SIAM J. MATRIX ANAL. APPL.
Vol. 40, No. 3, pp. 1047–1065

NUMERICAL COMPUTATION FOR ORTHOGONAL LOW-RANK APPROXIMATION OF TENSORS*

YU GUAN† AND DELIN CHU‡

Abstract. In this paper we study the orthogonal low-rank approximation problem of tensors in the general setting in the sense that more than one matrix factor is required to be mutually orthonormal, which includes the completely orthogonal low-rank approximation and semiorthogonal low-rank approximation as two special cases. It has been addressed in [L. Wang and M. T. Chu, *SIAM J. Matrix Anal. Appl.*, 35 (2014), pp. 1058–1072] that "the question of more than one semi-orthogonal factor matrix, except for the case of complete orthogonality, remains open." To deal with this open question we present an SVD-based algorithm. Our SVD-based algorithm updates two vectors simultaneously and maintains the required orthogonality conditions by means of the polar decomposition. The convergence behavior of our algorithm is analyzed for both objective function and iterates themselves and is illustrated by numerical experiments.

Key words. tensor, orthogonal low-rank approximation, singular value decomposition, polar decomposition

AMS subject classifications. 15A15, 15A29, 15A69, 65J10, 68W25

DOI. 10.1137/18M1208101

1. Introduction.

Tensor decompositions and approximations have received extensive attention in numerical linear algebra in recent years. Interest in tensor decompositions and approximations has expanded to various fields including signal processing [36], computer vision, chemometrics [3, 4], graph analysis, data mining and analysis [1, 2, 5, 10, 46], network analysis, scientific computing, telecommunications [11, 37, 38], independent component analysis (ICA) [12], neuroscience [33], Newton potential [21, 22], and stochastic PDEs [16, 49].

Three well-known low-rank approximations to high-order tensors include rank-1 approximation [14, 50], rank-$(r_1, r_2, \ldots, r_R)$ approximation with a full core and R orthogonal side-matrices (in the *Tucker fashion*) [13, 14, 45, 50], and approximations using certain outer-product terms (in the *CANDECOMP/PARAFAC Decomposition* (CP) fashion) [6, 23]. It is well known that the rank-1 tensor approximation is theoretically guaranteed to have a global optimum [15], orthogonal rank-R Tucker approximation exists, and the existence of the low-rank canonical and Tucker tensor decompositions for Cauchy, Hilbert, and Toeplitz types of tensors was rigorously proven in [25]. But the best low-rank-R tensor approximations with $R > 2$ in real space might not be unique or even exist [15, 28, 29, 30, 42, 43]. Hence, certain orthogonality conditions are required so that the resulting best orthogonal low-rank tensor approximations exist [9, 29]. Several orthogonal approximations of tensors named completely orthogonal low-rank approximation and semiorthogonal low-rank approximation have been investigated in literature [27, 28].

*Received by the editors August 17, 2018; accepted for publication (in revised form) by G. Beylkin June 4, 2019; published electronically September 10, 2019.
https://doi.org/10.1137/18M1208101
†Department of Mathematics, National University of Singapore, Singapore 119076 (ricky7guanyu@gmail.com).
‡Corresponding author. Department of Mathematics, National University of Singapore, Singapore 119076 (matchudl@nus.edu.sg).

A real-valued tensor of order-k is a k-way array of the form

$$T = \left[\tau_{i_1,\ldots,i_k}\right] \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k}$$

with elements $\tau_{i_1,\ldots,i_k}$ accessed via k indices. A tensor of the form

$$\bigotimes_{i=1}^{k} \mathbf{u}^{(i)} = \mathbf{u}^{(1)} \otimes \cdots \otimes \mathbf{u}^{(k)} := \left[u_{i_1}^{(1)}, \ldots, u_{i_k}^{(k)}\right],$$

where elements are the products of entries from vectors $\mathbf{u}^{(i)} \in \mathbb{R}^{I_i}$, $i = 1, \ldots, k$, is said to be of rank one. These vectors are referred to as components of the rank-1 tensor [27].

The low-rank approximation of tensor T is to minimize

$$\left\| T - \sum_{r=1}^{R} \lambda_r \underbrace{\bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)}}_{H_r} \right\|_F^2, \tag{1.1}$$

subject to the mutual orthogonality condition that

$$\langle H_{r_1}, H_{r_2} \rangle = \prod_{i=1}^{k} \left\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \right\rangle = \delta_{r_1 r_2} \quad \text{for all} \quad 1 \le r_1, r_2 \le R, \tag{1.2}$$

where $\mathbf{u}_r^{(i)}$ are unit vectors for $i = 1, \ldots, k$, $r = 1, \ldots, R$. To satisfy the constraints (1.2), one of the following orthogonality conditions has been imposed:

1. Complete orthogonality [9, 27, 28]: For all $i = 1, \ldots, k$,

$$\left\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \right\rangle = 0 \quad \forall 1 \le r_1 \ne r_2 \le R.$$

2. Orthogonality: For some $1 \le i_1 < \cdots < i_\mu \le k$,

$$\left\langle \mathbf{u}_{r_1}^{(i_1)}, \mathbf{u}_{r_2}^{(i_1)} \right\rangle = 0, \ldots, \left\langle \mathbf{u}_{r_1}^{(i_\mu)}, \mathbf{u}_{r_2}^{(i_\mu)} \right\rangle = 0 \quad \forall 1 \le r_1 \ne r_2 \le R.$$

3. Semiorthogonality [9, 41, 48]: For one i with $1 \le i \le k$ such that

$$\left\langle \mathbf{u}_{r_1}^{(i)}, \mathbf{u}_{r_2}^{(i)} \right\rangle = 0 \quad \forall 1 \le r_1 \ne r_2 \le R.$$

In this paper, approximation problem (1.1) with complete orthogonality constraint, orthogonality constraint, and semiorthogonality constraint are called completely orthogonal low-rank approximation, orthogonal low-rank approximation, and semiorthogonal low-rank approximation, respectively.

Completely orthogonal low-rank approximation and semiorthogonal low-rank approximation of tensors have been studied, for example, in [9, 35, 48]. In addition, imposing orthogonality has some applications. For example, orthogonal CP tensor decomposition is the key problem in signal processing in applications such as blind signal separation and identification and multiple access wireless communication systems [41]. More applications can be found in polarization sensitive array analysis, ICA, and image processing and data representation [41].

The "workhorse" algorithm for rank-1 approximation and semiorthogonal low-rank approximation of tensor has been the alternating least squares (ALS) method, which has been proved to be convergent for almost all tensors in [47, 48]. The efficient ALS-type algorithms for the rank-R CP approximation are commonly used in computational practice. In particular, the well-known HOSVD algorithm [26] is proven to be stable for the orthogonal canonical tensor decompositions. Theoretical analysis of the best R-term tensor approximation by using the orthogonal greedy-type algorithms have been considered in [44].

The completely orthogonal low-rank approximation of tensors is similar to the truncated SVD of matrices. Obviously, complete orthogonality implies semiorthogonality, but the converse does not hold. Orthogonality is just a bridge between complete orthogonality and semiorthogonality since orthogonality is exactly the complete orthogonality if $\mu = k$ and it reduces to semiorthogonality if $\mu = 1$. The more restricted constraint by adding extra orthogonal factor matrix might be useful for other purposes.

Because of the constraints (1.2), the base tensors H_r ($r = 1, \dots, R$) are mutually orthonormal, and the expression for the optimal scales λ_r in (1.1) can also be interpreted as the length of the projection of the "vector" T onto the "unit vector" H_r under the Frobenius inner product and are necessarily given by

$$\lambda_r = \lambda_r\left(\mathbf{u}_r^{(1)}, \dots, \mathbf{u}_r^{(k)}\right) = \left\langle T, \ \bigotimes_{i=1}^{k} \mathbf{u}_r^{(i)} \right\rangle, \quad r = 1, \dots, R. \tag{1.3}$$

Thus, the orthogonal low-rank approximation problem (1.1) can be reformulated as

- Orthogonal low-rank approximation:

$$\begin{cases} \max \sum_{r=1}^{R} \lambda_r^2, \\ \text{subject to the orthogonality constraint.} \end{cases} \tag{1.4}$$

In this paper, for orthogonal low-rank approximation problem (1.4), we assume without loss of generality that

$$\left\langle \mathbf{u}_{r_1}^{(k-\mu+1)}, \mathbf{u}_{r_2}^{(k-\mu+1)} \right\rangle = 0, \dots, \left\langle \mathbf{u}_{r_1}^{(k)}, \mathbf{u}_{r_2}^{(k)} \right\rangle = 0 \quad \forall 1 \le r_1 \ne r_2 \le R. \tag{1.5}$$

Orthogonal low-rank approximation of tensors has been highlighted in [48], and its ALS has been considered for this question. It has been pointed in [48] that "the technique employed in the preceding section might not be immediately generalizable because we need to prove the convergence of both sequences $\{\mathbf{v}_{r,[p]}^{(k-1)}\}$ and $\{\mathbf{v}_{r,[p]}^{(k)}\}$ simultaneously. More study is needed." Furthermore, it has also been addressed in [48] that "the question of more than one semi-orthogonal factor matrix, except for the case of complete orthogonality, remains open."

Since orthogonal low-rank approximation is an open question addressed in [48] and it includes the completely orthogonal low-rank approximation and semiorthogonal low-rank approximation as two special cases, we focus on this open question; i.e., we study orthogonal low-rank approximation of tensors in this paper.

