of tensors having TR ranks smaller than a given value is generally not a closed set (except in the case of TT), as shown in the next theorem, stated in a more general framework in [45, Theorem 9.10].

Theorem 1. *The set $TR(D; R_1, \dots, R_D; I_1, \dots, I_D)$ is not closed. In fact, there exists at least one D th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_D}$ with $D > 3$ and $I_d > 1$, such that $\mathcal{X} \in \overline{TR}(D; R'_1, \dots, R'_D; I_1, \dots, I_D)$ and $\mathcal{X} \notin TR(D; R_1, \dots, R_D; I_1, \dots, I_D)$.*

As shown above, a sequence of tensors with a given TR-rank may converge towards a limit with a larger TR-rank, which causes the divergence of at least two tensor factors. Therefore, TR decomposition is *unstable* and the algorithms used to compute it may get stuck in a false local minimum.

III. NONNEGATIVE TENSOR RING DECOMPOSITION

In this section, we will show that imposing nonnegative constraints on TR core tensors allows the set of tensors $TR(\cdot)$ to become closed, which will avoid instabilities. The nonnegative tensor ring (NTR) decomposition factorizes a nonnegative tensor into a sum of nonnegative rank-one tensors. For example, given a D th order tensor $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$, we wish to write it as

$$\mathcal{X} = \sum_{d=1}^D \sum_{r_d=1}^{R_d} \mathcal{G}_1(r_1, :, r_2) \otimes \mathcal{G}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{G}_D(r_D, :, r_1). \quad (7)$$

where $\mathcal{G}_d(r_d, :, r_{d+1}) \in \mathbb{R}_+^{I_d}$, $r_d = 1, \dots, R_d$, $d = 1, \dots, D$. Write $\mathcal{G}_d(r_d, :, r_{d+1}) = g_{r_d r_{d+1}} \mathcal{U}_d(r_d, :, r_{d+1})$ where $\mathcal{U}_d(r_d, :, r_{d+1})$ belongs to the nonnegative portion of the Euclidean unit sphere $\mathbb{S}_+^{I_d-1}$. Then \mathcal{X} can be rewritten as:

$$\mathcal{X} = \sum_{d=1}^D \sum_{r_d=1}^{R_d} z_{r_1, r_2, \dots, r_D} \mathcal{U}_1(r_1, :, r_2) \otimes \mathcal{U}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{U}_D(r_D, :, r_1). \quad (8)$$

where $z_{r_1, r_2, \dots, r_D} = g_{r_1 r_2} g_{r_2 r_3} \dots g_{r_D r_1}$, $r_d = 1, \dots, R_d$, $d = 1, \dots, D$. Denote \mathcal{Z} the tensor with entries z_{r_1, r_2, \dots, r_D} .

Theorem 2. *Given a D th order tensor $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$,*

$$\inf\{f(Q) | Q \in \mathbb{Q}\} \quad (9)$$

is attained, where $f(Q) = \|\mathcal{X} - \sum_{d=1}^D \sum_{r_d=1}^{R_d} z_{r_1, r_2, \dots, r_D} \mathcal{U}_1(r_1, :, r_2) \otimes \mathcal{U}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{U}_D(r_D, :, r_1)\|_F^2$, $Q = (\mathcal{Z}, \mathcal{U}_1, \dots, \mathcal{U}_D) \in \mathbb{Q}$, and $\mathbb{Q} = \mathbb{R}_+^{R_1 \times R_2 \times \dots \times R_D} \times (\mathbb{S}_+^{I_1-1})^{R_1 R_2} \times \dots \times (\mathbb{S}_+^{I_D-1})^{R_D R_1}$.

Proof. It is noted that \mathbb{Q} is closed but unbounded. The key idea of the proof is to show that the sublevel set of f restricted to \mathbb{Q} ,

$$\mathbb{P}_\alpha = \{Q \in \mathbb{Q} | f(Q) \leq \alpha\} \quad (10)$$

is compact for all α , so that the infimum of f on \mathbb{Q} must be attained. We will show $\mathbb{P}_\alpha = \mathbb{Q} \cap f^{-1}(0, \alpha]$ is closed and bounded, where f^{-1} is the inverse function of f .

Firstly, \mathbb{P}_α is closed since f is continuous.

Secondly, let us prove by contradiction that \mathbb{P}_α is bounded. Suppose there is a sequence $\{Q_t\}_{t=1}^\infty \in \mathbb{Q}$ with $\|Q_t\|_F \rightarrow \infty$ but $f(Q_t) \leq \alpha$ for all t . $\|Q_t\|_F \rightarrow \infty$ implies that at least one entry z_{p_1, p_2, \dots, p_D} , $p_d \in \{r_d, d = 1, \dots, D\} \rightarrow \infty$.

By Cauchy-Schwarz $\|\mathbf{A} - \mathbf{B}\|_F^2 \geq (\|\mathbf{A}\|_F - \|\mathbf{B}\|_F)^2$, we have

$$f(Q_d) \geq (\|\mathcal{X}\|_F - \|\sum_{d=1}^D \sum_{r_d=1}^{R_d} z_{r_1, r_2, \dots, r_D} \mathcal{U}_1(r_1, :, r_2) \otimes \mathcal{U}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{U}_D(r_D, :, r_1)\|_F)^2 \quad (11)$$

where $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$ and

$$\begin{aligned} & \|\sum_{d=1}^D \sum_{r_d=1}^{R_d} z_{r_1, r_2, \dots, r_D} \mathcal{U}_1(r_1, :, r_2) \otimes \mathcal{U}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{U}_D(r_D, :, r_1)\|_F \\ & \geq \|z_{p_1, p_2, \dots, p_D} \mathcal{U}_1(p_1, :, p_2) \otimes \mathcal{U}_2(p_2, :, p_3) \otimes \dots \otimes \mathcal{U}_D(p_D, :, p_1)\|_F \\ & = z_{p_1, p_2, \dots, p_D} \|\mathcal{U}_1(p_1, :, p_2) \otimes \mathcal{U}_2(p_2, :, p_3) \otimes \dots \otimes \mathcal{U}_D(p_D, :, p_1)\|_F \\ & = z_{p_1, p_2, \dots, p_D} \|\mathcal{U}_1(p_1, :, p_2)\|_F \|\mathcal{U}_2(p_2, :, p_3)\|_F \dots \|\mathcal{U}_D(p_D, :, p_1)\|_F \\ & = z_{p_1, p_2, \dots, p_D}, \end{aligned} \quad (12)$$

where the last two equalities follow from equation (1) and $\|\mathcal{U}_1(p_1, :, p_2)\|_F = \|\mathcal{U}_2(p_2, :, p_3)\|_F = \dots = \|\mathcal{U}_D(p_D, :, p_1)\|_F = 1$. Therefore, $\|Q_d\|_F \rightarrow \infty$, $z_{p_1, p_2, \dots, p_D} \rightarrow \infty$ and $f(Q_d) \rightarrow \infty$, and the assumption $f(Q_d) \leq \alpha$ for all n cannot be held. \square

