Computational Optimization and Applications (2023) 85:621–652
https://doi.org/10.1007/s10589-023-00471-6

Partially symmetric tensor structure preserving rank-*R* approximation via BFGS algorithm

Ciwen Chen[1] · **Guyan Ni**[1] · **Bo Yang**[1]

Received: 11 June 2022 / Accepted: 27 February 2023 / Published online: 20 March 2023

Abstract

It is known that many tensor data have symmetric or partially symmetric structure and structural tensors have structure preserving Candecomp/Parafac (CP) decompositions. However, the well-known alternating least squares (ALS) method cannot realize structure preserving CP decompositions of tensors. Hence, in this paper, we consider numerical problems of structure preserving rank-R approximation and structure preserving CP decomposition of partially symmetric tensors. For the problem of structure preserving rank-R approximation, we derive the gradient formula of the objective function, obtain BFGS iterative formulas, propose a BFGS algorithm for positive partially symmetric rank-R approximation, and discuss the convergence of the algorithm. For the problem of structure preserving CP decomposition, we give a necessary condition for partially symmetric tensors with even orders to have positive partially symmetric CP decompositions, and design a general partially symmetric rank-R approximation algorithm. Finally, some numerical examples are given. Through numerical examples, we find that if a tensor has a positive partially symmetric CP decomposition then its partially symmetric rank CP decomposition must be a positive CP decomposition. In addition, we compare the BFGS algorithm proposed in this paper with the standard CP-ALS method. Numerical examples show that the BFGS algorithm has better stability and faster computing speed than CP-ALS algorithm.

Keywords Partially symmetric tensors · Rank-R approximation · CP decomposition · BFGS method

✉ Guyan Ni
guyan-ni@163.com

Ciwen Chen
chen_cwen@163.com

Bo Yang
bo_yangemail@163.com

[1] Department of Mathematics, National University of Defense Technology, Changsha 410073, Hunan, China

1 Introduction

Tensor Candecomp/Parafac(CP) decomposition and tensor rank are the basic problems in tensor research and have lots of important applications in data analysis [1, 2], signal processing [3, 4] and others [5, 6]. Although the tensor decomposition problem is NP hard [7], there are many researchers still doing it without bored and tired. The decomposition of a tensor into an outer product of vectors and the corresponding notations were first proposed and studied by Frank L. Hitchcock in [8, 9]. The same decomposition was rediscovered in 1970s [20] in order to extend the data analysis model to multidimensional arrays.

Symmetric tensors are a class of tensors with many different decomposition properties and special applications [10–12], so many scholars have conducted in-depth research on symmetric tensors. Comon et al. [13] proved that any symmetric tensor can be decomposed into a linear combination of rank-1 tensor. Brachat et al. [14] presented an algebraic method for symmetric tensor decompositions, based on linear algebra computation with Hankel matrices by extending Sylvester's theorem to multiple variables in 2010. Kolda [15] considered real-valued decomposition of symmetric tensor, both nonnegative and sparse and presented a numerical optimization algorithm based on gradient in 2015. In 2017, Nie [16] devised a combination of polynomial optimization and numerical approaches for solving complex-valued symmetric tensor decompositions. In 2022, Liu [17] proposed an alternating gradient descent algorithm for solving symmetric tensor CP decomposition by minimizing a multiconvex optimization problem.

Partially symmetric tensor can be used to solve practical engineering problems, such as elastic material analysis [18] and quantum information theory [19]. Carroll and Chang [20] first considered the CP decomposition of partially symmetric tensors and proposed an alternating method of ignoring the symmetry, the decomposition obtained by this method does not have the property of structure preserving unless the decomposition satisfies the uniqueness condition [21]. In 2013, Li and Navasca [22] proposed the so called partial column-wise least squares method to obtain the CP decomposition of the partially symmetric tensor of the third and fourth order. Ni and Li [23] proved that any partially symmetric tensor has a partially symmetric CP decomposition and presented a semi-definite relaxation algorithm. However, the semi-definite relaxation is hard to calculate a partially symmetric CP decomposition if the tensor has higher order and higher dimension, since the number of moments will increase sharply with the increase of tensor order and dimension.

Motivated by the above, we consider the numerical algorithm of structure preserving rank-R approximation of partially symmetric tensors. We deduce the gradient formula of the structure preserving rank-R approximation loss function, propose a gradient descent algorithm and a BFGS algorithm, and analyze the convergence of the BFGS algorithm. The BFGS algorithm can be used to the structure preserving CP decomposition of partially symmetric tensors, completely symmetric tensors and non-symmetric tensors, respectively. Numerical examples show that the BFGS algorithm has good numerical performance.

The paper is structured as follows. Section 2 introduces some basic concepts and related properties, including matrix and tensor product, inner product and Frobenius norm of tensor, as well as partially symmetric tensor, etc. Section 3 deduces the gradient of the partially symmetric rank-R approximation loss function. In Sect. 4, we propose the BFGS algorithm and discuss its convergence. In Sect. 5, We derive the discrimination of partially symmetric tensors with the positive decomposition. Finally, numerical experiments are given in Sect. 6.

Notation. $\mathbb{N}_+$, $\mathbb{R}$ and $\mathbb{C}$ denote the set of positive integers, real field and complex field, respectively. A uppercase letter in calligraphic font denotes a tensor, e.g., $\mathcal{T}$. A uppercase letter represents a matrix, e.g., U. A boldface lowercase letter represents a vector, e.g., $\mathbf{v}$. A lowercase letter represents a scalar, e.g., x. The entry with row index i and column index j in a matrix U, i.e., $(U)_{ij}$, is symbolized by u_{ij} (also $(\mathbf{v})_i = v_i$ and $(\mathcal{T})_{i_1 i_2 \cdots i_N} = t_{i_1 i_2 \cdots i_N}$). Let $s > 0$ be an integer, denote $[s] := \{1, 2, \cdots, s\}$ as an integer set. $\mathfrak{S}_t$ denotes the set of all permutations of $\{1, 2, \cdots, m_t\}$, m_t is a positive integer, $t \in [s]$.

2 Preliminaries

2.1 Matrix and tensor products

A tensor can be regarded as a multiway array, which is a generalization of matrices. A tensor is represented with calligraphic script uppercase letters in this paper, e.g., $\mathcal{T}$. In this subsection, some basic concepts and products of matrix and tensor will be reviewed. For more details, please refer to the literature [24].

The Kronecker product of matrices $A \in \mathbb{R}^{I \times J}$ and $B \in \mathbb{R}^{K \times L}$, denoted as $A \otimes B$, is defined as

$$A \otimes B := \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1J}B \\ a_{21}B & a_{22}B & \cdots & a_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \cdots & a_{IJ}B \end{bmatrix} \in \mathbb{R}^{IK \times JL}.$$

The Khatri-Rao product of two matrices $A \in \mathbb{R}^{I \times K}$ and $B \in \mathbb{R}^{J \times K}$ is defined as

$$A \odot B := [\mathbf{a}_1 \otimes \mathbf{b}_1 \ \mathbf{a}_2 \otimes \mathbf{b}_2 \ \cdots \ \mathbf{a}_K \otimes \mathbf{b}_K] \in \mathbb{R}^{IJ \times K}.$$

If $\mathbf{a}$ and $\mathbf{b}$ are vectors, then the Khatri-Rao and Kronecker products are identical, i.e., $\mathbf{a} \otimes \mathbf{b} = \mathbf{a} \odot \mathbf{b}$.

The Hadamard product of two same-sized matrices $A,\ B \in \mathbb{R}^{I \times J}$, denoted as $A * B \in \mathbb{R}^{I \times J}$, is defined as

$$(A * B)_{ij} := a_{ij} b_{ij}.$$

We use "$\circ$" to represent the outer product of two arbitrary tensors $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$, $\mathcal{Y} \in \mathbb{R}^{J_1 \times J_2 \times \cdots \times J_n}$

$$\mathcal{X} \circ \mathcal{Y} = \left(x_{i_1 i_2 \cdots i_m} y_{j_1 j_2 \cdots j_n}\right) \in \mathbb{R}^{I_1 \times \cdots \times I_m \times J_1 \times \cdots \times J_n}. \tag{2.1}$$

Invoking the definition of tensor outer product as described in (2.1), we can get that for vectors $\mathbf{x}_k \in \mathbb{R}^{I_k}$, for $k \in [m]$, their tensor outer product $\mathbf{x}_1 \circ \mathbf{x}_2 \circ \cdots \circ \mathbf{x}_m \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ is an mth-order rank-1 tensor such that

$$(\mathbf{x}_1 \circ \mathbf{x}_2 \circ \cdots \circ \mathbf{x}_m)_{i_1 i_2 \cdots i_m} = (\mathbf{x}_1)_{i_1} (\mathbf{x}_2)_{i_2} \cdots (\mathbf{x}_m)_{i_m}.$$

In particular, if $\mathbf{x}_1 = \mathbf{x}_2 = \cdots \mathbf{x}_m = \mathbf{x} \in \mathbb{R}^n$, then $\mathbf{x}^{\circ m} \equiv \overbrace{\mathbf{x} \circ \mathbf{x} \circ \cdots \circ \mathbf{x}}^{m}$ is an mth-order n-dimensional symmetric rank-1 tensor and is denoted as $\mathbf{x}^m$ for simplicity.

Definition 2.1 Let $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ be a tensor and vectors $\mathbf{x}_k \in \mathbb{R}^{I_k}$ for $k \in [m]$. Define the contraction product of tensor $\mathcal{T}$ and a rank-1 tensor as follows

$$\mathcal{T} \cdot \circ_{i=1}^{m} \mathbf{x}_i := \sum_{i_1, \cdots, i_m = 1}^{I_1, \cdots, I_m} t_{i_1 \cdots i_m} (\mathbf{x}_1)_{i_1} \cdots (\mathbf{x}_m)_{i_m} \in \mathbb{R},$$

$$\mathcal{T} \cdot \circ_{i=1, i \neq k}^{m} \mathbf{x}_i := \left(\sum_{i_1, \cdots, i_{k-1}, i_{k+1}, \cdots, i_m = 1}^{I_1, \cdots, I_{k-1}, I_{k+1}, \cdots, I_m} t_{i_1 \cdots i_m} (\mathbf{x}_1)_{i_1} \cdots (\mathbf{x}_{k-1})_{i_{k-1}} (\mathbf{x}_{k+1})_{i_{k+1}} \cdots (\mathbf{x}_m)_{i_m} \right) \in \mathbb{R}^{I_k}.$$

Generally, the inner product of two tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$, denoted as $\mathcal{X} \cdot \mathcal{Y}$ or $\langle \mathcal{X}, \mathcal{Y} \rangle$, is defined as

$$\mathcal{X} \cdot \mathcal{Y} := \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_m=1}^{I_m} x_{i_1 i_2 \cdots i_m} y_{i_1 i_2 \cdots i_m}.$$

The Frobenius norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ is defined as

$$\|\mathcal{X}\|_F := \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_m=1}^{I_m} x_{i_1 i_2 \cdots i_m}^2}.$$

It follows immediately that $\langle \mathcal{X}, \mathcal{X} \rangle = \|\mathcal{X}\|_F^2$.

The mode-k matricization of a tensor $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ is denoted as $T_{(k)}$ and rearrange its elements into a matrix of size $I_k \times \prod_{i \neq k} I_i$. Element $(i_1, i_2, \cdots, i_m)$ maps to matrix entry (i_k, j), where

$$j = 1 + \sum_{t \neq k}^{m} (i_t - 1) \prod_{n \neq k}^{t-1} I_n.$$

Let $\mathcal{T} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_m}$ be an mth-order tensor. The CP decomposition of $\mathcal{T}$ is

$$\mathcal{T} = [\![A_1, A_2, \cdots, A_m]\!] \equiv \sum_{i=1}^{R} \mathbf{a}_i^{(1)} \circ \mathbf{a}_i^{(2)} \circ \cdots \circ \mathbf{a}_i^{(m)},$$

where $A_k = (\mathbf{a}_1^{(k)}\ \mathbf{a}_2^{(k)}\ \cdots\ \mathbf{a}_R^{(k)}) \in \mathbb{R}^{I_k \times R}$, $k \in [m]$ are called factor matrices. The mode-k matricization of tensor $\mathcal{T} = [\![A_1, A_2, \cdots, A_m]\!]$, $A_k \in \mathbb{R}^{I_k \times R}$, for $k \in [m]$ can be written in the form of factor matrix,

$$T_{(k)} = A_k (A_m \odot \cdots \odot A_{k+1} \odot A_{k-1} \odot \cdots \odot A_1)^T. \tag{2.2}$$

2.2 Partially symmetric tensor

Partially symmetric tensor is the main content of this paper, so its basic concepts and some properties are indispensable. Detailed introduction can be referred to the literature [23, 25–27].

Definition 2.2 [23] Let $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$. An $\mathbf{m}$th-order $\mathbf{n}$-dimensional tensor $\mathcal{T}$ is an array over the field $\mathbb{F}$ indexed by integer tuples $\left(i_1, \cdots, i_{m_1}, j_1, \cdots, j_{m_2}, \cdots, l_1, \cdots, l_{m_s}\right)$, i.e.,

$$\mathcal{T} = \left(t_{i_1,\cdots,i_{m_1},j_1,\cdots,j_{m_2},\cdots,l_1,\cdots,l_{m_s}}\right) \in \mathbb{F}^{n_1^{m_1} \times n_2^{m_2} \times \cdots \times n_s^{m_s}}$$

with $i_1, \cdots, i_{m_1} \in [n_1]$, $j_1, \cdots, j_{m_2} \in [n_2], \cdots, l_1, \cdots, l_{m_s} \in [n_s]$.

The space of all tensors over the filed $\mathbb{F}$ is denoted as $T\,[\mathbf{m}]\,\mathbb{F}\,[\mathbf{n}]$. $\mathcal{T}$ is called a square tensor, if all dimensions are equal, that is to say $n_1 = n_2 = \cdots = n_s$.

