SIAM J. MATRIX ANAL. APPL.
Vol. 41, No. 3, pp. 1416–1442

TENSOR RING DECOMPOSITION: OPTIMIZATION LANDSCAPE AND ONE-LOOP CONVERGENCE OF ALTERNATING LEAST SQUARES*

ZIANG CHEN†, YINGZHOU LI‡, AND JIANFENG LU§

Abstract. In this work, we study the tensor ring decomposition and its associated numerical algorithms. We establish a sharp transition of algorithmic difficulty of the optimization problem as the bond dimension increases: On one hand, we show the existence of spurious local minima for the optimization landscape even when the tensor ring format is much overparameterized, i.e., with bond dimension much larger than that of the true target tensor. On the other hand, when the bond dimension is further increased, we establish one-loop convergence for the alternating least squares algorithm for the tensor ring decomposition. The theoretical results are complemented by numerical experiments for both local minima and the one-loop convergence for the alternating least squares algorithm.

Key words. tensor ring decomposition, spurious local minima, alternating least squares, one-loop convergence

AMS subject classifications. 65D15, 33F05, 15A69

DOI. 10.1137/19M1270689

1. Introduction.

Tensors are ubiquitous especially for representing high-dimensional functions or operators. However, due to the curse of dimensionality, both the storage cost and the computational cost of vanilla tensors scale exponentially as the dimension increases.

To overcome the curse of dimensionality, tensor network decomposition has been widely used, in particular in the physics community, to represent specific families of high-dimensional tensors, with many fewer degrees of freedom than an arbitrary tensor. Perhaps the most famous tensor network is the *matrix product state* [1, 24]. The matrix product state forms the basis of the density matrix renormalization group algorithm [30, 31], which has been widely used in physics and chemistry literature and extremely successful for one-dimensional many-body physical systems (see, e.g., reviews [6, 11, 23, 32]). In fact, it has been shown rigorously that the ground state for a gapped one-dimensional many-body physical system can be efficiently represented by a matrix product state [2, 5, 12]. To go beyond and deal with more general physical systems, the matrix product state has also been extended to other tensor network decomposition formats, including PEPS [25] and MERA [27], just to name a few. See [18] for a recent review on tensor networks.

In the mathematics literature, the matrix product state is known as the *tensor train* (TT) format [19, 20], which is a special case of the hierarchical Tucker format [9,

*Received by the editors June 26, 2019; accepted for publication (in revised form) by J. Nie July 8, 2020; published electronically September 23, 2020.
https://doi.org/10.1137/19M1270689
Funding: The work of the authors was partially supported by the National Science Foundation grants OAC-1450280, DMS-1454939. The work of the first author was also partially supported by the elite undergraduate training program of the School of Mathematical Sciences in Peking University.
†School of Mathematical Sciences, Peking University, Beijing, China 100871 (chenziang19970214@pku.edu.cn).
‡Department of Mathematics, Duke University, Durham, NC 27708-0320 (yingzhou.li@duke.edu).
§Department of Mathematics, Department of Chemistry, and Department of Physics, Duke University, Durham, NC 27708-0320 (jianfeng@math.duke.edu).

10]. The study of tensor networks, including their algebraic and geometric structures and associated algorithms, has received increasing attention from the mathematics community as well (see, e.g., [7, 16, 21, 22, 33]). On the algorithmic side, most of the existing works focus on the TT format [4, 13, 19]. The convergence analysis for the construction and the compression algorithms of TT is established in [7, 22].

The *tensor ring* (TR) format extends the TT format to accommodate periodic boundary conditions and hence is suggested as the ansatz for periodic one-dimensional physical systems [26]. Unfortunately, the construction and the compression of TR turn out to be much more difficult than that of TT. Most algorithms working efficiently for TTs cannot be easily extended to TRs. The *alternating least squares* (ALS) algorithm is one exception, but it still relies on carefully designed sampling techniques and initial guesses [14] to be efficient. In addition to the difficulty in designing efficient algorithms, the inability to compute the exact minimal bond dimension TR decomposition was numerically demonstrated recently [3]. Although some success of TR is achieved in compressing tensors in practice [14, 28, 34, 35], TR decomposition and TR operations remain a challenging problem. The mathematical understanding of TR format and associated algorithms is still rather sparse. Motivated by such a gap, in this work, we analyze the TR decomposition. While TR is arguably the simplest tensor network beyond the TT format, our study may also shed some light on more complicated tensor network formats.

Let us also mention that there are other tensor decomposition formats besides tensor networks in the literature, such as tensor rank decomposition (often known as the CANDECOMP/PARAFAC (CP) decomposition) and Tucker decomposition. Many works have been devoted in designing efficient algorithms for finding near-optimal CP and Tucker decomposition (see the review article [15] and references therein).

1.1. Contribution. In this work, we analyze the optimization landscape of the TR decomposition and prove the existence of spurious local minima even if the TR format is overparameterized. More precisely, we propose a particular dth order tensor as the target tensor, which is of TR format with bond dimension $r+1$, and a spurious local minimum is identified in the space of TR with bond dimension r^{d-1}. Note that the bond dimension scales exponentially with d. Such a spurious local minimum casts trouble for an optimization problem associated with TR decomposition. Although the spurious local minimum identified might not be strict, we numerically validate that ALS in some sense cannot escape from a small neighborhood of the spurious local minimum.

Our second result establishes the one-loop convergence of the ALS algorithm for TR decomposition if we even further lift the bond dimension of the proposed TR space. More precisely, for any target dth order tensor of TR format with bond dimension r satisfying some full-rank conditions and starting from a random initial TR, ALS almost surely converges to the target tensor after one loop iteration, when the bond dimension of the proposed TR space is r^{d-1}.

Combining two results together, we establish a sharp transition between the triviality of ALS, i.e., the one-loop convergence, and the existence of spurious local minima. Up to some subtle differences, the results shown here for TR are similar to that for TT in [22]. Investigation of any of these subtle differences leads to the intrinsic difference between TT and TR, i.e., TT of fixed bond dimension is a closed set whereas TR of fixed bond dimension is not a closed set [16].

1.2. Organization. The rest of this paper is organized as follows. In section 2, we introduce some basics for the TR format. In section 3, we introduce TR decompo-

sition and analyze the optimization landscape. The existence of spurious local minima is established. In section 4, we introduce the ALS algorithm for TR decomposition and show the one-loop convergence. In section 5, we give some numerical validations of the theoretical results. The paper is concluded in section 6.

2. Preliminaries. In this section, we first introduce some tensor notation that will be used throughout the paper and then provide the precise definition of the TR format.

2.1. Tensor notation. While tensor is a powerful tool in many areas, the corresponding notation is somewhat complicated. In this section, we introduce some common notation in representing a tensor. A dth order tensor $\mathbf{X}$ is a d-dimensional array, i.e., $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, where $\vec{n} = (n_1, \ldots, n_d) \in \mathbb{N}_+^d$ is the size of the tensor. $\vec{n}$ is also called the external dimension of $\mathbf{X}$. Entries of $\mathbf{X}$ are denoted as $\mathbf{X}(x_1, \ldots, x_d)$, where $1 \le x_i \le n_i$ denotes the ith index of the tensor for $i = 1, 2, \ldots, d$. MATLAB colon (:) notation is powerful in representing contiguous entries of a tensor. For example, let $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a third order tensor: $\mathbf{X}(:, x_2, x_3)$ denotes a vector in $\mathbb{R}^{n_1}$ and $\mathbf{X}(:, x_2, :)$ denotes an $n_1 \times n_3$ matrix.

For two tensors of the same size $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, the inner product between $\mathbf{X}$ and $\mathbf{Y}$ is defined via

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{x_1=1}^{n_1} \sum_{x_2=1}^{n_2} \cdots \sum_{x_d=1}^{n_d} \mathbf{X}(x_1, \ldots, x_d) \mathbf{Y}(x_1, \ldots, x_d). \tag{2.1}$$

The Frobenius norm is used as a distance measure of $\mathbf{X}$, which is given as

$$\|\mathbf{X}\|_{\mathrm{F}} = \langle \mathbf{X}, \mathbf{X} \rangle^{1/2} = \left(\sum_{x_1=1}^{n_1} \sum_{x_2=1}^{n_2} \cdots \sum_{x_d=1}^{n_d} \mathbf{X}(x_1, \ldots, x_d)^2 \right)^{1/2}. \tag{2.2}$$

Suppose that $\mathbf{X} \in \mathbb{R}^{n_1 \times \cdots \times n_{d_1}}$ and $\mathbf{Y} \in \mathbb{R}^{m_1 \times \cdots \times m_{d_2}}$ are two tensors. The tensor product of $\mathbf{X}$ and $\mathbf{Y}$ is $\mathbf{X} \otimes \mathbf{Y} \in \mathbb{R}^{n_1 \times \cdots \times n_{d_1} \times m_1 \times \cdots \times m_{d_2}}$, with entries given by

$$(\mathbf{X} \otimes \mathbf{Y})(x_1, \ldots, x_{d_1}, y_1, \ldots, y_{d_2}) = \mathbf{X}(x_1, \ldots, x_{d_1}) \mathbf{Y}(y_1, \ldots, y_{d_2}), \tag{2.3}$$

where $x_i = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, d_1$, and $y_j = 1, 2, \ldots, m_j$, $j = 1, 2, \ldots, d_2$.

2.2. Tensor ring. A dth order tensor ring is a periodic consecutive product of d third order tensors, which can be viewed as a periodic version of a dth order tensor train. Before rigorously defining tensor ring, we first recall a periodic index i of periodicity d as $i = \mathrm{mod}(i-1, d) + 1$. Hence we have $i \in \{1, 2, \ldots, d\}$ for any integer i. A sequence $\vec{t} = (t_1, \ldots, t_d)$ is periodically indexed if t_i for $i < 1$ or $i > d$ is allowed and $t_i \equiv t_{\mathrm{mod}(i-1,d)+1}$, e.g., $t_0 \equiv t_d$, $t_{d+1} \equiv t_1$, etc. These periodic indices and periodically indexed sequences significantly reduce the redundancy of the notation in tensor ring. Define

$$\mathcal{U}_{\vec{r},\vec{n}}^d = \bigtimes_{i=1}^{d} \mathbb{R}^{r_i \times n_i \times r_{i+1}}, \tag{2.4}$$

where $\vec{n} = (n_1, \ldots, n_d)$ and $\vec{r} = (r_1, \ldots, r_d)$ denote the external and internal dimension, respectively. An element in $\mathcal{U}_{\vec{r},\vec{n}}^d$ is denoted as $\vec{\mathbf{u}} = \left(\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]}\right) \in \mathcal{U}_{\vec{r},\vec{n}}^d$, where $\mathbf{u}^{[i]} \in \mathbb{R}^{r_i \times n_i \times r_{i+1}}$ for $i = 1, 2, \ldots, d$. In order to shorten notation, we use two

notations for two unfolding tensors of $\mathbf{u}^{[i]}$, i.e., $\mathbf{u}^{[i]}_{k_1,k_2} = \mathbf{u}^{[i]}(k_1,:,k_2)$ is a vector in $\mathbb{R}^{n_i}$ and $\mathbf{u}^{[i]}(x_i) = \mathbf{u}^{[i]}(:,x_i,:)$ is a $r_i \times r_{i+1}$ matrix.

Let τ denote the mapping from $\mathcal{U}^d_{\vec{r},\vec{n}}$ to a dth order tensor of external dimension $\vec{n}$ as

$$\tau(\vec{\mathbf{u}}) = \sum_{k_1,\ldots,k_d} \mathbf{u}^{[1]}_{k_1,k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \otimes \cdots \otimes \mathbf{u}^{[d]}_{k_d,k_1}, \quad \vec{\mathbf{u}} \in \mathcal{U}^d_{\vec{r},\vec{n}}. \tag{2.5}$$

Throughout this paper, we abuse the mapping notation τ for any external and internal dimension $\vec{n}$ and $\vec{r}$ and order d. Elements of $\tau(\vec{\mathbf{u}})$ can be evaluated as

$$\tau(\vec{\mathbf{u}})(x_1,\ldots,x_d) = \operatorname{tr}\left(\mathbf{u}^{[1]}(x_1)\mathbf{u}^{[2]}(x_2)\cdots\mathbf{u}^{[d]}(x_d)\right) \tag{2.6}$$

for $x_i = 1,2,\ldots,n_i$, $i = 1,2,\ldots,d$.

A dth order tensor $\mathbf{T}$ of external dimension $\vec{n}$ has a TR format of internal dimension $\vec{r}$ if there exists a $\vec{\mathbf{u}} \in \mathcal{U}^d_{\vec{r},\vec{n}}$ such that $\mathbf{T} = \tau(\vec{\mathbf{u}})$. We denote the collection of such tensors as

$$\mathcal{R}^d_{\vec{r},\vec{n}} = \left\{\mathbf{T} \mid \mathbf{T} = \tau(\vec{\mathbf{u}}), \quad \vec{\mathbf{u}} \in \mathcal{U}^d_{\vec{r},\vec{n}}\right\}, \tag{2.7}$$

which includes all TRs with rank bounded by r. We call $\max_{1\le i\le d} r_i$ the *bond dimension* of the TR format. TR format is a special case of tensor networks, where the underlying network is a one-dimensional ring. More detailed discussions on tensor networks can be found in, e.g., [18].