The ALS method works on improving one factor a time, but it suffers from slow convergence and easy stagnation at a local solution. Hence, the SVD-based method, which updates two factors simultaneously, has been applied in our previous papers [19, 20] on the best rank-1 tensor approximation:

- The classical problem of finding the best rank-1 approximation to a tensor is revisited in [19]. The main focus is on providing a mathematical proof for the convergence of the iterates of an SVD-based algorithm.

- The problem of finding the best rank-1 approximation to a symmetric tensor is revisited in [20]. In contrast to the many long and lingering arguments in the literature, it offers a straightforward justification that generically the best rank-1 approximation to a symmetric tensor is symmetric. Furthermore, SVD-based algorithms which maintain symmetry are developed to the best rank-1 approximation for a symmetric tensor, and it is proved that not only the generalized Rayleigh quotients generated from the SVD-based algorithms enjoy monotone convergence but also that the iterates themselves converge.

It should be addressed that SVD-based algorithms in [19, 20] cannot be applied directly to our orthogonal low-rank approximation problem since the orthogonality constraint cannot be satisfied. We will apply polar decomposition to overcome this difficulty and ensure the orthogonality condition. The main contributions in this paper include the following:

- Inspired by the SVD-based algorithms for the rank-1 approximation [19, 20, 18] and algorithm in [9], we develop an SVD-based algorithm for orthogonal low-rank approximation of tensors which updates two factors simultaneously and maintains the required orthogonality conditions by means of the polar decomposition.
- The convergence of our SVD-based algorithm for both objective function and iterates themselves is established.
- Numerical experiments have been presented to illustrate the performances of our SVD-based algorithm.

The rest of this paper is organized as follows: Some preliminaries, including the basic definitions and notions of tensors, are provided in section 2. Then our SVD-based algorithm is developed in section 3, and the convergence is analyzed in section 4. Numerical experiments are presented in section 5. Conclusions are given in section 6.

2. Preliminaries.

Now we introduce some notations which are used in our paper. Let the symbol $[\![m]\!]$ denote henceforth the set of integers $\{1, \ldots, m\}$ for a given positive integer m. Suppose that the set $[\![k]\!]$ is partitioned as the union of two disjoint nonempty subsets $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_s\}$ and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_t\}$, where $s + t = k$. An element in the tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k}$ will be marked as $\tau^{(\boldsymbol{\alpha},\boldsymbol{\beta})}_{[\mathcal{I}|\mathcal{J}]}$, where $\mathcal{I} := (i_1, \ldots, i_s)$ and $\mathcal{J} := (j_1, \ldots, j_t)$ contain those indices at locations $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. Each index in the arrays $\mathcal{I}$ and $\mathcal{J}$ should be within the corresponding range of integers, e.g., $i_1 \in [\![I_{\alpha_1}]\!]$ and so on. If the reference to a specific partitioning $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is clear, then without causing ambiguity we abbreviate the element as $\tau_{[\mathcal{I}|\mathcal{J}]}$.

Given a fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, we regard an order-$k$ tensor $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$ as a "matrix representation" of a linear operator mapping order-s tensors to order-t tensors [48]. Specifically, we identify T with the linear map

$$\mathscr{T}_{\boldsymbol{\beta}} : \mathbb{R}^{I_{\alpha_1} \times \cdots \times I_{\alpha_s}} \to \mathbb{R}^{I_{\beta_1} \times \cdots \times I_{\beta_t}}, \tag{2.1}$$

where, for any $S \in \mathbb{R}^{I_{\alpha_1} \times \cdots \times I_{\alpha_s}}$, we have

$$\mathscr{T}_{\boldsymbol{\beta}}(S) := T \circledast_{\boldsymbol{\beta}} S = \left[\left\langle \tau_{[:|\ell_1,\ldots,\ell_t]}, S \right\rangle \right] \in \mathbb{R}^{I_{\beta_1} \times \cdots \times I_{\beta_t}}. \tag{2.2}$$

In the above, $\tau_{[:|\ell_1,\ldots,\ell_t]}$ denotes the $(\ell_1, \ldots, \ell_t)$th "slice" in the $\boldsymbol{\beta}$ direction of the tensor T; that is, ℓ_j occurs at the β_jth location in the array $[\![k]\!]$ and assumes an integer value in $[\![I_{\beta_\ell}]\!]$ for $\ell = 1, \ldots, t$, whereas the symbol ":" denotes a wild card at the $(\alpha_1, \ldots, \alpha_s)$ location to be summed over and

$$\left\langle \tau_{[:|\ell_1,\ldots,\ell_t]}, S \right\rangle := \sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_s=1}^{I_{\alpha_s}} \tau_{[i_1,\ldots,i_s|\ell_1,\ldots,\ell_t]} s_{i_1,\ldots,i_s} \tag{2.3}$$

is the Frobenius inner product generalized to multidimensional arrays.

LEMMA 2.1 (see [20]). *Given a tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k}$, a partition $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, and vectors $\mathbf{u}^{(i)} \in \mathbb{R}^{I_i}$, $i = 1, \ldots, k$, then it holds that*

$$\left\langle T, \bigotimes_{i=1}^{k} \mathbf{u}^{(i)} \right\rangle = \left\langle T \circledast_{\boldsymbol{\beta}} \bigotimes_{i=1}^{s} \mathbf{u}^{(\alpha_i)}, \bigotimes_{i=1}^{t} \mathbf{u}^{(\beta_i)} \right\rangle. \tag{2.4}$$

Given a rank-1 tensor $S \in \mathbb{R}^{I_{\alpha_1} \times \cdots \times I_{\alpha_{k-2}}}$, we may identify $T \circledast_{\boldsymbol{\beta}} S$ as a matrix in $\mathbb{R}^{I_{\beta_1} \times I_{\beta_2}}$. The following identity allows a convenient way to swap components with S.

LEMMA 2.2 (see [20]). *Given a tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_k}$, arbitrary vectors $\mathbf{u}^{(\alpha_i)} \in \mathbb{R}^{I_{\alpha_i}}$, $i = 1, \ldots, k-2$, and $\mathbf{v} \in \mathbb{R}^{I_{\beta_2}}$, then*

$$\left(T \circledast_{\{\beta_1,\beta_2\}} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\alpha_i)} \right) \mathbf{v} = \left(T \circledast_{\{\beta_1,\alpha_j\}} \bigotimes_{i=1}^{j-1} \mathbf{u}^{(\alpha_i)} \otimes \mathbf{v} \otimes \bigotimes_{i=j+1}^{k-2} \mathbf{u}^{(\alpha_i)} \right) \mathbf{u}^{(\alpha_j)} \tag{2.5}$$

for any $j \in [\![k-2]\!]$.

The following lemma is essentially the well-known polar decomposition [24] which reveals the trace maximizing property that will play an important role for the development in the next section.

LEMMA 2.3. *Let matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ have polar decomposition*

$$A = QS,$$

where $Q \in \mathbb{R}^{m \times n}$ is the column orthogonal polar factor and $S \in \mathbb{R}^{n \times n}$ is the symmetric positive semidefinite factor. Then

$$Q = \arg \max_{P \in \mathbb{R}^{m \times n},\ P^T P = I} \operatorname{Trace}\left(P^T A\right).$$

Moreover, if A is of full column rank, then Q above is unique.

The next Eckart–Young lemma provides the mechanism of our SVD-based algorithm in the next section for updating two factors simultaneously.

LEMMA 2.4 (see [17]). *Given a matrix $A \in \mathbb{R}^{m \times n}$, then the global maximum of the generalized Rayleigh quotient*

$$\max_{\substack{\mathbf{x} \in \mathbb{R}^m,\, \|\mathbf{x}\| = 1 \\ \mathbf{y} \in \mathbb{R}^n,\, \|\mathbf{y}\| = 1}} \langle \mathbf{x},\ A\mathbf{y} \rangle \tag{2.6}$$

is precisely the largest singular value σ_1 of A, where the global maximizer (x, y) consists of the corresponding left and right singular vectors of A, and thus the best rank-1 approximation to A is given by $\sigma_1 x^\top y$.

We close this section by one more lemma which will be used to prove the convergence of iterates in section 4.

Lemma 2.5 (see [20, 34]). *Let $\{a_n\}$ be a bounded sequence of real numbers. If the accumulation points of the sequence $\{a_n\}$ are isolated and for every subsequence $\{a_{k_j}\}$ converging to a accumulation point a^* there is an infinite subsequence $\{a_{k_{j_i}}\}$ such that $|a_{k_{j_i}+1} - a_{k_{j_i}}| \to 0$, then the whole sequence $\{a_n\}$ converges to a^*.*

3. Algorithms for orthogonal low-rank approximation.

Followed by [18] and our previous work [19, 20] for the rank-1 tensor approximation using SVD and [9] for the completely orthogonal low-rank tensor approximation, we apply the similar ideas to the orthogonal low-rank approximation problem (1.4) and update two factors simultaneously. There is an additional challenge to maintain the required orthogonality condition in (1.4); we handle this difficulty by means of the polar decomposition [24].

For any $\mathbf{u}_r^{(1)}, \ldots, \mathbf{u}_r^{(\ell)}, \mathbf{u}_r^{(\ell+1)}, \ldots, \mathbf{u}_r^{(k)},\ r = 1, \ldots, R$, it is important to know how we can update to obtain "better" ones.