Definition 2.3 [25] Let $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$. A tensor $\mathcal{S} \in T\,[\mathbf{m}]\,\mathbb{F}\,[\mathbf{n}]$ is called partially symmetric if

$$s_{i_1 \cdots i_{m_1} j_1 \cdots j_{m_2} \cdots l_1 \cdots l_{m_s}} = s_{\sigma_1(i_1)\cdots\sigma_1(i_{m_1})\sigma_2(j_1)\cdots\sigma_2(j_{m_2})\cdots\sigma_s(l_1)\cdots\sigma_s(l_{m_s})}$$

for every permutation $\sigma_t \in \mathfrak{S}_t$, where $\mathfrak{S}_t$ is the set of all permutations of $\{1, 2, \cdots, m_t\}$, $t \in [s]$.

The space of all $\mathbf{m}$th-order $\mathbf{n}$-dimensional partially symmetric tensors over the field $\mathbb{F}$ is denoted by $S[\mathbf{m}]\mathbb{F}[\mathbf{n}]$.

Definition 2.4 [23] Let vectors $\mathbf{u}_k^{(t)} \in \mathbb{F}^{n_t}$ for all $t \in [s]$. The result of the tensor outer product $\left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}$ is a rank-1 partially symmetric tensor. A tensor $\mathcal{S} \in S\,[\mathbf{m}]\,\mathbb{F}\,[\mathbf{n}]$ is said to have a partially symmetric CP decomposition if the tensor $\mathcal{S}$ has a decomposition as

$$\mathcal{S} = \sum_{k=1}^{R} \lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}, \lambda_k \in \mathbb{F}, \mathbf{u}_k^{(t)} \in \mathbb{F}^{n_t}, \|\mathbf{u}_k^{(t)}\|_2 = 1.$$

If $m_1, m_2, \cdots, m_s = 1$, then the tensor $\mathcal{T} \in T[\mathbf{m}]\mathbb{F}[\mathbf{n}]$ is an sth-order $n_1 \times n_2 \times \cdots \times n_s$-dimensional tensor, the partially symmetric CP decomposition of $\mathcal{T}$ is an usual nonsymmetric CP decomposition. If $s = 1$ and $n_1 = n_2 = \cdots = n_s = n$, let $m_1 = m$, then $\mathcal{T} \in T[m]\mathbb{F}[n]$ is an mth-order n-dimensional symmetric tensor.

Theorem 2.1 [23] *Let* $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$, *and* $\mathcal{S} \in S[\mathbf{m}]\mathbb{F}[\mathbf{n}]$ *be a partially symmetric tensor. Then there exist* $\lambda_k \in \mathbb{F}$, $\mathbf{u}_k^{(t)} \in \mathbb{F}^{n_t}$, $\|\mathbf{u}_k^{(t)}\|_2 = 1$ *for* $k \in [R]$ *and* $t \in [s]$ *such that*

$$\mathcal{S} = \sum_{k=1}^{R} \lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}. \tag{2.3}$$

i.e., any partially symmetric tensor (over any field) has a partially symmetric CP decomposition.

If $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ has a decomposition as (2.3) with $\lambda_k \geq 0$ for all $k \in [R]$, then we say that $\mathcal{S}$ has a positive partially symmetric CP decomposition. According to Theorem 2.1, there is the following corollary.

Corollary 2.1 *If there is* $i \in [s]$ *such that* m_i *is an odd number, then, any nonzero tensor* $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ *has a positive partially symmetric CP decomposition. However,* $\mathcal{S}$ *may not have positive partially symmetric CP decomposition, if* $m_1, \cdots, m_s$ *are even.*

Proof If there is $i \in [s]$ such that m_i is an odd number, without loss of generality, let m_1 be an odd number. Then, the m_1th real root of λ_k always exists for $k \in [R]$, so we can rewrite the item $\lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1}$ as

$$\lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1} = \left(\tilde{\mathbf{u}}_k^{(1)}\right)^{m_1}, \quad \text{where } \tilde{\mathbf{u}}_k^{(1)} = \sqrt[m_1]{\lambda}\mathbf{u}_k^{(1)}.$$

Then,

$$\mathcal{S} = \sum_{k=1}^{R} \left(\tilde{\mathbf{u}}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}.$$

That is, $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ has a positive partially symmetric CP decomposition. However, if all orders are even, then m_tth real root does not exist if $\lambda_k < 0$, for $t \in [s]$, so the scalar cannot be absorbed. In this case, $\mathcal{S}$ may not have positive partially symmetric CP decomposition. This completes the proof. □

If $\mathcal{S}$ has a positive partially symmetric CP decomposition as in (2.3), then we can write $\lambda_k \left(\mathbf{u}_k^{(t)}\right)^{m_t} = \left((\lambda_k)^{1/m_t}\mathbf{u}_k^{(t)}\right)^{m_t}$ for some $t \in [s]$. In this case, the decomposition (2.3) can be rewritten as follows

$$\mathcal{S} = \sum_{k=1}^{R} \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}. \tag{2.4}$$

Let $U_t = (\mathbf{u}_1^{(t)}, \mathbf{u}_2^{(t)}, \cdots, \mathbf{u}_R^{(t)}) \in \mathbb{R}^{n_t \times R}$ for $t \in [s]$. We call $\{U_1, U_2, \cdots, U_s\}$ as a tuple of factor matrices of partially symmetric tensor $\mathcal{S}$. The partially symmetric CP decomposition (2.4) can also be written as

$$\begin{aligned} \mathcal{S} &= [\![\overbrace{U_1, \cdots, U_1}^{m_1}, \overbrace{U_2, \cdots, U_2}^{m_2}, \cdots, \overbrace{U_s, \cdots, U_s}^{m_s}]\!] \\ &\equiv [\![U_1^{\times m_1}, U_2^{\times m_2}, \cdots, U_s^{\times m_s}]\!]. \end{aligned} \tag{2.5}$$

3 Positive partially symmetric rank-R approximation

Now, we study the positive partially symmetric rank-R approximation of partially symmetric tensors. Let s, R be two positive integers and $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$. Given a real partially symmetric tensor $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$, find positive scalars $\lambda_k > 0$ and unit-norm vectors $\mathbf{u}_k^{(t)} \in \mathbb{R}^{n_t}$ for $t \in [s]$ and $k \in [R]$ such that the rank-R tensor

$$\hat{\mathcal{S}} := \sum_{k=1}^{R} \lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}$$

minimizes the function

$$f(\hat{\mathcal{S}}) = \frac{1}{2} \|\mathcal{S} - \hat{\mathcal{S}}\|_F^2.$$

The positive partially symmetric rank-R approximation problem of tensors $\mathcal{S}$ is equivalent to solving the following optimization problem

$$\begin{aligned} &\min \frac{1}{2} \left\| \mathcal{S} - \sum_{k=1}^{R} \lambda_k \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s} \right\|_F^2 \\ &s.t. \lambda_k > 0, \mathbf{u}_k^{(t)} \in \mathbb{F}^{n_t}, \|\mathbf{u}_k^{(t)}\|_2 = 1, k \in [R], t \in [s]. \end{aligned} \tag{P1}$$

Since $\lambda_k > 0$ for all $k \in [R]$, we can rewrite $\hat{\mathcal{S}}$ as

$$\hat{\mathcal{S}} = \sum_{k=1}^{R} \left(\tilde{\mathbf{u}}_k^{(1)}\right)^{m_1} \circ \left(\tilde{\mathbf{u}}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\tilde{\mathbf{u}}_k^{(s)}\right)^{m_s}. \tag{3.1}$$

Then $\mathbf{u}_k^{(t)} = \tilde{\mathbf{u}}_k^{(t)} / \|\tilde{\mathbf{u}}_k^{(t)}\|_2$ and $\lambda_k = \prod_{t=1}^{s} \|\tilde{\mathbf{u}}_k^{(t)}\|_2^{m_t}$, for all $k \in [R]$ and $t \in [s]$.

The right hand side of (3.1) may also be written as the form of a tuple of factor matrices

$$\hat{\mathcal{S}} = [\![U_1^{\times m_1}, U_2^{\times m_2}, \cdots, U_s^{\times m_s}]\!],$$

where $U_t = \left(\tilde{\mathbf{u}}_1^{(t)}, \tilde{\mathbf{u}}_2^{(t)}, \cdots, \tilde{\mathbf{u}}_R^{(t)}\right) \in \mathbb{R}^{n_t \times R}$ is a matrix for all $t \in [s]$.

Let $\mathbf{U} \in \mathbb{R}^{(n_1+\cdots+n_s)\times r}$ be a block-matrix vector as

$$\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_s \end{pmatrix}.$$

Then the constrained optimization problem (P1) is turned to an unconstrained optimization

$$\min_{\mathbf{U}} f_{\mathcal{S}}(\mathbf{U}) = f_{\mathcal{S}}(U_1, U_2, \cdots, U_s) := \frac{1}{2} \left\| \mathcal{S} - [\![U_1^{\times m_1}, U_2^{\times m_2}, \cdots, U_s^{\times m_s}]\!] \right\|_F^2. \tag{3.2}$$

3.1 Gradient calculation of $f_{\mathcal{S}}(\mathbf{U})$

Lemma 3.1 *Let s and R be two positive integers, $\mathbf{m} = (m_1, m_2, \cdots, m_s), \mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$ and $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be a partially symmetric tensor. Let*

$$\hat{\mathcal{S}} = \sum_{k=1}^{R} \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}, \ \mathbf{u}_k^{(t)} \in \mathbb{R}^{n_t}.$$

Then

$$\frac{\partial \langle \mathcal{S}, \hat{\mathcal{S}} \rangle}{\partial \mathbf{u}_k^{(t)}} = m_t \ \mathcal{S} \cdot \left(\left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \cdots \circ \left(\mathbf{u}_k^{(t)}\right)^{m_t - 1} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s} \right) \in \mathbb{R}^{n_t}.$$

Proof Assume that $\mathcal{A} \in S[m]\mathbb{R}[n]$ be a symmetric tensor of m-order. Then,

$$\frac{\partial \mathcal{A} \cdot \mathbf{x}^m}{\partial \mathbf{x}} = m\mathcal{A} \cdot \mathbf{x}^{m-1}. \tag{3.3}$$

Define a function $F(\hat{\mathcal{S}}) = \langle \mathcal{S}, \hat{\mathcal{S}} \rangle$. Then,

$$\begin{aligned} F(\hat{\mathcal{S}}) &= \langle \mathcal{S}, \sum_{k=1}^{R} \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s} \rangle \\ &= \sum_{k=1}^{R} \mathcal{S} \cdot \left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \left(\mathbf{u}_k^{(2)}\right)^{m_2} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s}. \end{aligned}$$

According to the formula (3.3), we have that

$$\frac{\partial F(\hat{\mathcal{S}})}{\partial \mathbf{u}_k^{(t)}} = m_t \ \mathcal{S} \cdot \left(\left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \cdots \circ \left(\mathbf{u}_k^{(t)}\right)^{m_t - 1} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s} \right).$$

This completes proof. □

Theorem 3.1 *Let* $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$, $M_t := m_1 + \cdots + m_t$, $t \in [s]$. *Let* $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ *be a given partially symmetric tensor,* $U_t = (\mathbf{u}_1^{(t)}, \mathbf{u}_2^{(t)}, \cdots, \mathbf{u}_R^{(t)}) \in \mathbb{R}^{n_t \times R}$ *be a matrix for all* $k \in [R]$, $t \in [s]$ *and* $\mathbf{U} = (U_1^T, U_2^T, \cdots, U_s^T)^T$ *be a block-matrix vector. Then the first order partial derivative of the function* $f_{\mathcal{S}}(\mathbf{U})$ *defined in* (3.2) *with respect to* U_t *is*

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_t} = -m_t(S_{(M_t)} W_t - U_t W_t^T W_t),$$

where $S_{(M_t)}$ *is the mode-*M_t *matricization of the tensor* $\mathcal{S}$ *and*

$$W_t = \mathrm{U}_s{}^{\odot m_s} \odot \cdots \odot \mathrm{U}_t{}^{\odot m_t - 1} \odot \cdots \odot \mathrm{U}_1{}^{\odot m_1}. \tag{3.4}$$

Proof According to the definition of derivative, the first order partial derivative of the function with respect to U_t is

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_t} = \left(\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_1^{(t)}}, \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_2^{(t)}}, \cdots, \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_R^{(t)}} \right).$$

Let $\hat{\mathcal{S}} = [\![U_1^{\times m_1}, U_2^{\times m_2}, \cdots, U_s^{\times m_s}]\!]$. Then

$$\|\mathcal{S} - \hat{\mathcal{S}}\|^2 = \langle \mathcal{S}, \mathcal{S} \rangle - 2\langle \mathcal{S}, \hat{\mathcal{S}} \rangle + \langle \hat{\mathcal{S}}, \hat{\mathcal{S}} \rangle, \quad \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_k^{(t)}} = -\frac{\partial \langle \mathcal{S}, \hat{\mathcal{S}} \rangle}{\partial \mathbf{u}_k^{(t)}} + \frac{\partial \langle \hat{\mathcal{S}}, \hat{\mathcal{S}} \rangle}{2\partial \mathbf{u}_k^{(t)}}.$$

From the Lemma 3.1, for every $k \in [R]$, it is followed that

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_k^{(t)}} = -m_t \ (\mathcal{S} - \hat{\mathcal{S}}) \cdot \left(\left(\mathbf{u}_k^{(1)}\right)^{m_1} \circ \cdots \circ \left(\mathbf{u}_k^{(t)}\right)^{m_t - 1} \circ \cdots \circ \left(\mathbf{u}_k^{(s)}\right)^{m_s} \right).$$

According to definition of the mode-k matricization, we have that

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial \mathbf{u}_k^{(t)}} = -m_t(S - \hat{S})_{(M_t)} \left(\left(\mathbf{u}_k^{(s)}\right)^{\otimes m_s} \otimes \cdots \otimes \left(\mathbf{u}_k^{(t)}\right)^{\otimes m_t - 1} \otimes \cdots \otimes \left(\mathbf{u}_k^{(1)}\right)^{\otimes m_1} \right).$$

Thus,

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_t} = -m_t(S - \hat{S})_{(M_t)} \Big(U_s{}^{\odot m_s} \odot \cdots \odot U_t{}^{\odot m_t - 1} \odot \cdots \odot U_1{}^{\odot m_1} \Big).$$

Substitute (3.4) into the above equation, we get that

$$\frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_t} = -m_t(S_{(M_t)} W_t - U_t W_t^T W_t). \tag{3.5}$$

This completes the proof. □

If the gradient is calculated directly according to the formula (3.5), it will cost a lot of computational expenses. We can save the computational expenses by the following method. The first item $S_{(M_t)}W_t$ can be obtained by MTTKRP, which is a function in the tensor toolbox [29] to efficiently calculate matricized tensor times Khatri–Rao product. The second item can also avoid forming W_t since $W_t^T W_t$ is given by

$$W_t^T W_t = (U_s^T U_s)^{*m_s} * \cdots * (U_t^T U_t)^{*m_t - 1} * \cdots * (U_1^T U_1)^{*m_1}.$$

The following is the algorithm for calculating gradient value of $f_{\mathcal{S}}(\mathbf{U})$ defined as (3.2) at a given point.