Each element in $\mathcal{U}^d_{\vec{r},\vec{n}}$ corresponds to one tensor in $\mathcal{R}^d_{\vec{r},\vec{n}}$. However, the reverse is not true. Actually each element in $\mathcal{R}^d_{\vec{r},\vec{n}}$ corresponds to infinitely many elements in $\mathcal{U}^d_{\vec{r},\vec{n}}$. Let $\vec{\mathbf{u}}$ be an element in $\mathcal{U}^d_{\vec{r},\vec{n}}$ corresponding to $\mathbf{T} \in \mathcal{R}^d_{\vec{r},\vec{n}}$. Given any tuple of d invertible matrices, $\vec{A} = (A_1,\ldots,A_d)$, where $A_i \in \mathrm{GL}(r_i,\mathbb{R})$ is an $r_i \times r_i$ invertible matrix for $i = 1,2,\ldots,d$, we define $\theta_{\vec{A}}(\vec{\mathbf{u}}) = \vec{\mathbf{v}} = \left(\mathbf{v}^{[1]},\mathbf{v}^{[2]},\ldots,\mathbf{v}^{[d]}\right)$ via

$$\mathbf{v}^{[i]}(x_i) = A_i\mathbf{u}^{[i]}(x_i)A_{i+1}^{-1}, \quad x_i = 1,2,\ldots,n_i, \text{ and } i = 1,2,\ldots,d, \tag{2.8}$$

where periodic indexing is applied, i.e., $A_{d+1} = A_1$. It can be easily seen that $\tau(\theta_{\vec{A}}(\vec{\mathbf{u}})) = \tau(\vec{\mathbf{u}})$, i.e., $\theta_{\vec{A}}(\vec{\mathbf{u}})$ and $\vec{\mathbf{u}}$ correspond to the same dth order tensor. This is called *gauge freedom* or *gauge invariance*. The orbit

$$\mathcal{M}_{\vec{\mathbf{u}}} = \left\{\theta_{\vec{A}}(\vec{\mathbf{u}}) \mid \vec{A} \in \bigtimes_{i=1}^{d} \mathrm{GL}(r_i,\mathbb{R})\right\} \tag{2.9}$$

is called the manifold due to the gauge freedom.

3. Tensor ring decomposition and spurious local minima. A tensor ring provides an efficient representation of high order tensors, especially for those tensors with underlying physical geometry being a one-dimensional ring, and thus the periodicity becomes natural. However, finding such a TR representation of a given high order tensor is highly nontrivial in practice. In this section, we first cast the TR decomposition as a constrained optimization problem, which is widely used in the literature [14, 35]. Then an explicit spurious local minimum is constructed for the relaxed version of the constrained optimization. Such a spurious local minimum to some degree explains why TR decomposition is much more difficult than TT decomposition

and matrix factorization in practice. Recall that TT decomposition has one-loop convergence for exactly parameterized constraint set [19, 22] (the one-loop convergence for tensor ring will be discussed in section 4.2); and it is well known that matrix factorization including matrix eigenvalue decomposition [17] and low-rank factorization [8] does not have spurious local minima even in the underparameterized regime.

Let $\mathbf{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be the target dth order tensor. *TR decomposition* aims to find a $\vec{\mathbf{u}} \in \mathcal{U}^d_{\vec{r},\vec{n}}$ such that the distance between $\mathbf{T}$ and $\tau(\vec{\mathbf{u}})$ is minimized. If the tensor Frobenius norm is used as the distance, we can formulate the TR decomposition as the following constrained optimization problem:

$$\min_{\vec{\mathbf{u}} \in \mathcal{U}^d_{\vec{r},\vec{n}}} \frac{1}{2}\|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2. \tag{3.1}$$

Similar to TT decomposition and matrix factorization, the optimization (3.1) is a constrained nonconvex optimization problem.

Next, we would show that the TR decomposition optimization problem (3.1) has spurious local minima even in a relaxed constraint set. In the following, we will first construct an explicit spurious local minimum for a specific dth order tensor and then we remark that the specific tensor can be generalized to a set of tensors and the spurious local minimum exists for any tensor in the set.

To simplify the notation, we assume $\vec{r} = r$ and $\vec{n} = n = r^2 + 1$. Here and in the rest of the paper, we abuse notation $\vec{r} = r$ and $\vec{n} = n$, meaning that $\vec{r} = (r, \ldots, r)$ and $\vec{n} = (n, \ldots, n)$, respectively. The choice of $n = r^2 + 1$ comes from our specific construction of the target tensor: If we consider a TR format with the bond dimension being r, r^2 is a large enough external dimension since the dimension of the space spanned by $\mathbf{u}^{[i]}_{k_1,k_2}$, $k_1, k_2 = 1, 2, \ldots, r$, is smaller than or equal to r^2 for any $i = 1, 2, \ldots, d$. Then for constructing the target tensor, we add an additional orthogonal term, which enlarges the bond dimension from r to $r+1$ and the external dimension from r^2 to $n = r^2 + 1$. A discussion on $n \geq r^2 + 1$ cases is deferred to the end of this section.

Further, we denote the *lexicographical order* of a multituple, i.e.,

$$\pi(k_1, \ldots, k_d) := 1 + \sum_{i=1}^{d} (k_i - 1) r^{d-i} \tag{3.2}$$

for $1 \leq k_1, \ldots, k_d \leq r$. For example, for a 2-tuple, the lexicographical order function is $\pi(k_1, k_2) = (k_1 - 1)r + k_2$ for $1 \leq k_1, k_2 \leq r$.

The specific dth order tensor of bond dimension $r + 1$ is constructed as

$$\mathbf{T}_0 = \sum_{k_1,\ldots,k_d=1}^{r} \left(\bigotimes_{i=1}^{d} e_{\pi(k_{i+1},k_i)} \right) + \bigotimes_{i=1}^{d} e_n, \tag{3.3}$$

where e_j is an indicator vector of length n with one at jth entry and zero at every other entry. By the definition of the TR format, we can see that $\mathbf{T}_0 \in \mathcal{R}^d_{r+1,n}$.

Let us consider a relaxed optimization problem (3.1) for $\mathbf{T}_0$

$$\min_{\vec{\mathbf{u}} \in \mathcal{U}^d_{r^{d-1},n}} \frac{1}{2}\|\mathbf{T}_0 - \tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2, \tag{3.4}$$

where the constraint set $\mathcal{U}^d_{r+1,n}$ is relaxed to $\mathcal{U}^d_{r^{d-1},n}$, i.e., the bond dimension is increased from $r + 1$ to r^{d-1}, which is much larger. For simplicity of notation, we

denote the objective function as $f_0(\vec{\mathbf{u}}) = \frac{1}{2}\|\mathbf{T}_0 - \tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2$. Obviously, the objective function at a global minimum is 0.

We now take a specific point $\vec{\mathbf{u}}_0 \in \mathcal{U}^d_{r^{d-1},n}$ as

$$
\begin{aligned}
&\vec{\mathbf{u}}_0 = (\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \cdots, \mathbf{u}^{[d]}) \text{ with} \\
&\mathbf{u}^{[i]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \delta_{p_1 q_1} \cdots \delta_{p_{i-1} q_{i-1}} \delta_{p_{i+1} q_{i+1}} \cdots \delta_{p_{d-1} q_{d-1}} e_{\pi(p_i,q_i)} \text{ and} \\
&\mathbf{u}^{[d]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \delta_{p_2 q_1} \cdots \delta_{p_{d-1} q_{d-2}} e_{\pi(p_1,q_{d-1})},
\end{aligned} \tag{3.5}
$$

where $1 \le p_1, \ldots, p_{d-1}, q_1, \ldots, q_{d-1} \le r$ and $1 \le i \le d-1$. $\mathbf{T}_0$ and $\vec{\mathbf{u}}_0$ for $d = 3$ and $r = 2$ are illustrated explicitly in Appendix A.

We first prove a property which is essential in the discussions below.

PROPOSITION 3.1. *For $\vec{\mathbf{u}}_0$ defined in* (3.5)*, it holds that*

$$
\begin{aligned}
&\sum_{k_2,k_3,\ldots,k_{d-1}=1}^{m} \mathbf{u}^{[1]}_{\pi(p_1,\ldots,p_{d-1}),k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \\
&\qquad \otimes \cdots \otimes \mathbf{u}^{[d-2]}_{k_{d-2},k_{d-1}} \otimes \mathbf{u}^{[d-1]}_{k_{d-1},\pi(q_1,\ldots,q_{d-1})} = \bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)}
\end{aligned} \tag{3.6}
$$

for any $1 \le p_1, \ldots, p_{d-1}, q_1, \ldots, q_{d-1} \le r$.

Proof of Proposition 3.1. We can prove this proposition by the following direct computation:

$$
\begin{aligned}
&\sum_{k_2,k_3,\ldots,k_{d-1}=1}^{m} \mathbf{u}^{[1]}_{\pi(p_1,\ldots,p_{d-1}),k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \otimes \cdots \otimes \mathbf{u}^{[d-1]}_{k_{d-1},\pi(q_1,\ldots,q_{d-1})} \\
&= \sum_{k_3,\ldots,k_{d-1}=1}^{m} \sum_{p_1'=1}^{r} e_{\pi(p_1,p_1')} \otimes \mathbf{u}^{[2]}_{\pi(p_1',p_2,\ldots,p_d),k_d} \otimes \cdots \otimes \mathbf{u}^{[d-1]}_{k_{d-1},\pi(q_1,\ldots,q_{d-1})} \\
&= \cdots \\
&= \sum_{p_1',\ldots,p_{d-2}'=1}^{r} e_{\pi(p_1,p_1')} \otimes \cdots \otimes e_{\pi(p_{d-2},p_{d-2}')} \otimes \mathbf{u}^{[d-1]}_{\pi(p_1',\ldots,p_{d-2}',p_{d-1}),\pi(q_1,\ldots,q_{d-2},q_{d-1})} \\
&= \sum_{p_1',\ldots,p_{d-2}'=1}^{r} e_{\pi(p_1,p_1')} \otimes \cdots \otimes e_{\pi(p_{d-2},p_{d-2}')} \delta_{p_1' q_1} \cdots \delta_{p_{d-2}',q_{d-2}} e_{\pi(p_{d-1},q_{d-1})} \\
&= e_{\pi(p_1,q_1)} \otimes e_{\pi(p_2,q_2)} \otimes \cdots \otimes e_{\pi(p_{d-1},q_{d-1})}.
\end{aligned}
$$

∎

Theorem 3.3 below states that $\vec{\mathbf{u}}_0$ as in (3.5) is a local minimum for (3.4) with nonzero objective function, i.e., the problem (3.4) has a spurious local minimum. Both $\mathbf{T}_0$ in (3.3) and $\vec{\mathbf{u}}_0$ in (3.5) are abstract. The idea of $\vec{\mathbf{u}}_0$ is that we want to construct a TR format whose bond dimension is as large as possible and that has properties similar to (3.6), which is essential in the proof of Theorem 3.3. Then we come to $\vec{\mathbf{u}}_0$ and its corresponding bond dimension r^{d-1}. $\mathbf{T}_0$ is then constructed by adding an orthogonal term to $\tau(\vec{\mathbf{u}}_0)$.

We first provide the definition of local minimum used in this paper.

DEFINITION 3.2. *$\vec{\mathbf{u}}$ is a local minimum of a real-valued function $f(\cdot)$ if there exists a $\eta > 0$ such that for any $\|\vec{\mathbf{v}}\| < \eta$, $f(\vec{\mathbf{u}} + \vec{\mathbf{v}}) \ge f(\vec{\mathbf{u}})$.*

Due to the nonstrict inequality in Definition 3.2, the local minimum throughout this paper is also nonstrict. $\|\cdot\|$ can be any norm of a tensor ring since norms are equivalent in a finite-dimensional vector space. In the rest of this paper, the maximum norm of tensor ring $\|\cdot\|_{\max}$ is used for simplicity, i.e., $\|\vec{\mathbf{v}}\|_{\max}$ denotes the maximum of the absolute values of entries in $\vec{\mathbf{v}}$, i.e.,

$$\|\vec{\mathbf{v}}\|_{\max} := \max_{i,k,x,k'} \left|\mathbf{v}^{[i]}(k,x,k')\right|.$$

A tensor ring $\vec{\mathbf{u}}$ is a *spurious local minimum* of $f(\cdot)$ if $\vec{\mathbf{u}}$ is a local minimum and $f(\vec{\mathbf{u}}) > \min_{\vec{\mathbf{v}}} f(\vec{\mathbf{v}})$.

THEOREM 3.3. *For $d \geq 3$, $\vec{\mathbf{u}}_0 \in \mathcal{U}^d_{r^{d-1},n}$ as in* (3.5) *is a local minimum of problem* (3.4) *and $f_0(\vec{\mathbf{u}}_0) > 0$. Hence $\vec{\mathbf{u}}_0$ is a spurious local minimum of* (3.4).

