CONVERGENCE ANALYSIS OF AN SVD-BASED ALGORITHM FOR THE BEST RANK-1 TENSOR APPROXIMATION

YU GUAN* AND MOODY T. CHU†

Abstract. This paper revisits the classical problem of finding the best rank-1 approximation to a generical tensor. Conventionally, the approximation is carried out by the so called alternative least squares (ALS) method, also known as the high-order power method. The method works on one factor a time and suffers from slow convergence and easy stagnation at a local solution. Recently, an SVD-based that improves two factors simultaneously has been proposed in the literature and is reported to have better limiting behavior leading to better approximations. The main contribution of this paper is at providing a rigorous mathematical proof for the convergence of the iterates.

Key words. best rank-1 tensor approximation, singular value decomposition

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction. A real-valued tensor of order k can be represented by a k-way array

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

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

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

where elements are the products of entries from vectors $\mathbf{u}^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell = 1, \ldots, k$, is said to be of rank one. The problem of finding a best rank-1 approximation to T is to determine unit vectors $\mathbf{u}^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell = 1, \ldots k$, and a scalar λ such that the functional

$$f\left(\lambda, \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\right) := \left\| T - \lambda \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\|_F^2 = \sum_{i_1, i_2, \ldots, i_k} \left(t_{i_1,\ldots,i_k} - \lambda u_{i_1}^{(1)} \ldots u_{i_k}^{(k)} \right)^2 \tag{1.1}$$

is minimized. For any fixed unit vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$, the optimal value of λ for (1.1) is given precisely by the length of the projection of the "vector" T onto the direction of the "unit vector" $\bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k}$, i.e.,

$$\lambda = \lambda\left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\right) = \left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\rangle. \tag{1.2}$$

Thus, minimizing the orthogonal component of T, as is desired in (1.1), is equivalent to maximizing the length $|\lambda|$ of the parallel component. In [10], the expression (1.2) is called the generalized Rayleigh quotient of T relative to $\{\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\}$. Many approaches for finding the extreme values of (1.2) have been proposed in the literature. See, for example, [1, 3, 4, 5, 6, 7]. Switching the signs of the variables $\mathbf{u}^{(\ell)}$ if necessary, we may restrict our attention without loss of generality to the case that $\lambda > 0$ only. Enumerating the entries of T in different order if necessary, we shall also assume that $I_1 \geq I_2 \geq \ldots \geq I_k$ henceforth.

2. Basics. Tensors are multi-dimensional arrays. Thus, there are multiple ways to define tensor multiplications. Their appearance are often rather complex and perplexing. To facilitate the subsequence discussion, we first introduce a simple notation system that generalizes what we already know from the matrix theory. We also establish a few useful tools.

*Department of Mathematics, National Singapore University, Singapore 119076. (ricky7guanyu@gmail.com)

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu) This research was supported in part by the National Science Foundation under grant DMS-1316779.

Let the symbol $[\![m]\!]$ denote henceforth the set of integers $\{1, \dots m\}$ for a given positive integer m. Suppose that the set $[\![k]\!]$ is partitioned as the union of two disjoint nonempty subsets $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_s\}$ and $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_t\}$, where $s + t = k$. An element in the tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_k}$ will be marked as $\tau_{[\mathcal{I}|\mathcal{J}]}^{(\boldsymbol{\alpha},\boldsymbol{\beta})}$ where $\mathcal{I} := (i_1, \dots, i_s)$ and $\mathcal{J} := (j_1, \dots, j_t)$ 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 $\tau_{[\mathcal{I}|\mathcal{J}]}$. The representation $\tau_{[\mathcal{I}|\mathcal{J}]}$ is $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ specific, but there is no preference of $\boldsymbol{\alpha}$ over $\boldsymbol{\beta}$. The partition $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ may be regarded as generalizing the familiar notion of rows and columns for matrices which is easy for our perception. The concept of rows and columns is cumbersome to describe for multi-dimensional arrays.

Given a fixed partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, we shall regard an order-k tensor $T \in \mathbb{R}^{I_1 \times \dots \times I_k}$ as a "matrix representation" of a linear operator mapping order-s tensors to order-t tensors [9]. Specifically, we identify T with the linear map

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

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

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

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

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

is the Frobenius inner product generalized to multi-dimensional arrays. The tensor multiplication $\circledast_{\boldsymbol{\beta}}$ is a natural generalization of the usual matrix-vector multiplication in the sense that if an order-2 tensor $T \in \mathbb{R}^{m \times n}$ is regarded as a matrix and if we identify $\tau_{ij} = \tau_{[i|j]}^{(\{2\},\{1\})}$, then with respect to given column vectors $\mathbf{z} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ we can write

$$\begin{cases} T\mathbf{z} &= T \circledast_2 \mathbf{z}, \\ T^\top \mathbf{y} &= T \circledast_1 \mathbf{y}. \end{cases} \tag{2.4}$$

That is, the column is identified by $\boldsymbol{\alpha} = \{1\}$ and the row by $\boldsymbol{\beta} = \{2\}$.

We first describe two basic facts that are related to the associative laws.

