



Low Rank Tensor Decompositions and Approximations

Jiawang Nie¹ · Li Wang² · Zequn Zheng¹

Received: 15 April 2022 / Accepted: 15 January 2023
© The Author(s) 2023

Abstract

There exist linear relations among tensor entries of low rank tensors. These linear relations can be expressed by multi-linear polynomials, which are called generating polynomials. We use generating polynomials to compute tensor rank decompositions and low rank tensor approximations. We prove that this gives a quasi-optimal low rank tensor approximation if the given tensor is sufficiently close to a low rank one.

Keywords Tensor · Decomposition · Rank · Approximation · Generating polynomial

Mathematics Subject Classification 15A69 · 65F99

1 Introduction

Let m and n_1, \dots, n_m be positive integers. A tensor \mathcal{F} of order m and dimension $n_1 \times \dots \times n_m$ can be labeled such that

$$\mathcal{F} = (\mathcal{F}_{i_1, \dots, i_m})_{1 \leq i_1 \leq n_1, \dots, 1 \leq i_m \leq n_m}.$$

Let \mathbb{F} be a field (either the real field \mathbb{R} or the complex field \mathbb{C}). The space of all tensors of order m and dimension $n_1 \times \dots \times n_m$ with entries in \mathbb{F} is denoted as $\mathbb{F}^{n_1 \times \dots \times n_m}$. For vectors $v_i \in \mathbb{F}^{n_i}$, $i = 1, \dots, m$, their outer product $v_1 \otimes \dots \otimes v_m$ is the tensor in

Jiawang Nie is partially supported by the NSF grant DMS-2110780. Li Wang is partially supported by the NSF grant DMS-2009689.

✉ Jiawang Nie
njw@math.ucsd.edu

Li Wang
li.wang@uta.edu

Zequn Zheng
zez084@ucsd.edu

¹ Department of Mathematics, University of California San Diego, La Jolla, CA 92093, USA

² Department of Mathematics, University of Texas at Arlington, Arlington, TX 76019, USA

$\mathbb{F}^{n_1 \times \dots \times n_m}$ such that

$$(v_1 \otimes \dots \otimes v_m)_{i_1, \dots, i_m} = (v_1)_{i_1} \dots (v_m)_{i_m}$$

for all labels in the corresponding range. A tensor like $v_1 \otimes \dots \otimes v_m$ is called a rank-1 tensor. For every tensor $\mathcal{F} \in \mathbb{F}^{n_1 \times \dots \times n_m}$, there exist vector tuples $(v^{s,1}, \dots, v^{s,m})$, $s = 1, \dots, r$, $v^{s,j} \in \mathbb{F}^{n_j}$, such that

$$\mathcal{F} = \sum_{s=1}^r v^{s,1} \otimes \dots \otimes v^{s,m}. \tag{1}$$

The smallest such r is called the \mathbb{F} -rank of \mathcal{F} , for which we denote $\text{rank}_{\mathbb{F}}(\mathcal{F})$. When r is minimum, Eq. (1) is called a rank- r decomposition over the field \mathbb{F} . In the literature, this rank is sometimes referenced as the Candecomp/Parafac (CP) rank. We refer to [1, 2] for various notions of tensor ranks. Recent work for tensor decompositions can be found in [3–9]. Tensors are closely related to polynomial optimization [10–14]. Tensor decomposition has been widely used in temporal tensor analysis including discovering patterns [15], predicting evolution [16], and identifying temporal communities [17], and in multirelational data analysis including collective classification [18], word representation learning [19], and coherent subgraph learning [20]. Tensor approximation has been explored in signal processing applications [21] and multidimensional, multivariate data analysis [22]. Various other applications of tensors can be found in [23–25]. Throughout the paper, we use the Hilbert-Schmidt norm for tensors:

$$\|\mathcal{F}\| = \sqrt{\sum_{1 \leq i_j \leq n_j, 1 \leq j \leq m} |\mathcal{F}_{i_1, \dots, i_m}|^2}.$$

The low rank tensor approximation (LRTA) problem is to find a low rank tensor that is close to a given one. This is equivalent to a nonlinear least square optimization problem. For a given tensor $\mathcal{F} \in \mathbb{F}^{n_1 \times \dots \times n_m}$ and a given rank r , we look for r vector tuples $v^{(s)} := (v^{s,1}, \dots, v^{s,m})$, $s = 1, \dots, r$, such that

$$\mathcal{F} \approx \sum_{s=1}^r v^{s,1} \otimes \dots \otimes v^{s,m}, \quad v^{s,j} \in \mathbb{F}^{n_j}.$$

This requires to solve the following nonlinear least squares optimization

$$\min_{v^{s,j} \in \mathbb{F}^{n_j}, j=1, \dots, m} \left\| \mathcal{F} - \sum_{s=1}^r v^{s,1} \otimes \dots \otimes v^{s,m} \right\|^2. \tag{2}$$

When $r = 1$, the best rank-1 approximating tensor always exists and it is equivalent to computing the spectral norm [26, 27]. When $r > 1$, a best rank- r tensor approximation may not exist [28]. Classically used methods for solving low rank tensor approximations are the alternating least squares (ALS) method [29–31], higher-order

power iterations [32], semidefinite relaxations [10, 12], SVD-based methods [33], optimization-based methods [8, 34]. We refer to [3, 35] for recent work on low rank tensor approximations.

Contributions

In this paper, we extend the generating polynomial method in [7, 35] to compute tensor rank decompositions and low rank tensor approximations for nonsymmetric tensors. First, we estimate generating polynomials by solving linear least squares. Second, we find their approximately common zeros, which can be done by computing eigenvalue decompositions. Third, we get a tensor decomposition from their common zeros, by solving linear least squares. To find a low rank tensor approximation, we first apply the decomposition method to obtain a low rank approximating tensor and then use nonlinear optimization methods to improve the approximation. Our major conclusion is that if the tensor to be approximated is sufficiently close to a low rank one, then the obtained low rank tensor is a quasi-optimal low rank approximation. The proof is based on perturbation analysis of linear least squares and eigenvalue decompositions.