Theorem 2 implies that z_{r_1, r_2, \dots, r_D} is bounded in the NTR form, and there are no rank-1 components that cancel out each other, avoiding TR-core factor degeneracy discussed in Section II. Note that if the ℓ_1 norm is used instead of the Euclidean norm, Theorem 2 still holds true, and the proof is similar, by using the inequality $\|A - B\|_1 \geq |\|A\|_1 - \|B\|_1|$.

A. LS-NTR

In this section, we describe an algorithm that minimizes the fitting error of nonnegative tensor ring decomposition, which can be expressed as:

$$\min_{\{\mathcal{G}_d\}_{d=1}^D} \frac{1}{2} \|\mathcal{X} - \sum_{d=1}^D \sum_{r_d=1}^{R_d} \mathcal{G}_1(r_1, :, r_2) \otimes \mathcal{G}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{G}_D(r_D, :, r_1)\|_F^2, \quad (13)$$

where $\mathcal{G}_d(r_d, :, r_{d+1}) \in \mathbb{R}_+^{I_d \times R_d}$, $r_d = 1, \dots, R_d$, $d = 1, \dots, D$. The minimization can be conducted by a block coordinate descent algorithm, which splits the problem (13) into several subproblems, where one variable \mathcal{G}_d is updated in turns while the others are fixed. The subproblem of (13) with respect to \mathcal{G}_d can be written as

$$\min_{\mathcal{G}_d} \frac{1}{2} \|\mathbf{X}_{(d)} - \mathbf{G}_d(\mathbf{G}_{(2)}^{\#d})^T\|_F^2 \quad (14)$$

where $\mathbf{G}_d = \text{unfold}_2(\mathcal{G}_d, 2) \in \mathbb{R}_+^{I_d \times R_d R_{d+1}}$ and $\mathbf{G}_{(2)}^{\#d} = \text{unfold}_1(\mathcal{G}^{\#d}, 2) \in \mathbb{R}_+^{I_{d+1} \times \dots \times I_D, I_1, \dots, I_{d-1} \times R_d R_{d+1}}$ and $\mathcal{G}^{\#d}$ is a merging tensor, which is defined as:

$$\begin{aligned} \mathcal{G}^{\#d}(r_{d+1}, j, r_d) &= \mathcal{G}_{d+1}(r_{d+1}, i_{d+1}, :) \cdots \\ \mathcal{G}_D(:, i_D, :) \mathcal{G}_1(:, i_1, :) \cdots \mathcal{G}_{d-1}(:, i_{d-1}, r_d), \end{aligned} \quad (15)$$

where $j = \overline{i_{d+1}, \dots, i_D, i_1, \dots, i_{d-1}}$.

According to Lee-Seung multiplicative update formula, the update of \mathcal{G}_d can be obtained by

$$\mathbf{G}_d = \mathbf{G}_d \cdot * (\mathbf{X}_{(d)} \mathbf{G}_{(2)}^{\#d}) ./ (\mathbf{G}_d (\mathbf{G}_{(2)}^{\#d})^T \mathbf{G}_d), \quad (16)$$

where $\mathcal{G}_d = \text{fold}_2(\mathbf{G}_d, \text{size}(\mathcal{G}_d), 2)$, $\cdot *$ and $./$ are element-wise multiplication and division, respectively. We call this algorithm Lee-Seung nonnegative tensor ring approximation method (LS-NTR) and summarize it in Algorithm 1.

B. LS-NTR for tensor completion

In this section, we will apply LS-NTR to the tensor completion task with nonnegative missing data. It is a common fact that there are some missing entries in observed data during acquisition or transmission. The goal of tensor completion is to infer missing entries from their observed data. Given a D th order incomplete tensor $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$, the mathematical expression allowing nonnegative tensor ring completion is

$$\min_{\{\mathcal{G}_d\}_{d=1}^D} \frac{1}{2} \|\mathcal{P}_\circ(\mathcal{X} - \sum_{d=1}^D \sum_{r_d=1}^{R_d} \mathcal{G}_1(r_1, :, r_2) \otimes \mathcal{G}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{G}_D(r_D, :, r_1))\|_F^2, \quad (17)$$

Algorithm 1 LS-NTR

Input: A D th-order data tensor $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$, predefined nonnegative TR rank $[R_1, \dots, R_D]$, maximum number of iterations (Maxiter), and the threshold for stopping the algorithm $\text{tol} = 10^{-6}$