LEMMA 2.1. *Given a general tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_k}$, a partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$, and vectors $\mathbf{u}^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell = 1, \dots, k$, then it holds that*

$$\left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\rangle = \left\langle T \circledast_{\boldsymbol{\beta}} \bigotimes_{i=1}^{s} \mathbf{u}^{(\alpha_i)}, \bigotimes_{j=1}^{t} \mathbf{u}^{(\beta_j)} \right\rangle. \tag{2.5}$$

Proof. Based on the definition (2.2), the right hand side of (2.5) is simply a rearrangement of terms in the summation by the associative law. □

Given a rank-1 tensor $S \in \mathbb{R}^{I_{\alpha_1} \times \dots \times I_{\alpha_{k-2}}}$, let $\boldsymbol{\beta} = \{\beta_1, \beta_2\}$ be the complement of $\{\alpha_1, \dots, \alpha_{k-2}\}$. 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. *Given a general tensor $T \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_k}$, arbitrary vectors $\mathbf{u}^{(\alpha_i)} \in \mathbb{R}^{I_{\alpha_i}}$, $i \in [\![k-2]\!]$, $\mathbf{v} \in \mathbb{R}^{I_{\beta_2}}$, and $\mathbf{w} \in \mathbb{R}^{I_{\beta_1}}$, then*

$$\left(T \circledast_{\{\beta_1,\beta_2\}} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\alpha_i)}\right) \circledast_{\beta_1} \mathbf{v} = \left(T \circledast_{\{\beta_1,\alpha_j\}} \bigotimes_{i=1}^{j-1} \mathbf{u}^{(\alpha_i)} \otimes \mathbf{v} \otimes \bigotimes_{i=j+1}^{k-2} \mathbf{u}^{(\alpha_i)}\right) \circledast_{\beta_1} \mathbf{u}^{(\alpha_j)} \tag{2.6}$$

$$\left(T \circledast_{\{\beta_1,\beta_2\}} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\alpha_i)}\right) \circledast_{\beta_2} \mathbf{w} = \left(T \circledast_{\{\alpha_j,\beta_2\}} \bigotimes_{i=1}^{j-1} \mathbf{u}^{(\alpha_i)} \otimes \mathbf{w} \otimes \bigotimes_{i=j+1}^{k-2} \mathbf{u}^{(\alpha_i)}\right) \circledast_{\beta_2} \mathbf{u}^{(\alpha_j)} \tag{2.7}$$

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

Proof. The notion of a tensor entry $\tau_{[\mathcal{I}|\mathcal{J}]}^{(\boldsymbol{\alpha},\boldsymbol{\beta})}$ defined earlier will be informative here for tracking which index is being associated with which location. The μ-th entry of the vector $\left(T \circledast_{\{\beta_1,\beta_2\}} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\alpha_i)}\right) \circledast_{\beta_1} \mathbf{v}$ is given by

$$\begin{aligned}
&\sum_{\nu=1}^{I_{\beta_2}} \left(\sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_{k-2}=1}^{I_{\alpha_{k-2}}} \tau_{[i_1,\ldots,i_{k-2}|\mu,\nu]}^{(\boldsymbol{\alpha},\boldsymbol{\beta})} u_{i_1}^{(\alpha_1)} \ldots u_{i_{k-2}}^{(\alpha_{k-2})} \right) v_\nu \\
&= \sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_{k-2}=1}^{I_{\alpha_{k-2}}} \sum_{\nu=1}^{I_{\beta_2}} \tau_{[i_1,\ldots,i_{k-2},\nu|\mu]}^{(\boldsymbol{\alpha} \bigcup \{\beta_2\},\{\beta_1\})} u_{i_1}^{(\alpha_1)} \ldots u_{i_{k-2}}^{(\alpha_{k-2})} v_\nu \\
&= \sum_{i_j=1}^{I_{\alpha_j}} \left(\sum_{i_1=1}^{I_{\alpha_1}} \cdots \sum_{i_{k-2}=1}^{I_{\alpha_{k-2}}} \sum_{\nu=1}^{I_{\beta_2}} \tau_{[i_1,\ldots,i_{k-2},\nu|\mu,i_j]}^{(\boldsymbol{\alpha} \bigcup \{\beta_2\}-\{\alpha_j\},\{\beta_1,\alpha_j\})} u_{i_1}^{(\alpha_1)} \ldots u_{i_{k-2}}^{(\alpha_{k-2})} v_\nu \right) u_{i_j}^{(\alpha_j)},
\end{aligned}$$

where the last equality is obtained by the associative law so that the summation inside the parentheses contains no $u_{i_j}^{(\alpha_j)}$ terms. A similar argument can prove (2.7). □

For latter usage in our proof for convergence, we also need the following results from real analysis.

LEMMA 2.3. *Assume that a^* is an isolated accumulation point of a sequence $\{a_k\}$ such that for every subsequence $\{a_{k_j}\}$ converging to 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 $\{a_k\}$ converges to a^*.*

3. SVD-based best rank-1 approximation. We now investigate a possible application of the singular value decomposition (SVD) to the best rank-1 approximation of a generic tensor. We first restate the following fundamental fact in matrix theory, whence we explain the kind of generic property we need for a generic tensor. Our goal is to achieve the best rank-1 approximation by improving two components a time via the SVD.

LEMMA 3.1. *Given a matrix $A \in \mathbb{R}^{m \times n}$, then the global maximum of the generalized Rayleigh quotient*

$$\max_{\substack{\mathbf{y} \in \mathbb{R}^m,\, \|\mathbf{y}\| = 1 \\ \mathbf{z} \in \mathbb{R}^n,\, \|\mathbf{z}\| = 1}} \mathbf{y}^\top A \mathbf{z} \tag{3.1}$$

is precisely the largest singular value σ_1 of A, where the global maximizer $(\mathbf{y}_1, \mathbf{z}_1)$ consists of precisely the corresponding left and right singular vectors. The best rank-1 approximation to A is given by $\sigma_1 \mathbf{y}_1 \mathbf{z}_1^\top$.