The paper is organized as follows. In Sect. 2, we review some basic results about tensors. In Sect. 3, we introduce the concept of generating polynomials and study their relations to tensor decompositions. In Sect. 4, we give an algorithm for computing tensor rank decompositions for low rank tensors. In Sect. 5, we give an algorithm for computing low rank approximations. The approximation error analysis is also given. In Sect. 6, we present numerical experiments. Some conclusions are made in Sect. 7.

2 preliminary

Notation

The symbol \mathbb{N} (resp., \mathbb{R} , \mathbb{C}) denotes the set of nonnegative integers (resp., real, complex numbers). For an integer $r > 0$, denote the set $[r] := \{1, \dots, r\}$. Uppercase letters (e.g., A) denote matrices, A_{ij} denotes the (i, j) th entry of the matrix A , and Curl letters (e.g., \mathcal{F}) denote tensors, $\mathcal{F}_{i_1, \dots, i_m}$ denotes the (i_1, \dots, i_m) th entry of the tensor \mathcal{F} . For a complex matrix A , A^T denotes its transpose and A^* denotes its conjugate transpose. The Kruskal rank of A , for which we denote κ_A , is the largest number k such that every set of k columns of A is linearly independent. For a vector v , the $(v)_i$ denotes its i th entry and $\text{diag}(v)$ denotes the square diagonal matrix whose diagonal entries are given by the entries of v . The subscript $v_{s:t}$ denotes the subvector of v whose label is from s to t . For a matrix A , the subscript notation $A_{:,j}$ and $A_{i,:}$, respectively, denote its j th column and i th row. Similar subscript notation is used for tensors. For two matrices A, B , their classical Kronecker product is denoted as $A \otimes B$. For a set S , its cardinality is denoted as $|S|$.

For a tensor decomposition for \mathcal{F} such that

$$\mathcal{F} = \sum_{s=1}^r u^{s,1} \otimes u^{s,2} \otimes \dots \otimes u^{s,m}. \tag{3}$$

we denote the matrices

$$U^{(j)} = [u^{1,j}, \dots, u^{r,j}], \quad j = 1, \dots, m.$$

The $U^{(j)}$ is called the j th decomposing matrix for \mathcal{F} . For convenience of notation, we denote that

$$U^{(1)} \circ \dots \circ U^{(m)} = \sum_{i=1}^r (U^{(1)})_{:,i} \otimes \dots \otimes (U^{(m)})_{:,i}.$$

Then the above tensor decomposition is equivalent to $\mathcal{F} = U^{(1)} \circ \dots \circ U^{(m)}$.

For a matrix $V \in \mathbb{C}^{p \times n_t}$, define the matrix-tensor product

$$\mathcal{A} := V \times_t \mathcal{F}$$

is a tensor in $\mathbb{C}^{n_1 \times \dots \times n_{t-1} \times p \times n_{t+1} \times \dots \times n_m}$ such that the i th slice of \mathcal{A} is

$$\mathcal{A}_{i_1, \dots, i_{t-1}, :, i_{t+1}, \dots, i_m} = V \mathcal{F}_{i_1, \dots, i_{t-1}, :, i_{t+1}, \dots, i_m}.$$

2.1 Flattening matrices

We partition the dimensions n_1, n_2, \dots, n_m into two disjoint groups I_1 and I_2 such that the difference

$$\left| \prod_{i \in I_1} n_i - \prod_{j \in I_2} n_j \right|$$

is minimum. Up to a permutation, we write that $I_1 = \{n_1, \dots, n_k\}$, $I_2 = \{n_{k+1}, \dots, n_m\}$. For convenience, denote that

$$I = \{(t_1, \dots, t_k) : 1 \leq t_j \leq n_j, j = 1, \dots, k\},$$

$$J = \{(t_{k+1}, \dots, t_m) : 1 \leq t_j \leq n_j, j = k + 1, \dots, m\}.$$

For a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$, the above partition gives the flattening matrix

$$\text{Flat}(\mathcal{F}) := (\mathcal{F}_{i,J})_{i \in I, J \in J}. \tag{4}$$

This gives the most square flattening matrix for \mathcal{F} . Let σ_r denote the closure of all rank- r tensors in $\mathbb{C}^{n_1 \times \dots \times n_m}$, under the Zariski topology (see [36]). The set σ_r is

an irreducible variety of $\mathbb{C}^{n_1 \times \dots \times n_m}$. For a given tensor $\mathcal{F} \in \sigma_r$, it is possible that $\text{rank}(\mathcal{F}) > r$. This fact motivates the notion of border rank:

$$\text{rank}_B(\mathcal{F}) = \min \{r : \mathcal{F} \in \sigma_r\}. \tag{5}$$

For every tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$, one can show that

$$\text{rank Flat}(\mathcal{F}) \leq \text{rank}_B(\mathcal{F}) \leq \text{rank}(\mathcal{F}). \tag{6}$$

A property \mathbb{P} is said to hold *generically* on σ_r if \mathbb{P} holds on a Zariski open subset T of σ_r . For such a property \mathbb{P} , each $u \in T$ is called a generic point. Interestingly, the above three ranks are equal for generic points of σ_r for a range of values of r .

Lemma 1 *Let s be the smaller dimension of the matrix $\text{Flat}(\mathcal{F})$. For every $r \leq s$, the equalities*

$$\text{rank Flat}(\mathcal{F}) = \text{rank}_B(\mathcal{F}) = \text{rank}(\mathcal{F}) \tag{7}$$

hold for tensors \mathcal{F} in a Zariski open subset of σ_r .