(i) For any $1 \le \ell \le k - \mu - 1$ and $r = 1, 2, \ldots, R$, let $\beta_\ell = (\ell, \ell + 1)$ and

$$C_r^{(\ell)} = T \circledast_{\beta_\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \otimes \bigotimes_{i=\ell+2}^{k} \mathbf{u}_r^{(i)} \right).$$

Denote the dominant left and right singular vectors of $C_r^{(\ell)}$ corresponding to its largest singular value by $\tilde{\mathbf{u}}_r^{(\ell)}$ and $\tilde{\mathbf{u}}_r^{(\ell+1)}$, respectively. Then by Lemma 2.4, we have

$$\begin{aligned} \lambda_r &= \left\langle T,\ \bigotimes_{\ell=1}^{k} \mathbf{u}_r^{(\ell)} \right\rangle = \left\langle \mathbf{u}_r^{(\ell)},\ C_r^{(\ell)} \mathbf{u}_r^{(\ell+1)} \right\rangle \\ &\le \max_{\mathbf{x}^T\mathbf{x}=1,\ \mathbf{y}^T\mathbf{y}=1} \left\langle \mathbf{x},\ C_r^{(\ell)} \mathbf{y} \right\rangle = \left\langle \tilde{\mathbf{u}}_r^{(\ell)},\ C_r^{(\ell)} \tilde{\mathbf{u}}_r^{(\ell+1)} \right\rangle \\ &= \tilde{\lambda}_r. \end{aligned}$$

Obviously, we can update $\mathbf{u}_r^{(\ell)}$ by $\tilde{\mathbf{u}}_r^{(\ell)}$ and $\mathbf{u}_r^{(\ell+1)}$ by $\tilde{\mathbf{u}}_r^{(\ell+1)}$.

(ii) The Lagrangian for the optimization problem (1.4) is

$$\begin{aligned} \mathcal{L} := \sum_{r=1}^{R} \lambda_r^2 - \sum_{\ell=1}^{k} \sum_{r=1}^{R} \rho_r^{(\ell)} \left(\left\langle \mathbf{u}_r^{(\ell)},\ \mathbf{u}_r^{(\ell)} \right\rangle - 1 \right) \\ - \sum_{1 \le r_1 < r_2 \le R}\ \sum_{\ell=k-\mu+1}^{k} \alpha_{r_1 r_2}^{(\ell)} \left\langle \mathbf{u}_{r_1}^{(\ell)},\ \mathbf{u}_{r_2}^{(\ell)} \right\rangle, \end{aligned}$$

where λ_r is given by (1.3) and $\rho_r^{(\ell)}, \alpha_{r_1 r_2}^{(\ell)}$ are Lagrange multipliers. According to [32], the first-order optimality condition for a stationary point is to satisfy (3.1)

$$\lambda_r T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \otimes \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) = \rho_r^{(\ell)} \mathbf{u}_r^{(\ell)}, \quad \ell = 1, \ldots, k - \mu, \quad r = 1, \ldots, R,$$

and

$$\lambda_r T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \otimes \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right) \tag{3.2}$$

$$= \rho_r^{(\ell)} \mathbf{u}_r^{(\ell)} + \sum_{r_1<r} \frac{\alpha_{r_1 r}^{(\ell)}}{2} \mathbf{u}_{r_1}^{(\ell)}$$

$$+ \sum_{r<r_2} \frac{\alpha_{r r_2}^{(\ell)}}{2} \mathbf{u}_{r_2}^{(\ell)}, \ell = k-\mu+1, \ldots, k, r = 1, \ldots, R.$$

It follows from the orthogonality condition that

$$\rho_r^{(\ell)} = \lambda_r^2, \qquad \ell = 1, \ldots, k, \quad r = 1, \ldots, R,$$

and furthermore,

$$V^{(\ell)} \Lambda^{(\ell)} = U^{(\ell)} S^{(\ell)},\ S^{(\ell)} \text{ is symmetric}, \quad \ell = k-\mu+1, \ldots, k, \tag{3.3}$$

where

$$\mathbf{v}_r^{(\ell)} = T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_r^{(i)} \otimes \bigotimes_{i=\ell+1}^{k} \mathbf{u}_r^{(i)} \right), \quad \ell = k-\mu+1, \ldots, k, \quad r = 1, \ldots, R,$$

and

$$V^{(\ell)} = \left[\ \mathbf{v}_1^{(\ell)}, \ldots, \mathbf{v}_R^{(\ell)}\ \right], \quad U^{(\ell)} = \left[\ \mathbf{u}_1^{(\ell)}, \ldots, \mathbf{u}_R^{(\ell)}\ \right],$$

$$\Lambda^{(\ell)} = \begin{bmatrix} \lambda_1^{(\ell)} & & \\ & \ddots & \\ & & \lambda_R^{(\ell)} \end{bmatrix}.$$

It is from Lemma 2.3 and (3.3) that we can update $U^{(\ell)}$ by the orthogonal polar factor of the matrix $V^{(\ell)}\Lambda^{(\ell)}$: Let the polar decomposition of $V^{(\ell)}\Lambda^{(\ell)}$ be

$$V^{(\ell)} \Lambda^{(\ell)} = \tilde{U}^{(\ell)} \tilde{S}^{(\ell)}, \quad \ell = k-\mu+1, \ldots, k,$$

where $\tilde{U}^{(\ell)}$ is column orthogonal and $\tilde{S}^{(\ell)}$ is symmetric and positive semidefinite. Denote

$$\tilde{\lambda}_r^{(\ell)} = \left\langle \mathbf{v}_r^{(\ell)},\ \tilde{\mathbf{u}}_r^{(\ell)} \right\rangle, \quad \ell = k-\mu+1, \ldots, k, \quad r = 1, \ldots, R.$$

We have

$$\sum_{r=1}^{R} (\lambda_r^{(\ell)})^2 = \operatorname{Trace}\left((U^{(\ell)})^T V^{(\ell)} \Lambda^{(\ell)} \right) \leq \operatorname{Trace}\left((\tilde{U}^{(\ell)})^T V^{(\ell)} \Lambda^{(\ell)} \right)$$

$$= \sum_{r=1}^{R} \tilde{\lambda}_r^{(\ell)} \lambda_r^{(\ell)}, \quad \ell = k-\mu+1, \ldots, k.$$

Consequently, we have by using the Cauchy–Schwarz inequality that

$$\sum_{r=1}^{R}(\lambda_r^{(\ell)})^2 \leq \sum_{r=1}^{R}(\tilde{\lambda}_r^{(\ell)})^2, \quad \ell = k-\mu+1,\ldots,k, \tag{3.4}$$

and the equality holds if and only if

$$\lambda_r^{(\ell)} = \tilde{\lambda}_r^{(\ell)}, \quad \ell = k-\mu+1,\ldots,k, \quad r = 1,\ldots,R.$$

Hence, we update $U^{(\ell)}$ by $\tilde{U}^{(\ell)}$ for $\ell = k-\mu+1,\ldots,k$.

To update two factors as simultaneously as possible, we have to consider if $k-\mu$ is even or odd: For $r = 1,\ldots,R$,

- when $k-\mu$ is even, update $\mathbf{u}_r^{(\ell)}$ and $\mathbf{u}_r^{(\ell+1)}$ simultaneously by SVDs for $\ell = 1,3,\ldots,k-\mu-1$;
- when $k-\mu$ is odd, we can update $\mathbf{u}_r^{(\ell)}$ and $\mathbf{u}_r^{(\ell+1)}$ simultaneously by SVDs for $\ell = 1,3,\ldots,k-\mu-2$. Next update the new "$\mathbf{u}_r^{(k-\mu-1)}$" and $\mathbf{u}_r^{(k-\mu)}$ simultaneously by SVD (so, $\mathbf{u}_r^{(k-\mu-1)}$ will be updated one more time).

  Then we can update $U^{(\ell)}$ with $\ell = k-\mu+1,\ldots,k$ by polar decompositions given in (ii) above to ensure the orthogonality condition.

The idea above leads to our SVD-based algorithm named Algorithm 1 for orthogonal low-rank approximation of tensors.

Here we shortly discuss the complexity of Algorithm 1 per loop. For the SVD part, forming $\bigotimes_{i=1}^{\ell-1}\mathbf{u}_{r,[p+1]}^{(i)} \otimes \bigotimes_{i=\ell+2}^{k}\mathbf{u}_{r,[p]}^{(i)}$ requires $\prod_{i=1,i\neq\ell,i\neq\ell+1}^{k} I_i$ entry-to-entry multiplications. Then per β-product requires $\prod_{i=1}^{k} I_i$ to form the inner product. It is difficult to estimate the complexity of SVDs, which is much less than $O(\max(I_\ell I_{\ell+1}^2, I_{\ell+1}^3))$. Therefore, we estimate that the SVD part requires

$$\sum_{\ell=1,3,\ldots}^{k-\mu-1 \textbf{ or } k-\mu-2} O\left(\max(RI_\ell I_{\ell+1}^2, RI_{\ell+1}^3) + R\prod_{i=1,i\neq\ell,i\neq\ell+1}^{k} I_i + R\prod_{i=1}^{k} I_i\right)$$

scalar multiplications per update in total. In the polar part, it requires

$$\sum_{\ell=k-\mu+1}^{k} O\left(R\prod_{i=1,i\neq\ell}^{k} I_i + R\prod_{i=1}^{k} I_i + RI_\ell + R^2 I_\ell\right)$$

scalar multiplications per update.