Suppose that $\lambda \bigotimes_{\ell=1}^k \overline{\mathbf{u}}^{(\ell)}$ is the best rank-1 approximation to a given order-k tensor T. By (1.2), the generalized Rayleigh quotient $\lambda = \left\langle T, \bigotimes_{\ell=1}^k \overline{\mathbf{u}}^{(\ell)} \right\rangle$ is positive and maximal. Consider an arbitrary partitioning $[\![k]\!] = \boldsymbol{\alpha} \cup \boldsymbol{\beta}$ with the cardinality $|\boldsymbol{\beta}| = 2$. By Lemma 2.1, we can write

$$\lambda = \left\langle T \circledast_{\boldsymbol{\beta}} \bigotimes_{i=1}^{k-2} \overline{\mathbf{u}}^{(\alpha_i)}, \bigotimes_{j=1}^{2} \overline{\mathbf{u}}^{(\beta_j)} \right\rangle.$$

Algorithm 1 (Best rank-1 approximation via SVD updating with cyclic progression.)

Require: An order-k, n-dimensional tensor $T \in \mathbb{R}^{I_1 \times \cdots \times I_k}$ and k starting unit vectors $\mathbf{u}_{[0]}^{(\ell)} \in \mathbb{R}^{I_\ell}$, $\ell \in [\![k]\!]$
Ensure: A local best rank-1 approximation to T

1: **for** $p = 0, 1, \cdots$, **do**
2: **for** $\ell = 1, 2, \cdots, k-1$, **do**
3: $\boldsymbol{\beta}_\ell = (\ell, \ell+1)$
4: $C_{[p]}^{(\ell)} = T \circledast_{\boldsymbol{\beta}_\ell} \bigotimes_{i=1}^{\ell-1} \mathbf{u}_{[p+1]}^{(i)} \otimes \bigotimes_{i=\ell+2}^{k} \mathbf{u}_{[p]}^{(i)}$ {A matrix of size $I_\ell \times I_{\ell+1}$}
5: $[\mathbf{u}, s, \mathbf{v}] = \mathsf{svds}(C_{[p]}^{(\ell)}, 1)$ {Dominant singular value triplet via Matlab routine svds}
6: **if** $u_1 < 0$ **then**
7: $\mathbf{u} = -\mathbf{u}, \mathbf{v} = -\mathbf{v}$
8: **end if**
9: $\mathbf{u}_{[p+1]}^{(\ell)} := \mathbf{u}$ {If $\ell = 1$, this is $\widehat{\mathbf{u}}_{[p+1]}^{(1)}$; otherwise this is the second update $\mathbf{u}_{[p+1]}^{(\ell)}$, if $2 \le \ell < k$.}
10: $\widehat{\mathbf{u}}_{[p+1]}^{(\ell+1)} := \mathbf{v}$ {Skipping this step will not affect $C_{[p]}^{(\ell+1)}$ at Line 4.}
11: $\lambda_{[p+1]}^{(\ell)} := s$
12: **end for**
13: $\boldsymbol{\beta}_k = (1, k)$
14: $C_{[p]}^{(k)} = T \circledast_{\boldsymbol{\beta}_k} \bigotimes_{i=2}^{k-1} \mathbf{u}_{[p+1]}^{(i)}$ {A matrix of size $I_1 \times I_k$}
15: $[\mathbf{u}, s, \mathbf{v}] = \mathsf{svds}(C_{[p]}^{(k)}, 1)$ {Dominant singular value triplet via Matlab routine svds}
16: $\mathbf{u}_{[p+1]}^{(k)} := \mathbf{v}$ {After adjusting the signs of $\mathbf{u}$ and $\mathbf{v}$ properly as in Line 6.}
17: $\mathbf{u}_{[p+1]}^{(1)} := \mathbf{u}$
18: $\lambda_{[p+1]}^{(k)} := s$
19: **end for**

The product $C_{\boldsymbol{\beta}} := T \circledast_{\boldsymbol{\beta}} \bigotimes_{i=1}^{k-2} \overline{\mathbf{u}}^{(\alpha_i)}$ is a matrix in $\mathbb{R}^{I_{\beta_1} \times I_{\beta_2}}$. Since λ is the maximal generalized Rayleigh quotient, by Lemma 3.1, we conclude that $\overline{\mathbf{u}}^{(\beta_1)}$ and $\overline{\mathbf{u}}^{(\beta_2)}$ must be the left and the right singular vectors associated with the largest singular value λ of $C_{\boldsymbol{\beta}}$ for any $\boldsymbol{\beta}$. This is the SVD certification of the best rank-1 approximation to a given tensor T. We are thus motivated to formulate an SVD-based approach to calculate the best rank-1 approximation by iterations. The idea has been proposed earlier in [3], but so far as we know no convergence analysis has ever been established. The main contribution of this paper is to furnish the proof of convergence for generic tensors.

The most basic approach is outlined in Algorithm 1. For efficiency, we also propose a modification by random permutations in Algorithm 2. Two types of dynamics are involved in all algorithms. One is the dynamics of the objective values, of which the analysis is straightforward. The other is the dynamics of the iterates, which is much harder to characterize. We will discuss the convergence in the next section.