The proof of Theorem 3.3 consists of two parts. First we demonstrate that $\tau(\vec{\mathbf{u}}_0)$ is the first summation part of $\mathbf{T}_0$ as in (3.3). Hence $f_0(\vec{\mathbf{u}}_0) = \frac{1}{2} > 0$. Next, we prove that there exists a small constant η such that for any $\|\vec{\mathbf{v}}\|_{\max} < \eta$, we have $f_0(\vec{\mathbf{u}}_0 + \vec{\mathbf{v}}) \geq f_0(\vec{\mathbf{u}}_0)$. Therefore, $\vec{\mathbf{u}}_0$ is a local minimum in the topology deduced by the norm $\|\vec{\mathbf{v}}\|_{\max}$. Since $\mathbf{T}_0$ can be exactly represented by a tensor ring in $\mathcal{U}^d_{r^{d-1},n}$, $\vec{\mathbf{u}}_0$ is a spurious local minimum.

Proof of Theorem 3.3. For simplicity, we drop all subscripts 0 of $\mathbf{T}_0$, $\vec{\mathbf{u}}_0$, and f_0. Further, we denote the bond dimension of $\mathcal{U}^d_{r^{d-1},n}$ as $m = r^{d-1}$. Combining (3.6) together with the contraction of $\mathbf{u}^{[d]}$, we have

(3.7)

$$\begin{aligned}
\tau(\vec{\mathbf{u}}) &= \sum_{k_1,\ldots,k_d=1}^{m} \mathbf{u}^{[1]}_{k_1,k_2} \otimes \mathbf{u}^{[2]}_{k_2,k_3} \otimes \cdots \otimes \mathbf{u}^{[d]}_{k_d k_1} \\
&= \sum_{p_1,\ldots,p_{d-1},q_1,\ldots,q_{d-1}=1}^{r} e_{\pi(p_1,q_1)} \otimes \cdots \otimes e_{\pi(p_{d-1},q_{d-1})} \otimes \mathbf{u}^{[d]}_{\pi(q_1,\ldots,q_{d-1}),\pi(p_1,\ldots,p_{d-1})} \\
&= \sum_{q_1,\ldots,q_{d-1},p_{d-1}=1}^{r} e_{\pi(q_2,q_1)} \otimes e_{\pi(q_3,q_2)} \otimes \cdots \otimes e_{\pi(q_{d-1},q_{d-2})} \otimes e_{\pi(p_{d-1},q_{d-1})} \otimes e_{\pi(q_1,p_{d-1})} \\
&= \mathbf{T} - \bigotimes_{i=1}^{d} e_n.
\end{aligned}$$

Therefore, the objective function of $\vec{\mathbf{u}}$ is strictly positive,

$$f(\vec{\mathbf{u}}) = \frac{1}{2}\left\|\bigotimes_{i=1}^{d} e_n\right\|_{\mathrm{F}}^2 = \frac{1}{2} > 0. \tag{3.8}$$

One more point about $\tau(\vec{\mathbf{u}})$, which will become important later, is that $\tau(\vec{\mathbf{u}})$ has empty outer most layer, i.e.,

$$\tau(\vec{\mathbf{u}})(x_1,\ldots,x_d) = 0 \quad \text{if } x_1 = n, \text{or } x_2 = n, \text{or } \ldots, \text{or } x_d = n. \tag{3.9}$$

Hence $\tau(\vec{\mathbf{u}})$ is orthogonal to $\bigotimes_{i=1}^{d} e_n = \mathbf{T} - \tau(\vec{\mathbf{u}})$.

Next, we investigate the property of the neighborhood of $\vec{\mathbf{u}}$. For any point $\vec{\mathbf{v}} \in \mathcal{U}_{m,n}^d$, we have

$$
\begin{aligned}
f(\vec{\mathbf{u}}+\vec{\mathbf{v}}) - f(\vec{\mathbf{u}}) &= \frac{1}{2}\|\mathbf{T}-\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})\|_{\mathrm{F}}^2 - \frac{1}{2}\|\mathbf{T}-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 \\
&= \frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 - \langle \mathbf{T}-\tau(\vec{\mathbf{u}}), \tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\rangle \\
&= \frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 - \left\langle \bigotimes_{i=1}^{d} e_n, \tau(\vec{\mathbf{u}}+\vec{\mathbf{v}}) \right\rangle \\
&= \frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 - \sum_{k_1,\ldots,k_d=1}^{m} \prod_{i=1}^{d} \mathbf{v}^{[i]}(k_i, n, k_{i+1}) \\
&= \frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^3\right),
\end{aligned}
\tag{3.10}
$$

where the second equality is due to the definition of tensor Frobenius norm as (2.2), the third equality adopts the result in (3.7) and orthogonality between $\tau(\vec{\mathbf{u}})$ and $\mathbf{T}-\tau(\vec{\mathbf{u}})$, the fourth equality comes from the direct evaluation of the inner product, and $\|\vec{\mathbf{v}}\|_{\max,n}$ is defined as

$$
\|\vec{\mathbf{v}}\|_{\max,n} = \max_{j,k,k'} \left\{ \left| \mathbf{v}^{[j]}(k,n,k') \right| \right\} \le \|\vec{\mathbf{v}}\|_{\max}.
$$

In order to show that $\vec{\mathbf{u}}$ is a local minimum of f, we need to show that the second order term of $\frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2$ has a positive coefficient. Denote $\mathbf{V}$ as the difference between dth order tensor $\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})$ and $\tau(\vec{\mathbf{u}})$. For any fixed p_i, q_i, $i = 1, 2, \ldots, d-1$, the inner product of $\mathbf{V}$ with $(\bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)}) \otimes e_n$ yields

$$
\begin{aligned}
&\left\langle \mathbf{V}, \left(\bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)} \right) \otimes e_n \right\rangle \\
&= \left\langle \tau(\vec{\mathbf{u}}+\vec{\mathbf{v}}) - \tau(\vec{\mathbf{u}}), \left(\bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)} \right) \otimes e_n \right\rangle \\
&= \sum_{k_1,\ldots,k_d=1}^{m} \left\langle \bigotimes_{i=1}^{d} \left(\mathbf{u}_{k_i,k_{i+1}}^{[i]} + \mathbf{v}_{k_i,k_{i+1}}^{[i]} \right), \left(\bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)} \right) \otimes e_n \right\rangle \\
&= \sum_{k_1,\ldots,k_d=1}^{m} \mathbf{v}^{[d]}(k_d, n, k_1) \cdot \prod_{i=1}^{d-1} \left(\mathbf{u}^{[i]}(k_i, \pi(p_i,q_i), k_{i+1}) + \mathbf{v}^{[i]}(k_i, \pi(p_i,q_i), k_{i+1}) \right) \\
&= \sum_{k_1,\ldots,k_d=1}^{m} \mathbf{v}^{[d]}(k_d, n, k_1) \cdot \prod_{i=1}^{d-1} \mathbf{u}^{[i]}(k_i, \pi(p_i,q_i), k_{i+1}) + O\left(\|\vec{\mathbf{v}}\|_{\max,n} \cdot \|\vec{\mathbf{v}}\|_{\max} \right) \\
&=: S + O\left(\|\vec{\mathbf{v}}\|_{\max,n} \cdot \|\vec{\mathbf{v}}\|_{\max} \right),
\end{aligned}
\tag{3.11}
$$

where the third equality is due to (3.9) and the last equality defines S. The definition (3.5) implies that for any $i = 1, 2, \ldots, d-1$, $\mathbf{u}^{[i]}(k_i, \pi(p_i,q_i), k_{i+1}) \neq 0$ if and only if all coordinates of k_i and k_{i+1} are the same , except the ith one which is p_i and q_i for

k_i and k_{i+1}, respectively. Therefore, we know that

$$\prod_{i=1}^{d-1} \mathbf{u}^{[i]}(k_i, \pi(p_i, q_i), k_{i+1}) \neq 0$$

if and only if $k_i = \pi(q_1, \ldots, q_{i-1}, p_i, \ldots, p_{d-1})$ for all $i = 1, \ldots, d$. Hence the S part can be rewritten as

$$S = \mathbf{v}^{[d]}(\pi(q_1, \ldots, q_{d-1}), n, \pi(p_1, \ldots, p_{d-1})) = O\left(\|\vec{\mathbf{v}}\|_{\max,n}\right). \tag{3.12}$$

Substituting (3.12) into (3.11), we obtain a lower bound on the square of an element of $\mathbf{V}$:

$$\begin{aligned} &\mathbf{V}(\pi(p_1, q_1), \ldots, \pi(p_{d-1}, q_{d-1}), n)^2 \\ &= \left\langle \mathbf{V}, \left(\bigotimes_{i=1}^{d-1} e_{\pi(p_i,q_i)} \right) \otimes e_n \right\rangle^2 = S^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right) \\ &= \left|\mathbf{v}^{[d]}(\pi(q_1, \ldots, q_{d-1}), n, \pi(p_1, \ldots, p_{d-1}))\right|^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right). \end{aligned} \tag{3.13}$$

Hence, we have

$$\begin{aligned} &\sum_{x_1,\ldots,x_{d-1}=1}^{n-1} \mathbf{V}(x_1, \ldots, x_{d-1}, n)^2 \\ &= \sum_{p_1,\ldots,p_{d-1},q_1,\ldots,q_{d-1}=1}^{r} \mathbf{V}(\pi(p_1, q_1), \ldots, \pi(p_{d-1}, q_{d-1}), n)^2 \\ &= \sum_{p_1,\ldots,p_{d-1},q_1,\ldots,q_{d-1}=1}^{r} \left|\mathbf{v}^{[d]}(\pi(q_1, \ldots, q_{d-1}), n, \pi(p_1, \ldots, p_{d-1}))\right|^2 \\ &\qquad + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right) \\ &= \sum_{k_1,k_d=1}^{m} \left|\mathbf{v}^{[d]}(k_d, n, k_1)\right|^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right). \end{aligned} \tag{3.14}$$

In the derivation from (3.11) to (3.14), the only step that relies on index d is (3.12) and all other steps can be generalized to other index j with the notion of periodic index. Notice that with periodic indexing, we have

$$\begin{aligned} &\prod_{i=j+1}^{j-1} \mathbf{u}^{[i]}(k_i, \pi(p_i, q_i), k_{i+1}) \\ &= \prod_{i=j+1}^{j-1} \mathbf{u}^{[i]}(\pi(s_1^i, \ldots, s_{d-1}^i), \pi(p_i, q_i), \pi(s_1^{i+1}, \ldots, s_{d-1}^{i+1})) \\ &= \delta_{s_2^d, s_1^1} \cdots \delta_{s_{d-1}^d, s_{d-2}^1} \delta_{s_1^d, p_d} \delta_{s_{d-1}^1, q_d} \\ &\qquad \times \prod_{\substack{i=j+1 \\ i \neq d}}^{j-1} \delta_{s_1^i, s_1^{i+1}} \cdots \delta_{s_{i-1}^i, s_{i-1}^{i+1}} \delta_{s_{i+1}^i, s_{i+1}^{i+1}} \cdots \delta_{s_{d-1}^i, s_{d-1}^{i+1}} \delta_{s_i^i, p_i} \delta_{s_i^{i+1}, q_i}, \end{aligned} \tag{3.15}$$

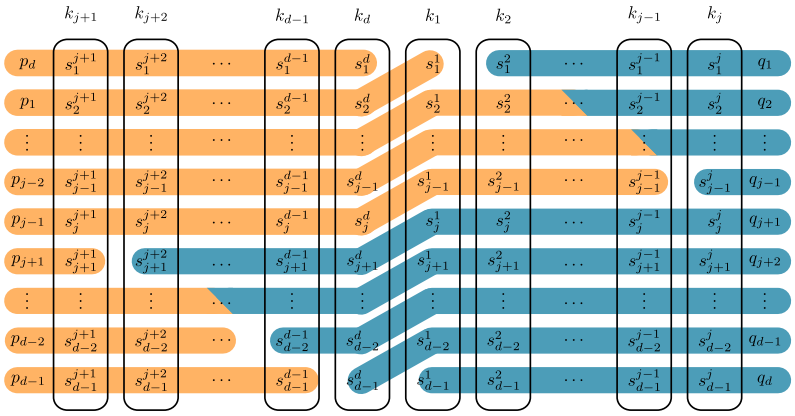

FIG. 1. *Derivation of generalized S term. Orange lines and blue lines indicate equivalent relation of p and q, respectively.*

where $\delta_{\cdot,\cdot}$ is the Kronecker delta, $k_i = \pi(s_1^i, \ldots, s_{d-1}^i)$, and any s is an integer in the interval $[1, r]$ for $j \in \{1, \ldots, d-1\}$. Through a careful derivation, (3.15) is nonzero if and only if $k_j = \pi(q_1, \ldots, q_{j-1}, q_{j+1}, \ldots, q_d)$, $k_{j+1} = \pi(p_d, p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_{d-1})$, and other k_i is given in Figure 1.