Measuring the theoretical number of floating-point operations is an objective means for gauging the computational complexity in itself. However, it should be pointed out that, given today's computing technologies where machines (even a desktop PC) are equipped with high-performance processors (vector or parallel, hyperthreading or multicores), simple flop counts are not valid any more for measuring the performance. A standard computer software library, such as the LAPACK, is often highly optimized, which muddles the flop counts. Furthermore, the memory latency to fetch anything not in the cache is much greater than the cost of a flop. For these reasons, MATLAB has removed one of its most famous commands flop since Version 6.

4. Convergence.

In this section we prove the convergence of objective function (1.1) first and then the convergence of iterates.

Algorithm 1 (Orthogonal low-rank approximation for tensors.)

Require: An order-k tensor $T \in \mathbb{R}^{I_1\times\cdots\times I_k}$ and starting unit vectors $\mathbf{u}_{r,[0]}^{(\ell)} \in \mathbb{R}^{I_\ell}, \ell = 1,\ldots,k, r = 1,\ldots,R$ and $\mathbf{u}_{r_1,[0]}^{(\ell)} \perp \mathbf{u}_{r_2,[0]}^{(\ell)}$ for $\ell = k-\mu+1,\ldots,k$, $1 \le r_1 \ne r_2 \le R$.
Ensure: An orthogonal rank-R approximation.

$\tau := k-\mu-1$
if $k-\mu$ is odd **then**
  $\tau := k-\mu-2$
end if
for $p = 0,1,\ldots,$ **do**
  for $\ell = 1,3,\ldots,\tau$ **do**
    $\beta_\ell = (\ell,\ell+1)$
    for $r = 1,2,\ldots,R,$ **do**
      $C_{r,[p+1]}^{(\ell)} = T \circledast_{\beta_\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_{r,[p+1]}^{(i)} \otimes \bigotimes_{i=\ell+2}^{k} \mathbf{u}_{r,[p]}^{(i)}\right)$ {A matrix of size $I_\ell \times I_{\ell+1}$}
      $[\mathbf{u}, s, \mathbf{v}] = \mathsf{svds}(C_{r,[p+1]}^{(\ell)}, 1)$ {Dominant singular value triplet via MATLAB routine svds; assume uniqueness}
      if $\mathbf{u}(1) < 0$ **then**
        $\mathbf{u} = -\mathbf{u}, \mathbf{v} = -\mathbf{v}$
      end if
      $\mathbf{u}_{r,[p+1]}^{(\ell)} := \mathbf{u}, \quad \mathbf{u}_{r,[p+1]}^{(\ell+1)} := \mathbf{v}$ {if $k-\mu$ is odd, use $\hat{\mathbf{u}}_{r,[p+1]}^{(k-\mu-1)} := \mathbf{v}$}
      $\lambda_{r,[p+1]}^{(\ell)} := s, \quad \lambda_{r,[p+1]}^{(\ell+1)} := s$ {if $k-\mu$ is odd, use $\hat{\lambda}_{r,[p+1]}^{(k-\mu-1)} := s$}
    end for
  end for
  if $\tau = k-\mu-2$ **then**
    $\beta_{k-\mu-1} = (k-\mu-1,\ k-\mu)$
    for $r = 1,2,\ldots,R,$ **do**
      $C_{r,[p+1]}^{(k-\mu-1)} = T \circledast_{\beta_{k-\mu-1}} \left(\bigotimes_{i=1}^{k-\mu-2} \mathbf{u}_{r,[p+1]}^{(i)} \otimes \bigotimes_{i=k-\mu+1}^{k} \mathbf{u}_{r,[p]}^{(i)}\right)$ {A matrix of size $I_{k-\mu-1} \times I_{k-\mu}$}
      $[\mathbf{u}, s, \mathbf{v}] = \mathsf{svds}(C_{r,[p+1]}^{(k-\mu-1)}, 1)$ {Dominant singular value triplet via MATLAB routine svds; assume uniqueness}
      if $\mathbf{u}_1 < 0$ **then**
        $\mathbf{u} = -\mathbf{u}, \mathbf{v} = -\mathbf{v}$
      end if
      $\mathbf{u}_{r,[p+1]}^{(k-\mu-1)} := \mathbf{u}, \quad \mathbf{u}_{r,[p+1]}^{(k-\mu)} := \mathbf{v}, \quad \lambda_{r,[p+1]}^{(k-\mu-1)} := s,\ \lambda_{r,[p+1]}^{(k-\mu)} := s$
    end for
  end if
  for $\ell = k-\mu+1,\ldots,k$ **do**
    for $r = 1,2,\ldots,R,$ **do**
      $\mathbf{v}_{r,[p+1]}^{(\ell)} = T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}_{r,[p+1]}^{(i)} \otimes \bigotimes_{i=\ell+1}^{k} \mathbf{u}_{r,[p]}^{(i)}\right)$ {define columns of $V_{[p+1]}^{(\ell)}$}
      $\hat{\lambda}_{r,[p+1]}^{(\ell)} := \langle \mathbf{v}_{r,[p+1]}^{(\ell)}, \mathbf{u}_{r,[p]}^{(\ell)} \rangle$ {define diagonals of $\Lambda_{[p+1]}^{(\ell)}$}
    end for
    $[U_{[p+1]}^{(\ell)}, S_{[p+1]}^{(\ell)}] = \mathsf{poldec}(V_{[p+1]}^{(\ell)} \Lambda_{[p+1]}^{(\ell)})$ {polar decomposition of $(V_{[p+1]}^{(\ell)} \Lambda_{[p+1]}^{(\ell)})$ is called here}
    for $r = 1,2,\ldots,R,$ **do**
      $\mathbf{u}_{r,[p+1]}^{(\ell)} := U_{[p+1]}^{(\ell)}(:,r)$
      $\lambda_{r,[p+1]}^{(\ell)} := S_{[p+1]}^{(\ell)}(r,r)(= \langle \mathbf{v}_{r,[p+1]}^{(\ell)}, \mathbf{u}_{r,[p+1]}^{(\ell)} \rangle)$
    end for
  end for
end for

4.1. Convergence of objective function. As discussed in (i) of section 3, Algorithm 1 ensures an inherent ascending property on $\lambda_{r,[p]} := \lambda_r\left(\mathbf{u}_{r,[p]}^{(1)},\dots,\mathbf{u}_{r,[p]}^{(k)}\right)$ in the sense that

$$\begin{aligned}\lambda_{r,[p]} \le \lambda_{r,[p+1]}^{(1)} = \lambda_{r,[p+1]}^{(2)} \le \lambda_{r,[p+1]}^{(3)}&\\ = \lambda_{r,[p+1]}^{(4)} \le \cdots \le \lambda_{r,[p+1]}^{(k-\mu-1)} = \lambda_{r,[p+1]}^{(k-\mu)},&\quad r=1,\dots,R.\end{aligned}$$

Therefore, we have

$$\sum_{r=1}^{R}(\lambda_{r,[p]})^2 \le \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(1)})^2 \le \cdots \le \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k-\mu)})^2.$$

Then, according to the discussion in (ii) of section 3, we also have

$$\begin{aligned}\sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k-\mu)})^2 \le \sum_{r=1}^{R}\lambda_{r,[p+1]}^{(k-\mu)}\lambda_{r,[p+1]}^{(k-\mu+1)} \le \cdots \le \sum_{r=1}^{R}\lambda_{r,[p+1]}^{(k-1)}\lambda_{r,[p+1]}^{(k)}&\\ \le \sum_{r=1}^{R}(\lambda_{r,[p+1]}^{(k)})^2 = \sum_{r=1}^{R}(\lambda_{r,[p+1]})^2.&\end{aligned}$$

Therefore, the sequence $\{\sum_{r=1}^{R}(\lambda_{r,[p]})^2\}$ is increasing. Since $\sum_{r=1}^{R}(\lambda_{r,[p]})^2 \le \|T\|_F^2$ for any $p=1,2,\dots$, the following convergence result is ready.

THEOREM 4.1. *Let the sequence $\{\lambda_{1,[p]},\dots,\lambda_{R,[p]}\}$ be generated in Algorithm* 1. *Then the objective value $\sum_{r=1}^{R}(\lambda_{r,[p]})^2$ converges.*

4.2. Convergence of iterates. Before proving the global convergence of the iterates $\{\mathbf{u}_{r,[p]}^{(\ell)}\}$, we show that generically accumulation points of Algorithm 1 are geometrically isolated. Denote the accumulation point generated by Algorithm 1 as

$$\{(\mathbf{u}_1^{(1)},\dots,\mathbf{u}_1^{(k)}),\dots,(\mathbf{u}_R^{(1)},\dots,\mathbf{u}_R^{(k)})\}. \tag{4.1}$$

Then, by (3.1) and (3.2), any accumulation point of Algorithm 1 necessarily satisfies the following system of nonlinear equations with $\beta_\ell = (\ell,\ell+1)$:

$$\begin{cases} T\circledast_{\beta_\ell}\left(\bigotimes_{i=1}^{\ell-1}\mathbf{u}_r^{(i)}\otimes\bigotimes_{i=\ell+2}^{k}\mathbf{u}_r^{(i)}\right)\mathbf{u}_r^{(\ell+1)} = \left\langle T,\ \bigotimes_{i=1}^{k}\mathbf{u}_r^{(i)}\right\rangle\mathbf{u}_r^{(\ell)},\\ \qquad\qquad \ell=1,\dots,k-\mu-1,\quad r=1,\dots,R,\\ T\circledast_{\ell}\left(\bigotimes_{i=1}^{\ell-1}\mathbf{u}_r^{(i)}\otimes\bigotimes_{i=\ell+1}^{k}\mathbf{u}_r^{(i)}\right) = \sum_{t=1}^{R}\left\langle T,\bigotimes_{i=1}^{\ell-1}\mathbf{u}_r^{(i)}\otimes\mathbf{u}_t^{(\ell)}\otimes\bigotimes_{i=\ell+1}^{k}\mathbf{u}_r^{(i)}\right\rangle\mathbf{u}_t^{(\ell)},\\ \qquad\qquad \ell=k-\mu+1,\dots,k,\quad r=1,\dots,R.\end{cases} \tag{4.2}$$

This is exactly the same as the accumulation point of ALS method proposed in [48] if $\mu=1$. Furthermore, this system is the same as the one characterizing the first-order optimality condition for a stationary point of the optimization problem (1.4). As a result, the following result in [48] also holds for Algorithm 1.