To convey the idea, it is convenient to adopt the subscript $_{[p]}$ in Algorithm 1 to indicate the quantity at the p-th iteration. Each sweep of p at Line 1 in Algorithm 1 involves k pairs of $\boldsymbol{\beta}$ ranging circularly from $(1,2), (2,3), \ldots, (k-1, k)$ and $(1, k)$. Thus, each $\mathbf{u}_{[p+1]}^{(\ell)}$ is updated twice. The first updates for $\ell = 2, \ldots, k$, denoted by $\widehat{\mathbf{u}}_{[p+1]}^{(\ell)}$ at Line 10, are not essential and can be completely removed from the algorithm, but its presence helps bridge the monotonicity. The update $\widehat{\mathbf{u}}_{[p+1]}^{(1)}$ is temporarily overwritten as $\mathbf{u}_{[p+1]}^{(1)}$ at Line 9 for the computation of $C_{[p]}^{(\ell)}$ at Line 4 for $\ell = 2, \ldots, k-1$, but will be updated again at Line 17. The switch of signs at Line 7 conditioned upon Line 6 is to ensure that the iterates are aligned in one direction and thus avoid discontinuous jumps. We register the intermediate values $\lambda_{[p+1]}^{(\ell)}$ as well, even though only $\lambda_{[p+1]}^{(k)}$ at the final stage is crucial.

The above algorithm is different from the alternating least squares (ALS) approach that has been popular for computing the best rank-1 approximation [2, 5, 10]. The most significant difference is that, since the dominant singular vector $\mathbf{u}_{[p+1]}^{(\ell)}$ of the matrix $C_{[p]}^{(\ell)}$ gives rise to the absolute maximal value $\lambda_{[p+1]}^{(\ell)}$ for the functional

$$g(\mathbf{x},\mathbf{y}) := \left\langle T, \bigotimes_{i=1}^{\ell-1} \mathbf{u}_{[p+1]}^{(i)} \otimes \mathbf{x} \otimes \mathbf{y} \otimes \bigotimes_{i=\ell+2}^{k} \mathbf{u}_{[p]}^{(i)} \right\rangle \tag{3.2}$$

among all possible vectors $\mathbf{x}$ and $\mathbf{y}$, the mechanism of updating $\mathbf{x}$ and $\mathbf{y}$ simultaneously in Algorithm 1 is going to increase the generalized Rayleigh quotient faster than the combination of two applications of ALS approach to $\mathbf{x}$ followed by $\mathbf{y}$. The gain is also better than the maximum of updating $\mathbf{x}$ or $\mathbf{y}$ separately as that discussed in [3, Preposition 4].

Algorithm 2 (Best rank-1 approximation via SVD updating with randomization.)

Require: An order-k, n-dimensional, generic tensor T and k starting unit vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)} \in \mathbb{R}^n$
Ensure: A local best rank-1 approximation to T

1: $t \leftarrow 0$
2: $\lambda_0 \leftarrow \left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\rangle$
3: **repeat**
4: $\quad t \leftarrow t+1$
5: $\quad \sigma \leftarrow$ random permutation of $\{1, \ldots, k\}$
6: $\quad \boldsymbol{\beta}_t \leftarrow (\sigma_{k-1}, \sigma_k)$
7: $\quad C_t \leftarrow T \circledast_{\boldsymbol{\beta}_t} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\sigma_i)}$
8: $\quad [\mathbf{u}_t, s_t, \mathbf{v}_t] = \mathsf{svds}(C_t, 1)$ {Dominant singular value triplet via Matlab routine svds}
9: $\quad$ **if** $(\mathbf{u}_t)_1 < 0$ **then**
10: $\quad\quad \mathbf{u} = -\mathbf{u}_t, \mathbf{v} = -\mathbf{v}_t$
11: $\quad$ **end if**
12: $\quad \lambda_t \leftarrow s_t$
13: $\quad \mathbf{u}^{(\sigma_{k-1})} \leftarrow \mathbf{u}_t, \mathbf{u}^{(\sigma_k)} \leftarrow \mathbf{v}_t$
14: **until** λ_t meets convergence criteria

The cyclic progression as is indicated by the ℓ-loop in Algorithm 1 is not necessary. An alternative way is to shuffle the columns $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$ by a random permutation σ and generate a matrix C for updating. This randomized procedure is modified at Line 7 in Algorithm 2. To avoid confusion with data generated from Algorithm 1, we employ a slightly different notation when describing this algorithm. For simplicity, we always choose to update the last two vectors $\mathbf{u}^{(\sigma_{k-1})}, \mathbf{u}^{(\sigma_k)}$ after the permutation. A classical result in probability theory asserts that the expected number of trials for a permutation to recur is $\frac{k(k-1)}{2}$. But by the time that a repetition occurs, the vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$ should have been updated. If the convergence is ever to happen, the effect of reshuffling will gradually diminish.

4. Convergence analysis. In this section, we analyze the convergence for the above algorithms. We first show the monotonicity of the generalized Rayleigh quotients. Then we argue that the iterates themselves also converge. The latter answers an open question on convergence analysis for the SVD-based methods, as it offers an understanding of what was declared as "we do not have a complete understanding when this will happen" in [3, Page 947]. We concentrate mainly on Algorithm 1. The analysis of Algorithm 2 is left to interested readers.

4.1. Convergence of objective values. Because the SVD is involved, where the dominant singular value and singular vectors are selected at each update, both algorithms enjoy the property that the corresponding sequences of the generalized Rayleigh quotients are bounded and monotone increasing.

LEMMA 4.1. *The scalars* $\left\{\lambda_{[p]}^{(\ell)}\right\}$ *generated in Algorithm 1 form a monotone convergent sequence for each* $\ell = 1, \ldots, k$ *and all converge to the same value.*

Proof. It suffices to prove the assertion for the case $\lambda_{[p]}^{(k)}$ because the following argument shows that all other cases are sandwiched in between.