Similar to (3.13), we can obtain a lower bound on the square of an element of $\mathbf{V}$

$$\begin{aligned}
&\mathbf{V}(\pi(p_1, q_1), \ldots, \pi(p_{j-1}, q_{j-1}), n, \pi(p_{j+1}, q_{j+1}), \ldots, \pi(p_d, q_d))^2 \\
&= \left|\mathbf{v}^{[j]}(\pi(q_1, \ldots, q_{j-1}, q_{j+1}, \ldots, q_d), n, \pi(p_d, p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_{d-1}))\right|^2 \\
&\quad + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right).
\end{aligned} \tag{3.16}$$

Hence, similarly, we have

$$\begin{aligned}
&\sum_{x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_d=1}^{n-1} \mathbf{V}(x_1, \ldots, x_{j-1}, n, x_{j+1}, \ldots, x_d)^2 \\
&= \sum_{k_j,k_{j+1}=1}^{m} \left|\mathbf{v}^{[j]}(k_j, n, k_{j+1})\right|^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right)
\end{aligned} \tag{3.17}$$

for any $j = 1, \ldots, d$.

It follows from (3.17) that

$$\begin{aligned}
\|\mathbf{V}\|_{\mathrm{F}}^2 &\geq \sum_{j=1}^{d} \sum_{x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_d=1}^{n-1} \mathbf{V}(x_1, \ldots, x_{j-1}, n, x_{j+1}, \ldots, x_d)^2 \\
&= \sum_{j=1}^{d} \sum_{k_j,k_{j+1}=1}^{m} \left|\mathbf{v}^{[j]}(k_j, n, k_{j+1})\right|^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^2 \cdot \|\vec{\mathbf{v}}\|_{\max}\right) \\
&\geq (1 + O(\|\vec{\mathbf{v}}\|_{\max})) \cdot \|\vec{\mathbf{v}}\|_{\max,n}^2.
\end{aligned} \tag{3.18}$$

Substituting (3.18) into (3.10), when η is sufficiently small, we have

$$
\begin{aligned}
f(\vec{\mathbf{u}}+\vec{\mathbf{v}})-f(\vec{\mathbf{u}}) &= \frac{1}{2}\|\tau(\vec{\mathbf{u}}+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2 - \sum_{k_1,\ldots,k_d=1}^{m}\prod_{i=1}^{d}\mathbf{v}^{[i]}(k_i,n,k_{i+1}) \\
&= \frac{1}{2}\|\mathbf{V}\|_{\mathrm{F}}^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^3\right) \\
&\geq \frac{1+O\left(\|\vec{\mathbf{v}}\|_{\max}\right)}{2}\cdot\|\vec{\mathbf{v}}\|_{\max,n}^2 + O\left(\|\vec{\mathbf{v}}\|_{\max,n}^3\right) \geq 0
\end{aligned} \tag{3.19}
$$

for sufficiently small $\|\vec{\mathbf{v}}\|_{\max}$. Since the minimum value of (3.4) is zero, $\vec{\mathbf{u}}$ is a spurious local minimum. ∎

Theorem 3.3 states that, for a particular dth order tensor as (3.3), when the target tensor is of TR format with bond dimension $r+1$, there exists at least one spurious local minima even when the restricted bond dimension of the problem (3.1) is r^{d-1}, much larger than $r+1$. Through a more detailed analysis, we can locally illustrate that when $f_0(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=f_0(\vec{\mathbf{u}}_0)$ the constructed tensors are identical, i.e., $\tau(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=\tau(\vec{\mathbf{u}}_0)$.

PROPOSITION 3.4. *Under the same assumptions as in Theorem* 3.3*, for a positive constant η which is small enough, any point $\vec{\mathbf{v}}\in\mathcal{U}_{m,n}^d$ with $\|\vec{\mathbf{v}}\|_{\max}<\eta$ satisfies $f_0(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=f_0(\vec{\mathbf{u}}_0)$ if and only if $\tau(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=\tau(\vec{\mathbf{u}}_0)$, i.e., $\vec{\mathbf{u}}_0$ and $\vec{\mathbf{u}}_0+\vec{\mathbf{v}}$ are TR formats of the same dth order tensor.*

Proof of Proposition 3.4. It can be easily seen that $\tau(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=\tau(\vec{\mathbf{u}}_0)$ implies $f_0(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=f_0(\vec{\mathbf{u}}_0)$.

Now we consider the reverse direction. Suppose that η is small enough. For any $\vec{\mathbf{v}}$ such that $\|\vec{\mathbf{v}}\|_{\max}<\eta$ and $f_0(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=f_0(\vec{\mathbf{u}}_0)$, the equality in (3.19) implies that $\|\vec{\mathbf{v}}\|_{\max,n}=0$, i.e.,

$$
\mathbf{v}^{[j]}(k_j,n,k_{j+1})=0, \quad \forall\, 1\leq j\leq d,\ 1\leq k_j,k_{j+1}\leq m, \tag{3.20}
$$

which leads to

$$
f_0(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})-f_0(\vec{\mathbf{u}}_0)=\frac{1}{2}\|\mathbf{V}\|_{\mathrm{F}}^2=\frac{1}{2}\|\tau(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})-\tau(\vec{\mathbf{u}}_0)\|_{\mathrm{F}}^2=0. \tag{3.21}
$$

Thus, we can conclude that $\tau(\vec{\mathbf{u}}_0+\vec{\mathbf{v}})=\tau(\vec{\mathbf{u}}_0)$. ∎

We remark that the same results as in Theorem 3.3 and Proposition 3.4 hold for $n\geq r^2+1$ if vectors e_k in the definition of $\mathbf{T}_0$ and $\vec{\mathbf{u}}_0$ are extended to be of length n with zero padding. Furthermore, Theorem 3.3 and Proposition 3.4 also hold for a generalized version of $\mathbf{T}_0$ and the associated $\vec{\mathbf{u}}_0$, where $\mathbf{T}_0$ is defined as

$$
\mathbf{T}_0=\sum_{k_1,\ldots,k_d=1}^{r}\left(\bigotimes_{i=1}^{d}\lambda_{\pi(k_{i+1},k_i)}^i g_{\pi(k_{i+1},k_i)}^i\right)+\bigotimes_{i=1}^{d}\lambda_{r^2+1}^i g_{r^2+1}^i \tag{3.22}
$$

with $\lambda_k^i>0$ and $\langle g_{k_1}^i, g_{k_2}^i\rangle=\delta_{k_1k_2}$ for any $1\leq k_1,k_2\leq r^2+1$, and $1\leq i\leq d$, and $\vec{\mathbf{u}}_0$ is given via replacing e_k in the definition of $\mathbf{u}^{[i]}$ in (3.5) by $\lambda_k^i g_k^i$ for $1\leq i\leq d$ and $1\leq k\leq r^2$. This is due to the orthogonal rotation invariant property of Frobenius norm and an observation that a scaling will not break the proof of Theorem 3.3 and Proposition 3.4 as long as $\eta>0$ is small enough.

Another remark is on the difficulty of TR decomposition. Theorem 3.3 and Proposition 3.4 show that TR decomposition is more difficult than TT decomposition, since TT decomposition has one-loop convergence [13, 22] if the restricted bond dimension equals the underlying bond dimension of the target TT format whereas local minima remain in TR decomposition even if the restricted bond dimension increases exponentially as in the size of the ring. When d is large, we may not expect a good landscape of TR decomposition even if it is very much overparameterized.

Finally, we also want to point out that if $\vec{\mathbf{u}} + \vec{\mathbf{v}} \in \mathcal{M}_{\vec{\mathbf{u}}}$, then $\tau(\vec{\mathbf{u}} + \vec{\mathbf{v}}) = \tau(\vec{\mathbf{u}})$, while it is not clear whether the reverse holds in general. This difficulty comes from the ring structure of TR format. In fact, for TT format, it can be proved that the gauge invariant and the orbit with the same whole tensors are equivalent if the TT format is full-rank [22].

4. Alternating least squares algorithm for tensor ring decomposition. In this section, we recall the ALS algorithm for computing the TR decomposition [14, 29] of a given dth order tensor. Some basic descriptions and properties are in section 4.1. ALS is a strictly monotonically descent algorithm unless a stationary point is found. In section 4.2, we establish the convergence analysis of ALS when the bond dimension is sufficiently higher than that of the target tensor. As will be shown, in such cases, ALS converges in one outer iteration, which is known as the *one-loop convergence*. Recall that in Theorem 3.3 we prove the existence of spurious local minima in the overparameterized case. The bond dimension required for the one-loop convergence is larger than that in Theorem 3.3 (much larger than that of the true tensor).

4.1. Algorithm. Note that the objective function (3.1) is not convex due to the multilinear mapping τ. Nevertheless, if we fix all but one of the third order tensors in $\vec{\mathbf{u}}$, e.g., $\mathbf{u}^{[i]}$, and consider the suboptimization problem with respect to $\mathbf{u}^{[i]}$,

$$\min_{\mathbf{u}^{[i]}} \frac{1}{2} \left\| \mathbf{T} - \tau(\mathbf{u}^{[1]}, \ldots, \mathbf{u}^{[i-1]}, \mathbf{u}^{[i]}, \mathbf{u}^{[i+1]}, \ldots, \mathbf{u}^{[d]}) \right\|_{\mathrm{F}}^2, \tag{4.1}$$

this gives a quadratic least squares problem in $\mathbf{u}^{[i]}$ and hence can be solved explicitly and efficiently.

In order to make the least squares formulation more explicit, we first define a sequence of matrices $\{B_i\}_{i=1}^d$ of unfolded $\mathbf{T}$ and two unfolding operators $\alpha(\cdot)$ and $\gamma(\cdot)$. In the following, the function π always denotes the lexicographical order. Given $1 \le i \le d$, B_i is an unfolding of $\mathbf{T}$ as a matrix of size $(\prod_{j \ne i} n_j) \times n_i$, i.e.,

$$B_i(\pi(x_{i+1}, \ldots, x_{i-1}), x_i) = B_i(\pi(x_{i+1}, \ldots, x_d, x_1, \ldots, x_{i-1}), x_i) := \mathbf{T}(x_1, \ldots, x_d), \tag{4.2}$$

where the periodic index convention is used and $1 \le x_j \le n_j$ for $j = 1, \ldots, d$. The unfolding operator $\alpha(\cdot)$ unfolds a sequence of third order tensors into a matrix as

$$A_i = \alpha(\mathbf{u}^{[i+1]}, \ldots, \mathbf{u}^{[i-1]}) \in \mathbb{R}^{(\prod_{j \ne i} n_j) \times m^2} \tag{4.3}$$

with entries being

$$A_i(\pi(x_{i+1}, \ldots, x_{i-1}), \pi(k_{i+1}, k_i)) = \sum_{k_{i+2}, \ldots, k_{i-1} = 1}^{m} \mathbf{u}^{[i+1]}(k_{i+1}, x_{i+1}, k_{i+2}) \cdots \mathbf{u}^{[i-1]}(k_{i-1}, x_{i-1}, k_i), \tag{4.4}$$

where m is the bond dimension, $1 \le x_j \le n_j$ for $j = 1, \ldots, d$, and $1 \le k_i, k_{i+1} \le m$. The unfolding operator $\gamma(\cdot)$ unfolds a third order tensor into a matrix with compatible indices of A_i and B_i, i.e.,

$$X_i = \gamma(\mathbf{u}^{[i]}) \in \mathbb{R}^{m^2 \times n_i} \tag{4.5}$$

with entries being

$$X_i(\pi(k_{i+1}, k_i), x_i) = \mathbf{u}^{[i]}(k_i, x_i, k_{i+1}), \tag{4.6}$$

where indices are in the same ranges as before. Since $\gamma(\cdot)$ acts on a single third tensor, the unfolding operator is invertible and the invert operator $\gamma^{-1}(\cdot)$ will be used in the later content.

With these unfolded matrices, (4.1) then can be rewritten as a standard least squares problem

$$\min_X \frac{1}{2}\|A_i X - B_i\|_{\mathrm{F}}^2, \tag{4.7}$$

where A_i and B_i are defined as (4.3) and (4.2), respectively, and the minimizer of (4.1) can be achieved from $\gamma^{-1}(X_i)$ for X_i being the minimizer of (4.7).

To simplify the notation, we denote the objective function in (4.1) as $f_{\mathbf{T}}$. A popular numerical approach for solving (3.1) is to solve (4.1) for each of $\mathbf{u}^{[1]}$ to $\mathbf{u}^{[d]}$ in a cyclic way. The corresponding algorithm is known as the ALS algorithm for TR decomposition. The pseudocode of ALS for TR decomposition is presented in Algorithm 4.1, with the subscript ℓ indicating the iteration number.

Algorithm 4.1 ALS for TR decomposition.

Input: Target dth order tensor $\mathbf{T}$ and initial tensor ring $\vec{\mathbf{u}}_0$.
Output: Converged tensor ring $\vec{\mathbf{u}}$.

1: **for** $\ell = 0, 1, 2, \ldots$ **do**
2: **for** $i = 1, 2, \ldots, d$ **do**
3: Perform an ALS microstep:

$$\mathbf{u}_{\ell+1}^{[i]} = \arg\min_{\mathbf{v}} \frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{u}_{\ell+1}^{[1]}, \ldots, \mathbf{u}_{\ell+1}^{[i-1]}, \mathbf{v}, \mathbf{u}_{\ell}^{[i+1]}, \ldots, \mathbf{u}_{\ell}^{[d]})\right\|_{\mathrm{F}}^2$$

4: **end for**
5: **end for**

The next lemma states that the objective function of $\vec{\mathbf{u}}_\ell$ in Algorithm 4.1 decreases monotonically until a stationary point is achieved.