Algorithm 1 The gradient $\mathbf{G}$ of the function $f_{\mathcal{S}}(\mathbf{U})$

Input: A tensor $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ and the factor matrix $\mathbf{U}^0 = (U_1^T, \cdots, U_s^T)^T$.
Output: The gradient $\mathbf{G} = (G_1^T, \cdots, G_s^T)^T$ of function $f_{\mathcal{S}}(\mathbf{U})$ at $\mathbf{U}^0$.
1: **for** $t = 1, \cdots, s$ **do**
2: Let $M_t = m_1 + m_2 + \cdots + m_t$.
3: Compute $V = (U_s^T U_s)^{*m_s} * \cdots * (U_t^T U_t)^{*m_t - 1} * \cdots * (U_1^T U_1)^{*m_1}$.
4: Compute $G_t = -m_t(\text{MTTKRP}(\mathcal{S}, [\![U_1^{\times m_1}, \cdots, U_s^{\times m_s}]\!], M_t) - U_t V)$.
5: **end for**
6: **return** $\mathbf{G} = (G_1^T, \cdots, G_s^T)^T$.

3.2 Gradient descent method

We introduce two linear search methods. Armijo inexact search rule adopts the backtracking strategy to obtain the step size. In other words, the step size is picked as $\alpha_k = \beta\gamma^{m_k}$, where m_k is the smallest nonnegative integer satisfying the following condition

$$f\left(\mathbf{u}^k + \beta\gamma^{m_k}\mathbf{d}^k\right) \leq f\left(\mathbf{u}^k\right) + \sigma\beta\gamma^{m_k}\langle \mathbf{g}^k, \mathbf{d}^k \rangle, \tag{3.6}$$

where $\beta > 0,\ \sigma,\ \gamma \in (0, 1)$.

Another inexact search rule choosing a step size $\alpha_k > 0$ is Armijo-Wolfe conditions [28],

$$\begin{cases} f\left(\mathbf{u}^k + \alpha\mathbf{d}^k\right) \leq f\left(\mathbf{u}^k\right) + c_1\alpha\langle \mathbf{g}^k, \mathbf{d}^k \rangle \\ \langle \nabla f\left(\mathbf{u}^k + \alpha\mathbf{d}^k\right), \mathbf{d}^k \rangle \geq c_2\langle \mathbf{g}^k, \mathbf{d}^k \rangle, \end{cases} \tag{3.7}$$

where $0 < c_1 < c_2 < 1$ are the parameters of line search algorithm, the above two condition are called the sufficient decrease condition and curvature condition respectively.

Therefore, the iterative scheme is

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \alpha_k\mathbf{d}^k,$$

where α_k is obtained from the above Armijo or Armijo-Wolfe rule, $\mathbf{d}^k$ is the search direction of the k-th iteration. If $\mathbf{d}^k = -\nabla f(\mathbf{u}^k)$, the iteration method is called gradient descent method.

The termination condition of optimization algorithm is that norm of gradient is less than a given criteria, i.e.,

$$\|\mathbf{g}^k\|_F < \epsilon.$$

For partially symmetric rank-R approximation problem of some partially symmetric tensors, the iterative formula of gradient descent method is that

$$\mathbf{U}^{k+1} = \mathbf{U}^k + \alpha_k \mathbf{G}^k,$$

where $\mathbf{G}^k = -\nabla f_{\mathcal{S}}(\mathbf{U}^k)$, α_k satisfies the Armijo or Armijo-Wolfe rule.

The following is gradient descent algorithm with Armijo rule (3.6) for the partially symmetric rank-R approximation of a partially symmetric tensor.

Algorithm 2 Positive partially symmetric rank-R approximation with gradient descent method

Input: A tensor $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ and a termination condition $\epsilon > 0$
Output: The factor matrix $\mathbf{U} = (U_1^T, \cdots, U_s^T)^T$ of $\hat{\mathcal{S}}$.
Initialization: Obtain an initial point $\mathbf{U}^1 = ({U_1^1}^T, \cdots, {U_s^1}^T)^T$, $k = 1$.
Step 1: Calculate the gradient $\mathbf{G}^k$ by Algorithm 1 with input $(\mathcal{S}, \mathbf{U}^k)$. If $\|\mathbf{G}^k\|_F < \epsilon$, then take $\mathbf{U} = \mathbf{U}^k$, and go to Step 4.
Step 2: Take the search direction $\mathbf{D}^k = -\mathbf{G}^k$, and pick the step size α_k with Armijo rule (3.6).
Step 3: Take $\mathbf{U}^{k+1} = \mathbf{U}^k + \alpha_k \mathbf{D}^k$ and $k = k + 1$. Turn to Step 1.
Step 4: Return $\mathbf{U} = (U_1^T, \cdots, U_s^T)^T$.

4 BFGS method

We now review the standard BFGS method with Armijo-Wolfe line search for solving the optimization problem $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$. For more details, please refer to [28]. The iterative formula of BFGS method for solving the problem $\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$ is

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k (B^k)^{-1}\mathbf{g}^k \text{ or } \mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k H^k \mathbf{g}^k.$$

where $\mathbf{g}^k = \nabla f(\mathbf{x}^k)$, α_k is step size satisfying Armijo-Wolfe rule. The initial point $\mathbf{x}^0$ of the iterative algorithm is usually given randomly; H^0 and B^0 must be positive definite matrices, usually the identity matrix.

Let $\mathbf{y}^k = \mathbf{g}^{k+1} - \mathbf{g}^k$ and $\mathbf{s}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$. The updated formula of B^k is that

$$B^{k+1} = B^k + \frac{\mathbf{y}^k(\mathbf{y}^k)^T}{(\mathbf{y}^k)^T\mathbf{s}^k} - \frac{B^k\mathbf{s}^k(\mathbf{s}^k)^T B^k}{(\mathbf{s}^k)^T B^k\mathbf{s}^k}.$$

The updated formula of H^k is that

$$H^{k+1} = \left(I - \frac{\mathbf{s}^k(\mathbf{y}^k)^T}{(\mathbf{y}^k)^T\mathbf{s}^k}\right) H^k \left(I - \frac{\mathbf{y}^k(\mathbf{s}^k)^T}{(\mathbf{y}^k)^T\mathbf{s}^k}\right) + \frac{\mathbf{s}^k(\mathbf{s}^k)^T}{(\mathbf{y}^k)^T\mathbf{s}^k}.$$

4.1 The BFGS method for partially symmetric rank-R approximation

Following the standard BFGS method, we will give the tensor BFGS method to solve the partially symmetric rank-R approximation of partially symmetric tensor. Let $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be a partially symmetric tensor, $\mathbf{U} = (U_1^T, \cdots, U_s^T)^T$ be a block-matrix vector, $U_t \in \mathbb{R}^{n_t \times r}$ for $t \in [s]$ and $n := n_1 + \cdots + n_s$. The function $f_{\mathcal{S}}(\mathbf{U})$ is defined as (3.2), then the gradient $\nabla f_{\mathcal{S}}(\mathbf{U})$ of is $f_{\mathcal{S}}(\mathbf{U})$ that

$$\nabla f_{\mathcal{S}}(\mathbf{U}) = \begin{pmatrix} \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_1} \\ \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_2} \\ \vdots \\ \frac{\partial f_{\mathcal{S}}(\mathbf{U})}{\partial U_s} \end{pmatrix} \in \mathbb{R}^{n \times R};$$

the Hessian is that

$$\nabla^2 f_{\mathcal{S}}(\mathbf{U}) = \left(\frac{\partial^2 f_{\mathcal{S}}(\mathbf{U})}{\partial u_{ij} \partial u_{kl}} \right) \in \mathbb{R}^{n \times R \times n \times R},$$

which is a fourth-order tensor.

In order to obtain the BFGS iterative formula for tensor, we define a multiplication of two fourth-order tensors and a product between a fourth-order tensor and a matrix. For any two tensors $\mathcal{A},\ \mathcal{B} \in \mathbb{R}^{K \times L \times K \times L}$, define $\mathcal{C} = \mathcal{A} \oslash \mathcal{B} \in \mathbb{R}^{K \times L \times K \times L}$ as

$$c_{i_1 i_2 i_3 i_4} = \sum_{j_1, j_2 = 1}^{K, L} a_{i_1 i_2 j_1 j_2} b_{j_1 j_2 i_3 i_4}. \tag{4.1}$$

For any $\mathcal{T} \in \mathbb{R}^{K \times L \times K \times L}$ and $A \in \mathbb{R}^{K \times L}$, define $B = \mathcal{T} A \in \mathbb{R}^{K \times L}$ as

$$b_{i_1 i_2} = \sum_{i_3, i_4 = 1}^{K, L} t_{i_1 i_2 i_3 i_4} a_{i_3 i_4}. \tag{4.2}$$

In the sense of the multiplication of two fourth-order tensors, we define the fourth-order identity tensor $\mathcal{I} \in \mathbb{R}^{K \times L \times K \times L}$ as

$$(\mathcal{I})_{k_1 l_1 k_2 l_2} = \begin{cases} 1 & \text{if } k_1 = k_2,\ l_1 = l_2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

According to the definition (4.3) and (4.2), for any matrix $A \in \mathbb{R}^{K \times L}$, we have $\mathcal{I} A = A$. The tensor $\mathcal{A} \in \mathbb{R}^{K \times L \times K \times L}$ is an invertible tensor, if there is a same-sized tensor $\mathcal{B}$ such that $\mathcal{A} \oslash \mathcal{B} = \mathcal{I}$ and $\mathcal{B} \oslash \mathcal{A} = \mathcal{I}$, and if so, then $\mathcal{B}$ is called the inverse of $\mathcal{A}$. The inverse of $\mathcal{A}$ is denoted by $\mathcal{A}^{-1}$.

Lemma 4.1 *Let fourth-order tensor $\mathcal{T} \in \mathbb{R}^{K\times L\times K\times L}$ be invertible and its elements satisfy $\mathcal{T}_{k_1l_1k_2l_2} = \mathcal{T}_{k_2l_2k_1l_1}$, and $A, B, C \in \mathbb{R}^{K\times L}$. Then*
(1) $\langle \mathcal{T}^{-1}A, \mathcal{T}B\rangle = \langle A, B\rangle$.
(2) $(A \circ B)C = A\langle B, C\rangle$.

Proof According to the multiplication defined as (4.2),

$$\begin{aligned}\langle \mathcal{T}^{-1}A, \mathcal{T}B\rangle &= \sum_{i,j}\left(\sum_{k_1,l_2}\mathcal{T}^{-1}{}_{ijk_1l_1}A_{k_1l_1}\right)\left(\sum_{k_1,l_2}\mathcal{T}_{ijk_2l_2}B_{k_2l_2}\right)\\ &= \sum_{k_1,l_1}\sum_{k_1,l_2}A_{k_1l_1}B_{k_2l_2}\sum_{i,j}\mathcal{T}^{-1}{}_{ijk_1l_1}\mathcal{T}_{ijk_2l_2}\end{aligned}$$

Since $\mathcal{T}_{k_1l_1k_2l_2} = \mathcal{T}_{k_2l_2k_1l_1}$, we have that

$$\begin{aligned}\langle \mathcal{T}^{-1}A, \mathcal{T}B\rangle &= \sum_{k_2,l_2}\sum_{k_1,l_1}\mathcal{I}_{k_2l_2k_1l_1}A_{k_1l_1}B_{k_2l_2}\\ &= \sum_{k_2,l_2}A_{k_2l_2}B_{k_2l_2}.\end{aligned}$$

Therefore, $\langle \mathcal{T}^{-1}A, \mathcal{T}B\rangle = \langle A, B\rangle$. According to the multiplication defined as (4.2) and tensor outer product defined as (2.1),

$$\begin{aligned}((A\circ B)C)_{ij} &= \sum_{k=1}^{K}\sum_{l=1}^{L}(A\circ B)_{ijkl}c_{kl}\\ &= \sum_{k=1}^{K}\sum_{l=1}^{L}a_{ij}b_{kl}c_{kl}\\ &= a_{ij}\sum_{k=1}^{K}\sum_{l=1}^{L}b_{kl}c_{kl}.\end{aligned}$$

Hence, $(A \circ B)C = A\langle B, C\rangle$. The proof is completed. □

Let $f_{\mathcal{S}}(\mathbf{U})$ be defined as (3.2), then the iterative formula of BFGS algorithm with Armijo-Wolfe line search for solving the optimization problem $\min_{\mathbf{U}} f_{\mathcal{S}}(\mathbf{U})$ is

$$\mathbf{U}^{k+1} = \mathbf{U}^k - \alpha_k(\mathcal{B}^k)^{-1}\mathbf{G}^k,$$

where $\mathbf{G}^k$ is gradient value of $f_{\mathcal{S}}(\mathbf{U})$ at $\mathbf{U}^k$, i.e., $\mathbf{G}^k = \nabla f_{\mathcal{S}}(\mathbf{U}^k)$, the step size α_k satisfies Armijo-Wolfe rule.