By applying Lemma 3.1 to each of the matrices $C_{[p]}^{(\ell)}$ consecutively, we observe that at any stage of p the chain of inequalities

$$\begin{aligned}
\lambda_{[p]}^{(k)} &= \left|\left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}_{[p]}^{(\ell)} \right\rangle\right| = \left|\left\langle T \circledast_{\boldsymbol{\beta}_1} \bigotimes_{\ell=3}^{k} \mathbf{u}_{[p]}^{(\ell)}, \mathbf{u}_{[p]}^{(1)} \otimes \mathbf{u}_{[p]}^{(2)} \right\rangle\right| \leq \left|\left\langle T \circledast_{\boldsymbol{\beta}_1} \bigotimes_{\ell=3}^{k} \mathbf{u}_{[p]}^{(\ell)}, \widehat{\mathbf{u}}_{[p+1]}^{(1)} \otimes \widehat{\mathbf{u}}_{[p+1]}^{(2)} \right\rangle\right| \\
&= \lambda_{[p+1]}^{(1)} = \left|\left\langle T \circledast_{\boldsymbol{\beta}_2} \widehat{\mathbf{u}}_{[p+1]}^{(1)} \otimes \bigotimes_{\ell=4}^{k} \mathbf{u}_{[p]}^{(\ell)}, \widehat{\mathbf{u}}_{[p+1]}^{(2)} \otimes \mathbf{u}_{[p]}^{(3)} \right\rangle\right| \leq \left|\left\langle T \circledast_{\boldsymbol{\beta}_2} \widehat{\mathbf{u}}_{[p+1]}^{(1)} \otimes \bigotimes_{\ell=4}^{k} \mathbf{u}_{[p]}^{(\ell)}, \mathbf{u}_{[p+1]}^{(2)} \otimes \widehat{\mathbf{u}}_{[p+1]}^{(3)} \right\rangle\right| \\
&= \lambda_{[p+1]}^{(2)} \leq \cdots \\
&\leq \lambda_{[p+1]}^{(k-1)} = \left|\left\langle T, \widehat{\mathbf{u}}_{[p+1]}^{(1)} \otimes \bigotimes_{\ell=2}^{k-1} \mathbf{u}_{[p+1]}^{(\ell)} \otimes \widehat{\mathbf{u}}_{[p+1]}^{(k)} \right\rangle\right| = \left|\left\langle T \circledast_{\boldsymbol{\beta}_k} \bigotimes_{\ell=2}^{k-1} \mathbf{u}_{[p+1]}^{(\ell)}, \widehat{\mathbf{u}}_{[p+1]}^{(k)} \otimes \widehat{\mathbf{u}}_{[p+1]}^{(1)} \right\rangle\right| \\
&\leq \left|\left\langle T \circledast_{\boldsymbol{\beta}_k} \bigotimes_{\ell=2}^{k-1} \mathbf{u}_{[p+1]}^{(\ell)}, \mathbf{u}_{[p+1]}^{(k)} \otimes \mathbf{u}_{[p+1]}^{(1)} \right\rangle\right| = \lambda_{[p+1]}^{(k)}
\end{aligned}$$

are always maintained. The monotone sequence $\left\{\lambda_{[p]}^{(k)}\right\}$ is bounded above by $\|T\|_F$, so it must converge. The inequalities sandwich the sequences one after another, so their limit points must be the same. □

LEMMA 4.2. *The scalars* $\{\lambda_t\}$ *generated in Algorithm 2 form a monotone convergent sequence.*

Proof. By applying Lemma 3.1, it is still true that at any stage of t we always have

$$\lambda_t = \left|\left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\rangle\right| = \left|\left\langle T \circledast_{\boldsymbol{\beta}_t} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\sigma_i)}, \mathbf{u}^{(\sigma_{k-1})} \otimes \mathbf{u}^{(\sigma_k)} \right\rangle\right| \leq \left|\left\langle T \circledast_{\boldsymbol{\beta}_t} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\sigma_i)}, \mathbf{u}_{t+1} \otimes \mathbf{v}_{t+1} \right\rangle\right| = \lambda_{t+1},$$

since $\mathbf{u}_{t+1}$ and $\mathbf{v}_{t+1}$ are the dominant singular vectors of the matrix $T \circledast_{\boldsymbol{\beta}_t} \bigotimes_{i=1}^{k-2} \mathbf{u}^{(\sigma_i)}$. Again, the monotone sequence $\{\lambda_t\}$ is bounded above by $\|T\|_F$, so it must converge. □

4.2. Convergence of Iterates. The above argument about the monotonicity of values $\lambda_{[p]}^{(k)}$ and λ_t is interesting and important, but certainly not enough, because the convergence of objective values does not guarantee the convergence of iterates to a global minimizer nor even to a stationary point [7]. What need be done for both algorithms is to argue that generically the iterates of vectors themselves also converge. We give a detailed account of the dynamical behavior for Algorithm 1 in this section.

The challenge of Algorithm 1 is at the complexity that $\mathbf{u}_{[p+1]}^{(\ell)}$ depends on both the new vectors $\left\{\mathbf{u}_{[p+1]}^{(i)}\right\}$, $i = 1, \ldots, \ell - 1$, and the old vectors $\left\{\mathbf{u}_{[p]}^{(i)}\right\}$, $i = \ell + 2, \ldots, k$. To characterize the collective behavior, observe first that for each fixed ℓ, it is always true that $\|\mathbf{u}_{[p]}^{(\ell)}\|_2 = 1$ for all p. Therefore it must have a convergent subsequence. There are only finitely many ℓ. Selecting a subsequence of a subsequence if necessary, we can find a common subset $\{p_j\}$ of nonnegative integers so that $\left\{\mathbf{u}_{[p_j]}^{(\ell)}\right\}$ converges simultaneously for all $\ell \in [\![k]\!]$.