Initialization: Initialize $\mathcal{G}_d \in \mathbb{R}_+^{R_d \times I_d \times R_{d+1}}$, $d = 1, \dots, D$

while $i \leq \text{Maxiter}$ **do**

for $d = 1 : D$ **do**

 obtain $\mathbf{X}_{(d)} = \text{unfold}_1(\mathcal{X}, d)$

 obtain $\mathbf{G}_2^{\#d} = \text{unfold}_1(\mathcal{G}^{\#d}, 2)$

 update \mathbf{G}_d via equation (16)

 obtain $\mathcal{G}_d = \text{fold}_2(\mathbf{G}_d, \text{size}(\mathcal{G}_d), 2)$

end for

$\hat{\mathcal{X}} = \mathfrak{R}(\mathcal{G}_1, \dots, \mathcal{G}_D)$

if $\frac{\|\hat{\mathcal{X}} - \mathcal{X}\|_F}{\|\mathcal{X}\|_F} \leq \text{tol}$ **then**

break

end if

$i = i + 1$;

end while

Output: \mathcal{G}_d , $d = 1, \dots, D$

where $\mathcal{G}_d(r_d, :, r_{d+1}) \in \mathbb{R}_+^{I_d \times R_d}$, $r_d = 1, \dots, R_d$, $d = 1, \dots, D$ and \mathcal{P}_\circ is a random sampling operator, which is defined as:

$$\mathcal{P}_\circ = \begin{cases} \mathcal{X}(i_1, i_2, \dots, i_D) & (i_1, i_2, \dots, i_D) \in \circ \\ 0 & \text{otherwise} \end{cases}$$

and \circ is the set of indices of observed entries in \mathcal{X} . We apply LS-NTR to solve the optimization problem (17). The update of \mathcal{G}_d in the tensor completion task is obtained by

$$\mathbf{G}_d(i_d, :) = \mathbf{G}_d \cdot * (\mathbf{X}_{(d)}(i_d, \circ_{i_d}) \mathbf{G}_{(2)}^{\#d}(\circ_{i_d}, :)) ./ (\mathbf{G}_d \mathbf{P}), \quad (18)$$

$\mathbf{P} = (\mathbf{G}_{(2)}^{\#d}(\circ_{i_d}, :))^T \mathbf{G}_{(2)}^{\#d}(\circ_{i_d}, :) \in \mathbb{R}_+^{R_d R_{d+1} \times R_d R_{d+1}}$, $i_d = 1, \dots, I_d$, \circ_{i_d} is the index set of known entries in $\mathbf{X}_{(d)}(i_d, :)$, and $\mathcal{G}_d = \text{fold}_2(\mathbf{G}_d, \text{size}(\mathcal{G}_d), 2)$. Note that the solution for \mathcal{G}_d is slightly different from that in equation (16) since we only use observed entries instead of all entries to update \mathcal{G}_d .

IV. EXPERIMENTAL RESULTS

Before conducting this experiment, we give two important definitions. For a D th-order tensor $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \dots \times I_D}$, the sampling ratio (SR) is defined by

$$\text{SR} = \frac{|\circ|}{\prod_{d=1}^D I_d}$$

where $|\circ|$ is the number of observed entries. The relative square error (RSE) is used to measure the recovery performance, which is defined as $\frac{\|\mathcal{X}(\circ) - \hat{\mathcal{X}}(\circ)\|_F}{\|\mathcal{X}(\circ)\|_F}$, where \mathcal{X} is the original tensor and $\hat{\mathcal{X}}$ is the recovered tensor. Note that the maximum value of RSE is 1 by definition; also note the algorithm definitely fails to recover the data if $\text{RSE} > 0.5$. All tests are conducted 100 times on simulated data and 10 times on real data. All tests are accomplished on a desktop computer with a 3.30GHz Intel(R) Xeon(R)(TM) CPU and 256GB RAM using MatLab 2018a.

A. Experimental results on simulated data

In this part, we conduct two groups of experiments to study the instability of TR decomposition under different conditions. We choose TR-ALS as our baseline algorithm and LS-NTR as a comparison algorithm. The original tensors are generated by core tensors $\mathcal{G}_d \in \mathbb{R}_+^{R_d \times I_d \times R_{d+1}}$, $d = 1, \dots, D$ whose entries are independently randomly drawn from half-normal distribution. For simplicity, we assume all TR ranks are the same, e.g., $R_1 = R_2 = \dots = R_d = R$.

1) *Recovery performance on different SR and TR ranks selection*: The goal of this group is to explore the relationship between the recovery performance and TR ranks selection on different SR conditions, where SR is chosen from {0.05, 0.1, 0.5, 1}, and the predefined TR ranks $R_1 = R_2 = \dots = R_d = R$ are varied in {2, 3, 4, 5, 6}.

The observed data is generated with $D = 3, R = 3, I = 30$ using different SRs. Fig. 2 shows RSE variations as a function of predefined TR ranks and SRs; in this figure, circular marks, red lines and pink asterisks are the mean, the median, and the outliers results of 100 repeated tests, respectively. From Fig. 2(a), we can see TR-ALS fails to recover the data with few observations when the predefined TR ranks are larger than or equal to the real ranks. Instead, the recovery performance of TR-ALS with underestimated TR ranks is better than that with overestimated ones. Interestingly, the change of RSE on LS-NTR is stable with different rank selections when $SR = 0.05$. With SR increased, LS-NTR still performs more steadily than TR-ALS. Specifically, as shown in Fig. 2(b), the RSE value of TR-ALS is larger than that of LS-NTR when the predefined TR ranks are overestimated and $SR=0.1$. Meanwhile, when the predefined TR ranks are underestimated with $SR=1$, the recovery performance of LS-NTR is superior to that of TR-ALS.

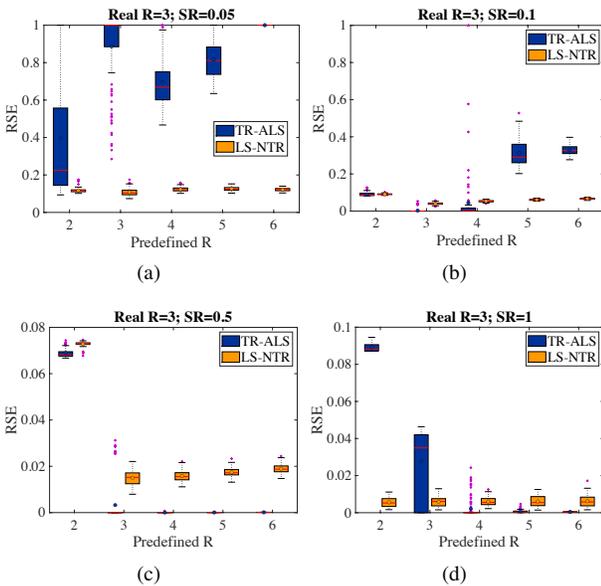


Fig. 2: Recovery performance on different SRs and predefined TR ranks

2) *Recovery performance on different SR and data structure*: The aim of this group is to investigate the relationship between the recovery performance and different data structures, e.g., $R^2 < I$, $R^2 = I$, $R^2 > I$. Three cases of experiments, namely $TR(3; 2, 2, 2; 10, 10, 10)$, $TR(3; 3, 3, 3; 9, 9, 9)$, and $TR(3; 4, 4, 4; 5, 5, 5)$, are performed with each one being repeated 100 times. Fig. 3 shows the variations of RSE along SRs using TR-ALS and LS-NTR. From Fig. 3, we can see LS-NTR almost outperforms TR-ALS on different relationships between R^2 and I . Meanwhile, the recovery performance of TR-ALS is sensitive to the high-rank structure of data according to the results shown in Fig.3(a) and Fig.3(c). For example, TR-ALS can recover the missing data when $R^2 < I$ and $SR > 0.2$, but it will fail if $R^2 > I$ and $SR < 1$. It may imply that TR decomposition is more unstable when the rank of data is large, e.g., $R^2 > I$.