The updated formula of $\mathcal{B}^k$ is that

$$\mathcal{B}^{k+1} = \mathcal{B}^k + \frac{\mathbf{Y}^k\circ\mathbf{Y}^k}{\langle\mathbf{Y}^k, \mathbf{S}^k\rangle} - \frac{\left(\mathcal{B}^k\mathbf{S}^k\right)\circ(\mathcal{B}^k\mathbf{S}^k)}{\langle\mathcal{B}^k\mathbf{S}^k, \mathbf{S}^k\rangle}. \tag{4.4}$$

where $\mathbf{Y}^k = \nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}) - \nabla f_{\mathcal{S}}(\mathbf{U}^k)$, $\mathbf{S}^k = \mathbf{U}^{k+1} - \mathbf{U}^k$.

Lemma 4.2 *Let $\mathcal{B}^{k+1}$ be defined as (4.4). Then, the inverse of $\mathcal{B}^{k+1}$ can be given by*

$$\mathcal{H}^{k+1} = \left(\mathcal{I} - \frac{\mathbf{S}^k \circ \mathbf{Y}^k}{\langle \mathbf{S}^k, \mathbf{Y}^k \rangle}\right) \oslash \mathcal{H}^k \oslash \left(\mathcal{I} - \frac{\mathbf{Y}^k \circ \mathbf{S}^k}{\langle \mathbf{S}^k, \mathbf{Y}^k \rangle}\right) + \frac{\mathbf{S}^k \circ \mathbf{S}^k}{\langle \mathbf{S}^k, \mathbf{Y}^k \rangle}. \tag{4.5}$$

Proof In order to prove that $\mathcal{H}^{k+1}$ is the inverse of $\mathcal{B}^{k+1}$, we just need to prove that, for any block-matrix $\mathbf{X}$, $\mathcal{B}^{k+1}(\mathcal{H}^{K+1}\mathbf{X}) = \mathbf{X}$. Let $\langle \mathbf{S}^k, \mathbf{Y}^k \rangle = \frac{1}{\mu}$, $\langle \mathcal{B}^k \mathbf{S}^k, \mathbf{S}^k \rangle = \frac{1}{\nu}$. From the formula (4.4), we have

$$\begin{aligned} \mathcal{B}^{k+1}(\mathcal{H}^{k+1}X) = \mathcal{B}^k(\mathcal{H}^{k+1}X) + \mu(\mathbf{Y}^k \circ \mathbf{Y}^k)(\mathcal{H}^{k+1}X) \\ -\nu(\mathcal{B}^k\mathbf{S}^k \circ \mathcal{B}^k\mathbf{S}^k)(\mathcal{H}^{k+1}X). \end{aligned} \tag{4.6}$$

According to the Lemma 4.1,

$$\begin{aligned} \mathcal{H}^{k+1}X = \mathcal{H}^k\mathbf{X} - \mu\mathcal{H}^k\mathbf{Y}^k\langle \mathbf{S}^k, \mathbf{X}\rangle - \mu\mathbf{S}^k\langle \mathbf{Y}^k, \mathcal{H}^k\mathbf{X}\rangle \\ +\mu^2\mathbf{S}^k\langle \mathbf{Y}^k, \mathcal{H}^k\mathbf{X}\rangle\langle \mathbf{S}^k, \mathbf{X}\rangle + \mu\mathbf{S}^k\langle \mathbf{S}^k, \mathbf{X}\rangle. \end{aligned} \tag{4.7}$$

By putting (4.7) into (4.6) and using the Lemma 4.1, we get $\mathcal{B}^{k+1}(\mathcal{H}^{K+1}\mathbf{X}) = \mathbf{X}$. The proof is completed. □ □

Now, we will give the BFGS algorithm with Armijo-Wolfe step size (3.7) for the positive partially symmetric rank-R approximation of a partially symmetric tensor as follows. The initial factor matrix $\mathbf{U}^1$ is obtained randomly and $\mathcal{H}^1$ is usually selected as fourth-order identity tensor $\mathcal{I}$ defined as (4.3).

Algorithm 3 Positive partially symmetric rank-R approximation with BFGS algorithm

Input: A tensor $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ and a termination condition $\epsilon > 0$
Output: The factor matrix $\mathbf{U} = (U_1^T, \cdots, U_s^T)^T$ of $\hat{\mathcal{S}}$.
Initialization: Obtain an initial point $\mathbf{U}^1 = ({U_1^1}^T, \cdots, {U_s^1}^T)^T$, and let $\mathcal{H}^1 = \mathcal{I}$.
Step 1: Calculate the gradient $\mathbf{G}^1$ by Algorithm 1 with input $(\mathcal{S}, \mathbf{U}^1)$. If $\|\mathbf{G}^1\|_F < \epsilon$, then take $\mathbf{U} = \mathbf{U}^1$, and go to Step 6.
Step 2: Take $\mathbf{D}^1 = -\mathbf{G}^1$ and $k = 1$.
Step 3: Pick the step size α_k with Armijo-Wolfe rule (3.7), and take $\mathbf{U}^{k+1} = \mathbf{U}^k + \alpha_k\mathbf{D}^k$.
Step 4: Calculate the gradient $\mathbf{G}^{k+1}$ by Algorithm 1 with input $(\mathcal{S}, \mathbf{U}^{k+1})$. If $\|\mathbf{G}^{k+1}\|_F < \epsilon$, then take $\mathbf{U} = \mathbf{U}^{k+1}$, and go to Step 6.
Step 5: Calculate $\mathcal{H}^{k+1}$ by the formula (4.5), and take the descent direction $\mathbf{D}^{k+1} = -\mathcal{H}^{k+1}\mathbf{G}^{k+1}$. Let $k = k + 1$, and turn to Step 3.
Step 6: Return $\mathbf{U} = (U_1^T, \cdots, U_s^T)^T$.

4.2 Convergence analysis of BFGS algorithm

In this subsection, the super-linear convergence of algorithm will be discussed. The proof of following results refers to [28]. First, we give the relevant assumptions.

Assumption 4.1 Let $s, r > 0$ be two integers and $\mathbf{m} = (m_1, m_2, \cdots, m_s)$, $\mathbf{n} = (n_1, n_2, \cdots, n_s) \in \mathbb{N}_+^s$. Let $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be a partially symmetric tensor, $n = n_1 + \cdots + n_s$. The function $f_{\mathcal{S}}(\mathbf{U})$ is defined as (3.2), where the block-matrix vector $\mathbf{U} = (U_1^T, U_2^T, \cdots, U_s^T)^T \in \mathbb{R}^{n\times r}$, $U_t \in \mathbb{R}^{n_t\times r}$ for all $t \in [s]$.

1) $\langle \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*)\mathbf{W}, \mathbf{W}\rangle > 0$ for any $\mathbf{W} \in \mathbb{R}^{n\times r}$, where $\mathbf{U}^*$ is a local minimum point of $f_{\mathcal{S}}(\mathbf{U})$;
2) there is a neighborhood $N(\mathbf{U}^*)$ of $\mathbf{U}^*$ and a constant $L > 0$ such that for any $\mathbf{U},\ \mathbf{V} \in N(\mathbf{U}^*)$

$$\|\nabla^2 f_{\mathcal{S}}(\mathbf{U}) - \nabla^2 f_{\mathcal{S}}(\mathbf{V})\|_F \le L\|\mathbf{U} - \mathbf{V}\|_F.$$

Lemma 4.3 *If the function $f_{\mathcal{S}}(\mathbf{U})$ satisfies the second condition of* Assumption *4.1, then for any $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^{n\times r}$,*

$$\|\nabla f_{\mathcal{S}}(\mathbf{V})-\nabla f_{\mathcal{S}}(\mathbf{W})-\nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F \le L\frac{\|\mathbf{U}-\mathbf{W}\|_F + \|\mathbf{U}-\mathbf{V}\|_F}{2}\|\mathbf{V}-\mathbf{W}\|_F. \tag{4.8}$$

Forthermore, if $\nabla^2 f_{\mathcal{S}}(\mathbf{U})$ is invertible and for every $\varepsilon \in (0, \frac{1}{L\|(\nabla^2 f_{\mathcal{S}}(\mathbf{U}))^{-1}\|_F})$, for any $\mathbf{V}$ and $\mathbf{W}$ that satisfy $\|\mathbf{U}-\mathbf{W}\|_F < \varepsilon$ and $\|\mathbf{U}-\mathbf{V}\|_F < \varepsilon$, respectively, then there are constants $\beta > \alpha > 0$ related to $\mathbf{U}$ such that

$$\alpha\|\mathbf{V}-\mathbf{W}\|_F \le \|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W})\|_F \le \beta\|\mathbf{V}-\mathbf{W}\|_F. \tag{4.9}$$

Proof For any $\mathbf{U}, \mathbf{V}, \mathbf{W} \in \mathbb{R}^{n\times r}$, according to the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W}) - \nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F\\
&= \left\|\int_0^1 (\nabla^2 f_{\mathcal{S}}(\mathbf{W} + \tau(\mathbf{V}-\mathbf{W})) - \nabla^2 f_{\mathcal{S}}(\mathbf{U}))(\mathbf{V}-\mathbf{W})d\tau\right\|_F\\
&\le \|\mathbf{V}-\mathbf{W}\|_F \int_0^1 L\|\mathbf{W} + \tau(\mathbf{V}-\mathbf{W}) - \mathbf{U}\|_F d\tau\\
&\le L\|\mathbf{V}-\mathbf{W}\|_F \int_0^1 (\tau\|\mathbf{V}-\mathbf{U}\|_F + (1-\tau)\|\mathbf{W}-\mathbf{U}\|_F)d\tau\\
&= L\|\mathbf{V}-\mathbf{W}\|_F \frac{\|\mathbf{V}-\mathbf{U}\|_F + \|\mathbf{W}-\mathbf{U}\|_F}{2}.
\end{aligned}$$

For every given $\varepsilon \in (0, \frac{1}{L\|[\nabla^2 f_{\mathcal{S}}(\mathbf{U})]^{-1}\|_F})$, any $\mathbf{U}$ and $\mathbf{V}$ satisfying

$$\max\{\|\mathbf{V}-\mathbf{U}\|_F, \|\mathbf{W}-\mathbf{U}\|_F\} < \varepsilon,$$

we have that

$$\begin{aligned}
&\|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W})\|_F\\
&\le \|\nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F + \|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W}) - \nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F
\end{aligned}$$

$$\leq \left(\|\nabla^2 f_{\mathcal{S}}(\mathbf{U})\|_F + \frac{L(\|\mathbf{V}-\mathbf{U}\|_F + \|\mathbf{W}-\mathbf{U}\|_F)}{2} \right) \|\mathbf{V}-\mathbf{W}\|_F$$
$$\leq (\|\nabla^2 f_{\mathcal{S}}(\mathbf{U})\|_F + L\varepsilon)\|\mathbf{V}-\mathbf{W}\|_F.$$

Let $\beta = \|\nabla^2 f_{\mathcal{S}}(\mathbf{U})\|_F + L\varepsilon$, then the second inequality of (4.9) holds. Using the properties of the norm, we have that

$$\begin{aligned}
&\|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W})\|_F \\
&\geq \|\nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F - \|\nabla f_{\mathcal{S}}(\mathbf{V}) - \nabla f_{\mathcal{S}}(\mathbf{W}) - \nabla^2 f_{\mathcal{S}}(\mathbf{U})(\mathbf{V}-\mathbf{W})\|_F \\
&\geq \left(1/\|(\nabla^2 f_{\mathcal{S}}(\mathbf{U}))^{-1}\|_F - \frac{L(\|\mathbf{V}-\mathbf{U}\|_F + \|\mathbf{W}-\mathbf{U}\|_F)}{2} \right) \|\mathbf{V}-\mathbf{W}\|_F \\
&\geq (\|1/\|(\nabla^2 f_{\mathcal{S}}(\mathbf{U}))^{-1}\|_F - L\varepsilon)\|\mathbf{V}-\mathbf{W}\|_F.
\end{aligned}$$

Let $\alpha = (\|1/\|(\nabla^2 f_{\mathcal{S}}(\mathbf{U}))^{-1}\|_F - L\varepsilon)$, then The first inequality of (4.9) holds. The proof is completed. □

Theorem 4.1 *Assume that the function $f_{\mathcal{S}}(\mathbf{U})$ satisfies the* Assumption 4.1, $\{\mathcal{B}^k\}$ *and* $\{\mathbf{U}^k\}$ *are obtained by* Algorithm 3 *without line search (that is, the step size α_k is uniformly 1), and $\{\mathbf{U}^k\}$ converges to a local minimum point $\mathbf{U}^*$ of $f_{\mathcal{S}}(\mathbf{U})$. Then the point sequence $\{\mathbf{U}^k\}$ super-linear converges to $\mathbf{U}^*$ if and only if*

$$\lim_{k\to\infty} \frac{\|(\mathcal{B}^k - \nabla^2 f(\mathbf{U}^*))(\mathbf{U}^{k+1}-\mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1}-\mathbf{U}^k\|_F} = 0. \tag{4.10}$$

Proof ($\Leftarrow$) First we prove the sufficiency. Let $\mathbf{S}^k = \mathbf{U}^{k+1} - \mathbf{U}^k$. Then

$$\begin{aligned}
(\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))\mathbf{S}^k &= -\nabla f_{\mathcal{S}}(\mathbf{U}^k) - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*)\mathbf{S}^k \\
&= \nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}) - \nabla f_{\mathcal{S}}(\mathbf{U}^k) - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*)\mathbf{S}^k - \nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}).
\end{aligned}$$

According to the triangle inequality and Lemma 4.3, for any $k \geq 0$, we have that

$$\begin{aligned}
0 &\leq \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\|\mathbf{S}^k\|_F} \\
&\leq \frac{\|(\mathcal{B}_k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))\mathbf{S}^k\|_F}{\|\mathbf{S}^k\|_F} + \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}) - \nabla f_{\mathcal{S}}(\mathbf{U}^k) - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*)\mathbf{S}^k\|_F}{\|\mathbf{S}^k\|_F} \\
&\leq \frac{\|(\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))\mathbf{S}^k\|_F}{\|\mathbf{S}^k\|_F} + \frac{L}{2}(\|\mathbf{U}^k - \mathbf{U}^*\|_F + \|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F).
\end{aligned} \tag{4.11}$$

If (4.10) holds, substituting (4.10) and the assumption $\lim_{k\to\infty} \mathbf{U}^k = \mathbf{U}^*$ into (4.11), we get

$$\lim_{k\to\infty} \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\|\mathbf{S}^k\|_F} = 0. \tag{4.12}$$

Because $\nabla^2 f_{\mathcal{S}}(\mathbf{U}^*)$ is invertible and $\mathbf{U}^k \to \mathbf{U}^*$, according to Lemma 4.3, there exist a real number $\alpha > 0$ and an integer $k_0 > 0$, for any integer $k \geq k_0$, such that

$$\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F = \|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}) - \nabla f_{\mathcal{S}}(\mathbf{U}^*)\|_F \geq \alpha \|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F.$$

Thus

$$\frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \geq \frac{\alpha \|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F + \|\mathbf{U}^k - \mathbf{U}^*\|_F} = \frac{\alpha \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F}}{1 + \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F}} > 0,$$

Since the left limit of the above inequality is zero by (4.12), so the right limit of the above inequality is also zero, i.e.,

$$\lim_{k \to \infty} \frac{\frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F}}{1 + \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F}} = 0.$$

It follows that

$$\lim_{k \to \infty} \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F} = 0,$$

which means that $\{\mathbf{U}^k\}$ super-linear converges to $\mathbf{U}^*$.