Proof Let ϕ_1, \dots, ϕ_ℓ be the $r \times r$ minors of the matrix

$$\text{Flat} \left(\sum_{i=1}^r x^{i,1} \otimes \dots \otimes x^{i,m} \right). \tag{8}$$

They are homogeneous polynomials in $x^{i,j}$ ($i = 1, \dots, r, j = 1, \dots, m$). Let x denote the tuple $(x^{1,1}, x^{1,2}, \dots, x^{r,m})$. Define the projective variety in $\mathbb{P}^{r(n_1 + \dots + n_m) - 1}$

$$Z = \{x : \phi_1(x) = \dots = \phi_\ell(x) = 0\}. \tag{9}$$

Then $Y := \mathbb{P}^{r(n_1 + \dots + n_m) - 1} \setminus Z$ is a Zariski open subset of full dimension. Consider the polynomial mapping $\pi : Y \rightarrow \sigma_r$,

$$(x^{1,1}, x^{1,2}, \dots, x^{r,m}) \mapsto \sum_{i=1}^r (x^{i,1}) \otimes \dots \otimes (x^{i,m}). \tag{10}$$

The image $\pi(Y)$ is dense in the irreducible variety σ_r . So, $\pi(Y)$ contains a Zariski open subset \mathcal{Y} of σ_r (see [37]). For each $\mathcal{F} \in \mathcal{Y}$, there exists $u \in Y$ such that $\mathcal{F} = \pi(u)$. Because $u \notin Z$, at least one of $\phi_1(u), \dots, \phi_\ell(u)$ is nonzero, and hence $\text{rank Flat}(\mathcal{F}) \geq r$. By (6), we know (7) holds for all $\mathcal{F} \in \mathcal{Y}$ since $\text{rank}(\mathcal{F}) \leq r$. Since \mathcal{Y} is a Zariski open subset of σ_r , the lemma holds.

By Lemma 1, if $r \leq s$ and \mathcal{F} is a generic tensor in σ_r , we can use $\text{rank Flat}(\mathcal{F})$ to estimate $\text{rank}(\mathcal{F})$. However, for a generic $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ such that $\text{rank Flat}(\mathcal{F}) = r$, we cannot conclude $\mathcal{F} \in \sigma_r$.

2.2 Reshaping of tensor decompositions

A tensor \mathcal{F} of order greater than 3 can be reshaped to another tensor $\widehat{\mathcal{F}}$ of order 3. A tensor decomposition of $\widehat{\mathcal{F}}$ can be converted to a decomposition for \mathcal{F} under certain conditions. In the following, we assume a given tensor \mathcal{F} has decomposition (3). Suppose the set $\{1, \dots, m\}$ is partitioned into 3 disjoint subsets

$$\{1, \dots, m\} = I_1 \cup I_2 \cup I_3.$$

Let $p_i = |I_i|$ for $i = 1, 2, 3$. For the reshaped vectors

$$\begin{cases} w^{s,1} = u^{s,i_1} \otimes \dots \otimes u^{s,i_{p_1}} & \text{for } I_1 = \{i_1, \dots, i_{p_1}\}, \\ w^{s,2} = u^{s,j_1} \otimes \dots \otimes u^{s,j_{p_2}} & \text{for } I_2 = \{j_1, \dots, j_{p_2}\}, \\ w^{s,3} = u^{s,k_1} \otimes \dots \otimes u^{s,k_{p_3}} & \text{for } I_3 = \{k_1, \dots, k_{p_3}\}, \end{cases} \tag{11}$$

we get the following tensor decomposition:

$$\widehat{\mathcal{F}} = \sum_{s=1}^r w^{s,1} \otimes w^{s,2} \otimes w^{s,3}. \tag{12}$$

Conversely, for a decomposition like (12) for $\widehat{\mathcal{F}}$, if all $w^{s,1}, w^{s,2}, w^{s,3}$ can be expressed as rank-1 products as in (11), then Eq. (12) can be reshaped to a tensor decomposition for \mathcal{F} as in (3). When the flattened tensor $\widehat{\mathcal{F}}$ satisfies some conditions, the tensor decomposition of $\widehat{\mathcal{F}}$ is unique. For such a case, we can obtain a tensor decomposition for \mathcal{F} through decomposition (12). A classical result about the uniqueness is the Kruskal’s criterion [38].

Theorem 1 (Kruskal’s criterion, [38]) *Let $\mathcal{F} = U^{(1)} \circ U^{(2)} \circ U^{(3)}$ be a tensor with each $U^{(i)} \in \mathbb{C}^{n_i \times r}$. Let κ_i be the Kruskal rank of $U^{(i)}$, for $i = 1, 2, 3$. If*

$$2r + 2 \leq \kappa_1 + \kappa_2 + \kappa_3,$$

then \mathcal{F} has a unique rank- r tensor decomposition.

The Kruskal’s criterion can be generalized for more range of r as in [39]. Assume the dimension $n_1 \geq n_2 \geq n_3 \geq 2$ and the rank r is such that

$$2r + 2 \leq \min(n_1, r) + \min(n_2, r) + \min(n_3, r),$$

or equivalently, for $\delta = n_2 + n_3 - n_1 - 2$, r is such that

$$r \leq n_1 + \min\left\{\frac{1}{2}\delta, \delta\right\}.$$

If \mathcal{F} is a generic tensor of rank r as above in the space $\mathbb{C}^{n_1 \times n_2 \times n_3}$, then \mathcal{F} has a unique rank- r decomposition. The following is a uniqueness result for reshaped tensor decompositions.

Theorem 2 (Reshaped Kruskal Criterion, [39, Theorem 4.6]) *For the tensor space $\mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$ with $m \geq 3$, let $I_1 \cup I_2 \cup I_3 = \{1, 2, \dots, m\}$ be a union of disjoint sets and let*

$$p_1 = \prod_{i \in I_1} n_i, \quad p_2 = \prod_{j \in I_2} n_j, \quad p_3 = \prod_{k \in I_3} n_k.$$

Suppose $p_1 \geq p_2 \geq p_3$ and let $\delta = p_2 + p_3 - p_1 - 2$. Assume

$$r \leq p_1 + \min\left\{\frac{1}{2}\delta, \delta\right\}. \tag{13}$$

If \mathcal{F} is a generic tensor of rank r in $\mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$, then the reshaped tensor $\widehat{\mathcal{F}} \in \mathbb{C}^{p_1 \times p_2 \times p_3}$ as in (12) has a unique rank- r decomposition.