LEMMA 4.2 (see [48]). *For almost all tensors $T\in\mathbb{R}^{I_1\times\cdots\times I_k}$, except for an affine algebraic set of codimension one, the accumulation points of any sequence generated by Algorithm* 1 *and the stationary points of the optimization problem* (1.4) *are necessarily isolated.*

Assumption A. We say that a given tensor $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$ satisfies Assumption A if for every convergent subsequence $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$ generated by Algorithm 1, the dominant singular value of the limiting point $C^{(\ell)}_r$ of the corresponding subsequence $\{C^{(\ell)}_{r,[p_j]}\}$ are simple for all $\ell = 1, \ldots, k-\mu$, $r = 1, \ldots, R$. Moreover, the limiting point $V^{(\ell)}\Lambda^{(\ell)}$ of the matrix $V^{(\ell)}_{[p_j]}\Lambda^{(\ell)}_{[p_j]}$ for $\ell = k-\mu+1, \ldots, k$ are of full column rank.

Assumption A is to ensure the uniqueness of singular value decompositions and polar decompositions in Algorithm 1. Since vectors $\mathbf{u}^{(\ell)}_{r,[p]}$ defined in Algorithm 1 are unit vectors, the sequence $\{\mathbf{u}^{(\ell)}_{r,[p]}\}$ must have a convergent subsequence. We select the common subset $\{r, [p_j]\}$ of nonnegative integers so that $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$ converges for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$.

Lemma 4.3. *For all $\ell = 1, \ldots, k$, $r = 1, \ldots, R$, if subsequences $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$ generated by Algorithm* 1 *converge, then subsequences $\{\mathbf{u}^{(\ell)}_{r,[p_j+1]}\}$ also converge. Furthermore, under Assumption* A*, $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$ and $\{\mathbf{u}^{(\ell)}_{r,[p_j+1]}\}$ converge to the same limiting point for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$.*

Proof. The convergence of $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$ for $\ell = 3, \ldots, k$, $r = 1, \ldots, R$ implies that the subsequence $\{C^{(1)}_{r,[p_j+1]}\}$ converges. By the continuity inherited in the SVD, $\{\mathbf{u}^{(1)}_{r,[p_j+1]}\}$ and $\{\mathbf{u}^{(2)}_{r,[p_j+1]}\}$ converge since we have aligned the direction of dominant left and right singular vector in Algorithm 1. This results in $\{C^{(3)}_{r,[p_j+1]}\}$ and $\{\mathbf{u}^{(3)}_{r,[p_j+1]}\}$ and $\{\mathbf{u}^{(4)}_{r,[p_j+1]}\}$ converging. Then the convergence of $\{\mathbf{u}^{(\ell+2)}_{r,[p_j]}\}, \ldots, \{\mathbf{u}^{(k)}_{r,[p_j]}\}$ and $\{\mathbf{u}^{(1)}_{r,[p_j+1]}\}$, $\ldots$, $\{\mathbf{u}^{(\ell-1)}_{r,[p_j+1]}\}$ ensures that $\{C^{(\ell)}_{r,[p_j+1]}\}$ converges, and consequently $\{\mathbf{u}^{(\ell)}_{r,[p_j+1]}\}$ and $\{\mathbf{u}^{(\ell+1)}_{r,[p_j+1]}\}$ converge for $\ell = 1, 3, \ldots, k-\mu-1$, $r = 1, \ldots, R$. One thing should be mentioned is that there are additional convergent $\{\hat{\mathbf{u}}^{(k-\mu-1)}_{r,[p_j+1]}\}$ if $k-\mu$ is odd. Furthermore, $V^{(\ell)}_{[p_j+1]}\Lambda^{(\ell)}_{[p_j+1]}$ converges by definition, and thus, by continuity of the polar decomposition, $\{\mathbf{u}^{(\ell)}_{r,[p_j+1]}\}$ converges for $\ell = k-\mu+1, \ldots, k$, $r = 1, \ldots, R$.

Let the limiting points of $\{\lambda_{r,[p_j]}\}$, $\{\lambda^{(\ell)}_{r,[p_j+1]}\}$, $\{\mathbf{u}^{(\ell)}_{r,[p_j]}\}$, $\{\mathbf{u}^{(\ell)}_{r,[p_j+1]}\}$, and $\{C^{(\ell)}_{r,[p_j+1]}\}$ be λ_r, $\tilde{\lambda}^{(\ell)}_r$, $\mathbf{u}^{(\ell)}_r$, $\tilde{\mathbf{u}}^{(\ell)}_r$, and $\tilde{C}^{(\ell)}_r$, respectively. Now we prove that $\mathbf{u}^{(\ell)}_r = \tilde{\mathbf{u}}^{(\ell)}_r$ for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$. Here we just prove $k-\mu$ is even, and the odd case can be proved in similar way.

First, we have for $r = 1, \ldots, R$,

$$
\begin{aligned}
&\lambda_{r,[p_j]} = \left\langle \mathbf{u}^{(1)}_{r,[p_j]},\ C^{(1)}_{r,[p_j+1]}\mathbf{u}^{(2)}_{r,[p_j]} \right\rangle \\
&\le \lambda^{(1)}_{r,[p_j+1]} = \left\langle \mathbf{u}^{(1)}_{r,[p_j+1]},\ C^{(1)}_{r,[p_j+1]}\mathbf{u}^{(2)}_{r,[p_j+1]} \right\rangle = \left\langle \mathbf{u}^{(3)}_{r,[p_j]},\ C^{(3)}_{r,[p_j+1]}\mathbf{u}^{(4)}_{r,[p_j]} \right\rangle = \lambda^{(2)}_{r,[p_j+1]} \\
&\le \cdots \\
&\le \lambda^{(\ell)}_{r,[p_j+1]} = \left\langle \mathbf{u}^{(\ell)}_{r,[p_j+1]},\ C^{(\ell)}_{r,[p_j+1]}\mathbf{u}^{(\ell+1)}_{r,[p_j+1]} \right\rangle = \left\langle \mathbf{u}^{(\ell+2)}_{r,[p_j]},\ C^{(\ell+2)}_{r,[p_j+1]}\mathbf{u}^{(\ell+3)}_{r,[p_j]} \right\rangle \\
&\qquad\qquad = \lambda^{(\ell+1)}_{r,[p_j+1]} \quad (\ell = 3, 5, \ldots) \\
&\le \cdots \\
&\le \lambda^{(k-\mu-1)}_{r,[p_j+1]} = \left\langle \mathbf{u}^{(k-\mu-1)}_{r,[p_j+1]},\ C^{(k-\mu-1)}_{r,[p_j+1]}\mathbf{u}^{(k-\mu)}_{r,[p_j+1]} \right\rangle = \lambda^{(k-\mu)}_{r,[p_j+1]};
\end{aligned}
$$

then we have, by taking the limits, that

$$
\begin{aligned}
&\lambda_r = \left\langle \mathbf{u}_r^{(1)},\ \tilde{C}_r^{(1)}\mathbf{u}_r^{(2)} \right\rangle \\
&\le \tilde{\lambda}_r^{(1)} = \left\langle \tilde{\mathbf{u}}_r^{(1)},\ \tilde{C}_r^{(1)}\tilde{\mathbf{u}}_r^{(2)} \right\rangle = \left\langle \mathbf{u}_r^{(3)},\ \tilde{C}_r^{(3)}\mathbf{u}_r^{(4)} \right\rangle = \tilde{\lambda}_r^{(2)} \\
&\le \cdots \\
&\le \tilde{\lambda}_r^{(\ell)} = \left\langle \tilde{\mathbf{u}}_r^{(\ell)},\ \tilde{C}_r^{(\ell)}\tilde{\mathbf{u}}_r^{(\ell+1)} \right\rangle = \left\langle \mathbf{u}_r^{(\ell+2)},\ \tilde{C}_r^{(\ell+2)}\mathbf{u}_r^{(\ell+3)} \right\rangle = \tilde{\lambda}_r^{(\ell+1)} \quad (\ell = 3, 5, \ldots) \\
&\le \cdots \\
&\le \tilde{\lambda}_r^{(k-\mu-1)} = \left\langle \tilde{\mathbf{u}}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)}\tilde{\mathbf{u}}_r^{(k-\mu)} \right\rangle = \tilde{\lambda}_r^{(k-\mu)}.
\end{aligned}
$$