($\Rightarrow$) Next we prove the necessity. By Lemma 4.3, there exist a real number $\beta > 0$ and an integer $k_0 > 0$, for any integer $k \geq k_0$, such that

$$\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F = \|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1}) - \nabla f_{\mathcal{S}}(\mathbf{U}^*)\|_F \leq \beta \|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F.$$

Because $\{\mathbf{U}^k\}$ super-linear converges $\mathbf{U}^*$, we have that

$$\begin{aligned} 0 &= \lim_{k \to \infty} \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^*\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F} \\ &\geq \lim_{k \to \infty} \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\beta \|\mathbf{U}^k - \mathbf{U}^*\|_F} \\ &= \lim_{k \to \infty} \frac{1}{\beta} \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \frac{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F}{\|\mathbf{U}^k - \mathbf{U}^*\|_F}. \end{aligned}$$

Then

$$\lim_{k \to \infty} \frac{\|\nabla f_{\mathcal{S}}(\mathbf{U}^{k+1})\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0.$$

Thus,the equation (4.10) holds. This completes the proof. □

Based on the Theorem 4.1, we give a theorem of super-linear convergence of the Algorithm 3.

Theorem 4.2 *Let* $\{\mathcal{B}^k\}$ *and* $\{\mathbf{U}^k\}$ *be obtained by* Algorithm 3. *Suppose* $f_{\mathcal{S}}(\mathbf{U})$ *satisfies the* Assumption 4.1, $\{\mathbf{U}^k\}$ *converges to local minimum point* $\mathbf{U}^*$, *and*

$$\lim_{k\to\infty} \frac{\|(\mathcal{B}^k - \nabla^2 f(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0. \tag{4.13}$$

Then, $\{\mathbf{U}^k\}$ *converges to* $\mathbf{U}^*$ *super-linearly if and only if step size* $\{\alpha_k\}$ *converges to 1.*

Proof First, we prove the sufficiency. From the trigonometric inequality, we get

$$\begin{aligned}
&\frac{\|(\alpha_k^{-1}\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \\
&\le \frac{\|(\mathcal{B}^k - \nabla^2 f(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} + \frac{\|(\alpha_k^{-1} - 1)\mathcal{B}^k(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F}.
\end{aligned}$$

If (4.13) holds, and $\{\alpha_k\}$ converges to 1, then the limit on the right side of the above inequality is equal to 0 when $k \to \infty$. It follows that

$$\lim_{k\to\infty} \frac{\|(\alpha_k^{-1}\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0. \tag{4.14}$$

By the Theorem 4.1 and (4.14), we can get that $\{\mathbf{U}^k\}$ superlinearly converges to $\mathbf{U}^*$.

Next, we prove the necessity. If $\{\mathbf{U}^k\}$ converges to $\mathbf{U}^*$ super-linearly, by the Theorem 4.1, it is followed that (4.14) holds. Since $\mathcal{B}^k(\mathbf{U}^{k+1} - \mathbf{U}^k) = -\alpha_k \nabla f_{\mathcal{S}}(\mathbf{U}^k)$, by (4.13), we have that

$$\begin{aligned}
\lim_{k\to\infty} \frac{\|(\alpha_k - 1)\nabla f_{\mathcal{S}}(\mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} &= \lim_{k\to\infty} \frac{\|(\alpha_k^{-1} - 1)\mathcal{B}^k(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \\
\le \lim_{k\to\infty} &\frac{\|(\alpha_k^{-1}\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \\
&+ \lim_{k\to\infty} \frac{\|(\mathcal{B}^k - \nabla^2 f_{\mathcal{S}}(\mathbf{U}^*))(\mathbf{U}^{k+1} - \mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0.
\end{aligned}$$

Hence,

$$\lim_{k\to\infty} \frac{\|(\alpha_k - 1)\nabla f_{\mathcal{S}}(\mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0. \tag{4.15}$$

By Assumption 4.1 and Lemma 4.3, there exist a real number $\alpha > 0$ and an integer $k_0 > 0$, for any integer $k \ge k_0$, such that

$$\alpha\|\mathbf{U}^k - \mathbf{U}^*\|_F \le \|\nabla f_{\mathcal{S}}(\mathbf{U}^k) - \nabla f_{\mathcal{S}}(\mathbf{U}^*)\|_F = \|\nabla f_{\mathcal{S}}(\mathbf{U}^k)\|_F.$$

So

$$\alpha|\alpha_k - 1| \frac{\|\mathbf{U}^k - \mathbf{U}^*\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} \le \frac{\|(\alpha_k - 1)\nabla f_{\mathcal{S}}(\mathbf{U}^k)\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F}. \tag{4.16}$$

By (4.15) and (4.16), we have that

$$\lim_{k\to\infty} \alpha|\alpha_k - 1| \frac{\|\mathbf{U}^k - \mathbf{U}^*\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 0.$$

Since again $\{\mathbf{U}^k\}$ super-linearly converges to $\mathbf{U}^*$, so

$$\lim_{k\to\infty} \frac{\|\mathbf{U}^k - \mathbf{U}^*\|_F}{\|\mathbf{U}^{k+1} - \mathbf{U}^k\|_F} = 1.$$

Hence $\lim\limits_{k\to\infty} \alpha|\alpha_k - 1| = 0$, i.e., $\lim\limits_{k\to\infty} \alpha_k = 1$. This completes the proof. □

5 Discrimination of positive partially symmetric CP decompositions

Let $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$. When all orders $m_1, m_2, \cdots, m_s$ are even, according to the Theorem 2.1, the tensor $\mathcal{S}$ may not have positive partially symmetric CP decomposition. In this case, we cannot obtain its partially symmetric CP decomposition by Algorithm 3. Therefore, we discuss how to identify whether even order partially symmetric tensors have positive partially symmetric CP decompositions in real number field.

Theorem 5.1 *Let* $\mathbf{m} = (m_1, m_2, \cdots, m_s)$*,* $\mathbf{n} = (n_1, n_2, \cdots, n_s)$*, and* $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ *be a partially symmetric tensor. For any vectors* $\mathbf{x}^{(i)} \in \mathbb{R}^{n_i}$ *and* $\mathbf{y}^{(j)} \in \mathbb{R}^{n_j}$*,* $i, j \in [s]$*, denote*

$$P^t_{\mathbf{m}}(m_t - k, k) := (\mathbf{x}^{(1)})^{m_1} \circ \cdots \circ (\mathbf{x}^{(t)})^{m_t - k} \circ (\mathbf{y}^{(t)})^k \circ \cdots \circ (\mathbf{x}^{(s)})^{m_s},$$

where $k \in \{0, 1, \cdots, m_t\}$*. Then,*

$$\langle \mathcal{S}, (\mathbf{x}^{(1)})^{m_1} \circ \cdots \circ (\mathbf{x}^{(t)} + \mathbf{y}^{(t)})^{m_t} \circ \cdots \circ (\mathbf{x}^{(s)})^{m_s} \rangle = \sum_{k=0}^{m_t} \binom{m_t}{k} \langle \mathcal{S}, P^t_{\mathbf{m}}(m_t - k, k) \rangle.$$

Proof Let $I_k := \{\mathbf{i}_t = (i_{t,1}, i_{t,2}, \cdots, i_{t,m_t}) | i_{t,1}, \cdots, i_{t,m_t} \in [n_t]\}$, $t \in [s]$. For any vectors $\mathbf{x}^{(i)} \in \mathbb{R}^{n_i}$ and $\mathbf{y}^{(j)} \in \mathbb{R}^{n_j}$, $i, j \in [s]$,

$$\begin{aligned}
&\langle \mathcal{S}, (\mathbf{x}^{(1)})^{m_1} \circ \cdots \circ (\mathbf{x}^{(t)} + \mathbf{y}^{(t)})^{m_t} \circ \cdots \circ (\mathbf{x}^{(s)})^{m_s} \rangle \\
= &\sum_{\mathbf{i}_t \in I_t, t \in [s]} \mathcal{S}_{\mathbf{i}_1 \cdots \mathbf{i}_t \cdots \mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots \\
&\cdot (\mathbf{x}^{(t)} + \mathbf{y}^{(t)})_{i_{t,1}} \cdots (\mathbf{x}^{(t)} + \mathbf{y}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s} \\
= &\sum_{\mathbf{i}_t \in I_t, t \in [s]} \mathcal{S}_{\mathbf{i}_1 \cdots \mathbf{i}_t \cdots \mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots (\mathbf{x}^{(t)})_{i_{t,1}} \cdots (\mathbf{x}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s} \\
&+ \binom{m_t}{1} \sum_{\mathbf{i}_t \in I_t, t \in [s]} \mathcal{S}_{\mathbf{i}_1 \cdots \mathbf{i}_t \cdots \mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots
\end{aligned}$$

$$
\begin{aligned}
&\cdot(\mathbf{x}^{(t)})_{i_{t,1}} \cdots (\mathbf{x}^{(t)})_{i_{t,m_t-1}} (\mathbf{y}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s} \\
&+\binom{m_t}{2} \sum_{\mathbf{i}_t \in I_t, t\in[s]} \mathcal{S}_{\mathbf{i}_1\cdots\mathbf{i}_t\cdots\mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots \\
&\cdot(\mathbf{x}^{(t)})_{i_{t,1}} \cdots (\mathbf{x}^{(t)})_{i_{t,m_t-2}} (\mathbf{y}^{(t)})_{i_{t,m_t-1}} (\mathbf{y}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s} \\
&+\cdots+\binom{m_t}{m_t} \sum_{\mathbf{i}_t \in I_t, t\in[s]} \mathcal{S}_{\mathbf{i}_1\cdots\mathbf{i}_t\cdots\mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots \\
&\cdot(\mathbf{y}^{(t)})_{i_{t,1}} \cdots (\mathbf{y}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s}.
\end{aligned}
$$

For $k = 0, 1, 2, \cdots, m_t$, it is obtained that

$$
\begin{aligned}
\langle \mathcal{S}, P^t_{\mathbf{m}}(m_t - k, k)\rangle = &\sum_{\mathbf{i}_t \in I_t, t\in[s]} \mathcal{S}_{\mathbf{i}_1\cdots\mathbf{i}_t\cdots\mathbf{i}_s} ((\mathbf{x}^{(1)})^{m_1})_{\mathbf{i}_1} \cdots (\mathbf{x}^{(t)})_{i_{t,1}} \cdots \\
&\cdot(\mathbf{x}^{(t)})_{i_{t,m_t-k}} (\mathbf{y}^{(t)})_{i_{t,m_t-k+1}} \cdots (\mathbf{y}^{(t)})_{i_{t,m_t}} \cdots ((\mathbf{x}^{(s)})^{m_s})_{\mathbf{i}_s}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\langle \mathcal{S}, (\mathbf{x}^{(1)})^{m_1} \circ \cdots \circ (\mathbf{x}^{(t)} + \mathbf{y}^{(t)})^{m_t} \circ \cdots \circ (\mathbf{x}^{(s)})^{m_s}\rangle \\
= &\sum_{k=0}^{m_t} \binom{m_t}{k} \langle \mathcal{S}, P^t_{\mathbf{m}}(m_t - k, k)\rangle.
\end{aligned}
$$

This completes the proof. □

Theorem 5.2 *Let $\mathcal{S}_1, \mathcal{S}_2 \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be two partially symmetric tensors. Assume that all orders $m_1, m_2, \cdots, m_s$ are even. Then $\langle \mathcal{S}_1, \mathcal{S}_2\rangle \geq 0$ if $\mathcal{S}_1$ and $\mathcal{S}_2$ both have positive partially symmetric CP decomposition.*

Proof Since $\mathcal{S}_1, \mathcal{S}_2 \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$, and they have positive partially symmetric CP decomposition. Hence, $\mathcal{S}_1$ and $\mathcal{S}_2$ can be represented as the following form

$$
\begin{aligned}
\mathcal{S}_1 &= \sum_{k=1}^{R_1} (\mathbf{u}_1^{(k)})^{m_1} \circ \cdots \circ (\mathbf{u}_s^{(k)})^{m_s}, \\
\mathcal{S}_2 &= \sum_{l=1}^{R_2} (\mathbf{v}_1^{(l)})^{m_1} \circ \cdots \circ (\mathbf{v}_s^{(l)})^{m_s},
\end{aligned}
$$

where $\mathbf{u}_t^{(k)}, \mathbf{v}_t^{(l)} \in \mathbb{R}^{n_t}$, $k \in [R_1]$, $l \in [R_2]$, $t \in [s]$.