3 Generating Polynomials

Generating polynomials can be used to compute tensor rank decompositions. We consider tensors whose ranks r are not bigger than the highest dimension, say, $r \leq n_1$ where n_1 is the biggest of n_1, \dots, n_m . Denote the indeterminate vector variables

$$\mathbf{x}_1 = (x_{1,1}, \dots, x_{1,n_1}), \dots, \mathbf{x}_m = (x_{m,1}, \dots, x_{m,n_m}).$$

A tensor in $\mathbb{C}^{n_1 \times \dots \times n_m}$ can be labeled by monomials like $x_{1,i_1} x_{2,i_2} \dots x_{m,i_m}$. Let

$$\mathbb{M} := \{x_{1,i_1} \dots x_{m,i_m} \mid 1 \leq i_j \leq n_j\}. \tag{14}$$

For a subset $J \subseteq \{1, 2, \dots, m\}$, denote that

$$\begin{aligned} J^c &:= \{1, 2, \dots, m\} \setminus J, \\ \mathbb{M}_J &:= \{x_{1,i_1} \dots x_{m,i_m} \mid x_{j,i_j} = 1 \ \forall j \in J^c\}, \\ \mathcal{M}_J &:= \text{span}\{\mathbb{M}_J\}. \end{aligned} \tag{15}$$

A label tuple (i_1, \dots, i_m) is uniquely determined by the monomial $x_{1,i_1} \dots x_{m,i_m}$. So a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ can be equivalently labeled by monomials such that

$$\mathcal{F}_{x_{1,i_1} \dots x_{m,i_m}} := \mathcal{F}_{i_1, \dots, i_m}. \tag{16}$$

With the new labeling by monomials, define the bi-linear product

$$\left\langle \sum_{\mu \in \mathbb{M}} c_\mu \mu, \mathcal{F} \right\rangle := \sum_{\mu \in \mathbb{M}} c_\mu \mathcal{F}_\mu. \tag{17}$$

In the above, each c_μ is a scalar and \mathcal{F} is labeled by monomials as in (16).

Definition 1 ([40, 41]) For a subset $J \subseteq \{1, 2, \dots, m\}$ and a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$, a polynomial $p \in \mathcal{M}_J$ is called a generating polynomial for \mathcal{F} if

$$\langle pq, \mathcal{F} \rangle = 0 \quad \text{for all } q \in \mathbb{M}_{J^c}. \tag{18}$$

The following is an example of generating polynomials.

Example 1 Consider the cubic order tensor $\mathcal{F} \in \mathbb{C}^{3 \times 3 \times 3}$ given as

$$[\mathcal{F}_{:, :, 1} | \mathcal{F}_{:, :, 2} | \mathcal{F}_{:, :, 3}] = \left[\begin{array}{ccc|ccc} 11 & 20 & 10 & 15 & 24 & 12 & 7 & 10 & 5 \\ 7 & 10 & 5 & 15 & 18 & 9 & 9 & 10 & 5 \\ 12 & 18 & 9 & 24 & 30 & 15 & 14 & 16 & 8 \end{array} \right].$$

The following is a generating polynomial for \mathcal{F} :

$$p := (2x_{1,1} - x_{1,2})(2x_{2,1} - x_{2,2}).$$

Note that $p \in \mathcal{M}_{\{1,2\}}$ and for each $i_3 = 1, 2, 3$,

$$p \cdot x_{3,i_3} = (2x_{1,1} - x_{1,2})(2x_{2,1} - x_{2,2})x_{3,i_3}.$$

One can check that for each $i_3 = 1, 2, 3$,

$$4\mathcal{F}_{1,1,i_3} - 2\mathcal{F}_{1,2,i_3} - 2\mathcal{F}_{2,1,i_3} + \mathcal{F}_{2,2,i_3} = 0.$$

This is because $[4 \ -2 \ -2 \ 1]$ is orthogonal to

$$[\mathcal{F}_{1,1,i_3} \ \mathcal{F}_{1,2,i_3} \ \mathcal{F}_{2,1,i_3} \ \mathcal{F}_{2,2,i_3}]$$

for each $i_3 = 1, 2, 3$.

Suppose the rank $r \leq n_1$ is given. For convenience of notation, denote the label set

$$J := \{(i, j, k) : 1 \leq i \leq r, 2 \leq j \leq m, 2 \leq k \leq n_j\}. \tag{19}$$

For a matrix $G \in \mathbb{C}^{[r] \times J}$ and a triple $\tau = (i, j, k) \in J$, define the bi-linear polynomial

$$\phi[G, \tau](x) := \sum_{\ell=1}^r G(\ell, \tau)x_{1,\ell}x_{j,1} - x_{1,i}x_{j,k} \in \mathcal{M}_{\{1,j\}}. \tag{20}$$

The rows of G are labeled by $\ell = 1, 2, \dots, r$, and the columns of G are labeled by $\tau \in J$. We are interested in G such that $\phi[G, \tau]$ is a generating polynomial for a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$. This requires that

$$\langle \phi[G, \tau] \cdot \mu, \mathcal{F} \rangle = 0 \quad \text{for all } \mu \in \mathbb{M}_{\{1,j\}^c}.$$

The above is equivalent to the equation (\mathcal{F} labeled as in (16))

$$\sum_{\ell=1}^r G(\ell, \tau) \mathcal{F}_{x_{1,\ell} \cdot \mu} = \mathcal{F}_{x_{1,i} x_{j,k} \cdot \mu}. \tag{21}$$

Definition 2 ([40, 41]) When (21) holds for all $\tau \in J$, the G is called a generating matrix for \mathcal{F} .