LEMMA 4.1. *Let $\{\vec{\mathbf{u}}_\ell\}_{\ell=0}^{\infty}$ be a sequence of tensor ring generated by Algorithm* 4.1. *For any ℓ, if $\vec{\mathbf{u}}_\ell$ is not a stationary point of $f_{\mathbf{T}}$, i.e., $\nabla f_{\mathbf{T}}(\vec{\mathbf{u}}_\ell) \neq 0$, then $f_{\mathbf{T}}(\vec{\mathbf{u}}_{\ell+1}) < f_{\mathbf{T}}(\vec{\mathbf{u}}_\ell)$.*

Proof. Since $\nabla f_{\mathbf{T}}(\vec{\mathbf{u}}_\ell) \neq 0$, there exists a set of indices $\mathcal{J}$ such that $\nabla_{\mathbf{u}^{[j]}} f_{\mathbf{T}}(\vec{\mathbf{u}}_\ell) \neq 0$ for $j \in \mathcal{J}$. Let i be the smallest index in $\mathcal{J}$. The ith microstep solves a least squares problem with nonzero gradient. Hence the objective value strictly decreases. For all later $(i+1)$th to dth microsteps the objective value is nonincreasing due to the nature of least square solutions. Therefore, we conclude that $f_{\mathbf{T}}(\vec{\mathbf{u}}_{\ell+1}) < f_{\mathbf{T}}(\vec{\mathbf{u}}_\ell)$ if $\vec{\mathbf{u}}_\ell$ is not a stationary point of $f_{\mathbf{T}}$. ∎

We would hope to get a stronger result than Lemma 4.1 showing that ALS converges to a stationary point. However, while ALS has the monotonic descent, the convergence to a stationary point is still open due to the following difficulties.

First, the boundedness of $\{\vec{\mathbf{u}}_\ell\}_{\ell=0}^{\infty}$ cannot be ensured. Thus, it is hard to say that $\{\vec{\mathbf{u}}_\ell\}_{\ell=0}^{\infty}$ has an accumulation point. To bypass the unboundedness, we can instead consider the accumulation point of the sequence of manifolds $\mathcal{M}_{\vec{\mathbf{u}}_\ell}$. But it is also not clear that whether there exists a sequence of gauge $\{\vec{A}_\ell\}_{\ell=0}^{\infty}$ such that $\{\theta_{\vec{A}_\ell}(\vec{\mathbf{u}}_\ell)\}_{\ell=0}^{\infty}$ is bounded. If we consider the sequence of the whole tensor, it can be proved that $\{\tau(\vec{\mathbf{u}}_\ell)\}_{\ell=0}^{\infty}$ is bounded. An accumulation point of $\{\tau(\vec{\mathbf{u}}_\ell)\}_{\ell=0}^{\infty}$ may not be located in $\mathcal{R}_{\vec{r},\vec{n}}^{d}$ though since it is known that the set of tensors in TR format with a fixed bond dimension is not closed [16] due to the underlying ring structure.

Second, even if $\{\vec{\mathbf{u}}_\ell\}_{\ell=0}^{\infty}$ or $\{\theta_{\vec{A}_\ell}(\vec{\mathbf{u}}_\ell)\}_{\ell=0}^{\infty}$ has an accumulation point $\vec{\mathbf{u}}$, it is hard to say that the rank of $\vec{\mathbf{u}}$ is equal to the rank of $\vec{\mathbf{u}}_\ell$ when ℓ is large enough. If equality does not hold, some continuity properties do not hold at $\vec{\mathbf{u}}$, which leads to difficulties when analyzing the limiting behavior of $\{\vec{\mathbf{u}}_\ell\}_{\ell=0}^{\infty}$ or $\{\theta_{\vec{A}_\ell}(\vec{\mathbf{u}}_\ell)\}_{\ell=0}^{\infty}$.

In fact, if the boundedness and the equality of rank are assumed, convergence to the stationary point of ALS can be ensured, similar to the analysis in [7]. It is an interesting future research direction to establish these conditions for the TR decomposition.

4.2. One-loop convergence. Even though the general convergence result without the assumptions above is still open, we can prove one-loop convergence of the Algorithm 4.1 in an extremely overparameterized case, which means that Algorithm 4.1 converges in d microsteps. In this section, we will prove the one-loop convergence of Algorithm 4.1 under mild assumptions on the target tensor.

Let us consider the case that the target tensor admits a TR decomposition as

$$\mathbf{T} = \tau(\vec{\mathbf{w}}) = \sum_{k_1,\dots,k_d=1}^{r} \mathbf{w}_{k_1,k_2}^{[1]} \otimes \mathbf{w}_{k_2,k_3}^{[2]} \otimes \cdots \otimes \mathbf{w}_{k_d,k_1}^{[d]} \in \mathcal{R}_{r,\vec{n}}^{d} \tag{4.8}$$

with bond dimension r and TR components $\vec{\mathbf{w}} = \left(\mathbf{w}^{[1]}, \mathbf{w}^{[2]}, \dots, \mathbf{w}^{[d]}\right) \in \mathcal{U}_{r,\vec{n}}^{d}$. From Algorithm 4.1, we notice that if both the target tensor and the initial tensor ring are multiplied by an orthogonal matrix on an external dimension, all iterators of the algorithm remain the same up to the orthogonal transformation on the corresponding external dimension. Hence we claim Algorithm 4.1 is invariant under orthogonal transformations on external dimensions and so is the related analysis.

We assume that the external dimension $\vec{n}$ is large enough with $n_i \geq r^2$ for $i = 1, \dots, d$ and we consider the problem $\min \frac{1}{2}\|\mathbf{T} - \tau(\vec{\mathbf{u}})\|_{\mathrm{F}}^2$, where $\vec{\mathbf{u}} \in \mathcal{U}_{m,\vec{n}}^{d}$ with bond dimension $m = r^{d-1}$. Since the dimension of span $\left\{\mathbf{w}_{k_1,k_2}^{[i]} : 1 \leq k_1, k_2 \leq r\right\}$ is bounded above by r^2 and the invariant property of Algorithm 4.1, without loss of generality, we assume that $\vec{\mathbf{w}}$ is located in a small subspace of $\mathcal{U}_{r,\vec{n}}^{d}$:

$$\mathcal{W}_{r,\vec{n}}^{d} := \left\{\vec{\mathbf{w}} \in \mathcal{U}_{r,\vec{n}}^{d} \mid \mathbf{w}_{k_1,k_2}^{[i]}(s) = 0 \quad \forall\ 1 \leq i \leq d,\ 1 \leq k_1, k_2 \leq r,\ s \geq r^2 + 1\right\}, \tag{4.9}$$

which is because that orthogonal transformations do not change the Frobenius norm and implies that $\mathbf{T}(x_1, \dots, x_d) = 0$ as long as one of $x_1, \dots, x_d$ is greater than or equal to $r^2 + 1$. Since the Frobenius norm is invariant under the orthogonal rotations in each outer dimension of the target tensor, $\mathcal{W}_{r,\vec{n}}^{d}$ can be generalized to $\mathcal{U}_{r,\vec{n}}^{d}$ under orthogonal rotations. Our convergence result is the following theorem, in which μ

denotes the proper Lebesgue measure. The theorem guarantees one-loop convergence for a typical target tensor and initial guess, when the bond dimension is large enough.

THEOREM 4.2. *There exists* $\Omega_1 \subseteq \mathcal{W}^d_{r,\vec{n}}$ *with* $\mu(\Omega_1) = 0$, *such that for any* $\vec{\mathbf{w}} \in \mathcal{W}^d_{r,\vec{n}} \setminus \Omega_1$ *and* $\mathbf{T} = \tau(\vec{\mathbf{w}})$, *there exists* $\Omega_2 \subseteq \mathcal{U}^d_{m,\vec{n}}$ *with* $\mu(\Omega_2) = 0$, *such that Algorithm* 4.1 *converges to the global minimum in* d *microsteps as long as the initial point* $\vec{\mathbf{u}}_0 \notin \Omega_2$.

Remark 4.3. In the proof of Theorem 4.2, the technical part is to characterize two zero-measure sets Ω_1 and Ω_2. Once these two sets are settled, the remaining proof is straightforward. We postpone the proof of Theorem 4.2 toward the end of this section. In the following lemmas, we prove that a set is of zero measure through establishing the equivalence between this set and the set of roots of a polynomial, since the Lebesgue measure of the root set of a nonzero polynomial is zero.

We first focus on the characterization Ω_2, which will be defined in Lemma 4.6; Ω_1 will be characterized along the analysis and will be defined in Lemma 4.8. Both Ω_1 and Ω_2 are constructed somewhat implicitly. Intuitively, Ω_2 is the set of $\vec{\mathbf{u}}_0$ which leads to some degeneracy in the first d microsteps of ALS. We show that Ω_2 is zero-measure following the argument sketched in Remark 4.3, which would require some assumptions on $\mathbf{T}$. Then we denote Ω_1 as the set of $\vec{\mathbf{w}}$ such that at least one of those assumptions is violated and prove that Ω_1 is also of zero-measure.

For simplicity, in the rest of this section we denote $\vec{\mathbf{u}} = \vec{\mathbf{u}}_0$ and $\vec{\mathbf{v}} = \vec{\mathbf{u}}_1$ as the initial tensor vector and the tensor vector after one macrostep (one loop). Each microstep of Algorithm 4.1 solves a least squares problem as (4.7). The full-column-rankness of A_i leads to the uniqueness of the solution, which is crucial for the one-loop convergence. Thus, we lay down these natural assumptions for $j = 1, 2, \ldots, d$:

(A.j) $\qquad A_j = \alpha(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]})$ has full column rank.

Once (A.j) is satisfied, (4.7) has a unique minimizer and $\mathbf{v}^{[j]}$ can be uniquely determined.

All later proofs rely on a homogeneity property as defined in Definition 4.4.

DEFINITION 4.4. *A multivariable function mapping from* d *Euclidean spaces to an Euclidean space, i.e.,*

$$\begin{aligned} f:\ \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \times \ldots, \times \mathbb{R}^{p_d} &\to \quad \mathbb{R}^q, \\ (x_1, x_2, \ldots, x_d) &\mapsto f(x_1, x_2, \ldots, x_d) \end{aligned}$$

is multi-homogeneous-poly *if for any* $\lambda \in \mathbb{R}$ *and index* $j \in \{1, 2, \ldots, d\}$,

$$f(x_1, \ldots, x_{j-1}, \lambda x_j, x_{j+1}, \ldots, x_d) = \lambda^{s_j} f(x_1, \ldots, x_d), \tag{4.10}$$

where s_j *is a* j*-dependent nonnegative integer, and each entry of* $f(x_1, \ldots, x_d)$ *is a polynomial of entries of* $x_1, x_2, \ldots, x_d$.

We emphasize that all $p_1, \ldots, p_d$ and q in Definition 4.4 are a multi-index notation, e.g., $q = 2 \times 3$. This means both the input x_i and the output $f(x_1, \ldots, x_d)$ could be scalars, matrices, or tensors.

Next, we list two properties of multi-homogeneous-poly function without detailed proof. Concrete examples for both properties are given in Appendix B.

- (Productivity) The product of two multi-homogeneous-poly functions is a multi-homogeneous-poly. This product includes entrywise product as well as compatible tensor contractions.

- (Composition) The composition of a multi-homogeneous-poly function with a multi-homogeneous-poly function is multi-homogeneous-poly.

Obviously, unfolding operator $\alpha(\cdot)$ is a multi-homogeneous-poly function. In Lemma 4.5, we show that under condition (A.j) and with proper scaling, the function mapping from $(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]})$ to $\mathbf{v}^{[j]}$ can be described by a multi-homogeneous-poly function. In Lemma 4.6, we characterize Ω_2 and show the existence of a multi-homogeneous-poly function whose root set equals Ω_2.

LEMMA 4.5. *There exists a multi-homogeneous-poly function*

$$G_j : \bigtimes_{i=j+1}^{j-1} \mathbb{R}^{m\times n_i\times m} \to \mathbb{R}^{m\times n_j\times m},$$

such that for any $(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]})$ *with*

$$A_j = \alpha(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]})$$

satisfying condition (A.j)*, it holds that*

$$\mathbf{v}^{[j]} = \frac{1}{\det\{A_j^\top A_j\}} G_j(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]}) \tag{4.11}$$

for $\mathbf{v}^{[j]}$ *being the solution of* (4.1) *at the* j*th microstep.*

Proof of Lemma 4.5. As $\alpha(\cdot)$ and $\alpha^\top(\cdot)$ are multi-homogeneous-poly functions, $A_j^\top A_j$ is a multi-homogeneous-poly function, due to the productivity property of multi-homogeneous-poly functions. According to the definition of adjugate operation, we have $\mathrm{adj}(\lambda A) = \lambda^{m^2-1}\mathrm{adj}(A)$, for all $\lambda\in\mathbb{R}$, $A\in\mathbb{R}^{m^2\times m^2}$, and each entry of $\mathrm{adj}(A)$ is a polynomial of entries of A. Hence, due to the composition property, $\mathrm{adj}(A_j^\top A_j)$ is a multi-homogeneous-poly function. The folding operator γ^{-1} is also multi-homogeneous-poly, i.e., $\gamma^{-1}(\lambda X) = \lambda\gamma^{-1}(X)$ for any λ and X. Applying the productivity property and composition property of the multi-homogeneous-poly function again, we have that

$$G_j(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]}) = \gamma^{-1}(\mathrm{adj}(A_j^\top A_j)A_j^\top B_j) \tag{4.12}$$

is a multi-homogeneous-poly function.