To conclude, the instability of TR is related to the sample ratio SR and the data structure. Especially, when SR is smaller or when low-rank structure of missing data is not obvious, the recovery performance of TR-ALS will be worse. Instead, LS-NTR performs well in these cases.

B. Application on image completion

In this section, we consider eight color images for completion, as shown in Fig. 4. Two unconstrained TR-based methods including TR-ALS [22] and TRWOPT [26] are taken as baseline methods. In addition, some improved TR-based methods, which are robust to TR ranks selection, including LRSTR [23], TRLRF [40] and TR-VBI [37], are used for comparison. The initialization of core tensors in all TR-based methods are the same and tensor ring ranks $R_d, n = 1, 2, 3$ are chosen in {2,4,6,8,10,12,14,16,18,20}.

In addition, to verify whether the poor performance of TR decomposition on image completion is caused by the instability of TR decomposition or by a parameter overfitting, we add a TT-based algorithm (TTWOPT [53]) since TT decomposition is stable. In this case, TT ranks are determined by the parameters of the TRWOPT algorithm. For instance, TT ranks can be determined according to the same parameters of TTWOPT and TRWOPT with respect to different TR ranks. The observations are chosen from $SR = \{0.03, 0.05, 0.1, 0.15, 0.2\}$, and each test is repeated 10 times. Two kinds of performance metrics are employed to evaluate the estimation accuracy, which are Peak Signal-to-Noise Ratio (PSNR) and RSE.

Fig. 5, Fig. 6 and Fig. 7 show the comparison results with different TR ranks using different methods with a few observations, e.g. $SR = 0.03$, $SR = 0.05$ and $SR = 0.1$. It can be observed that TR-ALS and TRWOPT are sensitive to the TR ranks, especially, when few observations are available: the larger R , the poorer the recovery performance. Instead, other improved TR-methods including LRSTR, LS-NTR, TRLRF and TR-VBI show the advantage of a robust TR rank selection. Among them, the recovery results of LS-NTR and TRLRF perform better when $SR=0.03$. In addition, the recovery performance of LRSTR, LS-NTR and TR-VBI is superior to others in terms of image resolution when $SR=0.05$ and $SR=0.1$. Interestingly, TTWOPT and TRWOPT have the

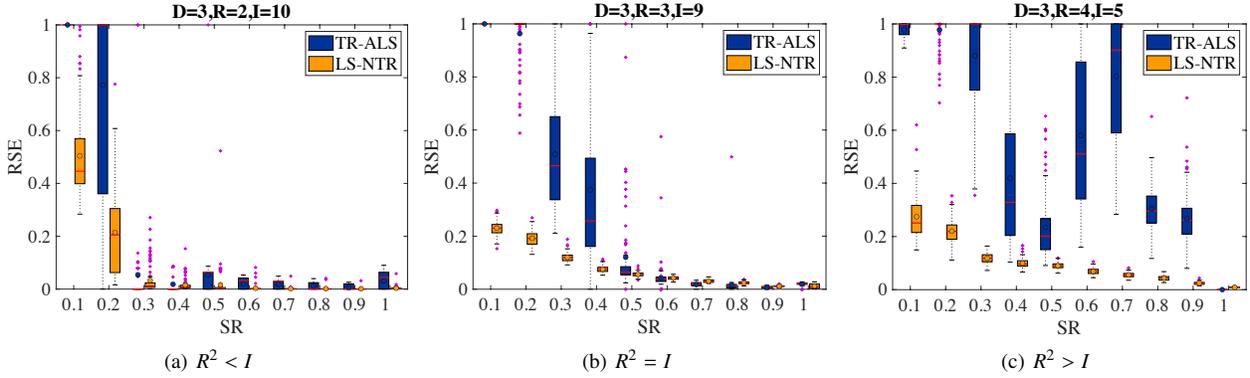


Fig. 3: Recovery performance on different relationships between R^2 and I .

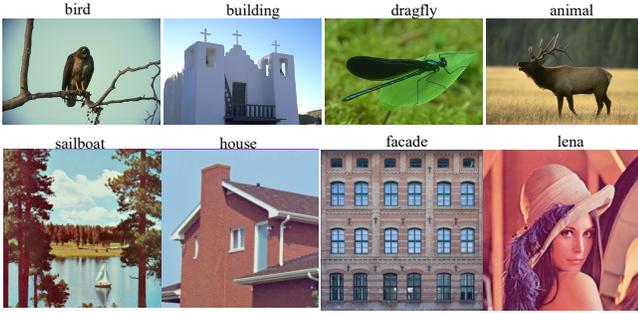


Fig. 4: Testing images.

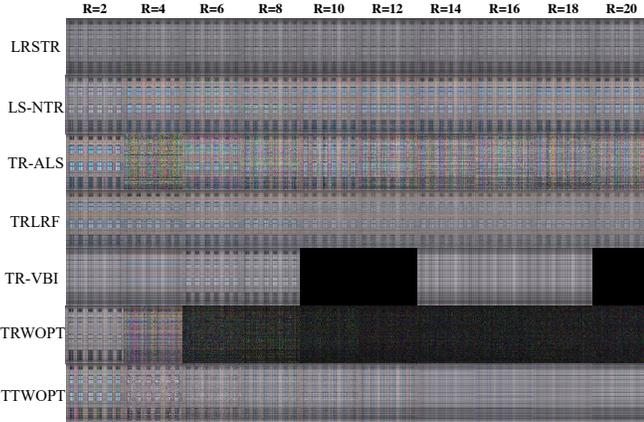


Fig. 5: Examples on image completion using different methods when $SR=0.03$.

same parameters, but TTWOPT outperforms TRWOPT when TR ranks are large with few samples.