Then, the inner product of $\mathcal{S}_1$ and $\mathcal{S}_2$ is

$$
\langle \mathcal{S}_1, \mathcal{S}_2\rangle = \langle \sum_{k=1}^{R_1} (\mathbf{u}_1^{(k)})^{m_1} \circ \cdots \circ (\mathbf{u}_s^{(k)})^{m_s}, \sum_{l=1}^{R_2} (\mathbf{v}_1^{(l)})^{m_1} \circ \cdots \circ (\mathbf{v}_s^{(l)})^{m_s}\rangle
$$

$$
\begin{aligned}
&= \sum_{k=1}^{R_1}\sum_{l=1}^{R_2}\langle (\mathbf{u}_1^{(k)})^{m_1}\circ\cdots\circ(\mathbf{u}_s^{(k)})^{m_s}, (\mathbf{v}_1^{(l)})^{m_1}\circ\cdots\circ(\mathbf{v}_s^{(l)})^{m_s}\rangle \\
&= \sum_{k=1}^{R_1}\sum_{l=1}^{R_2}\left(\langle \mathbf{u}_1^{(k)}, \mathbf{v}_1^{(l)}\rangle\right)^{m_1}\cdots\left(\langle \mathbf{u}_s^{(k)}, \mathbf{v}_s^{(l)}\rangle\right)^{m_s}.
\end{aligned}
$$

Since orders $m_1, m_2, \cdots, m_s$ are even, $\left(\langle \mathbf{u}_t^{(k)}, \mathbf{v}_t^{(l)}\rangle\right)^{m_t} \geq 0$, for all $k \in [R_1], l \in [R_2], t \in [s]$. Hence, it is followed that $\langle \mathcal{S}_1, \mathcal{S}_2\rangle \geq 0$. The proof is completed. □

Assume that $\{U_1, U_2, \cdots, U_s\}$ is a tuple of factor matrices of $\hat{\mathcal{S}} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$, i.e.,

$$
\hat{\mathcal{S}} = [\![U_1^{\times m_1}, U_2^{\times m_2}, \cdots, U_s^{\times m_s}]\!].
$$

In the next, we see the object function $F_{\mathcal{S}}(\hat{\mathcal{S}}) = \|\mathcal{S} - \hat{\mathcal{S}}\|_F^2$ as a function of factor matrix $U_t, t \in [s]$, i.e.,

$$
F_{\mathcal{S}}(U_t) = F_{\mathcal{S}}(\hat{\mathcal{S}}) = \|\mathcal{S} - \hat{\mathcal{S}}\|_F^2. \tag{5.1}
$$

For convenience, we denote

$$
\begin{aligned}
&\hat{\mathcal{S}}[U_t^{\times k}, \Delta U_t{}^{\times(m_t-k-l)}, U_t^{\times l}] \\
&:= [\![U_1^{\times m_1}, \cdots, U_{t-1}^{\times m_{t-1}}, U_t^{\times k}, \Delta U_t{}^{\times(m_t-k-l)}, U_t^{\times l}, U_{t+1}^{\times m_{t+1}}, \cdots, U_s^{\times m_s}]\!].
\end{aligned}
$$

Next, in order to study the convexity of the function $F_{\mathcal{S}}(U_t)$, we derive a second-order Taylor formula of $F_{\mathcal{S}}(U_t)$.

Theorem 5.3 *Let $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be a partially symmetric tensor and $F_{\mathcal{S}}(U_t)$ is defined as (5.1). Let*

$$
L_1 := -2\binom{m_t}{1}\langle \mathcal{S} - \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\rangle, \tag{5.2}
$$

$$
\begin{aligned}
L_2 :=& -2\binom{m_t}{2}\langle \mathcal{S} - \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle \\
&+\|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(k-1)}] + \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(k-2)}] + \cdots + \hat{\mathcal{S}}[U_t^{\times(k-1)}, \Delta U_t]\|_F^2.
\end{aligned} \tag{5.3}
$$

Then, the second-order Taylor formula of the function $F_{\mathcal{S}}(U_t)$ is as follows

$$
F_{\mathcal{S}}(U_t + \Delta U_t) = F_{\mathcal{S}}(U_t) + L_1 + L_2 + o(\|\Delta U_t\|_F^2). \tag{5.4}
$$

Proof From (5.1), $F_{\mathcal{S}}(U_t + \Delta U_t) = \|\mathcal{S} - \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\|_F^2$. Then,

$$
F_{\mathcal{S}}(U_t + \Delta U_t) = \langle \mathcal{S}, \mathcal{S}\rangle - 2\langle \mathcal{S}, \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\rangle
$$

$$+\langle \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}], \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\rangle. \tag{5.5}$$

Now, we compute the first degree term and the second degree term of $\triangle U_t$ in the right hand side of (5.5). Since,

$$\langle \mathcal{S}, \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\rangle = \langle \mathcal{S}, \sum_{k=1}^{R} (\mathbf{u}_1^{(k)})^{m_1} \circ \cdots \circ (\mathbf{u}_t^{(k)} + \Delta \mathbf{u}_t^{(k)})^{m_t} \circ \cdots \circ (\mathbf{u}_s^{(k)})^{m_s}\rangle.$$

According to the Theorem 5.1, we can get that

$$\begin{aligned}
&\langle \mathcal{S}, \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\rangle \\
&= \langle \mathcal{S}, \hat{\mathcal{S}}\rangle + \binom{m_t}{1}\langle \mathcal{S}, \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\rangle \\
&\quad + \binom{m_t}{2}\langle \mathcal{S}, \hat{\mathcal{S}}[\Delta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle + o(\|\Delta U_t\|_F^2).
\end{aligned} \tag{5.6}$$

Since,

$$\begin{aligned}
&\|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}] + \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(m_t-2)}] + \cdots + \hat{\mathcal{S}}[U_t^{\times(m_t-1)}, \Delta U_t]\|_F^2 \\
&= \binom{m_t}{1}\binom{m_t-1}{1}\langle \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}], \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(m_t-2)}]\rangle \\
&\quad + \binom{m_t}{1}\|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\|_F^2.
\end{aligned} \tag{5.7}$$

And, for all $k, l \in [R]$ and $t \in [s]$,

$$\begin{aligned}
&\langle (\mathbf{u}_t^{(k)} + \Delta \mathbf{u}_t^{(k)})^{m_t}, (\mathbf{u}_t^{(l)} + \Delta \mathbf{u}_t^{(l)})^{m_t}\rangle \\
&= \left(\langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle + \langle \Delta \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle + \langle \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle + \langle \Delta \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle\right)^{m_t} \\
&= \langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^{m_t} + \binom{m_t}{1}\left(\langle \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle + \langle \Delta \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle\right)\langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^{m_t-1} \\
&\quad + \binom{m_t}{1}\langle \Delta \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle\langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^{m_t-1} \\
&\quad + \binom{m_t}{1}\binom{m_t-1}{1}\langle \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle\langle \Delta \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle\langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^{m_t-2} \\
&\quad + \binom{m_t}{2}\left(\langle \mathbf{u}_t^{(k)}, \Delta \mathbf{u}_t^{(l)}\rangle^2 + \langle \Delta \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^2\right)\langle \mathbf{u}_t^{(k)}, \mathbf{u}_t^{(l)}\rangle^{m_t-2} \\
&\quad + o\left(\|\Delta \mathbf{u}_t^{(l)}\|_F^2 + \|\Delta \mathbf{u}_t^{(k)}\|_F^2\right).
\end{aligned} \tag{5.8}$$

Let

$$\mathcal{P}_k := (\mathbf{u}_1^{(k)})^{m_1} \circ \cdots \circ (\mathbf{u}_{t-1}^{(k)})^{m_{t-1}}, k \in [R],$$

$$\mathcal{Q}_k := (\mathbf{u}_{t+1}^{(k)})^{m_{t+1}} \circ \cdots \circ (\mathbf{u}_s^{(k)})^{m_s}, k \in [R].$$

Then, by (5.7) and (5.8), we have that

$$\begin{aligned}
&\langle \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}], \hat{\mathcal{S}}[(U_t + \Delta U_t)^{\times m_t}]\rangle \\
&= \sum_{k=1}^{R}\sum_{l=1}^{R} \langle \mathcal{P}_k \circ (\mathbf{u}_t^{(k)} + \Delta \mathbf{u}_t^{(k)})^{m_t} \circ \mathcal{Q}_k, \mathcal{P}_l \circ (\mathbf{u}_t^{(l)} + \Delta \mathbf{u}_t^{(l)})^{m_t} \circ \mathcal{Q}_l\rangle \\
&= \sum_{k=1}^{R}\sum_{l=1}^{R} \langle \mathcal{P}_k, \mathcal{P}_l\rangle \langle (\mathbf{u}_t^{(k)} + \Delta \mathbf{u}_t^{(k)})^{m_t}, (\mathbf{u}_t^{(l)} + \Delta \mathbf{u}_t^{(l)})^{m_t}\rangle \langle \mathcal{Q}_k, \mathcal{Q}_l\rangle \\
&= \langle \hat{\mathcal{S}}, \hat{\mathcal{S}}\rangle + 2\binom{m_t}{1}\langle \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\rangle + \binom{m_t}{1}\|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\|_F^2 \\
&\quad + \binom{m_t}{1}\binom{m_t-1}{1}\langle \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}], \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(m_t-2)}]\rangle \\
&\quad + 2\binom{m_t}{2}\langle \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle + o(\|\Delta U_t\|_F^2). \\
&= \langle \hat{\mathcal{S}}, \hat{\mathcal{S}}\rangle + 2\binom{m_t}{1}\langle \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\rangle + 2\binom{m_t}{2}\langle \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle \\
&\quad + \|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(k-1)}] + \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(k-2)}] + \cdots + \hat{\mathcal{S}}[U_t^{\times(k-1)}, \Delta U_t]\|_F^2 \\
&\quad + o(\|\Delta U_t\|_F^2).
\end{aligned} \tag{5.9}$$

Hence, by (5.6) and (5.9), the first degree term and the second degree term of ΔU_t in the right hand side of (5.5) are as follows, respectively.

$$\begin{aligned}
L_1 &= -2\binom{m_t}{1}\langle \mathcal{S} - \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t, U_t^{\times(m_t-1)}]\rangle, \\
L_2 &= -2\binom{m_t}{2}\langle \mathcal{S} - \hat{\mathcal{S}}, \hat{\mathcal{S}}[\Delta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle \\
&\quad + \|\hat{\mathcal{S}}[\Delta U_t, U_t^{\times(k-1)}] + \hat{\mathcal{S}}[U_t, \Delta U_t, U_t^{\times(k-2)}] + \cdots + \hat{\mathcal{S}}[U_t^{\times(k-1)}, \Delta U_t]\|_F^2.
\end{aligned}$$

This completes the proof. □

It is well known that (1) if U_t is a stationary point of function $F_{\mathcal{S}}(U_t)$, then $L_1 = 0$ for any ΔU_t; (2) the function $F_{\mathcal{S}}(U_t)$ is convex at the point U_t if and only if $L_2 \geq 0$ for any ΔU_t. From this, we will discuss the discrimination of partially symmetric tensors with the positive decomposition.

Theorem 5.4 *Assume that orders $m_1, \cdots, m_s$ are even and $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ has positive partially symmetric CP decomposition. Then, zero tensor $\mathcal{O} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ is the global minimum point of the function $F_{-\mathcal{S}}(\hat{\mathcal{S}})$, and the global minimum value is equal to $F_{-\mathcal{S}}(\mathcal{O}) = \|\mathcal{S}\|_F^2$.*

Proof Since $F_{-\mathcal{S}}(U_t) = \|\mathcal{S} + \hat{\mathcal{S}}\|_F^2$. By Theorem 5.3, L_1 and L_2 of $F_{-\mathcal{S}}(U_t)$ are as follows

$$\begin{aligned}L_1 &= 2\binom{m_t}{1}\langle \mathcal{S} + \hat{\mathcal{S}}, \hat{\mathcal{S}}[\varDelta U_t, U_t^{\times(m_t-1)}]\rangle,\\ L_2 &= 2\binom{m_t}{2}\langle \mathcal{S} + \hat{\mathcal{S}}, \hat{\mathcal{S}}[\varDelta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle\\ &\quad +\|\hat{\mathcal{S}}[\varDelta U_t, U_t^{\times(k-1)}] + \hat{\mathcal{S}}[U_t, \varDelta U_t, U_t^{\times(k-2)}] + \cdots + \hat{\mathcal{S}}[U_t^{\times(k-1)}, \varDelta U_t]\|_F^2.\end{aligned}$$

Since $\mathcal{S}$ has positive partially symmetric CP decomposition, and orders $m_1, m_2, \cdots, m_s$ are even. According to Theorem 5.2, if $U_i \neq \mathbf{0}$ for all $i \in [s],\ i \neq t$, then $\langle \mathcal{S}, \hat{\mathcal{S}}[\varDelta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle$ is nonnegative and $\langle \hat{\mathcal{S}}, \hat{\mathcal{S}}[\varDelta U_t{}^{\times 2}, U_t^{\times(m_t-2)}]\rangle$ is positive if $U_t \neq \mathbf{0}$. Hence, L_2 is positive at U_t for any $\varDelta U_t$, i.e., the function $F_{-\mathcal{S}}(U_t)$ is convex for every $t \in [s]$ if $U_i \neq \mathbf{0}$ for all $i \in [s],\ i \neq t$. Since $\hat{\mathcal{S}} = \mathcal{O}$ if and only if there exists $U_i = \mathbf{0}$ for some $i \in [s]$. Hence, $\hat{\mathcal{S}} = \mathcal{O}$ is the global minimum point of the function $F_{-\mathcal{S}}(\hat{\mathcal{S}})$ and the global minimum value is $F_{-\mathcal{S}}(\mathcal{O}) = \|\mathcal{S}\|_F^2$. This completes the proof. □

Assume that $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ be a partially symmetric tensor and its orders are even. From the Theorem 5.4, $U_t = \mathbf{0}$ is the global minimum point of the function $F_{-\mathcal{S}}(U_t)$, and the global minimum value is equal to $\|\mathcal{S}\|_F^2$, if partially symmetric tensor $\mathcal{S}$ has positive CP decomposition. According to Theorem 5.4, we have the corollary in the following.