When A_j has full column rank, i.e., $\det\{A_j^\top A_j\}\neq 0$, the unique minimizer of (4.7) can be written as

$$X_j = (A_j^\top A_j)^{-1}A_j^\top B_j = \frac{1}{\det\{A_j^\top A_j\}}\mathrm{adj}(A_j^\top A_j)A_j^\top B_j, \tag{4.13}$$

where $\mathrm{adj}(\cdot)$ denotes the adjugate. Hence,

$$\mathbf{v}^{[j]} = \gamma^{-1}(X_j) = \frac{1}{\det\{A_j^\top A_j\}} G_j(\mathbf{u}^{[j+1]},\dots,\mathbf{u}^{[d]},\mathbf{v}^{[1]},\dots,\mathbf{v}^{[j-1]}) \tag{4.14}$$

proves the lemma. ∎

LEMMA 4.6. *Denote* $\Omega_2 = \{(\mathbf{u}^{[1]},\dots,\mathbf{u}^{[d]}) \mid$ *at least one of* (A.j) *fails for* $j = 1,2,\dots,d\}$*. There exists a multi-homogeneous-poly function* $F : \mathcal{U}^d_{m,\vec{n}} \to \mathbb{R}$*, such that* Ω_2 *is the root set of* F*.*

Proof of Lemma 4.6. Lemma 4.6 is proven by induction from d down to 1. We first define a sequence of set $\Omega^{[j]}$ for $j = 1, 2, \ldots, d$ via

$$\Omega^{[j]} = \{(\mathbf{u}^{[j]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]}) \mid \text{at least one of (A.i) fails for } i = j, \ldots, d\}.$$

Notice that $\Omega_2 = \Omega^{[1]}$.

First consider $\Omega^{[d]}$. If (A.d) fails, i.e., $A_d = \alpha(\mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[d-1]})$ does not have full column rank, then $\det\{A_d^\top A_d\} = 0$ defines a multi-homogeneous-poly function

$$F_d(\mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[d-1]}) = \det\{A_d^\top A_d\} \tag{4.15}$$

with $\mathbf{u}^{[d]}$ being a dummy variable, such that the root set of F_d equals $\Omega^{[d]}$.

Now, we take induction step. Assume for $j+1$ there exists a multi-homogeneous-poly function $F_{j+1}(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j]})$ such that the root set of F_{j+1} equals $\Omega^{[j+1]}$. The construction of F_j can be divided into two scenarios: (A.j) fails and (A.j) holds.

When (A.j) fails, i.e., $A_j = \alpha(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]})$ does not have full column rank, then $\det\{A_j^\top A_j\} = 0$ defines a multi-homogeneous-poly function

$$H_j^1(\mathbf{u}^{[j]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]}) = \det\{A_j^\top A_j\} \tag{4.16}$$

with $\mathbf{u}^{[j]}$ being dummy variable, such that the root set of H_j^1 equals $\Omega^{[j]}$ given (A.j) fails.

When (A.j) holds, according to Lemma 4.5, there exists a multi-homogeneous-poly function G_j such that

$$\mathbf{v}^{[j]} = \frac{1}{\det\{A_j^\top A_j\}} G_j(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]}). \tag{4.17}$$

Since (A.j) holds, at least one of (A.i) for $i = j+1, \ldots, d$ fails and, hence, F_{j+1} exists. Substituting (4.17) into F_{j+1} and multiplying by $\det\{A_j^\top A_j\}^s$, where s is the homogeneity degree of the last variable, gives a multi-homogeneous-poly function

$$\begin{aligned} &H_j^2(\mathbf{u}^{[j]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]}) \\ &= F_{j+1}(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]}, G_j(\mathbf{u}^{[j+1]}, \ldots, \mathbf{u}^{[d]}, \mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[j-1]})). \end{aligned} \tag{4.18}$$

The root set of H_j^2 equals $\Omega^{[j]}$ given (A.j) holds.

Combining two scenarios together, we define the multi-homogeneous-poly function

$$F_j = H_j^1 \cdot H_j^2 \tag{4.19}$$

with root set equals $\Omega^{[j]}$.

Finally, setting $F = F_1$ completes the proof. ∎

Next we prove that the multi-homogeneous-poly function in Lemma 4.6 is not constantly zero, which is the second step of the strategy described in Remark 4.3. We need some mild assumptions on the target tensor $\mathbf{T}$. Let $\mathbf{T}^j \in \mathbb{R}^{r^{d-2} \times r^2 \times r^d}$ be a

reshape of nonzeros of $\mathbf{T}$ for $j = 1, 2, \ldots, d-1$ satisfying

$$\mathbf{T}^j(\pi(p_1, \ldots, p_{j-1}, q_{j+1}, \ldots, q_{d-1}), :, \pi(q_1, \ldots, q_j, p_j, \ldots, p_{d-1})) = \mathbf{T}(\pi(p_1, q_1), \ldots, \pi(p_{d-1}, q_{d-1}), 1 : r^2), \tag{4.20}$$

where $1 \le p_i, q_i \le r$ and $1 \le i \le d-1$. The mild assumptions state as

$$\begin{aligned} &\mathbf{T}^j(:, :, \pi(q_1, \ldots, q_j, p_j, \ldots, p_{d-1})),\ 1 \le q_1, \ldots, q_j, p_j, \ldots, p_{d-1} \le r, \\ &\text{are linearly independent,} \end{aligned} \tag{B.j}$$

for $j = 1, 2, \ldots, d-1$. Ω_1 is the set of tensors violating these assumption. Later we will prove that Ω_1 has zero measure in Lemma 4.8.

Given these assumptions, we can show that the multi-homogeneous-poly function in Lemma 4.6 is not constantly zero.

LEMMA 4.7. *Suppose* $\vec{\mathbf{w}} \in \mathcal{W}^d_{r,\vec{n}}$ *and* (B.j) *holds for* $j = 1, 2, \ldots, d-1$. *The multi-homogeneous-poly function* F *in Lemma* 4.6 *is not constant zero and* Ω_2 *has zero Lebesgue measure.*

Proof of Lemma 4.7. First, F is a polynomial of entries of $\mathbf{u}^{[1]}, \ldots, \mathbf{u}^{[d]}$. Then showing a polynomial is not constantly zero, it is sufficient to show that there exists a $\vec{\mathbf{u}}$ such that $F(\vec{\mathbf{u}}) \neq 0$, which is equivalent to showing that for this $\vec{\mathbf{u}}$ the condition (A.j) holds for any $j = 1, 2, \ldots, d$.

Let each third order tensor of $\vec{\mathbf{u}} = (\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]}) \in \mathcal{U}^d_{m,\vec{n}}$ be

$$\mathbf{u}^{[i]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \delta_{p_1 q_1} \cdots \delta_{p_{i-1} q_{i-1}} \delta_{p_{i+1} q_{i+1}} \cdots \delta_{p_{d-1} q_{d-1}} e_{\pi(p_i, q_i)} \tag{4.21}$$

for $1 \le i \le d-1$, and

$$\mathbf{u}^{[d]}_{\pi(p_1,\ldots,p_{d-1}),\pi(q_1,\ldots,q_{d-1})} = \mathbf{T}(\pi(q_1, p_1), \pi(q_2, p_2), \ldots, \pi(q_{d-1}, p_{d-1}), :), \tag{4.22}$$

where $1 \le p_1, \ldots, p_{d-1}, q_1, \ldots, q_{d-1} \le r$.

Since $\mathbf{u}^{[i]}$ is defined identically to that in (3.5) for $1 \le i \le d-1$, then (3.6) holds here as well. Hence, it is easy to verify that $\tau(\vec{\mathbf{u}}) = \mathbf{T}$. Since $\vec{\mathbf{u}}$ is already a minimizer, we have $\vec{\mathbf{v}} = \vec{\mathbf{u}}$ if (A.j) holds for all j, i.e., the minimizer in each microstep is unique. Thus, to prove that (A.j) holds is equivalent to saying that

$$\sum_{k_{j+2},\ldots,k_{j-1}=1}^{m} \mathbf{u}^{[j+1]}_{k_{j+1},k_{j+2}} \otimes \cdots \otimes \mathbf{u}^{[j-1]}_{k_{j-1},k_j}, \tag{4.23}$$

$k_{j+1}, k_j = 1, 2, \ldots, m$, are linearly independent, since each column of A_j is an unfolding of (4.23). Due to (3.6), (A.d) holds directly.

Consider a fixed $j \in \{1, 2, \ldots, d-1\}$, and denote $k_j = \pi(p_1, p_2, \ldots, p_{d-1})$ and $k_{j+1} = \pi(q_1, q_2, \ldots, q_{d-1})$. With a careful index check, (4.23) equals

$$\begin{aligned} &\sum_{q'_{j+1},\ldots,q'_{d-1},p'_1,\ldots,p'_{j-1}=1}^{r} e_{\pi(q_{j+1},q'_{j+1})} \otimes \cdots \otimes e_{\pi(q_{d-1},q'_{d-1})} \\ &\otimes \mathbf{T}(\pi(p'_1, q_1), \ldots, \pi(p'_{j-1}, q_{j-1}), \pi(p_j, q_j), \pi(p_{j+1}, q'_{j+1}), \ldots, \pi(p_{d-1}, q'_{d-1}), :) \\ &\otimes e_{\pi(p'_1,p_1)} \otimes \cdots \otimes e_{\pi(p'_{j-1},p_{j-1})}. \end{aligned} \tag{4.24}$$

Hence (A.j) holding for $\vec{\mathbf{u}}$ is equivalent to (B.j). We have showed that $F(\vec{\mathbf{u}}) \neq 0$.

Since Ω_2 is the root set of F, which is a nonzero polynomial, the measure of Ω_2 is zero. ∎

If we merge the first and the second index of $\mathbf{T}^j$ together, then $\mathbf{T}^j$ becomes an $r^d \times r^d$ matrix, denoted by $\widetilde{\mathbf{T}}^j$. Assumption (B.j) is equivalent to saying that the matrix is full-rank. In random matrix theory, we know that the measure of degenerate matrices is zero. Here, Lemma 4.8 points out that a similar conclusion holds for $\mathbf{T}^j$ if $\mathbf{w}^{[1]}, \ldots, \mathbf{w}^{[d]}$ are generated randomly.

LEMMA 4.8. *Let* Ω_1 *be the set of failure of* (B.j)*, i.e.,*

$$\Omega_1 = \{\vec{\mathbf{w}} \in \mathcal{W}_{r,\vec{n}}^d \mid \textit{at least one of } \text{(B.j)} \textit{ is not satisfied for } j = 1, 2, \ldots, d-1\}.$$

Then Ω_1 *has Lebesgue measure* 0*.*

Proof of Lemma 4.8. Since d is a finite number, it is sufficient to prove that for any $j = 1, 2, \ldots, d-1$,

$$\mu(\{\vec{\mathbf{w}} \mid \text{(B.j) is not satisfied}\}) = 0. \tag{4.25}$$

Consider a fixed $j \in \{1, 2, \ldots, d-1\}$. For any matrix $X \in \mathbb{R}^{r^d \times r^d}$, X is column-rank-deficient if and only if $\det\{X^T X\} = 0$. Similar to the proof of Lemma 4.6, it can be shown that $f_j = \det\{(\widetilde{\mathbf{T}}^j)^T \widetilde{\mathbf{T}}^j\}$ is a multi-homogeneous-poly polynomial. The condition in Lemma 4.8 can be rewritten in terms of a multi-homogeneous-poly polynomial, i.e., the set in (4.25) can be restated as

$$\{\vec{\mathbf{w}} \mid \text{(B.j) is not satisfied}\} = \{\vec{\mathbf{w}} \mid f_j(\vec{\mathbf{w}}) = 0\}. \tag{4.26}$$

If f_j is not a zero polynomial, then the set of its roots has Lebesgue measure zero.

The rest of the proof states that f_j is not a zero polynomial. Consider the point

$$\mathbf{w}_{k_1,k_2}^{[i]} = e_{\pi(k_2,k_1)}, \quad 1 \le k_1, k_2 \le r, 1 \le i \le d, \tag{4.27}$$

which results in $\mathbf{T} = \sum_{k_1,\ldots,k_d=1}^{r} \bigotimes_{i=1}^{d} e_{\pi(k_{i+1},k_i)}$ and the reshaped tensor,

$$\begin{aligned} &\mathbf{T}^j(\pi(p_1, \ldots, p_{j-1}, q_{j+1}, \ldots, q_{d-1}), :, \pi(q_1, \ldots, q_j, p_j, \ldots, p_{d-1})) \\ &\qquad = \delta_{p_1 q_2} \cdots \delta_{p_{j-1} q_j} \delta_{q_{j+1} p_j} \cdots \delta_{q_{d-1} p_{d-2}} e_{\pi(q_1, p_{d-1})} \in \mathbb{R}^{r^2}, \end{aligned} \tag{4.28}$$

whose corresponding $r^d \times r^d$ matrix $\widetilde{\mathbf{T}}^j$ is an identity matrix. Since (B.j) is satisfied for this specific $\vec{\mathbf{w}}$, f_j is not a zero polynomial. ∎

Finally, with the above technical lemmas, we now prove Theorem 4.2.