For given $G, j \in \{2, \dots, m\}$ and $k \in \{2, \dots, n_j\}$, we denote the matrix

$$M^{j,k}[G] := \begin{bmatrix} G(1, (1, j, k)) & G(2, (1, j, k)) & \dots & G(r, (1, j, k)) \\ G(1, (2, j, k)) & G(2, (2, j, k)) & \dots & G(r, (2, j, k)) \\ \vdots & \vdots & \ddots & \vdots \\ G(1, (r, j, k)) & G(2, (r, j, k)) & \dots & G(r, (r, j, k)) \end{bmatrix}. \tag{22}$$

For each (j, k) , define the matrix/vector

$$\begin{cases} A[\mathcal{F}, j] := \left(\mathcal{F}_{x_{1,\ell} \cdot \mu} \right)_{\mu \in \mathbb{M}_{\{1,j\}^c}, 1 \leq \ell \leq r}, \\ b[\mathcal{F}, j, k] := \left(\mathcal{F}_{x_{1,\ell} x_{j,k} \cdot \mu} \right)_{\mu \in \mathbb{M}_{\{1,j\}^c}, 1 \leq \ell \leq r}. \end{cases} \tag{23}$$

Equation (21) is then equivalent to

$$A[\mathcal{F}, j] (M^{j,k}[G])^\top = b[\mathcal{F}, j, k]. \tag{24}$$

The following is a useful property for the matrices $M^{j,k}[G]$.

Theorem 3 ([40, 41]) Suppose $\mathcal{F} = \sum_{s=1}^r u^{s,1} \otimes \dots \otimes u^{s,m}$ for vectors $u^{s,j} \in \mathbb{C}^{n_j}$. If $r \leq n_1, (u^{s,2})_1 \dots (u^{s,m})_1 \neq 0$, and the first r rows of the first decomposing matrix

$$U^{(1)} := [u^{1,1} \ \dots \ u^{r,1}]$$

are linearly independent, then there exists a G satisfying (24) and satisfying (for all $j \in \{2, \dots, m\}, k \in \{2, \dots, n_j\}$ and $s = 1, \dots, r$)

$$M^{j,k}[G] \cdot (u^{s,1})_{1:r} = (u^{s,j})_k \cdot (u^{s,1})_{1:r}. \tag{25}$$

4 Low Rank Tensor Decompositions

Without loss of generality, assume the dimensions are decreasing as

$$n_1 \geq n_2 \geq \dots \geq n_m.$$

We discuss how to compute tensor decomposition for a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ when the rank r is not bigger than the highest dimension, i.e., $r \leq n_1$. As in Theorem 3, the

decomposing vectors $(u^{s,1})_{1:r}$ are common eigenvectors of the matrices $M^{j,k}[G]$, with $(u^{s,j})_k$ being the eigenvalues, respectively. This implies that the matrices $M^{j,k}[G]$ are simultaneously diagonalizable. This property can be used to compute tensor decompositions.

Suppose G is a matrix such that (24) holds and $M^{j,k}[G]$ are simultaneously diagonalizable. That is, there is an invertible matrix $P \in \mathbb{C}^{r \times r}$ such that all the products $P^{-1}M^{j,k}[G]P$ are diagonal for all $j = 2, \dots, m$ and for all $k = 2, \dots, n_j$. Suppose $M^{j,k}[G]$ are diagonalized such that

$$P^{-1}M^{j,k}[G]P = \text{diag}[\lambda_{j,k,1}, \lambda_{j,k,2}, \dots, \lambda_{j,k,r}] \tag{26}$$

with the eigenvalues $\lambda_{j,k,s}$. For each $s = 1, \dots, r$ and $j = 2, \dots, m$, denote the vectors

$$u^{s,j} := (1, \lambda_{j,2,s}, \dots, \lambda_{j,n_j,s}). \tag{27}$$

When \mathcal{F} is rank- r , there exist vectors $u^{1,1}, \dots, u^{r,1} \in \mathbb{C}^{n_1}$ such that

$$\mathcal{F} = \sum_{s=1}^r u^{s,1} \otimes u^{s,2} \otimes \dots \otimes u^{s,m}. \tag{28}$$

The vectors $u^{s,1}$ can be found by solving linear equations after $u^{s,j}$ are obtained for $j = 2, \dots, m$ and $s = 1, \dots, r$. The existence of vectors $u^{s,1}$ satisfying tensor decomposition (28) is shown in the following theorem.

Theorem 4 *Let $\mathcal{F} = V^{(1)} \circ \dots \circ V^{(m)}$ be a rank- r tensor, for matrices $V^{(i)} \in \mathbb{C}^{n_i \times r}$, such that the first r rows of $V^{(1)}$ are linearly independent. Suppose G is a matrix satisfying (24) and $P \in \mathbb{C}^{r \times r}$ is an invertible matrix such that all matrix products $P^{-1} \cdot M^{j,k}[G] \cdot P$ are simultaneously diagonalized as in (26). For $j = 2, \dots, m$ and $s = 1, \dots, r$, let $u^{s,j}$ be vectors given as in (27). Then, there must exist vectors $u^{1,1}, \dots, u^{r,1} \in \mathbb{C}^{n_1}$ such that tensor decomposition (28) holds.*

Proof Since the matrix $P = [p_1 \ \dots \ p_r]$ is invertible, there exist scalars $c_1, \dots, c_r \in \mathbb{C}$ such that

$$\mathcal{F}_{1:r,1,\dots,1} = c_1 p_1 + c_2 p_2 + \dots + c_r p_r. \tag{29}$$

Consider the new tensor

$$\mathcal{H} := \sum_{s=1}^r c_s p_s \otimes u^{s,2} \otimes \dots \otimes u^{s,m}.$$

In the following, we show that $\mathcal{F}_{1:r,;\dots,;} = \mathcal{H}$ and there exist vectors $u^{1,1}, \dots, u^{r,1} \in \mathbb{C}^{n_1}$ satisfying equation (28).