Corollary 5.1 *Assume that $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ and its orders are even. The function $F_{-\mathcal{S}}(\hat{\mathcal{S}})$ is defined as in (5.1). Then, the partially symmetric tensor $\mathcal{S}$ has no positive partially symmetric CP decomposition, if zero tensor $\mathcal{O}$ is not the global minimum point of the function $F_{-\mathcal{S}}(\hat{\mathcal{S}})$.*

Proof These results can be obtained directly form Theorem 5.4. □

When dealing with partially symmetric rank-R approximation of $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ with even orders, i.e., m_t for all $t \in [s]$ are even, we construct a new tensor $\tilde{\mathcal{S}} \in S[1, \mathbf{m}]\mathbb{R}[1, \mathbf{n}]$ satisfying

$$\tilde{\mathcal{S}}_{1 i_1 \cdots i_{m_1} j_1 \cdots j_{m_2} \cdots l_1 \cdots l_{m_s}} = \mathcal{S}_{i_1 \cdots i_{m_1} j_1 \cdots j_{m_2} \cdots l_1 \cdots l_{m_s}}.$$

We compute the partially symmetric rank-R approximation of $\tilde{\mathcal{S}}$ by the Algorithm 3 as

$$\tilde{\mathcal{S}} \approx [\![\tilde{U}_0, \tilde{U}_1^{\times m_1}, \cdots, \tilde{U}_s^{\times m_s}]\!],$$

where $\tilde{U}_t = \left(\tilde{\mathbf{u}}_t^{(1)}, \tilde{\mathbf{u}}_t^{(2)}, \cdots, \tilde{\mathbf{u}}_t^{(R)}\right)$, $t \in [s]$. Compute $\mathbf{u}_k^{(t)}$ and λ_k, $t \in [s]$, $k \in [R]$ by the following formula

$$\lambda_k = \tilde{u}_k^{(0)} \prod_{t=1}^{s} \|\tilde{\mathbf{u}}_k^{(t)}\|_2^{m_t}, \text{ and } \mathbf{u}_k^{(t)} = \tilde{\mathbf{u}}_k^{(t)} / \|\tilde{\mathbf{u}}_k^{(t)}\|_2.$$

So, we obtain that

$$\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \cdots, \boldsymbol{\lambda}_R)^T, \; U_t = \left(\mathbf{u}_t^{(1)}, \mathbf{u}_t^{(2)}, \cdots, \mathbf{u}_t^{(R)}\right), \; t \in [s].$$

That is, the general partially symmetric rank-R approximation of $\mathcal{S}$ is

$$\mathcal{S} \approx [\![\boldsymbol{\lambda}; U_1^{\times m_1}, \cdots, U_s^{\times m_s}]\!].$$

We give an algorithm to calculate general partially symmetric rank-R approximation of partially symmetric tensors with BFGS algorithm as follows.

Algorithm 4 General partially symmetric rank-R approximation of partially symmetric tensors

Input: A partially symmetric tensor $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$. A positive integer R.
Output: A general partially symmetric rank-R approximation of $\mathcal{S}$
Step 1: Construct a new tensor $\tilde{\mathcal{S}} \in S[1, \mathbf{m}]\mathbb{R}[1, \mathbf{n}]$ with

$$\tilde{\mathcal{S}}_{1 i_1 \cdots i_{m_1} j_1 \cdots j_{m_2} \cdots l_1 \cdots l_{m_s}} = \mathcal{S}_{i_1 \cdots i_{m_1} j_1 \cdots j_{m_2} \cdots l_1 \cdots l_{m_s}}.$$

Step 2: Compute the partially symmetric rank-R approximation of $\tilde{\mathcal{S}}$ by Algorithm 3,

$$\tilde{\mathcal{S}} \approx [\![\tilde{U}_0, \tilde{U}_1^{\times m_1}, \cdots, \tilde{U}_s^{\times m_s}]\!],$$

Step 3: Compute $\mathbf{u}_k^{(t)}$ and λ_k, $t \in [s]$, $k \in [R]$, by

$$\lambda_k = \tilde{u}_k^{(0)} \prod_{t=1}^{s} \|\tilde{\mathbf{u}}_k^{(t)}\|_2^{m_t}, \;\; \mathbf{u}_k^{(t)} = \tilde{\mathbf{u}}_k^{(t)} / \|\tilde{\mathbf{u}}_k^{(t)}\|_2.$$

Step 4: Obtain $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_R)^T, \; U_t = \left(\mathbf{u}_t^{(1)}, \mathbf{u}_t^{(2)}, \cdots, \mathbf{u}_t^{(R)}\right), \; t \in [s]$.
Step 5: Return $\{\lambda, U_1, \ldots, U_s\}$, the general partially symmetric rank-R approximation

$$\mathcal{S} \approx [\![\boldsymbol{\lambda}; U_1^{\times m_1}, \cdots, U_s^{\times m_s}]\!].$$

Table 1 The minimum values of $F_{-\mathcal{S}}(\hat{\mathcal{S}})$ for $R = 1, 2, \cdots, 6$

	$R=1$	$R=2$	$R=3$	$R=4$	$R=5$	$R=6$
$\min F_{-\mathcal{S}}(\hat{\mathcal{S}})$	44.7896	39.1948	38.7092	38.7092	38.092	38.7092
$\Vert\hat{\mathcal{S}}\Vert_F$	2.3677	3.3274	3.3996	3.3996	3.3996	3.3996
run-time	0.1357	0.1799	0.2175	0.2222	0.2348	0.2989

Table 2 The positive partially symmetric rank-R approximation of $\mathcal{S}$

	$R=1$	$R=2$	$R=3$	$R=4$	$R=5$	$R=6$
$\min F_{\mathcal{S}}(\hat{\mathcal{S}})$	26.6158	14.9132	14.2719	14.1711	14.1711	14.1711
run-time	0.2049	0.3430	0.6942	0.5885	0.3417	0.2971

6 Numerical examples

We give some examples in this section. All experiments on a laptop with an Intel(R) Core(TM) i7-8550U CPU and 8.00 GB of RAM, using MATLAB 2018b on a Microsoft Win 10. All tensor computations use the tensor toolbox for MATLAB, Version [29].

Example 6.1 (A random tensor) Consider a partially symmetric CP decomposition of a tensor $\mathcal{S} \in S[2, 2]\mathbb{R}[3, 2]$. It is obtained randomly as follows

$$\mathcal{S}(:, :, 1, 1) = \begin{pmatrix} 1.4897 & 0.6715 & 1.6302 \\ 0.6715 & 1.2075 & 0.4889 \\ 1.6302 & 0.4889 & 1.0347 \end{pmatrix},$$

$$\mathcal{S}(:, :, 1, 2) = \begin{pmatrix} 0.7296 & 0.7873 & -1.0689 \\ 0.7873 & 0.8884 & -0.8095 \\ -1.0689 & -0.8095 & -2.9443 \end{pmatrix},$$

$$\mathcal{S}(:, :, 2, 2) = \begin{pmatrix} 1.4384 & 1.3703 & -0.2412 \\ 1.3703 & -1.7115 & 0.3192 \\ -0.2412 & 0.3192 & 0.3129 \end{pmatrix}.$$

Firstly, we discuss whether there is a positive partially symmetric CP decomposition of tensor $\mathcal{S}$. We compute the minimum value of $F_{-\mathcal{S}}(\hat{\mathcal{S}})$ defined as in (5.1) for all $R = 1, 2 \cdots, 6$ by Algorithm 3 and obtain the minimum values $\min F_{-\mathcal{S}}(\hat{\mathcal{S}})$, norms of the minimum points $\Vert\hat{\mathcal{S}}\Vert_F$ and running time as in Table 1. From the datum in Table 1, it can be seen that the minimum points $\hat{\mathcal{S}}$ are not zero tensor. Hence, tensor $\mathcal{S}$ has no positive partially symmetric CP decomposition by Corollary 5.1.

Secondly, we compute the positive partially symmetric rank-R approximation of $\mathcal{S}$ by the Algorithm 3 for all $R = 1, 2 \cdots, 6$. We obtain in the minimum values min $F_{\mathcal{S}}(\hat{\mathcal{S}})$ and running time as in Table 2. It can be seen from Table 2 that the minimum values of the objective function are always greater than 14.

Finally, we consider the general partially symmetric CP decomposition of tensor $\mathcal{S}$. We compute the partially symmetric rank-R approximation of the tensor $\mathcal{S}$ by the

Table 3 The general partially symmetric rank-R approximation of $\mathcal{S}$

	$R=1$	$R=2$	$R=3$	$R=4$	$R=5$	$R=6$
$\min F_{\mathcal{S}}(\hat{\mathcal{S}})$	26.6158	14.9132	7.3878	0.3930	0.0116	2.6001e−15
run-time	0.2496	0.2999	0.6163	0.5746	0.5476	0.9072

Table 4 A general partially symmetric CP decomposition $\hat{\mathcal{S}}$

λ_k	−4.3145	−2.4363	−1.1755	0.7307	5.4125	5.5547
$\mathbf{u}_k$	0.2957	−0.0615	−0.1442	0.6882	−0.3700	−0.6367
	0.1253	−0.8871	0.9895	−0.6924	−0.1581	−0.6216
	−0.9470	0.4575	0.0088	−0.2169	−0.9155	0.4563
$\mathbf{v}_k$	0.8793	−0.0408	−0.0831	0.9564	0.8834	−0.6275
	0.4763	−0.9992	−0.9965	0.2919	−0.4685	−0.7787

Algorithm 4 for $R = 1, 2, \cdots, 6$ and obtain the minimum values $\min F_{\mathcal{S}}(\hat{\mathcal{S}})$ and running time as in Table 3.

Therefore, we obtain a general partially symmetric CP decomposition of the tensor $\mathcal{S}$ with $R = 6$ as

$$\hat{\mathcal{S}} = \sum_{k=1}^{6} \lambda_k (\mathbf{u}_k)^2 \circ (\mathbf{v}_k)^2,$$

where $\lambda_k \in \mathbb{R}$, $\mathbf{u}_k \in \mathbb{R}^3$ and $\mathbf{v}_k \in \mathbb{R}^2$ for $k = 1, 2, \cdots, 6$ are as in the Table 4

Example 6.2 (Comparison of convergence speed between the gradient descent method and the BFGS method) Partially symmetric tensor $\mathcal{S} \in S[1, 2]\mathbb{R}[n_1, n_2]$ is given by $\mathcal{S} = [\![U_1, U_2^{\times 2}]\!]$, where $U_1 \in \mathbb{R}^{n_1 \times R}$, $U_2 \in \mathbb{R}^{n_2 \times R}$ are randomly obtained. We use Algorithm 2 and Algorithm 3 to calculate the rank-R approximation of tensor $\mathcal{S}$, that is, its structure preserving CP decomposition. The termination condition of both methods are that the norm of gradient is less 10^{-5}, or the number of iteration reaches 5000. We calculate four types of tensors: $(n_1, n_2, R) = (5, 5, 3), (5, 5, 5), (10, 10, 5), (10, 10, 10)$. For each type, we calculate 10 times. The results of numerical calculation are shown in Table 5, where ‘Iter’ denotes the average number of iterations, ‘Time’ denotes average running time and ‘Error’ denotes average value of $\|\mathcal{S} - \hat{\mathcal{S}}\|_F$, respectively. Form Table 5, we see that the running time of BFGS method is less than that of gradient descent method.

Example 6.3 (Rank partially symmetric CP decomposition) Assume orders $\mathbf{m}$ of $\mathcal{S} \in S[\mathbf{m}]\mathbb{R}[\mathbf{n}]$ are even. If the tensor $\mathcal{S}$ has a positive partially symmetric CP decomposition, must its rank partially symmetric CP decomposition be a positive CP decomposition? Here, we assume that the tensor $\mathcal{S} \in S[2, 2, 2]\mathbb{R}[3, 4, 5]$ has the following form

$$\mathcal{S} = \sum_{k=1}^{r} (\mathbf{u}_k)^2 \circ (\mathbf{v}_k)^2 \circ (\mathbf{w}_k)^2,$$

Table 5 Comparison of gradient descent method and BFGS method

	BFGS Method			Gradient Descent Method		
(n_1, n_2, R)	Iter	Time	Error	Iter	Time	Error
(5,5,3)	88	0.4958	2.8621e−06	1784	13.9672	7.8444e−06
(5,5,5)	139	0.8539	2.3775e−06	4327	34.3292	1.1041e−05
(10,10,5)	196	1.2396	1.8571e−06	1567	14.6899	2.9783e−06
(10,10,10)	302	1.9261	1.6014e−06	3024	28.3224	4.2299e−06

Table 6 A random tensor $\mathcal{S}$

$\mathbf{u}_k$	0.5879	0.0573	−0.5386	−0.4456	0.2703
	0.2569	0.6998	0.2485	0.8666	−0.2906
	−0.7671	−0.7121	−0.8051	−0.2247	−0.9178
$\mathbf{v}_k$	−0.5227	0.4239	0.4971	−0.1866	−0.8625
	−0.6135	−0.8625	0.0352	0.2793	−0.0620
	0.2197	0.1632	0.8410	0.6191	0.0810
	−0.5496	−0.2233	−0.2105	−0.7099	0.4957
$\mathbf{w}_k$	−0.1873	−0.0872	0.0549	0.3834	−0.3450
	0.0660	−0.6735	0.4485	−0.1169	−0.5523
	0.4818	0.1916	−0.2199	0.1492	−0.3000
	−0.4444	−0.3929	0.4977	−0.3105	−0.3327
	−0.7286	−0.5897	0.7070	0.8489	−0.6126

Table 7 The positive partially symmetric rank-R approximation of $\mathcal{S}$

	$R = 1$	$R = 2$	$R = 3$	$R = 4$	$R = 5$
$\min F_{\mathcal{S}}(\hat{\mathcal{S}})$	3.8551	2.8202	1.7086	0.7745	3.8329e−14
run-time	0.6164	0.8543	1.2143	3.6154	3.1265

where $\mathbf{u}_k \in \mathbb{R}^3$, $\mathbf{v}_k \in \mathbb{R}^4$ and $\mathbf{w}_k \in \mathbb{R}^5$ are generated randomly unit vectors for all $k \in [r]$.