LEMMA 4.3. *If subsequences* $\left\{\mathbf{u}_{[p_j]}^{(\ell)}\right\}$ *converge simultaneously for all* $\ell \in [\![k]\!]$*, then so do subsequences* $\left\{C_{[p_j]}^{(\ell)}\right\}$ *and* $\left\{\mathbf{u}_{[p_j+1]}^{(\ell)}\right\}$.

Proof. The simultaneous convergence of $\left\{\mathbf{u}_{[p_j]}^{(\ell)}\right\}$ for $\ell = 3, \dots k$ implies that the subsequence $\left\{C_{[p_j]}^{(1)}\right\}$ converges. By continuity, the subsequence of the left singular vectors $\left\{\widehat{\mathbf{u}}_{[p_j+1]}^{(1)}\right\}$ of $C_{[p_j]}^{(1)}$ converges also. But then by definition, $\left\{C_{[p_j]}^{(2)}\right\}$ converges and, thus, so does $\left\{\mathbf{u}_{[p_j+1]}^{(2)}\right\}$. We can repeat this argument by cycling through the ℓ-loop in Algorithm 1. At the end, the matrices $\left\{C_{[p_j]}^{(k)}\right\}$ together with the corresponding left singular vectors $\left\{\mathbf{u}_{[p_j+1]}^{(1)}\right\}$ and the right singular vectors $\left\{\mathbf{u}_{[p_j+1]}^{(k)}\right\}$ must also converge. □

Denote the respective limit points by

$$\left\{\begin{array}{lcl} \lim_{j\to\infty} \mathbf{u}_{[p_j]}^{(\ell)} & = & u_{\natural}^{(\ell)}, \\ \lim_{j\to\infty} C_{[p_j]}^{(\ell)} & = & C^{(\ell)}, \\ \lim_{j\to\infty} \widehat{\mathbf{u}}_{[p_j+1]}^{(\ell)} & = & \widehat{\mathbf{u}}_{\sharp}^{(\ell)}, \\ \lim_{j\to\infty} \mathbf{u}_{[p_j+1]}^{(\ell)} & = & \mathbf{u}_{\sharp}^{(\ell)}, \end{array}\right. \qquad \ell \in [\![k]\!], \tag{4.1}$$

where the subscript $_\sharp$ is a handy way to distinguish, at least for now, that the limit points corresponding to the subsequence $_{[p_j+1]}$ might be different from those, denoted by the subscript $_\natural$, of the original subsequence $_{[p_j]}$. Obviously,

$$C^{(\ell)} = T\circledast_{\boldsymbol{\beta}_\ell}\left(\bigotimes_{i=1}^{\ell-1}\mathbf{u}_{\sharp}^{(i)} \otimes \bigotimes_{i=\ell+2}^{k}\mathbf{u}_{\natural}^{(i)}\right) \tag{4.2}$$

and that, by continuity, all matrices $C^{(\ell)}$, $\ell \in [\![k]\!]$, share the same dominant singular value $\widetilde{\lambda}$. It only remains to explore the relationships among dominant singular vectors of all $C^{(\ell)}$ matrices. We shall assume that the generic condition that $C^{(\ell)} \in \mathbb{R}^{I_\ell \times I_{\ell+1}}$ is of full column rank and that $\widetilde{\lambda}$ is simple for all ℓ.

For convenience, we employ the abbreviations $\circledast_1$ and $\circledast_2$ to indicate, respectively, the row-matrix and matrix-column multiplications delineated in (2.4). In detail, be aware that we have to deal with multiplications of matrices and vectors of different sizes.

By using Lemma 2.2, we obtain the equalities

$$\begin{aligned} C_{[p_j]}^{(\ell)}\circledast_1 \mathbf{u}_{[p_j+1]}^{(\ell)} &= \lambda_{[p_j+1]}^{(\ell)}\widehat{\mathbf{u}}_{[p_j+1]}^{(\ell+1)} \\ &= \left(T\circledast_{\boldsymbol{\beta}_\ell}\left(\widehat{\mathbf{u}}_{[p_j+1]}^{(1)} \otimes \bigotimes_{i=2}^{\ell-1}\mathbf{u}_{[p_j+1]}^{(i)} \otimes \bigotimes_{i=\ell+2}^{k}\mathbf{u}_{[p_j]}^{(i)}\right)\right)\circledast_1 \mathbf{u}_{[p_j+1]}^{(\ell)} \\ &= \left(T\circledast_{\boldsymbol{\beta}_{\ell+1}}\left(\widehat{\mathbf{u}}_{[p_j+1]}^{(1)} \otimes \bigotimes_{i=2}^{\ell}\mathbf{u}_{[p_j+1]}^{(i)} \otimes \bigotimes_{i=\ell+3}^{k}\mathbf{u}_{[p_j]}^{(i)}\right)\right)\circledast_2 \mathbf{u}_{[p_j]}^{(\ell+2)} \\ &= C_{[p_j]}^{(\ell+1)}\circledast_2 \mathbf{u}_{[p_j]}^{(\ell+2)}, \quad \ell = 2 \dots k-2; \end{aligned} \tag{4.3}$$