By Theorem 3, one can see that the generating matrix G for \mathcal{F} is also a generating matrix for \mathcal{H} , so it holds that

$$\langle \phi[G, \tau]p, \mathcal{F} \rangle = \langle \phi[G, \tau]p, \mathcal{H} \rangle = 0, \quad \text{for all } p \in \mathbb{M}_{\{1,j\}^c}. \tag{30}$$

Therefore, we have

$$\langle \phi[G, \tau]_p, \mathcal{H} - \mathcal{F} \rangle = 0, \quad \text{for all } p \in \mathbb{M}_{\{1, j\}^c}. \tag{31}$$

By (29), one can see that

$$(\mathcal{H} - \mathcal{F})_{1:r, 1, \dots, 1} = 0. \tag{32}$$

In (31), for each $\tau = (i, 2, k) \in J$ and $p = 1$, we can get

$$(\mathcal{H} - \mathcal{F})_{1:r, :, 1, \dots, 1} = 0.$$

Similarly, for $\tau = (i, 2, k) \in J$ and $p = x_{3, j_3}$, we can get

$$(\mathcal{H} - \mathcal{F})_{1:r, :, :, 1, \dots, 1} = 0.$$

Continuing this, we can eventually get $\mathcal{H} = \mathcal{F}_{1:r, :, :, \dots, :}$. Since the matrix $(V^{(1)})_{1:r, :}$ is invertible, there exists a matrix $W \in \mathbb{C}^{n_1 \times r}$ such that $V^{(1)} = W(V^{(1)})_{1:r, :}$. Observe that

$$\mathcal{F} = W \times_1 \mathcal{F}_{1:r, :, :, \dots, :} = W \times_1 \mathcal{H}.$$

Let $u^{s, 1} = W \cdot (c_s p_s)$ for $s = 1, \dots, r$. Then tensor decomposition (28) holds.

4.1 An Algorithm for Computing Tensor Decompositions

Consider a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$ with a given rank r . Recall that the dimensions are ordered such that $n_1 \geq n_2 \geq \dots \geq n_m$. We discuss how to compute a rank- r tensor decomposition for \mathcal{F} . Recall $A[\mathcal{F}, j]$, $b[\mathcal{F}, j, k]$ as in (23), for $j > 1$. Note that $A[\mathcal{F}, j]$ has the dimension $N_j \times r$, where

$$N_j := \frac{n_2 \cdots n_m}{n_j}. \tag{33}$$

If $r \leq N_j$, then the matrices $A[\mathcal{F}, j]$ have full column rank for generic cases. For instance, $r \leq N_3$ if $m = 3$ and $r \leq n_2$. Since N_2 is the smallest, we often use the matrices $A[\mathcal{F}, j]$ for $j \geq 3$. For convenience, denote the label set

$$\Upsilon := \{(j, k) : 3 \leq j \leq m, 2 \leq k \leq n_j\}. \tag{34}$$

In the following, we consider the case that $r \leq N_3$. For each pair $(j, k) \in \Upsilon$, linear system (24) has a unique solution, for which we denote

$$Y^{j, k} = M^{j, k}[G].$$

For $j = 2$, equation (24) may not have a unique solution if $r > N_2$. In the following, we show how to get the tensor decomposition without using the matrices $M^{2,k}[G]$. By Theorem 3, the matrices $Y^{j,k}$ are simultaneously diagonalizable, that is, there is an invertible matrix $P \in \mathbb{C}^{r \times r}$ such that all products $P^{-1}Y^{j,k}P$ are diagonal for every $(j, k) \in \Upsilon$. Suppose they are diagonalized as

$$P^{-1}Y^{j,k}P = \text{diag}[\lambda_{j,k,1}, \lambda_{j,k,2}, \dots, \lambda_{j,k,r}] \tag{35}$$

with the eigenvalues $\lambda_{j,k,s}$. Write P in the column form

$$P = [p_1 \cdots p_r].$$

For each $s = 1, \dots, r$ and $j = 3, \dots, m$, let

$$v^{s,j} := (1, \lambda_{j,2,s}, \dots, \lambda_{j,n_j,s}). \tag{36}$$

Suppose \mathcal{F} has a rank- r decomposition

$$\mathcal{F} = \sum_{s=1}^r u^{s,1} \otimes \cdots \otimes u^{s,m}.$$

Under the assumptions of Theorem 3, linear system (24) has a unique solution for each pair $(j, k) \in \Upsilon$. For every $j \in \{3, \dots, m\}$, there exist scalars $c_{s,j}, c_{s,1}$ such that

$$u^{s,j} = c_{s,j}v^{s,j}, \quad u^{s,1} = c_{s,1}p_s.$$

Then, we consider the sub-tensor equation in the vector variables $y_1, \dots, y_r \in \mathbb{C}^{n_2}$

$$\mathcal{F}_{1:r, :, \dots, :} = \sum_{s=1}^r p_s \otimes y_s \otimes v^{s,3} \otimes \cdots \otimes v^{s,m}. \tag{37}$$

There are $rn_2 \cdots n_m$ equations and rn_2 unknowns. This overdetermined linear system has solutions such that

$$y_s = c_{s,2}u^{s,2}, \text{ for some } c_{s,2} \in \mathbb{C}.$$

After all y_s are obtained, we solve the linear equation in $z_1, \dots, z_r \in \mathbb{C}^{n_1-r}$

$$\mathcal{F}_{r+1:n_1, :, \dots, :} = \sum_{s=1}^r z_s \otimes y_s \otimes v^{s,3} \otimes \cdots \otimes v^{s,m}. \tag{38}$$

After all y_s, z_s are obtained, we choose the vectors ($s = 1, \dots, r$)

$$v^{s,1} = \begin{bmatrix} p_s \\ z_s \end{bmatrix}, \quad v^{s,2} = y_s.$$

Then we get the tensor decomposition

$$\mathcal{F} = \sum_{s=1}^r v^{s,1} \otimes v^{s,2} \otimes \dots \otimes v^{s,m}. \tag{39}$$

Summarizing the above, we get the following algorithm for computing tensor decompositions when $r \leq n_1$ and $r \leq N_3$. Suppose the dimensions are ordered such that $n_1 \geq n_2 \geq \dots \geq n_m$.