In addition, Fig. 8 illustrates the average change performance of all methods on eight color images along the TR ranks in terms of RSE with $SR=0.05$. It can be observed TR-ALS and TRWOPT almost fail to recover the image when $R > 4$. Fortunately, the recovery performance of improved TR methods is stable with the change of TR rank. Among them, LS-NTR performs better with large TR ranks in most cases. Meanwhile, TTWOPT performs better than TRWOPT under identical parameter conditions.

Fig. 9, Fig. 10 and Fig. 11 provide the recovery performance of all methods with different TR ranks on eight color images

along different SRs in terms of RSE, PSNR and CPU time, respectively. In these figures, the circular marks and red lines are the average and the median results of different TR ranks and 10 repeated tests, respectively. We can see the RSE of LRSTR, LS-NTR and TRLRF remain almost unchanged with respect to different TR ranks when $SR=0.03$, as shown in Fig. 9. Meanwhile, constrained TR-based methods outperform TR-ALS in all cases. In addition, we can see when $SR=0.2$, the recovery performance of TR-ALS is greatly influenced by the predefined TR ranks. From Fig. 10, it can be observed that the PSNR of the improved TR-based methods is stable with different predefined TR ranks and different SRs. Interestingly, the recovery performance of TRWOPT is superior to that of TR-ALS when $SR=0.2$. It can be seen from the data in Fig. 11 that the computational complexity of TT-based methods is smaller than that of TR-based ones. Among all constrained TR-based methods, the CPU Time of LS-NTR and TRLRF cost less than that of LRSTR and TR-VBI.

C. Discussion

1) *Analysis of the causes of instability:* To further explore the reason for the instability of TR decomposition, the changes of stability metric σ of TR-ALS and LS-NTR are shown in Fig. 12(a) and Fig. 12(b). The ratio $\sigma = \frac{\|\mathcal{G}\|}{\|\mathcal{X}\|_F}$ is used to measure the stability [24], where $\mathcal{X} = \sum_{d=1}^D \sum_{r_d}^{R_d} \mathcal{Z}$, $\mathcal{Z} = \mathcal{G}_1(r_1, :, r_2) \otimes \mathcal{G}_2(r_2, :, r_3) \otimes \dots \otimes \mathcal{G}_D(r_D, :, r_1)$ is a rank-one tensor, and $\|\mathcal{G}\| = \sum_{d=1}^D \sum_{r_d}^{R_d} \|\mathcal{Z}\|_F = \sum_{d=1}^D \sum_{r_d}^{R_d} (\prod_{d=1}^D \|\mathcal{G}_d(r_d, :, r_{d+1})\|_F)$. Note that the negative data in Fig. 12 are ignored. From Fig. 12(a-b), we could observe that the changing trend of σ is consistent with that of RSE in Fig. 2 and Fig. 3. The larger the RSE, the more unstable the TR decomposition. In addition, Fig. 12 (a-b) further testify that the instability of TR decomposition is mainly caused by tensor factor degeneracy, e.g. $\|\mathcal{Z}\|_F$ is very large.

2) *The link to existing works:* These results on image completion indicate that TR decomposition easily exhibits instability, especially when few observations are available. Instead, adding some constraints, including adding the norm-regularized (LRSTR, TRLRF) or nonnegative (LS-NTR) on TR-cores, can improve the recovery performance.

To explore the reason, we plot the change in the mean of σ on comparison recovery methods along different TR-ranks in Fig. 12(c). We can see that the σ of TR-ALS and TR-WOPT

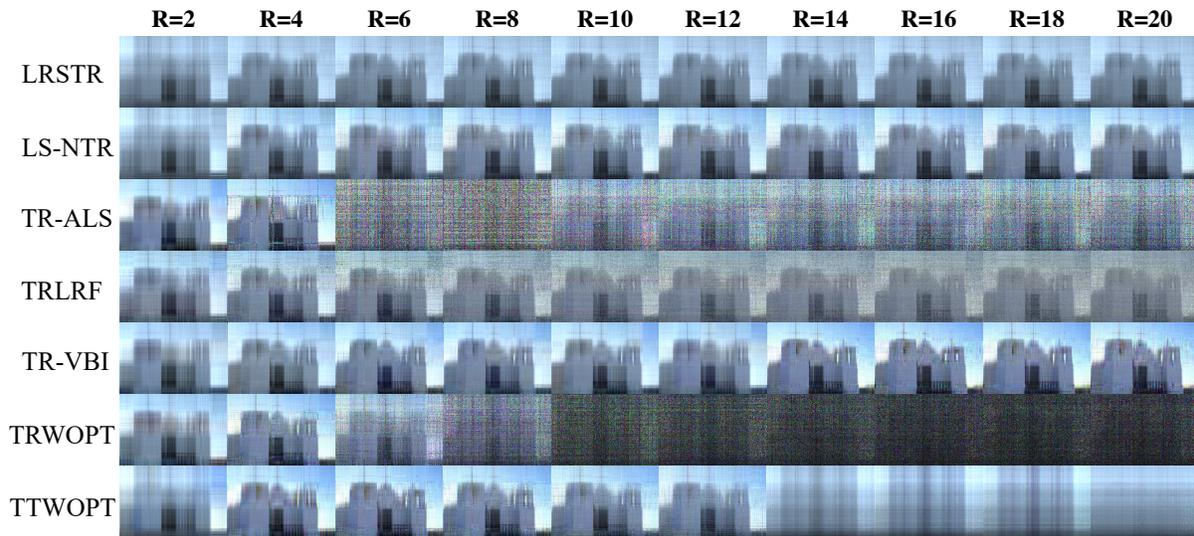


Fig. 6: Examples on image completion using different methods when $SR=0.05$.



Fig. 7: Examples on image completion using different methods when $SR=0.1$.