$$\begin{aligned} C_{[p_j]}^{(1)}\circledast_1 \widehat{\mathbf{u}}_{[p_j+1]}^{(1)} &= \lambda_{[p_j+1]}^{(1)}\widehat{\mathbf{u}}_{[p_j+1]}^{(2)} \\ &= \left(T\circledast_{\boldsymbol{\beta}_1}\left(\bigotimes_{i=3}^{k}\mathbf{u}_{[p_j]}^{(i)}\right)\right)\circledast_1 \widehat{\mathbf{u}}_{[p_j+1]}^{(1)} \\ &= \left(T\circledast_{\boldsymbol{\beta}_2}\left(\widehat{\mathbf{u}}_{[p_j+1]}^{(1)} \otimes \bigotimes_{i=4}^{k}\mathbf{u}_{[p_j]}^{(i)}\right)\right)\circledast_2 \mathbf{u}_{[p_j]}^{(3)} = C_{[p_j]}^{(2)}\circledast_2 \mathbf{u}_{[p_j]}^{(3)}. \end{aligned} \tag{4.4}$$

We also have the identities

$$
\begin{aligned}
C^{(k)}_{[p_j-1]} \circledast_2 \mathbf{u}^{(k)}_{[p_j]} &= \lambda^{(k)}_{[p_j]} \mathbf{u}^{(1)}_{[p_j]} \\
&= \left(T \circledast_{\boldsymbol{\beta}_k} \left(\bigotimes_{i=2}^{k-1} \mathbf{u}^{(i)}_{[p_j]} \right) \right) \circledast_2 \mathbf{u}^{(k)}_{[p_j]} \\
&= \left(T \circledast_{\boldsymbol{\beta}_1} \left(\bigotimes_{i=3}^{k} \mathbf{u}^{(i)}_{[p_j]} \right) \right) \circledast_2 \mathbf{u}^{(2)}_{[p_j]} = C^{(1)}_{[p_j]} \circledast_2 \mathbf{u}^{(2)}_{[p_j]}; \qquad (4.5) \\
C^{(k-1)}_{[p_j]} \circledast_1 \mathbf{u}^{(k-1)}_{[p_j+1]} &= \lambda^{(k-1)}_{[p_j+1]} \widehat{\mathbf{u}}^{(k)}_{[p_j+1]} \\
&= \left(T \circledast_{\boldsymbol{\beta}_{k-1}} \left(\widehat{\mathbf{u}}^{(1)}_{[p_j+1]} \otimes \bigotimes_{i=2}^{k-2} \mathbf{u}^{(i)}_{[p_j+1]} \right) \right) \circledast_1 \mathbf{u}^{(k-1)}_{[p_j+1]} \\
&= \left(T \circledast_{\boldsymbol{\beta}_k} \left(\bigotimes_{i=2}^{k-1} \mathbf{u}^{(i)}_{[p_j+1]} \right) \right) \circledast_1 \widehat{\mathbf{u}}^{(1)}_{[p_j+1]} = C^{(k)}_{[p_j]} \circledast_1 \widehat{\mathbf{u}}^{(1)}_{[p_j+1]}. \qquad (4.6)
\end{aligned}
$$

Taking the limits, then it follows by construction and continuity that we have the relationships:

$$
\begin{aligned}
\widetilde{\lambda} \mathbf{u}^{(1)}_{\natural} &= C^{(1)} \circledast_2 \mathbf{u}^{(2)}_{\natural}, \quad \text{(by (4.5))} && (4.7) \\
C^{(1)} \circledast_1 \widehat{\mathbf{u}}^{(1)}_{\sharp} = \widetilde{\lambda} \widehat{\mathbf{u}}^{(2)}_{\sharp} &= C^{(2)} \circledast_2 \mathbf{u}^{(3)}_{\natural}, \quad \text{(by (4.4))} && (4.8) \\
C^{(\ell)} \circledast_1 \mathbf{u}^{(\ell)}_{\sharp} = \widetilde{\lambda} \widehat{\mathbf{u}}^{(\ell+1)}_{\sharp} &= C^{(\ell+1)} \circledast_2 \mathbf{u}^{(\ell+2)}_{\natural}, \quad \ell = 2, \ldots, k-2, \quad \text{(by (4.3))} && (4.9) \\
C^{(k-1)} \circledast_1 \mathbf{u}^{(k-1)}_{\sharp} = \widetilde{\lambda} \widehat{\mathbf{u}}^{(k)}_{\sharp} &= C^{(k)} \circledast_1 \widehat{\mathbf{u}}^{(1)}_{\sharp}, \quad \text{(by (4.6))} && (4.10) \\
C^{(k)} \circledast_1 \mathbf{u}^{(1)}_{\sharp} = \widetilde{\lambda} \mathbf{u}^{(k)}_{\sharp}. &\quad \text{f(by Lines 16-17 in Algorithm 1)} && (4.11)
\end{aligned}
$$

Under the assumption that $\widetilde{\lambda}$ is simple, the corresponding singular vector is unique up to a sign change. However, because in Lines 6 to 8 of Algorithm 1 we have required that the first entry of the dominant singular vector be positive, such a sign change does not exist. The best rank-1 approximation to the matrix $C^{(\ell)}$ therefore is unique. Recursively, the above relationships imply that the best rank-1 approximation to the matrix $C^{(\ell)}$ can be expressed in two ways:

$$
\left\{ \begin{array}{rcll}
C^{(1)} & \approx & \widetilde{\lambda} \mathbf{u}^{(1)}_{\natural} \otimes \mathbf{u}^{(2)}_{\natural} & = \widetilde{\lambda} \widehat{\mathbf{u}}^{(1)}_{\sharp} \otimes \widehat{\mathbf{u}}^{(2)}_{\sharp}, \\
C^{(\ell)} & \approx & \widetilde{\lambda} \widehat{\mathbf{u}}^{(\ell)}_{\sharp} \otimes \mathbf{u}^{(\ell+1)}_{\natural} & = \widetilde{\lambda} \mathbf{u}^{(\ell)}_{\sharp} \otimes \widehat{\mathbf{u}}^{(\ell+1)}_{\sharp}, \quad \ell = 2, \ldots, k-1, \\
C^{(k)} & \approx & \widetilde{\lambda} \widehat{\mathbf{u}}^{(1)}_{\sharp} \otimes \widehat{\mathbf{u}}^{(k)}_{\sharp} & = \widetilde{\lambda} \mathbf{u}^{(1)}_{\sharp} \otimes \mathbf{u}^{(k)}_{\sharp}.
\end{array} \right. \qquad (4.12)
$$