Algorithm 1 (Rank- r tensor decomposition)

Input A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ with rank $r \leq \min(n_1, N_3)$.

Step 1 For each pair $(j, k) \in \Upsilon$, solve the matrix equation for the solution $Y^{j,k}$:

$$A[\mathcal{F}, j]Y^{j,k} = b[\mathcal{F}, j, k]. \tag{40}$$

Step 2 Choose generic scalars $\xi_{j,k}$. Then compute the eigenvalue decomposition $P^{-1}YP = D$ for the matrix

$$Y := \frac{1}{\sum_{(j,k) \in \Upsilon} \xi_{j,k}} \sum_{(j,k) \in \Upsilon} \xi_{j,k} Y^{j,k}.$$

Step 3 For $s = 1, \dots, r$ and $j \geq 3$, let $v^{s,j}$ be the vectors as in (36).

Step 4 Solve linear system (37) for vectors y_1, \dots, y_r .

Step 5 Solve linear system (38) for vectors z_1, \dots, z_r .

Step 6 For each $s = 1, \dots, r$, let $v^{s,1} = \begin{bmatrix} p_s \\ z_s \end{bmatrix}$ and $v^{s,2} = y_s$.

Output A tensor rank- r decomposition as in (39).

The correctness of Algorithm 1 is justified as follows.

Theorem 5 Suppose $n_1 \geq n_2 \geq \dots \geq n_m$ and $r \leq \min(n_1, N_3)$ as in (33). For a generic tensor \mathcal{F} of rank- r , Algorithm 1 produces a rank- r tensor decomposition for \mathcal{F} .

Proof This can be implied by Theorem 4.

4.2 Tensor Decompositions Via Reshaping

A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ can be reshaped as a cubic order tensor $\widehat{\mathcal{F}}$ as in (12). One can apply Algorithm 1 to compute tensor decomposition (12) for $\widehat{\mathcal{F}}$. If the decomposing vectors $w^{s,1}, w^{s,2}, w^{s,3}$ can be reshaped to rank-1 tensors, then we can convert (12) to a tensor decomposition for \mathcal{F} . This is justified by Theorem 2, under some assumptions. A benefit for doing this is that we may be able to compute tensor decompositions for the case that

$$N_3 < r \leq p_2,$$

with the dimension p_2 as in Theorem 2. This leads to the following algorithm for computing tensor decompositions.

Algorithm 2 (Tensor decompositions via reshaping.) Let p_1, p_2, p_3 be dimensions as in Theorem 2.

- Input A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ with rank $r \leq p_2$.
- Step 1 Reshape the tensor \mathcal{F} to a cubic tensor $\widehat{\mathcal{F}} \in \mathbb{C}^{p_1 \times p_2 \times p_3}$ as in (12).
- Step 2 Use Algorithm 1 to compute the tensor decomposition

$$\widehat{\mathcal{F}} = \sum_{s=1}^r w^{s,1} \otimes w^{s,2} \otimes w^{s,3}. \tag{41}$$

- Step 3 If all $w^{s,1}, w^{s,2}, w^{s,3}$ can be expressed as outer products of rank-1 tensors as in (11), then output the tensor decomposition as in (3). If one of $w^{s,1}, w^{s,2}, w^{s,3}$ cannot be expressed as in (11), then the reshaping does not produce a tensor decomposition for \mathcal{F} .

Output A tensor decomposition for \mathcal{F} as in (3).

For Algorithm 2, we have a similar conclusion like Theorem 5. For cleanness of the paper, we do not repeat it here.

5 Low Rank Tensor Approximations

When a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times \dots \times n_m}$ has the rank bigger than r , the linear systems in Algorithm 1 may not be consistent. However, we can find linear least squares solutions for them. This gives an algorithm for computing low rank tensor approximations. Recall the label set Υ as in (34). The following is the algorithm.

Algorithm 3 (Rank- r tensor approximation.)

- Input A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$ and a rank $r \leq \min(n_1, N_3)$.
- Step 1 For each pair $(j, k) \in \Upsilon$, solve the linear least squares problem

$$\min_{Y^{j,k} \in \mathbb{C}^{r \times r}} \left\| A[\mathcal{F}, j](Y^{j,k})^T - b[\mathcal{F}, j, k] \right\|^2. \tag{42}$$

Let $\widehat{Y}^{j,k}$ be an optimizer.

- Step 2 Choose generic scalars $\xi_{j,k}$ and let

$$\widehat{Y}[\xi] = \frac{1}{\sum_{(j,k) \in \Upsilon} \xi_{j,k}} \sum_{(j,k) \in \Upsilon} \xi_{j,k} \widehat{Y}^{j,k}.$$

Compute the eigenvalue decomposition $\widehat{P}^{-1} \widehat{Y}[\xi] \widehat{P} = \Lambda$ such that $\widehat{P} = [\widehat{p}_1 \dots \widehat{p}_r]$ is invertible and Λ is diagonal.

- Step 3 For each pair $(j, k) \in \Upsilon$, select the diagonal entries

$$\text{diag}[\widehat{\lambda}_{j,k,1} \widehat{\lambda}_{j,k,2} \dots \widehat{\lambda}_{j,k,r}] = \text{diag}(\widehat{P}^{-1} \widehat{Y}^{j,k} \widehat{P}).$$

For each $s = 1, \dots, r$ and $j = 3, \dots, m$, let

$$\widehat{v}^{s,j} = (1, \widehat{\lambda}_{j,2,2}, \dots, \widehat{\lambda}_{j,n_j,s}).$$

Step 4 Let $(\hat{y}_1, \dots, \hat{y}_r)$ be an optimizer for the following least squares:

$$\min_{(y_1, \dots, y_r)} \left\| \mathcal{F}_{1:r, :, \dots, :} - \sum_{s=1}^r \hat{p}_s \otimes y_s \otimes \hat{v}^{s,3} \otimes \dots \otimes \hat{v}^{s,m} \right\|^2. \tag{43}$$

Step 5 Let $(\hat{z}_1, \dots, \hat{z}_r)$ be an optimizer for the following least squares:

$$\min_{(z_1, \dots, z_r)} \left\| \mathcal{F}_{r+1:n_1, :, \dots, :} - \sum_{s=1}^r z_s \otimes \hat{y}_s \otimes \hat{v}^{s,3} \otimes \dots \otimes \hat{v}^{s,m} \right\|^2. \tag{44}$$

Step 6 Let $\hat{v}^{s,1} = \begin{bmatrix} \hat{p}_s \\ \hat{z}_s \end{bmatrix}$ and $\hat{v}^{s,2} = \hat{y}_s$ for each $s = 1, \dots, r$.