Proof of Theorem 4.2. Let Ω_1 and Ω_2 be the measure-zero sets as defined in Lemmas 4.8 and 4.6, respectively. We consider points $\vec{\mathbf{w}} \notin \Omega_1$ and $\vec{\mathbf{u}} \notin \Omega_2$. Set

$$\mathbb{X}_i := \operatorname{span}\left\{\mathbf{w}_{k_1,k_2}^{[i]} : k_1, k_2 = 1, 2, \ldots, r\right\}$$

and denote $\mathbb{P}_i$ as the orthogonal projector onto $\mathbb{X}_i$ for $i = 1, 2, \ldots, d$.

According to (4.7), we have

$$\frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{x}^{[1]}, \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]})\right\|_{\mathrm{F}}^2 \ge \frac{1}{2}\left\|\mathbf{T} - \tau(\gamma^{-1}(\mathbb{P}_1(\gamma(\mathbf{x}^{[1]}))), \mathbf{u}^{[2]}, \ldots, \mathbf{u}^{[d]})\right\|_{\mathrm{F}}^2$$

for any $\mathbf{x}^{[1]}$, which implies that $\mathbf{v}_{k_1,k_2}^{[1]} \in \mathbb{X}_1,\ \forall\ k_1, k_2 = 1, 2, \ldots, m$. Similarly, we have

$$\mathbf{v}_{k_1,k_2}^{[i]} \in \mathbb{X}_i \quad \forall\ k_1, k_2 = 1, 2, \ldots, m,$$

for $i = 1, 2, \ldots, d-1$.

Now we consider the last microstep

$$\mathbf{v}^{[d]} = \arg\min_{\mathbf{x}^{[d]}} \frac{1}{2}\left\|\mathbf{T} - \tau(\mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[d-1]}, \mathbf{x}^{[d]})\right\|_{\mathrm{F}}^2. \tag{4.29}$$

The condition (A.d) implies that

$$\sum_{k_2,\ldots,k_{d-1}=1}^{m} \mathbf{v}_{k_1,k_2}^{[1]} \otimes \cdots \otimes \mathbf{v}_{k_{d-1},k_d}^{[d-1]}, \quad k_1, k_d = 1, \ldots, m,$$

are linearly independent. Combining the linear independence with

$$m^2 = r^{2(d-1)} \geq \prod_{k=1}^{d-1} \dim(\mathbb{X}_k) = \dim(\mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1}) \tag{4.30}$$

and

$$\sum_{k_2,\ldots,k_{d-1}=1}^{m} \mathbf{v}_{k_1,k_2}^{[1]} \otimes \cdots \otimes \mathbf{v}_{k_{d-1},k_d}^{[d-1]} \in \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1} \quad \forall\ k_1, k_d = 1, \ldots, m, \tag{4.31}$$

we obtain that

$$\operatorname{span}\left\{ \sum_{k_2,\ldots,k_{d-1}=1}^{m} \mathbf{v}_{k_1,k_2}^{[1]} \otimes \cdots \otimes \mathbf{v}_{k_{d-1},k_d}^{[d-1]},\ k_1, k_d = 1, \ldots, m \right\} = \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1}. \tag{4.32}$$

Noticing that $\mathbf{T} \in \mathbb{X}_1 \otimes \cdots \otimes \mathbb{X}_{d-1} \otimes \mathbb{R}^{n_d}$, there exists $\mathbf{v}^{[d]}$ such that

$$\tau(\mathbf{v}^{[1]}, \ldots, \mathbf{v}^{[d-1]}, \mathbf{v}^{[d]}) = \mathbf{T}, \tag{4.33}$$

which proves the theorem. ∎

5. Numerical results. In this section, we present two sets of numerical results to validate and further support Theorems 3.3 and 4.2, respectively. For Theorem 3.3, in section 3, we identify a nonstrict spurious local minimum for a target tensor of bond dimension $r+1$ as in (3.3) given the optimization problem in $\mathcal{U}^d_{r^{d-1},n}$. In this section, we numerically validate that ALS algorithm in some sense cannot escape from the spurious local minimum in Theorem 3.3 though Proposition 3.4 suggests that this local minimum might not be strict, whereas for Theorem 4.2, in section 4, the one-loop convergence is proven for target tensors with bond dimension r given the optimization problem in $\mathcal{U}^d_{r^{d-1},n}$ solved via ALS. We numerically test the tightness of the bond dimension r^{d-1} and show that the one-loop convergence does not hold when the TR space is reduced to $\mathcal{U}^d_{r^{d-1}-1,n}$. All numerical results in this section are generated from codes implemented and executed with MATLAB.

5.1. The stability of the spurious local minimum. Theorem 3.3 shows that for the given target tensor as (3.3), there exists a carefully designed local minimum. Due to the intrinsic difficulty of the TR format, the theorem does not characterize the neighborhood of all tensor rings with an equivalent format. Hence we numerically demonstrate that the designed local minimum is somewhat a numerically inescapable local minimum for ALS.

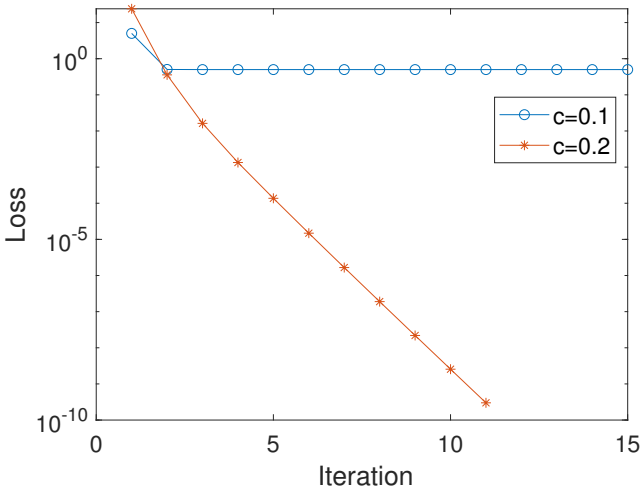

FIG. 2. *Typical convergence behavior of a "trapping" and an "escaping" iteration with c being* 0.1 *and* 0.2*, respectively.*

We construct the target tensor as (3.3) with $d = 3$, $r = 3$, and $n = r^2 + 1 = 10$ and the local minimum $\vec{\mathbf{u}}_0$ as (3.5) is constructed accordingly. For the purpose of this section, we apply ALS with a initial tensor ring being a perturbation of $\vec{\mathbf{u}}_0$. The perturbation is added as follows. Given a perturbation size $c \geq 0$, we add independent random numbers, uniformly distributed on $[-c, c]$, on each entry of $\vec{\mathbf{u}}_0$. We select 1000 choices of c between 0 and 0.3. For each c, 10^5 perturbations are tested via ALS and the least squares problems therein are solved via MATLAB backslash. When the objective functions of a converged iteration stay above $\frac{1}{2}$, we claim it is trapped by the local minimum. Otherwise, it escapes from the local minimum. Figure 2 shows two typical convergence behaviors of a "trapping" and an "escaping" iteration. Figure 3 plots the phase transition of the empirical probability that ALS is trapped at a TR format which leads to the same whole tensor as the local minimum $\vec{\mathbf{u}}_0$.

The curve in Figure 3 decreases monotonically, which indicates that ALS is more likely to escape from the set $\{\vec{\mathbf{u}} : \tau(\vec{\mathbf{u}}) = \tau(\vec{\mathbf{u}}_0)\}$ if it starts at a point further away from $\vec{\mathbf{u}}_0$, while, if the perturbation size is small enough, i.e., ALS starts at a point close to local minimum $\vec{\mathbf{u}}_0$, then ALS cannot escape empirically. In Figure 2, we plot two curves of loss corresponding to $c = 0.1$ and $c = 0.2$, respectively. We can see that ALS traps when $c = 0.1$ and escapes when $c = 0.2$. This provides numerical evidence that the local minimum $\vec{\mathbf{u}}_0$ is somewhat stable, i.e., $\vec{\mathbf{u}}_0$ is a interior point of the basin of attraction of the ALS and the basin has positive measure.

5.2. One-loop convergence. In this section, we numerically show that Theorem 4.2 holds in practice and the bond dimension given in the theorem is tight, i.e., one-loop convergence fails if $m = r^{d-1} - 1$. We present the results for different choices of $r \geq 3$, $d \geq 3$, and $n \geq r^2$. Two choices of bond dimensions, $m = r^{d-1}$ and $m = r^{d-1} - 1$, are tested. For a given r, d, n, and m, we randomly generate an initial tensor ring with each entry being a standard Gaussian random and then perform ALS for d microsteps. The final function value $f(\vec{\mathbf{u}}_1)$ is reported as the result. Such an experiment is repeated for 100 times for every given r, d, n, and the statistics of $f(\vec{\mathbf{u}}_1)$s are reported in Table 1.

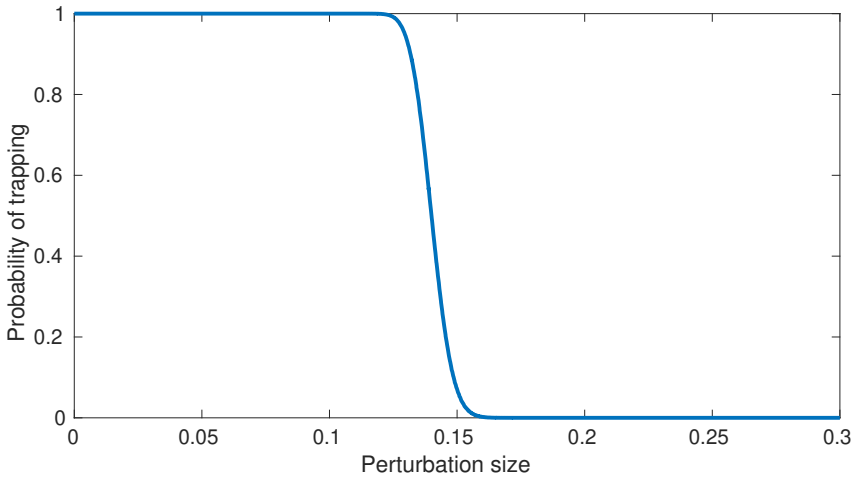

FIG. 3. *The stability of the spurious local minimum. The probability of trapping at* $\vec{\mathbf{u}}_0$ *via ALS against the perturbation size* c.

TABLE 1
One-loop convergence

d	r	n	m	$\max f(\vec{\mathbf{u}}_1)$	$\min f(\vec{\mathbf{u}}_1)$
3	3	10	9	1.79×10^{-13}	1.01×10^{-22}
			8	7.37×10^{1}	5.29×10^{0}
4	3	10	27	4.45×10^{-9}	6.69×10^{-19}
			26	1.06×10^{2}	2.44×10^{0}
3	4	16	16	8.58×10^{-5}	5.10×10^{-17}
			15	5.61×10^{1}	1.63×10^{0}

Table 1 shows that $\vec{\mathbf{u}}_1$ definitely does not converge in one loop for bond dimension $m = r^{d-1} - 1$, since $f(\vec{\mathbf{u}}_1)$ is far away from zero. Hence Theorem 4.2 is numerically verified and so is its tightness. We notice that the $\max f(\vec{\mathbf{u}}_1)$ is not close to machine accuracy after one loop when $m = r^{d-1}$, which is due to the initial random tensor ring being ill-conditioned and the numerical inverse being significantly polluted by the numerical error.

6. Conclusion. In this paper, we investigate a sharp transition for the optimization landscape associated with the TR decomposition. Considering least squares fitting of a target tensor $\mathbf{T}$ by dth order TR format with bond dimension r^{d-1}, or equivalently solving the optimization problem (3.1) with bond dimension r^{d-1}, if $\mathbf{T}$ is in TR format with bond dimension r, the problem is trivial, i.e., one-loop convergence holds. However, if $\mathbf{T}$ is in TR format with bond dimension $r + 1$, the landscape of (3.1) might be quite bad, i.e., there may exist some spurious local minima. These results tell us that even in the overparameterized case, we may not expect a good optimization landscape of TR decomposition, which in some sense shows the difficulty in numerical algorithms for TR decomposition.

Appendix A. Visualization of the target tensor and the local minimum in Theorem 3.3.

In the case that $d = 3$ and $r = 2$, the target tensor $\mathbf{T}_0 \in \mathbb{R}^{5\times5\times5}$ is explicitly given by

$$\mathbf{T}_0(:,:,1) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{T}_0(:,:,2) = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{T}_0(:,:,3) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{T}_0(:,:,4) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{T}_0(:,:,5) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

We emphasize that the last columns and rows of the first four slices are all zeros. In the last slice, $\mathbf{T}_0(:,:,5)$ is all zero except the last entry is one, which comes from the extra term in (3.3).