Combined with the discussion in (ii) of section 3,

$$
\begin{aligned}
&\sum_{r=1}^{R}(\lambda_r)^2 \le \sum_{r=1}^{R}(\tilde{\lambda}_r^{(1)})^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r^{(2)})^2 \\
&\le \cdots \\
&\le \sum_{r=1}^{R}(\tilde{\lambda}_r^{(\ell)})^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r^{(\ell+1)})^2 \quad (\ell = 3, 5, \ldots) \\
&\le \cdots \\
&\le \sum_{r=1}^{R}(\tilde{\lambda}_r^{(k-\mu-1)})^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r^{(k-\mu)})^2 \le \sum_{r=1}^{R}\tilde{\lambda}_r^{(k-\mu)}\tilde{\lambda}_r^{(k-\mu+1)} \\
&\le \sum_{r=1}^{R}(\tilde{\lambda}_r^{(k-\mu+1)})^2 \le \cdots \le \sum_{r=1}^{R}(\tilde{\lambda}_r^{(k)})^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r)^2.
\end{aligned}
$$

In addition, we also know by Theorem 4.1 that $\sum_{r=1}^{R}(\lambda_r)^2 = \sum_{r=1}^{R}(\tilde{\lambda}_r)^2$. Hence, we obtain for $r = 1, \ldots, R$, $\lambda_r = \tilde{\lambda}_r^{(1)} = \cdots = \tilde{\lambda}_r^{(k-\mu-1)} = \tilde{\lambda}_r^{(k-\mu)}$, which, in return, gives

$$
\tilde{\lambda}_r^{(1)} = \left\langle \tilde{\mathbf{u}}_r^{(1)},\ \tilde{C}_r^{(1)}\tilde{\mathbf{u}}_r^{(2)} \right\rangle = \left\langle \mathbf{u}_r^{(1)},\ \tilde{C}_r^{(1)}\mathbf{u}_r^{(2)} \right\rangle = \lambda_r, \tag{4.3}
$$

$$\vdots$$

$$
\tilde{\lambda}_r^{(\ell)} = \left\langle \tilde{\mathbf{u}}_r^{(\ell)},\ \tilde{C}_r^{(\ell)}\tilde{\mathbf{u}}_r^{(\ell+1)} \right\rangle = \left\langle \mathbf{u}_r^{(\ell)},\ \tilde{C}_r^{(\ell)}\mathbf{u}_r^{(\ell+1)} \right\rangle = \tilde{\lambda}_r^{(\ell-1)} \quad (\ell = 3, 5, \ldots), \tag{4.4}
$$

$$\vdots$$

$$
\begin{aligned}
\tilde{\lambda}_r^{(k-\mu-1)} &= \left\langle \tilde{\mathbf{u}}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)}\tilde{\mathbf{u}}_r^{(k-\mu)} \right\rangle \\
&= \left\langle \mathbf{u}_r^{(k-\mu-1)},\ \tilde{C}_r^{(k-\mu-1)}\mathbf{u}_r^{(k-\mu)} \right\rangle = \tilde{\lambda}_r^{(k-\mu-2)}.
\end{aligned} \tag{4.5}
$$

Since $\lambda_{r,[p_j+1]}^{(1)}, \ldots, \lambda_{r,[p_j+1]}^{(k-\mu-1)}$ are the largest singular values of $C_{r,[p_j+1]}^{(1)}, \ldots, \lambda_{r,[p_j+1]}^{(k-\mu-1)}$, respectively, $\tilde{\lambda}_r^{(1)}, \ldots, \tilde{\lambda}_r^{(k-\mu-1)}$ are the largest singular values of $\tilde{C}_r^{(1)}, \ldots, \tilde{C}_r^{(k-\mu-1)}$, respectively.

Furthermore, under Assumption A, the dominant left and right singular vectors of $\tilde{C}_r^{(1)}, \ldots, \tilde{C}_r^{(k-\mu-1)}$ are unique. Therefore, we have for $r = 1, \ldots, R$,

$$\begin{aligned}
&\tilde{\mathbf{u}}_r^{(1)} = \mathbf{u}_r^{(1)},\ \tilde{\mathbf{u}}_r^{(2)} = \mathbf{u}_r^{(2)}, \quad \text{(by (4.3))}\\
&\ \vdots\\
&\tilde{\mathbf{u}}_r^{(\ell)} = \mathbf{u}_r^{(\ell)},\ \tilde{\mathbf{u}}_r^{(\ell+1)} = \mathbf{u}_r^{(\ell+1)}\ (\ell = 3, 5, \ldots), \quad \text{(by (4.4))}\\
&\ \vdots\\
&\tilde{\mathbf{u}}_r^{(k-\mu-1)} = \mathbf{u}_r^{(k-\mu-1)}, \tilde{\mathbf{u}}_r^{(k-\mu)} = \mathbf{u}_r^{(k-\mu)}; \quad \text{(by (4.5))}
\end{aligned}$$

that is, $\mathbf{u}_r^{(\ell)} = \tilde{\mathbf{u}}_r^{(\ell)}$ hold for $\ell = 1, \ldots, k-\mu$, $r = 1, \ldots, R$. Finally, since

$$\tilde{V}^{(\ell)} = \lim V_{[p_j+1]}^{(\ell)} = \lim V_{[p_j]}^{(\ell)} = V^{(\ell)}, \quad \tilde{\Lambda}^{(\ell)} = \lim \Lambda_{[p_j+1]}^{(\ell)} = \lim \Lambda_{[p_j]}^{(\ell)} = \Lambda^{(\ell)},$$

we have with Assumption A that

$$\tilde{U}^{(\ell)} = \lim U_{[p_j+1]}^{(\ell)} = \lim U_{[p_j]}^{(\ell)} = U^{(\ell)} \quad \forall \ell = k-\mu+1, \ldots, k. \qquad \square$$

Now we are ready to prove the convergence of the whole sequence $\{\mathbf{u}_{r,[p]}^{(\ell)}\}$ for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$.

THEOREM 4.4. *For almost all tensors $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$ satisfying Assumption* A, *except for an affine algebraic set of codimension one, the sequence $\{\mathbf{u}_{r,[p]}^{(\ell)}\}$ generated in Algorithm* 1 *converges for all $\ell = 1, \ldots, k$, $r = 1, \ldots, R$.*

Proof. Suppose that $\{\mathbf{u}_{r,[p_j]}^{(\ell)}\}$ is any subsequence converging to a limiting point $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$. By Lemma 4.2, $\mathbf{u}_r^{(\ell)}$ is isolated, and by Lemma 4.3, the subsequence $\{\mathbf{u}_{r,[p_j+1]}^{(\ell)}\}$ also converges to $\mathbf{u}_r^{(\ell)}$ and $\|\mathbf{u}_{r,[p_j+1]}^{(\ell)} - \mathbf{u}_{r,[p_j]}^{(\ell)}\| \to 0$ for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$. It follows from Lemma 2.5 that the whole sequence $\{\mathbf{u}_{r,[p_j+1]}^{(\ell)}\}$ converges to $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \ldots, k$, $r = 1, \ldots, R$. $\square$

5. Numerical experiments. In this section, we present numerical experiments to illustrate the convergence of Algorithm 1 by measuring

- objective value $\sum_{r=1}^R \lambda_r^2$;
- iterate error $\sum_{\ell=1}^k \sum_{r=1}^R \|\mathbf{u}_{r,[p+1]}^{(\ell)} - \mathbf{u}_{r,[p]}^{(\ell)}\|_2^2$.

In our experiments, $\mu = 2$, $R = 5$, and the size of all tensors is $\mathbb{R}^{20\times16\times10\times32}$.

For unstructured tensor data, it is ideal to compute the full rank decomposition. But this is a difficult task in general, so the low-rank tensor approximation is solved instead. Therefore, in our experiments we first test some unstructured tensors including random tensor, stochastic tensor, Cauchy tensor, Hilbert tensor, and Toeplitz tensor in order to test the the convergence of Algorithm 1:

- Random tensor [9]: randomly generate.
- Stochastic tensor [31]: $\tau_{i_1,i_2,i_3,i_4} = \begin{cases} c, & i_1 \neq i_2, i_2 \neq i_3, i_3 \neq i_4, \\ 0, & i_1 = i_2, i_2 \neq i_3, i_3 \neq i_4, \\ 1/20, & \text{otherwise,} \end{cases}$
  where c is randomly in $(0,1)$ by the homogenous distribution such as $\sum_{i_1 \in [\![20]\!]} \tau_{i_1,i_2,i_3,i_4} = 1$ with $i_j \neq i_{j+1}, j = 1, 2, 3$.
- Cauchy tensor [7]: $\tau_{i_1,i_2,i_3,i_4} = \frac{1}{c(i_1)+c(i_2)+c(i_3)+c(i_4)}$, where c is a random vector with size 32.
- Hilbert tensor [39]: $\tau_{i_1,i_2,i_3,i_4} = \frac{1}{i_1+i_2+i_3+i_4-3}$.

- Toeplitz tensor [8]: $\tau_{i_1+j,i_2+j,i_3+j,i_4+j} = \tau_{i_1,i_2,i_3,i_4}$ for $j \in [\![\min(20-i_1, 16-i_2, 10-i_3, 32-i_4)]\!]$.

Then we also test two simple, synthetic tensors for which low-rankness is guaranteed to test if Algorithm 1 provides solutions to the ground-truth solutions:

- Tensor 1: Randomly generate a rank-5 order-4 tensor satisfying the related orthogonality constraint and add a noise tensor which is generated by $10^{-6} *$ randn$(20, 16, 10, 32)$.
- Tensor 2: Randomly generate a rank-5 order-4 tensor satisfying the related orthogonality constraint and add a noise tensor which is generated by $10^{-5} *$ randn$(20, 16, 10, 32)$.