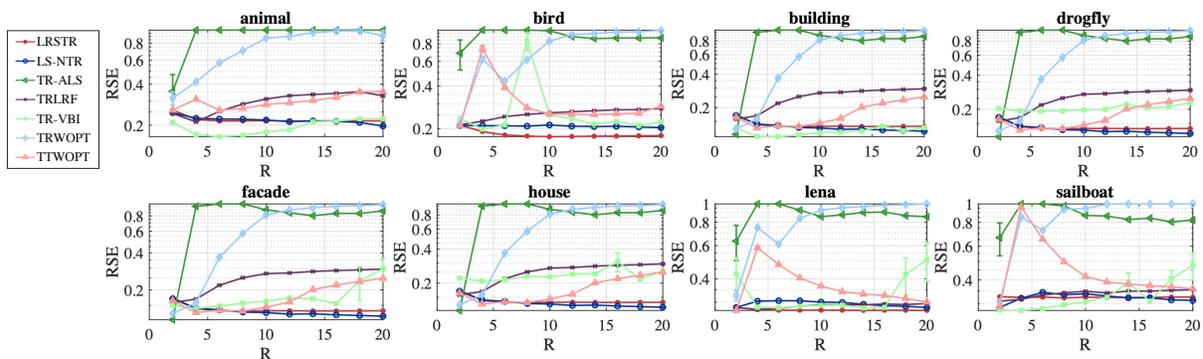


Fig. 8: Comparison recovery performance on eight color images using different methods when $SR=0.05$ (RSE vs. R).

are much larger than that of others, which demonstrates that the instability of TR decomposition is caused by the TR-core factor divergence. The norm or non-negative constraint

is imposed on the TR-core factors to make them bounded, which can further improve the stability of the TR decomposition. Furthermore, when comparing the recovery results of

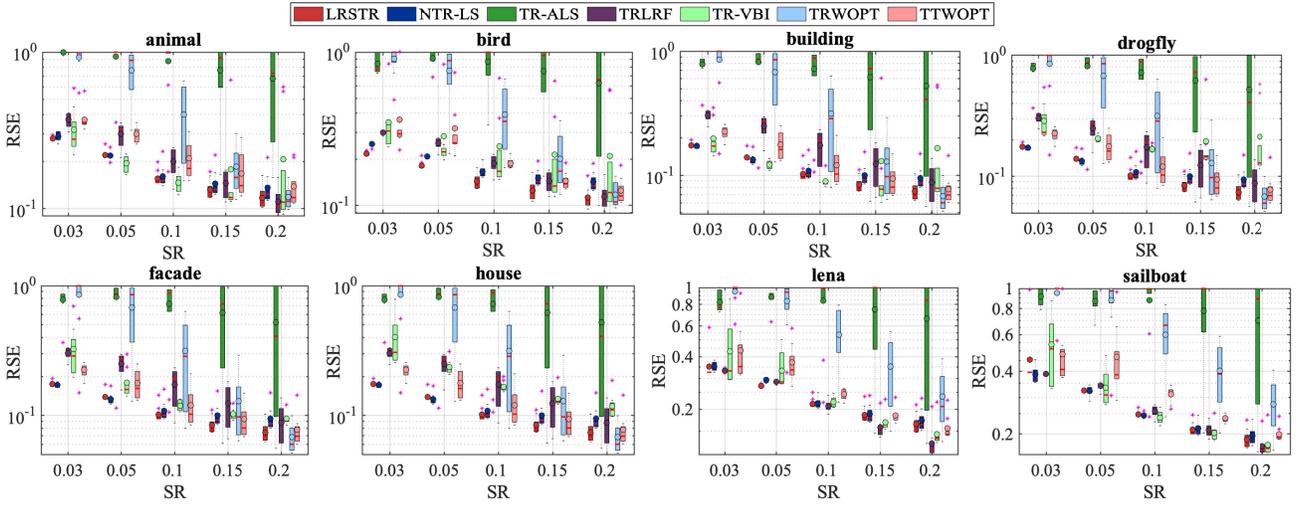


Fig. 9: The change of RSE along different SRs on eight color images using different methods.

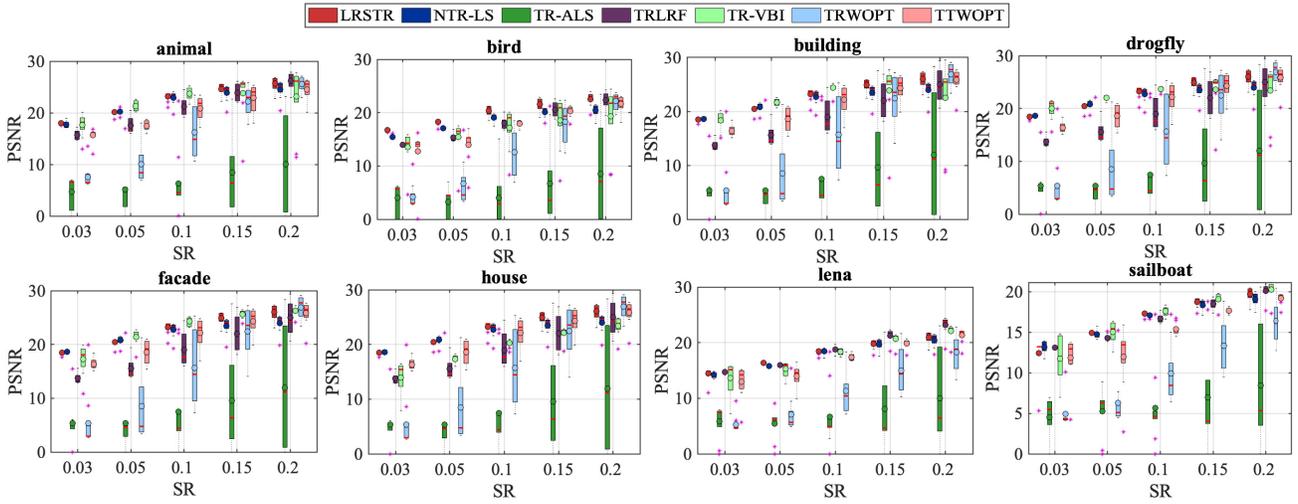


Fig. 10: The change of PSNR along different SRs on eight color images using different methods.