Let $r = 5$. The tensor $\mathcal{S}$ is obtained randomly as in Table 6.

We first compute the partially symmetric rank of $\mathcal{S}$ by Algorithm 4. We compute the general partially symmetric rank-R approximation by the Algorithm 4 for $R = 1, 2, \cdots, 5$ as in Table 7. It is clear that the partially symmetric rank of tensor $\mathcal{S}$ is equal to 5.

Furthermore, we obtain a general partially symmetric rank decomposition of the tensor $\mathcal{S}$ with $R = 5$ as in Table 8. It is observed that the rank partially symmetric CP decomposition is also a positive CP decomposition. When we take $r = 10$ or 15, the results are the same. Hence, we have the result that if a tensor $\mathcal{S}$ has a positive partially symmetric CP decomposition, then its rank partially symmetric CP decomposition must be a positive CP decomposition.

Table 8 A rank partially symmetric CP decomposition of $\mathcal{S}$

λ_k	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathbf{u}_k$	0.5879	0.0573	−0.5386	−0.4456	0.2703
	0.2569	0.6998	0.2485	0.8666	−0.2906
	−0.7671	−0.7121	−0.8051	−0.2247	−0.9178
$\mathbf{v}_k$	−0.5227	0.4239	0.4971	−0.1866	−0.8625
	−0.6135	−0.8625	0.0352	0.2793	−0.0620
	0.2197	0.1632	0.8410	0.6191	0.0810
	−0.5496	−0.2233	−0.2105	−0.7099	0.4957
$\mathbf{w}_k$	−0.1873	−0.0872	0.0549	0.3834	−0.3450
	0.0660	−0.6735	0.4485	−0.1169	−0.55233
	0.4818	0.1916	−0.2199	0.1492	−0.3000
	−0.4444	−0.3929	0.4977	−0.3105	−0.3327
	−0.7286	−0.5897	0.7070	0.8489	−0.6126

Example 6.4 (Comparison with ALS method) Given a random nonsymmetric tensor $\mathcal{S} \in S[1, 1, 1]\mathbb{R}[3, 4, 5]$ as in the following. We compare the compute efficient of the ALS method and the BFGS method.

$$\mathcal{S}(1, :, :) = \begin{pmatrix} 0.8147 & 0.9572 & 0.6787 & 0.6948 & 0.7094 \\ 0.9134 & 0.1419 & 0.3922 & 0.0344 & 0.6797 \\ 0.2785 & 0.7922 & 0.7060 & 0.7655 & 0.1189 \\ 0.9649 & 0.0357 & 0.0462 & 0.4898 & 0.3403 \end{pmatrix},$$

$$\mathcal{S}(2, :, :) = \begin{pmatrix} 0.9058 & 0.4854 & 0.7577 & 0.3171 & 0.7547 \\ 0.6324 & 0.4218 & 0.6555 & 0.4387 & 0.6551 \\ 0.5469 & 0.9595 & 0.0318 & 0.7952 & 0.4984 \\ 0.1576 & 0.8491 & 0.0971 & 0.4456 & 0.5853 \end{pmatrix},$$

$$\mathcal{S}(3, :, :) = \begin{pmatrix} 0.1270 & 0.8003 & 0.7431 & 0.9502 & 0.2760 \\ 0.0975 & 0.9157 & 0.1712 & 0.3816 & 0.1626 \\ 0.9575 & 0.6557 & 0.2769 & 0.1869 & 0.9597 \\ 0.9706 & 0.9340 & 0.8235 & 0.6463 & 0.2238 \end{pmatrix}.$$

We compute the CP-rank $\operatorname{rank}(\mathcal{S}) = 7$ by Algorithm 3. Hence, we take $R = 7$ to compute its CP decomposition ten times by the ALS method and the BFGS method, respectively. For each time, the initial point is obtained randomly. For the ALS method, the program code "cp_als" comes from the tensor toolbox [29]. The termination condition of both algorithms is that the iterative steps reaches 10,000 or the absolute error $\|\mathcal{S} - \hat{\mathcal{S}}\|_F$ is less than 10^{-5}. We obtain the number of iteration 'it-number', running time and the absolute error $\|\mathcal{S} - \hat{\mathcal{S}}\|_F$ in Table 9. The average running time of Algorithm 3 and ALS algorithm is 1.4743 seconds and 8.8574 seconds, respectively. The numerical results show that the BFGS method is more effective than the ALS method in terms of calculation accuracy and calculation speed.

Table 9 Comparison between ALS method and BFGS method

BFGS Method			ALS Method		
it-number	run-time	$\|\mathcal{S}-\hat{\mathcal{S}}\|_F$	it-number	run-time	$\|\mathcal{S}-\hat{\mathcal{S}}\|_F$
196	1.5016	8.6112e−06	1412	2.4816	9.9639e−06
288	2.1853	9.7341e−06	10,000	15.1429	0.0105
104	0.7118	7.9662e−06	6192	9.2558	9.9998e−06
112	0.8293	7.4256e−06	8478	12.3807	9.9960e−06
116	0.9340	7.1936e−06	10,000	14.8388	0.0016
123	0.9170	9.6007e−06	10,000	14.9394	0.0168
526	3.9134	4.9603e−06	10,000	14.9196	0.0130
141	1.0186	6.0423e−06	1288	1.9106	9.9485e−06
187	1.4230	6.9865e−06	315	0.4857	9.9445e−06
174	1.3086	8.3163e−06	1495	2.2186	9.9702e−06

Table 10 Comparison with stability between ALS method and BFGS method

Tensor types			BFGS		ALS	
m	**n**	R	success	run-time	success	run-time
4	6	12	10	14.3239	3	67.7969
4	10	12	10	21.3027	7	41.4185
2, 2	6, 6	12	10	31.1316	6	63.1839
2, 4	6, 5	12	8	350.2533	1	383.3560
2, 2, 2	6, 6, 6	12	9	368.0565	4	311.4647
2, 2, 4	6, 4, 4	8	10	738.7009	2	1027.4193

Example 6.5 (Comparison with ALS method) We compare the stability and computing speed of the BFGS algorithm and the ALS algorithm. The BFGS algorithm is the structure preserving CP decomposition method proposed in this paper, while the ALS algorithm is the usual CP decomposition method and does not have the structure preserving property. In the numerical example, partially symmetric tensors have the following form

$$\mathcal{S}=\sum_{k=1}^{R}(\mathbf{u}_k)^{m_1}\circ(\mathbf{v}_k)^{m_2}\circ(\mathbf{w}_k)^{m_3},$$

where $\mathbf{u}_k\in\mathbb{R}^{n_1}$ and $\mathbf{v}_k\in\mathbb{R}^{n_2}$ and $\mathbf{w}_k\in\mathbb{R}^{n_3}$ are generated randomly, for all $k\in[R]$.

For each parameter tuple $(\mathbf{m},\mathbf{n},R)$, we randomly generate tensor $\mathcal{S}$ ten times, and use Algorithm 3 and the ALS algorithm to calculate their CP decomposition respectively, where the program code "cp_als" comes from the tensor toolbox [29]. The termination condition of both algorithms is that the iterative steps reaches 10,000 or the relative error $\|\mathcal{S}-\hat{\mathcal{S}}\|_F/\|\mathcal{S}\|_F$ is less than 10^{-5}. We say a run is successful, if the relative error is less than 10^{-5}. We obtain the total running time and success times of each tuple $(\mathbf{m},\mathbf{n},R)$ in Table 10. The numerical results show that the BFGS method is more stable and faster than the ALS method.

7 Conclusion

In this paper, we study the numerical problem of structure preserving rank-R approximation and structure preserving CP decomposition of partially symmetric tensors. For the problem of structure preserving rank-R approximation, we deduce the gradient formula of the objective function, obtain the BFGS iterative formula with tensor form, propose a BFGS algorithm for positive partially symmetric rank-R approximation, and discuss the convergence of the algorithm. For the problem of structure preserving CP decomposition, we give a necessary condition for partially symmetric tensors with even orders to have positive partially symmetric CP decomposition, and design a general partially symmetric rank-R algorithm to obtain structure preserving CP decomposition. Finally, some numerical examples are given. We compute the partially symmetric CP decomposition of the random partially symmetric tensors. By some numerical examples, we find that if a tensor has a positive partially symmetric CP decomposition then its partially symmetric rank CP decomposition must be a positive CP decomposition. Meanwhile, in some numerical examples, we compare the BFGS algorithm proposed in this paper with the standard CP-ALS method.

When $m_1, m_2, \ldots, m_s$ are all even numbers, it is difficult to judge and obtain the positive partially symmetric CP decomposition of $\mathcal{S}$. In particular, when $m_1 = m_2 = \ldots = m_s = 2$, the tensor $\mathcal{S}$ can be regarded as a real Hermitian tensor, and $\mathcal{S}$ has real Hermitian separability if and only if $\mathcal{S}$ has a positive partially symmetric CP decomposition, see reference [30].

Acknowledgements The authors would like to thank two anonymous referees for their valuable suggestions, which helped us to improve this manuscript. This work is supported by the National Natural Science Foundation of China (No. 11871472).

Data Availability We do not analyse or generate any datasets, because our work proceeds within a theoretical and mathematical approach.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Feng, J., Yang, L., Zhu, Q., Choo, K.: Privacy preserving tensor decomposition over encrypted data in a federated cloud environment. IEEE T. Depend. Secure. **17**, 857–868 (2020)
2. Song, Q., Ge, H., Caverlee, J., Xu, X.: Tensor completion algorithms in big data analytics. ACM T. Knowl. Discov. D. **13**, 1–48 (2019)
3. Cichocki, A., Mandic, D.P., Lathauwer, L.D., Zhou, G., Zhao, Q., Caiafa, C.F.: Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Proc. Mag. **32**, 145–163 (2015)
4. Sidiropoulos, N., Lathauwer, L.D., Fu, X., Huang, K., Papalexakis, E.E., Faloutsos, C.: Tensor decomposition for signal processing and machine learning. IEEE Trans. Signal Proces. **65**, 3551–3582 (2017)
5. Veganzones, M.A., Cohen, J.E., Cabral, F.R., Chanussot, J., Comon, P.: Nonnegative tensor CP decomposition of hyperspectral data. IEEE T. Geosci. Remote. **54**, 2577–2588 (2016)

6. Xu, Y., Wu, Z., Chanussot, J., Comon, P., Wei, Z.: Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion. IEEE T. Geosci. Remote. **58**, 348–362 (2020)
7. Hillar, C., Lim, L.-H.: Most tensor problems are NP-hard. J. ACM. **60**(6) (2013)
8. Hitchcock, F.L.: The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. **6**, 164–189 (1927)
9. Hitchcock, F.L.: Multiple invariants and generalized rank of a p-way matrix or tensor. J. Math. Phys. **7**, 39–79 (1927)
10. Boyer, R., Lathauwer, L.D., Meraim, K.A.: Higher order tensor-based method for delayed exponential fitting. IEEE T. Signal Proces. **55**, 2795–2809 (2007)
11. Singh, S., Vidal, G.: Global symmetries in tensor network states: symmetric tensors versus minimal bond dimension. Phys. Rev. B. **88**, 115–147 (2013)
12. Uemura, W., Sugino, O.: Symmetric tensor decomposition description of fermionic many-body wave functions. Phys. Rev. Lett. **109**, 253001 (2012)
13. Comon, P., Golub, G.H., Lim, L.-H., Mourrain, B.: Symmetric tensors and symmetric tensor rank. SIAM J. Matrix Anal. Appl. **30**, 1254–1279 (2008)
14. Brachat, J., Comom, P., Mourrain, B., Tsigaridas, E.: Symmetric tensor decomposition. Linear Algebra Appl. **433**, 1851–1872 (2010)
15. Kolda, T.G.: Numerical optimization for symmetric tensor decomposition. Math. Program. **151**, 225–248 (2015)
16. Nie, J.: Generating polynomials and symmetric tensor decompositions. Found. Comput. Math. **17**, 423–465 (2017)
17. Liu, H.: Symmetric tensor decomposition by alternating gradient descent. Numer. Linear Algera Appl. **29**, e2406 (2022)
18. Li, S., Li, C., Li, Y.: M-eigenvalue inclusion intervals for a fourth-order partially symmetric tensor. J. Comput. Appl. Math. **356**, 391–401 (2019)
19. Ballico, E., Bernardi, A., Christandl, M., Gesmundo, F.: On the partially symmetric rank of tensor products of w-states and other symmetric tensors. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur **30**, 93–124 (2019)
20. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. Psychometrika **35**, 283–319 (1970)
21. Sidiropoulos, N., Bro, R.: On the uniqueness of multilinear decomposition of n-way arrays. J. Chemometr. **14**, 229–239 (2000)
22. Li, N., Navasca, C., Glenn, C.: Iterative methods for symmetric outer product tensor decomposition. Electron. Trans. Numer. Anal. **44**, 124–139 (2015)
23. Ni, G., Li, Y.: A semidefinite relaxation method for partially symmetric tensor decomposition. Math. Oper. Res. (2022). https://doi.org/10.1287/moor.2021.1231
24. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2009)
25. Wang, X., Che, M., Wei, Y.: Best rank-one approximation of fourth-order partially symmetric tensors by neural network. Numer. Math. Theor. Meth. Appl. **11**, 673–700 (2018)
26. Gesmundo, F., Oneto, A., Ventura, E.: Partially symmetric variants of Comon's problem via simultaneous rank. SIAM J. Matrix Anal. Appl. **40**, 1453–1477 (2019)
27. Landsberg, J.: Tensors: Geometry and Applications. Graduate Student Mathematics, vol. 128. American Mathematical Society, Providence (2011)
28. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering, 2nd edn. Springer, New York (2006)
29. Bader, B. W., Kolda, T. G.: Matlab tensor toolbox version 2.5. http://www.sandia.gov/~tgkolda/TensorToolbox/ (2012)
30. Nie, J., Yang, Z.: Hermitian tensor decompositions. SIAM J. Matrix Anal. Appl. **41**(3), 1115–1144 (2020)