The initial vectors are chosen by four different ways:

- "Random Initial"—Unit vectors $\mathbf{u}_r^{(\ell)}$ for $\ell = 1, \dots, k$ and $r = 1, \dots, R$ are generated randomly to satisfy orthogonality constraint with $\mu = 2$.
- "Identity Initial"—Each $[\mathbf{u}_1^{(\ell)}, \dots, \mathbf{u}_R^{(\ell)}]$ for $\ell = 1, \dots, k$ are taken as the first R columns of identity matrices.
- "Orthogonal Initial"—Each $[\mathbf{u}_1^{(\ell)}, \dots, \mathbf{u}_R^{(\ell)}]$ for $\ell = 1, \dots, k$ are taken as the first R columns of random orthonormal matrices.
- "Singular Value Initial"—The major left singular vectors of the unfoldings of the tensors are used as initials.

Numerical results for the first 150 iterations are shown in Figures 5.1–5.7, in which the vertical scales on the left subfigures denote objective value $\sum_{r=1}^{R} \lambda_r^2$ and the vertical scales on the right subfigures denote the iterate errors$\sum_{\ell=1}^{k} \sum_{r=1}^{R} \|\mathbf{u}_{r,[p+1]}^{(\ell)} - \mathbf{u}_{r,[p]}^{(\ell)}\|_2^2$.

Figures 5.1–5.7 lead to the following observations:

- Objective value satisfies the monotone increasing property for each iteration as proved in the previous section.
- For different initial vectors, the approximated objective values may be different for the same test tensor; that is, iterates may converge to different limit points. Hence, the computed result is only optimal in a local neighborhood for each initial vector. As addressed in [9], it is interesting to study for what tensors or what initial guesses Algorithm 1 converges to the global optimum.
- When it comes to the qualities of the final approximation, among four different initial vectors, no single one offers obvious advantages. It is challenging to study for what initial guesses Algorithm 1 converges faster in terms of both objective values and iterate errors.

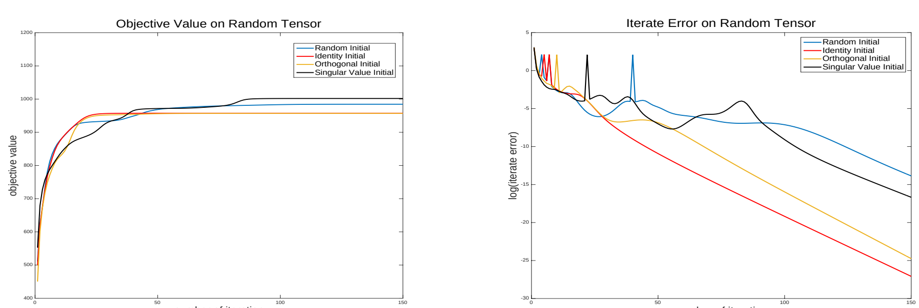

FIG. 5.1. *Comparison of random tensor.*

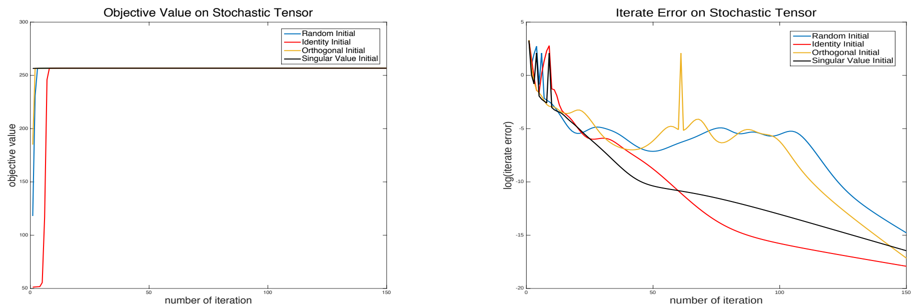

FIG. 5.2. *Comparison of stochastic tensor.*

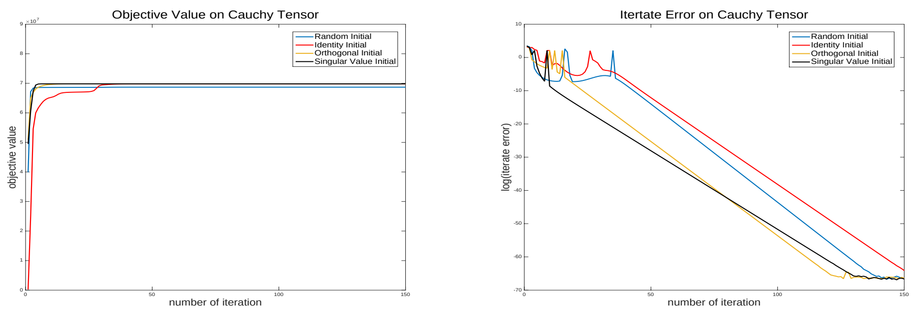

FIG. 5.3. *Comparison of Cauchy tensor.*

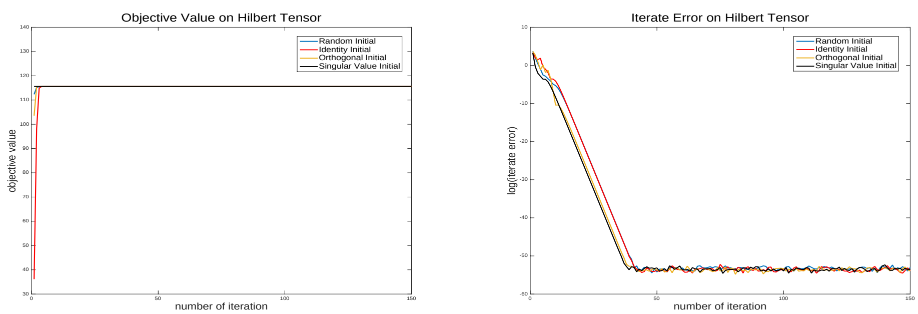

FIG. 5.4. *Comparison of Hilbert tensor.*

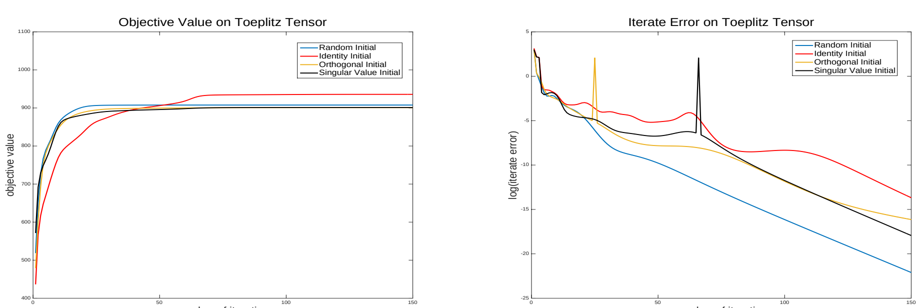

FIG. 5.5. *Comparison of Toeplitz tensor.*

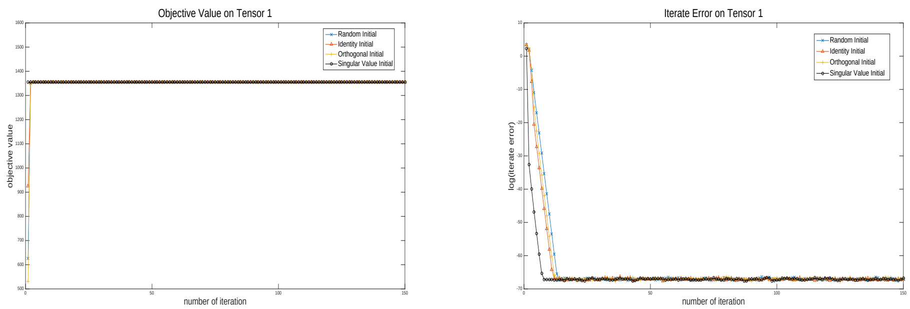

FIG. 5.6. *Comparison of tensor* 1.

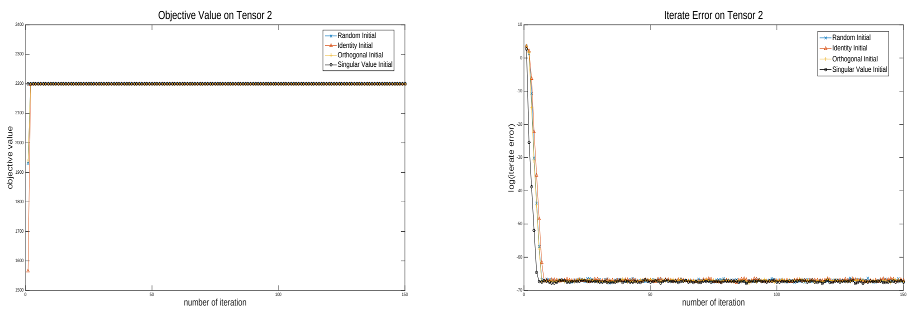

FIG. 5.7. *Comparison of tensor* 2.