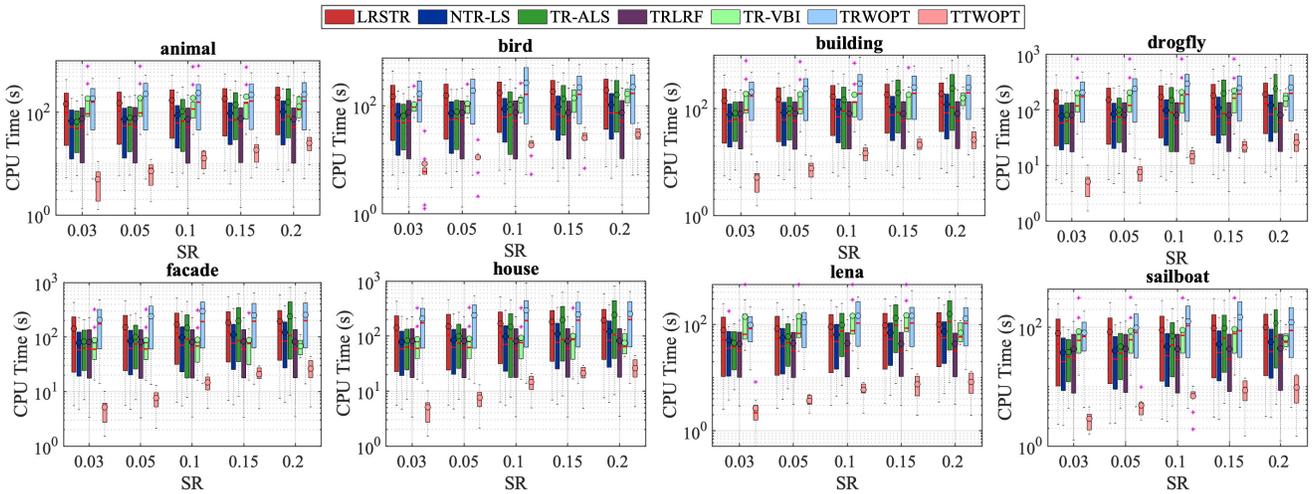


Fig. 11: The change of CPU time along different SRs on eight color images using different methods.

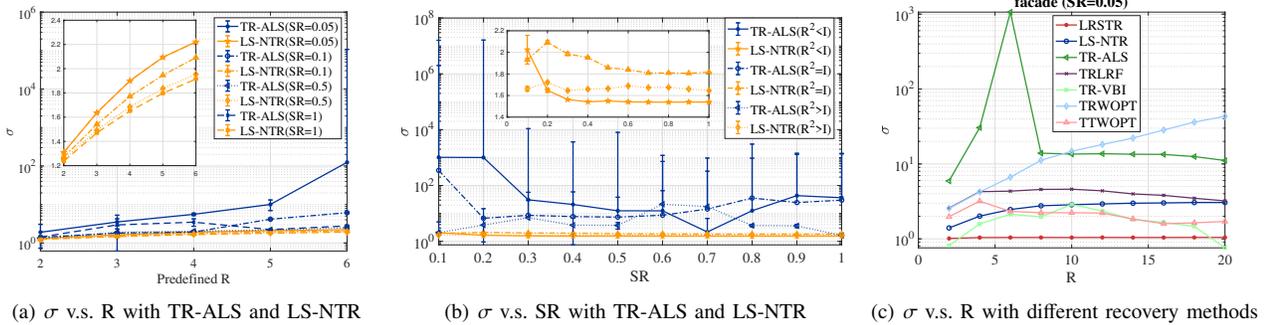


Fig. 12: The change of stability metric under different conditions.

TTWOPT and TRWOPT with similar parameters, the recovery performance of TTWOPT is significantly better than that of TRWOPT, especially when there are few observations. This means that the reason for the poor recovery results of TRWOPT may mostly originate from TR decomposition instability.

V. CONCLUSION

This paper provides an understandable explanation from the perspective of mathematical analysis that TR-core factor degeneracy leads to instability of the TR decomposition, thus making the TR-based methods perform worse. On the other hand, we prove that the optimal low-rank NTR approximation exists. Moreover, NTR decomposition is applied to image completion. Simulation experiments show that the TR decomposition is more prone to instability on few observations, or when the missing data do not actually have a low rank, but the NTR decomposition remains stable in these cases. In addition, experimental results on image completion show the proposed method performs well in terms of recovery accuracy and time efficiency. At the same time, experiments confirm that the TR-core factor divergence is one of the reasons that affect the instability of the TR decomposition.