The local minimum $\vec{\mathbf{u}}_0 = (\mathbf{u}^{[1]}, \mathbf{u}^{[2]}, \mathbf{u}^{[3]})$ defined in (3.5) is

$$\mathbf{u}^{[1]}_{\pi(p_1,p_2),\pi(q_1,q_2)} = \begin{cases} e_{\pi(p_1,q_1)} & \text{if } p_2 = q_2, \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

$$\mathbf{u}^{[2]}_{\pi(p_1,p_2),\pi(q_1,q_2)} = \begin{cases} e_{\pi(p_2,q_2)} & \text{if } p_1 = q_1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.2}$$

and

$$\mathbf{u}^{[3]}_{\pi(p_1,p_2),\pi(q_1,q_2)} = \begin{cases} e_{\pi(p_1,q_2)} & \text{if } p_2 = q_1, \\ 0 & \text{otherwise.} \end{cases} \tag{A.3}$$

In the setting $r = 2$, $\mathbf{u}^{[1]}$, $\mathbf{u}^{[2]}$, and $\mathbf{u}^{[3]}$ are all $4 \times 5 \times 4$-tensors with

$$\mathbf{u}^{[1]}(:,1,:) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[1]}(:,2,:) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{u}^{[1]}(:,3,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[1]}(:,4,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{u}^{[1]}(:,5,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$\mathbf{u}^{[2]}(:,1,:) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[2]}(:,2,:) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{u}^{[2]}(:,3,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[2]}(:,4,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{u}^{[2]}(:,5,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

and

$$\mathbf{u}^{[3]}(:,1,:) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[3]}(:,2,:) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\mathbf{u}^{[3]}(:,3,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \mathbf{u}^{[3]}(:,4,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{u}^{[3]}(:,5,:) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The last slide for any $\mathbf{u}^{[k]}$ are empty.

The tensor $\tau(\vec{\mathbf{u}}_0) \in \mathbb{R}^{5\times5\times5}$ is then the same as $\mathbf{T}_0$ except the last slice, i.e., $\tau(\vec{\mathbf{u}}_0)(:,:,5)$ is an all zero matrix.

Appendix B. Examples of multi-homogeneous-poly. We first give an example for the entrywise productivity property. Let $f(x_1, x_2)$ and $g(x_1, x_3)$ be two multi-homogeneous-poly with output being in $\mathbb{R}^{2\times1}$, i.e.,

$$f(x_1,x_2) = \begin{pmatrix} f_1(x_1,x_2) \\ f_2(x_1,x_2) \end{pmatrix} \text{ and } g(x_1,x_3) = \begin{pmatrix} g_1(x_1,x_3) \\ g_2(x_1,x_3) \end{pmatrix}. \tag{B.1}$$

Then the entrywise product of f and g, denoted as

$$h(x_1,x_2,x_3) = \begin{pmatrix} f_1(x_1,x_2)g_1(x_1,x_3) \\ f_2(x_1,x_2)g_2(x_1,x_3) \end{pmatrix}, \tag{B.2}$$

is still an entrywise polynomial of entries of x_1, x_2, x_3. The homogeneity can be justified as

$$\begin{aligned} &h(\lambda x_1,x_2,x_3) = \lambda^{s_1^f+s_1^g} h(x_1,x_2,x_3), \\ &h(x_1,\lambda x_2,x_3) = \lambda^{s_2^f} h(x_1,x_2,x_3), \text{ and} \\ &h(x_1,x_2,\lambda x_3) = \lambda^{s_3^g} h(x_1,x_2,x_3), \end{aligned} \tag{B.3}$$

where s_1^f, s_2^f, s_1^g, and s_3^g are the homogeneity degree of f and g, respectively. Hence h is also multi-homogeneous-poly.

We then give an example for the tensor contraction (matrix product) productivity property. Let $f(x_1, x_2)$ and $g(x_1, x_3)$ be two multi-homogeneous-poly with output being in $\mathbb{R}^{2\times 2}$, i.e.,

$$\begin{aligned} f(x_1,x_2) &= \begin{pmatrix} f_{11}(x_1,x_2) & f_{12}(x_1,x_2) \\ f_{21}(x_1,x_2) & f_{22}(x_1,x_2) \end{pmatrix} \text{ and} \\ g(x_1,x_3) &= \begin{pmatrix} g_{11}(x_1,x_3) & g_{12}(x_1,x_3) \\ g_{21}(x_1,x_3) & g_{22}(x_1,x_3) \end{pmatrix}. \end{aligned} \tag{B.4}$$

The matrix product of f and g, denoted as

$$h(x_1,x_2,x_3) = \begin{pmatrix} f_{11}g_{11} + f_{12}g_{21} & f_{11}g_{12} + f_{12}g_{22} \\ f_{21}g_{11} + f_{22}g_{21} & f_{21}g_{12} + f_{22}g_{22} \end{pmatrix}, \tag{B.5}$$

is still an entrywise polynomial of entries of x_1, x_2, x_3. The homogeneity can be justified as

$$\begin{aligned} h(\lambda x_1, x_2, x_3) &= \lambda^{s_1^f + s_1^g} h(x_1, x_2, x_3), \\ h(x_1, \lambda x_2, x_3) &= \lambda^{s_2^f} h(x_1, x_2, x_3), \text{ and} \\ h(x_1, x_2, \lambda x_3) &= \lambda^{s_3^g} h(x_1, x_2, x_3), \end{aligned} \tag{B.6}$$

where s_1^f, s_2^f, s_1^g, and s_3^g are the homogeneity degree of f and g, respectively. Hence h is also multi-homogeneous-poly.

Finally, we give an example for the composition property. Let $f(x_1, x_2, y)$ and $g(x_1, x_3)$ be two multi-homogeneous-poly and g is of the same dimension as y. The composition of f and g, denoted as

$$h(x_1,x_2,x_3) = f(x_1,x_2,g(x_1,x_3)), \tag{B.7}$$

is still an entrywise polynomial of entries of x_1, x_2, x_3. The homogeneity can be justified as

$$\begin{aligned} h(\lambda x_1, x_2, x_3) &= \lambda^{s_1^f + s_y^f s_1^g} h(x_1, x_2, x_3), \\ h(x_1, \lambda x_2, x_3) &= \lambda^{s_2^f} h(x_1, x_2, x_3), \text{ and} \\ h(x_1, x_2, \lambda x_3) &= \lambda^{s_y^f s_3^g} h(x_1, x_2, x_3), \end{aligned} \tag{B.8}$$

where s_1^f, s_2^f, s_y^f, s_1^g, and s_3^g are the homogeneity degree of f and g, respectively. Hence h is also multi-homogeneous-poly.

REFERENCES

[1] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, *Valence bond ground states in isotropic quantum antiferromagnets*, Comm. Math. Phys., 115 (1988), pp. 477–528, https://doi.org/10.1007/BF01218021.

[2] I. Arad, Z. Landau, U. Vazirani, and T. Vidick, *Rigorous RG algorithms and area laws for low energy eigenstates in* 1D, Comm. Math. Phys., 356 (2017), pp. 65–105, https://doi.org/10.1007/s00220-017-2973-z.

[3] K. Batselier, *The Trouble with Tensor Ring Decompositions*, preprint, https://arxiv.org/abs/1811.03813, 2018.

[4] D. Bigoni, A. P. Engsig-Karup, and Y. M. Marzouk, *Spectral tensor-train decomposition*, SIAM J. Sci. Comput., 38 (2016), pp. A2405–A2439, https://doi.org/10.1137/15M1036919.
[5] F. G. S. L. Brandão and M. Horodecki, *Exponential decay of correlations implies area law*, Comm. Math. Phys., 333 (2015), pp. 761–798, https://doi.org/10.1007/s00220-014-2213-8.
[6] G. K.-L. Chan and S. Sharma, *The density matrix renormalization group in quantum chemistry*, Annu. Rev. Phys. Chem., 62 (2011), pp. 465–481, https://doi.org/10.1146/annurev-physchem-032210-103338.
[7] M. Espig, W. Hackbusch, and A. Khachatryan, *On the Convergence of Alternating Least Squares Optimisation in Tensor Format Representations*, preprint, https://arxiv.org/abs/1506.00062, 2015.
[8] R. Ge, C. Jin, and Y. Zheng, *No spurious local minima in nonconvex low rank problems: A unified geometric analysis*, in Proceedings of the 34th International Conference on Machine Learning, PMLR 70, 2017, pp. 1233–1242, http://proceedings.mlr.press/v70/ge17a.html.
[9] L. Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054, https://doi.org/10.1137/090764189.
[10] W. Hackbusch and S. Kühn, *A new scheme for the tensor representation*, J. Fourier Anal. Appl., 15 (2009), pp. 706–722, https://doi.org/10.1007/s00041-009-9094-9.
[11] K. A. Hallberg, *New trends in density matrix renormalization*, Adv. Phys., 55 (2006), pp. 477–526.
[12] M. B. Hastings, *An area law for one-dimensional quantum systems*, J. Stat. Mech. Theory Exp., 2007 (2007), 08024, https://doi.org/10.1088/1742-5468/2007/08/P08024.
[13] S. Holtz, T. Rohwedder, and R. Schneider, *The alternating linear scheme for tensor optimization in the tensor train format*, SIAM J. Sci. Comput., 34 (2012), pp. A683–A713, https://doi.org/10.1137/100818893.
[14] Y. Khoo, J. Lu, and L. Ying, *Efficient Construction of Tensor Ring Representations from Sampling*, preprint, https://arxiv.org/abs/1711.00954, 2017.
[15] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.
[16] J. M. Landsberg, Y. Qi, and K. Ye, *On the geometry of tensor network states*, Quantum Inf. Comput., 12 (2012), pp. 346–354.
[17] Y. Li, J. Lu, and Z. Wang, *Coordinatewise descent methods for leading eigenvalue problem*, SIAM J. Sci. Comput., 41 (2019), pp. A2681–A2716, https://doi.org/10.1137/18M1202505.
[18] R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Ann. Phys. (NY), 349 (2014), pp. 117–158, https://doi.org/10.1016/j.aop.2014.06.013.
[19] I. V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317, https://doi.org/10.1137/090752286.
[20] I. V. Oseledets and E. E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759, https://doi.org/10.1137/090748330.
[21] E. Robeva and A. Seigal, *Duality of Graphical Models and Tensor Networks*, preprint, https://arxiv.org/abs/1710.01437, 2017.
[22] T. Rohwedder and A. Uschmajew, *On local convergence of alternating schemes for optimization of convex problems in the tensor train format*, SIAM J. Numer. Anal., 51 (2013), pp. 1134–1162, https://doi.org/10.1137/110857520.
[23] U. Schollwöck, *The density-matrix renormalization group*, Rev. Modern Phys., 77 (2005), 259.
[24] U. Schollwöck, *The density-matrix renormalization group in the age of matrix product states*, Ann. Phys. (NY), 326 (2011), pp. 96–192, https://doi.org/10.1016/J.AOP.2010.09.012.
[25] F. Verstraete and J. I. Cirac, *Renormalization Algorithms for Quantum-Many Body Systems in Two and Higher Dimensions*, preprint, https://arxiv.org/abs/cond-mat/0407066, 2004.
[26] F. Verstraete, D. Porras, and J. I. Cirac, *Density matrix renormalization group and periodic boundary conditions: A quantum information perspective*, Phys. Rev. Lett., 93 (2004), 227205, https://doi.org/10.1103/PhysRevLett.93.227205.
[27] G. Vidal, *Entanglement renormalization*, Phys. Rev. Lett., 99 (2007), 220405, https://doi.org/10.1103/PhysRevLett.99.220405.
[28] W. Wang, V. Aggarwal, and S. Aeron, *Efficient low rank tensor ring completion*, in Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2017, pp. 5698–5706, https://doi.org/10.1109/ICCV.2017.607.
[29] W. Wang, V. Aggarwal, and S. Aeron, *Efficient low rank tensor ring completion*, in Proceeding of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 5698–5706, https://doi.org/10.1109/ICCV.2017.607.

[30] S. R. White, *Density matrix formulation for quantum renormalization groups*, Phys. Rev. Lett., 69 (1992), pp. 2863–2866, https://doi.org/10.1103/PhysRevLett.69.2863.
[31] S. R. White, *Density-matrix algorithms for quantum renormalization groups*, Phys. Rev. B, 48 (1993), pp. 10345–10356, https://doi.org/10.1103/PhysRevB.48.10345.
[32] S. Wouters and D. Van Neck, *The density matrix renormalization group for ab initio quantum chemistry*, Eur. Phys. J. D, 68 (2014), 272.
[33] K. Ye and L.-H. Lim, *Tensor Network Ranks*, preprint, https://arxiv.org/abs/1801.02662, 2018.
[34] Q. Zhao, M. Sugiyama, L. Yuan, and A. Cichocki, *Learning efficient tensor representations with ring structure networks*, in Proceedings of the International Conference on Learning Representatives, 2018.
[35] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, *Tensor Ring Decomposition*, preprint, https://arxiv.org/abs/1606.05535, 2016.