- The objective values often "flatten out" quickly, whereas iterates will eventually converge, but they are not monotone in each step, as the iterate errors keep on changing for much longer. This is consistent with the well-known fact that tensors are known to demonstrate poor orthogonal approximation properties (i.e., the large rank parameter), and so in usual practice either the CP or the orthogonal Tucker decomposition is applied in the case of moderate-size tensors. Thus, the phenomenon that iterates converge much slower than the objective values is very intriguing and worthy of further investigation.
- We have checked that in all our examples the largest singular values of $\{C^{(\ell)}_{r,[p]}\}$ are simple and the matrix $V^{(\ell)}_{[p]}\Lambda^{(\ell)}_{[p]}$ are of full column rank. Hence, Assumption A holds in our numerical examples. Consequently, it is not clear why the iterate errors may increase suddenly. Hence, it is important to study the ways to eliminate the sudden jump of the iterate errors.
- For Tensors 1 and 2, the approximated objective values converge to the optimal values quickly. In this sense, Algorithm 1 provides solutions to the ground-truth solutions.

6. Conclusion. In this paper an SVD-based algorithm—Algorithm 1—has been presented for the orthogonal low-rank approximation problem (1.4) of tensors, which includes the completely orthogonal low-rank approximation [9] and semiorthogonal low-rank approximation [48] as two special cases. The convergence of Algorithm 1 for both objective function and iterates themselves has been analyzed and also illustrated by some numerical examples.

Some important issues are worthy of further study:

- The optimizer of our approximation problem is not unique since an optimal solution can be changed to another by applying appropriate orthogonal transformations on the components of the lower-rank tensor; that is, the objective values are constant on the orbits of certain group transformation. The fact that the objective value levels out probably indicates that the optimization problem is stuck within an orbit of such a group action, but then why should the iterates converge since all points in the group action give the same objective value, which means that any point in the orbit is equally valid as an optimizer? This is an important task for our future research.
- It has been proved that Algorithm 1 is convergent. But it is not clear how the rank parameter affects the convergence rate; this is still under investigation.
- Our convergence analysis on iterates in Algorithm 1 is based on Assumption A, which applies to the sequence of iterations generated by Algorithm 1 but not to the target tensor. Hence, this assumption could not be verified before the full run of Algorithm 1. Therefore, it is important to develop a mathematical theory to verify if Assumption A holds or propose some conditions on the target tensor which can be verified numerically and ensure that Assumption A holds. This is a subject for our future work.

Acknowledgments. The authors would like to thank Professor Gregory Beylkin, and the anonymous referees for their valuable comments and suggestions which have helped to improve the overall presentation of the paper. Both authors are grateful to Professor Moody T. Chu for his kind advice, comments, and suggestions on the present work.

REFERENCES

[1] E. Acar, S. A. Çamtepe, M. S. Krishnamoorthy, and B. Yener, *Modeling and multiway analysis of chatroom tensors*, in International Conference on Intelligence and Security Informatics, Springer, Berlin, 2005, pp. 256–268.

[2] E. Acar, S. A. Çamtepe, and B. Yener, *Collective sampling and analysis of high order tensors for chatroom communications*, in International Conference on Intelligence and Security Informatics, Springer, Berlin, 2006, pp. 213–224.

[3] C. M. Andersen and R. Bro, *Practical aspects of parafac modeling of fluorescence excitation-emission data*, J. Chemometrics, 17 (2003), pp. 200–215.

[4] C. J. Appellof and E. R. Davidson, *Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents*, Anal. Chem., 53 (1981), pp. 2053–2056.

[5] B. W. Bader, M. W. Berry, and M. Browne, *Discussion tracking in Enron email using Parafac*, in Survey of Text Mining II, Springer, Berlin, 2008, pp. 147–163.

[6] J. Carroll and J.-J. Chang, *Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart–Young decomposition*, Psychometrika, 35 (1970), pp. 283–319.

[7] H. Chen and L. Qi, *Positive definiteness and semi-definiteness of even order symmetric Cauchy tensors*, JEM, 11 (2015), pp. 1263–1274.

[8] Z. Chen and L. Qi, *Circulant tensors with applications to spectral hypergraph theory and stochastic process*, JEM, 12 (2016), pp. 1227–1247.

[9] J. Chen and Y. Saad, *On the tensor SVD and the optimal low rank orthogonal approximation of tensors*, SIAM J. Matrix Anal. Appl., 30 (2008/09), pp. 1709–1734.

[10] L. Cheng, Y.-C. Wu, and H. V. Poor, *Probabilistic tensor canonical polyadic decomposition with orthogonal factors*, IEEE Trans. Signal Process., 65 (2017), pp. 663–676.

[11] L. De Lathauwer and J. Castaing, *Tensor-based techniques for the blind separation of DS–CDMA signals*, Signal Process., 87 (2007), pp. 322–336.

[12] L. De Lathauwer and J. Castaing, *Blind identification of underdetermined mixtures by simultaneous matrix diagonalization*, IEEE Trans. Signal Process., 56 (2008), pp. 1096–1105.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
[15] V. de Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[16] A. Doostan, G. Iaccarino, and N. Etemadi, *A least-squares approximation of high-dimensional uncertain systems*, Annual Research Briefs, Center for Turbulence Research, Stanford University, (2007), pp. 121–132.
[17] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.
[18] S. Friedland, V. Mehrmann, R. Pajarola, and S. K. Suter, *On best rank one approximation of tensors*, Numer. Linear Algebra Appl., 20 (2013), pp. 942–955.
[19] Y. Guan, M. T. Chu, and D. Chu, *Convergence analysis of an SVD-based algorithm for the best rank-1 tensor approximation*, Linear Algebra Appl., 555 (2018), pp. 53–69.
[20] Y. Guan, M. T. Chu, and D. Chu, *SVD-based algorithms for the best rank-1 approximation of a symmetric tensor*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1095–1115.
[21] W. Hackbusch and B. N. Khoromskij, *Tensor-product approximation to operators and functions in high dimensions*, J. Complexity, 23 (2007), pp. 697–714.
[22] W. Hackbusch, B. N. Khoromskij, and E. E. Tyrtyshnikov, *Hierarchical Kronecker Tensor-Product Approximations*, J. Numer. Math., 13 (2005), pp. 119–156.
[23] R. Harshman, *Foundations of the Parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970).
[24] N. J. Higham, *Computing the polar decomposition—with applications*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 1160–1174.
[25] B. N. Khoromskij, *Tensor Numerical Methods in Scientific Computing. Research Monograph*, De Gruyter Verlag, Berlin, 2018.
[26] B. N. Khoromskij and V. Khoromskaia, *Multigrid tensor approximation of function related arrays*, SIAM J. Sci. Comput., 31 (2009), pp. 3002–3026.
[27] T. G. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255.
[28] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[29] W. P. Krijnen, T. K. Dijkstra, and A. Stegeman, *On the non-existence of optimal solutions and the occurrence of "degeneracy" in the CANDECOMP/PARAFAC model*, Psychometrika, 73 (2008), pp. 431–439.
[30] D. Leibovici and R. Sabatier, *A singular value decomposition of a k-way array for a principal component analysis of multiway data, PTA-k*, Linear Algebra Appl., 269 (1998), pp. 307–329.
[31] W. Li, D. Liu, M. K. Ng, and S.-W. Vong, *The uniqueness of multilinear PageRank vectors*, Numer. Linear Algebra Appl., 24 (2017), e2017.
[32] L.-H. Lim, *Singular values and eigenvalues of tensors: A variational approach*, in Proceedings of the 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005, pp. 129–132.
[33] J. Mocks, *Topographic components model for event-related potentials and some biophysical considerations*, IEEE Trans. Biomed. Eng., 35 (1988), pp. 482–484.
[34] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.
[35] P. A. Regalia, *Monotonically convergent algorithms for symmetric tensor approximation*, Linear Algebra Appl., 438 (2013), pp. 875–890.
[36] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, *Parallel factor analysis in sensor array processing*, IEEE Trans. Signal Process., 48 (2000), pp. 2377–2388.
[37] N. D. Sidiropoulos and R. S. Budampati, *Khatri-rao space-time codes*, IEEE Trans. Signal Process., 50 (2002), pp. 2396–2407.
[38] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, *Blind PARAFAC receivers for DS-CDMA systems*, IEEE Trans. Signal Process., 48 (2000), pp. 810–823.
[39] Y. Song and L. Qi, *Infinite and finite dimensional Hilbert tensors*, Linear Algebra Appl., 451 (2014), pp. 1–14.
[40] A. J. Sommese and C. W. Wampler, II, *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*, World Scientific, Hackensack, NJ, 2005.

[41] M. Sørensen, L. D. Lathauwer, P. Comon, S. Icart, and L. Deneire, *Canonical polyadic decomposition with a columnwise orthonormal factor matrix*, SIAM J. Matrix Anal. Appl., 33 (2012), pp. 1190–1213.
[42] A. Stegeman, *Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p+1$ or higher*, Psychometrika, 71 (2006), pp. 483–501.
[43] A. Stegeman, *Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 988–1007.
[44] V. N. Temlyakov, *Greedy algorithms and M-term approximation with regard to redundant dictionaries*, J. Approx. Theory, 98 (1999), pp. 117–145.
[45] L. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
[46] C. F. Van Loan, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123 (2000), pp. 85–100.
[47] L. Wang and M. T. Chu, *On the global convergence of the alternating least squares method for rank-one tensor approximation*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1058–1072.
[48] L. Wang, M. T. Chu, and B. Yu, *Orthogonal low rank tensor approximation: Alternating least squares method and its global convergence*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 1–19.
[49] E. Zander and H. Matthies, *Tensor product methods for stochastic problems*, PAMM, 7 (2007), pp. 2040067–2040068.
[50] T. Zhang and G. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.