REFERENCES

- [1] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of eeg signals: a brief review," *Journal of neuroscience methods*, vol. 248, pp. 59–69, 2015.
- [2] L.-H. Lim and P. Comon, "Multirray signal processing: Tensor decomposition meets compressed sensing," *Comptes Rendus Mecanique*, vol. 338, no. 6, pp. 311–320, 2010.
- [3] V. Hore, A. Viñuela, A. Buil, J. Knight, M. I. McCarthy, K. Small, and J. Marchini, "Tensor decomposition for multiple-tissue gene expression experiments," *Nature genetics*, vol. 48, no. 9, pp. 1094–1100, 2016.
- [4] S. Rabanser, O. Schur, and S. Günnemann, "Introduction to tensor decompositions and their applications in machine learning," *arXiv preprint arXiv:1711.10781*, 2017.
- [5] A. Tjandra, S. Sakti, and S. Nakamura, "Tensor decomposition for compressing recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2018.
- [6] T. Lacroix, N. Usunier, and G. Obozinski, "Canonical tensor decomposition for knowledge base completion," in *International Conference on Machine Learning*, pp. 2863–2872, PMLR, 2018.
- [7] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis*. Chichester UK: Wiley, 2004.
- [8] C. M. Andersen and R. Bro, "Practical aspects of parafac modeling of fluorescence excitation-emission data," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 4, pp. 200–215, 2003.
- [9] P. Comon, "Tensors: a brief introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 44–53, 2014.
- [10] M. A. Veganzones, J. E. Cohen, R. C. Farias, J. Chanussot, and P. Comon, "Nonnegative tensor cp decomposition of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2577–2588, 2015.
- [11] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [12] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [13] Y.-D. Kim and S. Choi, "Nonnegative Tucker decomposition," in *2007 IEEE conference on computer vision and pattern recognition*, pp. 1–8, IEEE, 2007.
- [14] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Annals of physics*, vol. 349, pp. 117–158, 2014.
- [15] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. P. Mandic, "Tensor networks for dimensionality reduction and large-scale optimization: Part I low-rank tensor decompositions," *Foundations and Trends® in Machine Learning*, vol. 9, no. 4-5, pp. 249–429, 2016.
- [16] A. Cichocki, "Era of big data processing: A new approach via tensor networks and tensor decompositions," *arXiv preprint arXiv:1403.2048*, 2014.
- [17] E. Stoudenmire and D. J. Schwab, "Supervised learning with tensor networks," *Advances in Neural Information Processing Systems*, vol. 29, 2016.
- [18] I. V. Oseledets, "Tensor-train decomposition," *SIAM Journal on Scientific Computing*, vol. 33, no. 5, pp. 2295–2317, 2011.
- [19] J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2466–2479, 2017.
- [20] M. A. Nielsen and I. Chuang, "Quantum computation and quantum information," 2002.
- [21] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016.
- [22] W. Wang, V. Aggarwal, and S. Aeron, "Efficient low rank tensor ring completion," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5697–5705, 2017.
- [23] J. Yu, G. Zhou, W. Sun, and S. Xie, "Robust to rank selection: Low-rank sparse tensor-ring completion," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [24] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42. Springer, 2019. 2nd edition.
- [25] J. Landsberg, Y. Qi, and K. Ye, "On the geometry of tensor network states," *Quantum Inf. Comput.*, vol. 1, no. 3, pp. 346–354, 2012.
- [26] L. Yuan, J. Cao, X. Zhao, Q. Wu, and Q. Zhao, "Higher-dimension tensor completion via low-rank tensor ring decomposition," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1071–1076, IEEE, 2018.
- [27] H. Huang, Y. Liu, Z. Long, and C. Zhu, "Robust low-rank tensor ring completion," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1117–1126, 2020.
- [28] W. Wang, Y. Sun, B. Eriksson, W. Wang, and V. Aggarwal, "Wide compression: Tensor ring nets," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9329–9338, 2018.
- [29] Y. Pan, J. Xu, M. Wang, J. Ye, F. Wang, K. Bai, and Z. Xu, "Compressing recurrent neural networks with tensor ring for action recognition," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 4683–4690, 2019.

- [30] A. Obukhov, M. Rakhuba, S. Georgoulis, M. Kanakis, D. Dai, and L. Van Gool, "T-basis: a compact representation for neural networks," in *International Conference on Machine Learning*, pp. 7392–7404, PMLR, 2020.
- [31] J. Liu, C. Zhu, Z. Long, H. Huang, and Y. Liu, "Low-rank tensor ring learning for multi-linear regression," *Pattern Recognition*, vol. 113, p. 107753, 2021.
- [32] M. Kuznetsov, D. Polykovskiy, D. P. Vetrov, and A. Zhebrak, "A prior of a googol gaussians: a tensor ring induced prior for generative models," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [33] M. Wang, C. Zhang, Y. Pan, J. Xu, and Z. Xu, "Tensor ring restricted boltzmann machines," in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2019.
- [34] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 11, pp. 4747–4760, 2020.
- [35] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, "Nonlocal tensor-ring decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1348–1362, 2019.
- [36] W. He, N. Yokoya, L. Yuan, and Q. Zhao, "Remote sensing image reconstruction using tensor ring completion and total variation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 8998–9009, 2019.
- [37] Z. Long, C. Zhu, J. Liu, and Y. Liu, "Bayesian low rank tensor ring for image recovery," *IEEE Transactions on Image Processing*, vol. 30, pp. 3568–3580, 2021.
- [38] J. Liu, C. Zhu, and Y. Liu, "Smooth compact tensor ring regression," *IEEE Transactions on Knowledge and Data Engineering*, 2020.
- [39] F. Sedighin, A. Cichocki, and A.-H. Phan, "Adaptive rank selection for tensor ring decomposition," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 454–463, 2021.
- [40] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 9151–9158, 2019.
- [41] K. Batselier, "The trouble with tensor ring decompositions," *arXiv preprint arXiv:1811.03813*, 2018.
- [42] S. Handschuh, "Numerical methods in tensor networks," *PhD thesis, Faculty of mathematics and Informatics, University Leipzig*, 2015.
- [43] J. M. Landsberg, "Tensors: geometry and applications," *Representation theory*, vol. 381, no. 402, p. 3, 2012.
- [44] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42. Springer, 2012.
- [45] K. Ye and L.-H. Lim, "Tensor network ranks," *arXiv preprint arXiv:1801.02662*, 2018.
- [46] L.-H. Lim and P. Comon, "Nonnegative approximations of nonnegative tensors," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 432–441, 2009.
- [47] P. Comon, X. Luciani, and A. L. De Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 23, no. 7-8, pp. 393–405, 2009.
- [48] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 1260–1280, 2013.
- [49] D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, 2000.
- [50] Y. Yu, G. Zhou, N. Zheng, Y. Qiu, S. Xie, and Q. Zhao, "Graph-regularized non-negative tensor-ring decomposition for multiway representation learning," *IEEE Transactions on Cybernetics*, 2022.
- [51] X. Zhao, Y. Yu, G. Zhou, Q. Zhao, and W. Sun, "Fast hypergraph regularized nonnegative tensor ring decomposition based on low-rank approximation," *Applied Intelligence*, pp. 1–24, 2022.
- [52] C. Harris, M. Michalek, and E. C. Sertoz, "Computing images of polynomial maps," *Advances Comput. Math.*, vol. 45, pp. 2845–2865, 2019.
- [53] L. Yuan, Q. Zhao, and J. Cao, "Completion of high order tensor data with missing entries via tensor-train decomposition," in *International Conference on Neural Information Processing*, pp. 222–229, Springer, 2017.