By uniqueness, we obtain the following important result.

LEMMA 4.4. *If subsequences* $\left\{ \mathbf{u}^{(\ell)}_{[p_j]} \right\}$ *converge simultaneously for all* $\ell \in [\![k]\!]$*, then the limit points defined in (4.1) are such that*

$$
\left\{ \begin{array}{rcl}
\mathbf{u}^{(\ell)}_{\natural} & = & \widehat{\mathbf{u}}^{(\ell)}_{\sharp} = \mathbf{u}^{(\ell)}_{\sharp}, \\
C^{(\ell)} & = & T \circledast_{\boldsymbol{\beta}_\ell} \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}^{(i)}_{\natural} \otimes \bigotimes_{i=\ell+2}^{k} \mathbf{u}^{(i)}_{\natural} \right),
\end{array} \right. \qquad \ell \in [\![k]\!]. \qquad (4.13)
$$

LEMMA 4.5. *For almost all tensors* $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$*, the accumulation points of the sequence* $\{\mathbf{u}_t\}$ *generated by Algorithm 1 are geometrically isolated.*

Proof. By the way the iteration is defined in Algorithm 1, a stationary point necessarily satisfies the system of equations

$$T \circledast_\ell \left(\bigotimes_{i=1}^{\ell-1} \mathbf{u}^{(i)} \otimes \bigotimes_{i=\ell+1}^{k} \mathbf{u}^{(i)} \right) = \left\langle T, \bigotimes_{\ell=1}^{k} \mathbf{u}^{(\ell)} \right\rangle \mathbf{u}^{(\ell)}, \quad \ell \in [\![k]\!]. \tag{4.14}$$

The equation (4.14) is a polynomial system in the unknowns $\left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\right) \in \mathbb{R}^{I_1} \times \ldots \times \mathbb{R}^{I_k}$ with leading coefficients from entries of T. By the theory of parameter continuation [8, Theorem 7.1.1], we know that for almost all tensor $T \in \mathbb{C}^{I_1 \times \ldots \times I_k}$, except for an affine algebraic subset of codimension one in $\mathbb{C}^{I_1} \times \ldots \times \mathbb{C}^{I_k}$, the solutions to (4.14) are isolated. Together with the fact that $\mathbb{R}$ is dense in $\mathbb{C}$ under the Zariski topology, the real solutions are also isolated. □

THEOREM 4.6. *For almost all tensors $T \in \mathbb{R}^{I_1 \times \ldots \times I_k}$ and arbitrary starting points, the sequences $\left(\mathbf{u}_{[p]}^{(1)}, \ldots, \mathbf{u}_{[p]}^{(k)}\right)$ generated in Algorithm 1 converge to a local maximizer of the generalized Rayleigh quotient $\lambda\left(\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}\right)$ defined in (1.2).*

Proof. Let $\left\{\mathbf{u}_{[p_j]}^{(\ell)}\right\}$ be any simultaneously convergent subsequences for $\ell \in [\![k]\!]$. By Lemma 4.3, the subsequences $\left\{\mathbf{u}_{[p_j+1]}^{(\ell)}\right\}$ also converges. Indeed, by Lemma 4.4, both subsequences converges to the same limit point for all $\ell \in [\![k]\!]$. Thus $\|\mathbf{u}_{[p_j+1]}^{(\ell)} - \mathbf{u}_{[p_j]}^{(\ell)}\| \to 0$. On the other hand, by Lemma 4.5 we assume that the limit point is geometrically isolated. The convergence of the entire sequence $\left\{\mathbf{u}_{[p_j]}^{(\ell)}\right\}$ to the same limit point follows from Lemma 2.3. □

REFERENCES

[1] J. BRACHAT, P. COMON, B. MOURRAIN, AND E. TSIGARIDAS, *Symmetric tensor decomposition*, Linear Algebra Appl., 433 (2010), pp. 1851–1872.

[2] P. COMON, G. GOLUB, L.-H. LIM, AND B. MOURRAIN, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279.

[3] 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.

[4] M. ISHTEVA, P.-A. ABSIL, AND P. VAN DOOREN, *Jacobi algorithm for the best low multilinear rank approximation of symmetric tensors*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 651–672.

[5] E. KOFIDIS AND P. A. REGALIA, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 863–884.

[6] T. G. KOLDA, *Numerical optimization for symmetric tensor decomposition*, Math. Program., 151 (2015), pp. 225–248.

[7] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.

[8] A. J. SOMMESE AND C. W. WAMPLER, II, *The numerical solution of systems of polynomials arising in engineering and science*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[9] L. WANG AND M. T. CHU, *On the global convergence of the alternating least squares method for rank-one approximation to generic tensors*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 1058–1072.

[10] T. ZHANG AND G. H. GOLUB, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550 (electronic).