Output A rank- r approximation tensor

$$\mathcal{X}^{gp} := \sum_{s=1}^r \hat{v}^{s,1} \otimes \hat{v}^{s,2} \otimes \dots \otimes \hat{v}^{s,m}. \tag{45}$$

If \mathcal{F} is sufficiently close to a rank- r tensor, then \mathcal{X}^{gp} is expected to be a good rank- r approximation. Mathematically, the tensor \mathcal{X}^{gp} produced by Algorithm 3 may not be a best rank- r approximation. However, in computational practice, we can use (45) as a starting point to solve the nonlinear least squares optimization

$$\min_{(u^{s,1}, \dots, u^{s,m})} \left\| \mathcal{F} - \sum_{s=1}^r u^{s,1} \otimes u^{s,2} \otimes \dots \otimes u^{s,m} \right\|^2. \tag{46}$$

to improve the approximation quality. Let \mathcal{X}^{opt} be a rank- r approximation tensor

$$\mathcal{X}^{opt} := \sum_{s=1}^r u^{s,1} \otimes u^{s,2} \otimes \dots \otimes u^{s,m} \tag{47}$$

which is an optimizer to (46) obtained by nonlinear optimization methods with \mathcal{X}^{opt} as the initial point.

5.1 Approximation Error Analysis

Suppose the tensor \mathcal{F} has a best (or nearly best) rank- r approximation

$$\mathcal{X}^{bs} := \sum_{s=1}^r (x^{s,1}) \otimes (x^{s,2}) \otimes \dots \otimes (x^{s,m}). \tag{48}$$

Let \mathcal{E} be the tensor such that

$$\mathcal{F} = \mathcal{X}^{bs} + \mathcal{E}. \tag{49}$$

We analyze the approximation performance of \mathcal{X}^{gp} when the distance $\epsilon = \|\mathcal{E}\|$ is small. For a generating matrix G and a generic $\xi = (\xi_{j,k})_{(j,k) \in \Upsilon}$, denote that

$$M[\xi, G] := \frac{1}{\sum_{(j,k) \in \Upsilon} \xi_{j,k}} \sum_{(j,k) \in \Upsilon} \xi_{j,k} M^{j,k}[G]. \tag{50}$$

Recall the $A[\mathcal{F}, j], b[\mathcal{F}, j, k]$ as in (23). Note that

$$\begin{aligned} A[\mathcal{F}, j] &= A[\mathcal{X}^{bs}, j] + A[\mathcal{E}, j], \\ b[\mathcal{F}, j, k] &= b[\mathcal{X}^{bs}, j, k] + b[\mathcal{E}, j, k]. \end{aligned} \tag{51}$$

Suppose $(x^{s,j})_1 \neq 0$ for $j = 2, \dots, m$.

Theorem 6 *Let \mathcal{X}^{gp} be produced by Algorithm 3. Let $\mathcal{F}, \mathcal{X}^{bs}, \mathcal{X}^{opt}, \mathcal{E}, x^{s,j}, \xi_{j,k}$ be as above. Assume the following conditions hold:*

- (i) *The subvectors $(x^{1,1})_{1:r}, \dots, (x^{r,1})_{1:r}$ are linearly independent.*
- (ii) *All matrices $A[\mathcal{F}, j]$ and $A[\mathcal{X}^{bs}, j]$ ($3 \leq j \leq m$) have full column rank.*
- (iii) *The first entry $(x^{s,j})_1 \neq 0$ for all $j = 2, \dots, m$.*
- (iv) *The following scalars are pairwise distinct*

$$\sum_{(j,k) \in \Upsilon} \xi_{j,k} (x^{1,j})_k, \dots, \sum_{(j,k) \in \Upsilon} \xi_{j,k} (x^{r,j})_k. \tag{52}$$

If the distance $\epsilon = \|\mathcal{F} - \mathcal{X}^{bs}\|$ is sufficiently small, then

$$\|\mathcal{X}^{bs} - \mathcal{X}^{gp}\| = O(\epsilon) \quad \|\mathcal{F} - \mathcal{X}^{gp}\| = O(\epsilon). \tag{53}$$

where the constants in the above $O(\cdot)$ only depend on \mathcal{F} and ξ .

Proof By conditions (i) and (iii) and by Theorem 3, there exists a generating matrix G^{bs} for \mathcal{X}^{bs} such that

$$A[\mathcal{X}^{bs}, j](M^{j,k}[G^{bs}])^T = b[\mathcal{X}^{bs}, j, k] \tag{54}$$

for all $j \in \{2, \dots, m\}$ and $k \in \{2, \dots, n_j\}$. Note that $Y^{j,k}$ is the least squares solution to (42), so for each $(j, k) \in \Upsilon$,

$$Y^{j,k} = A[\mathcal{F}, j]^\dagger \cdot b[\mathcal{F}, j, k], \quad M^{j,k}[G_0^{bs}] = A[\mathcal{X}^{bs}, j]^\dagger \cdot b[\mathcal{X}^{bs}, j, k].$$

(The superscript \dagger denotes the pseudoinverse of a matrix.) By (49), for $j = 2, \dots, m$, we have

$$\begin{aligned} \|A[\mathcal{F}, j] - A[\mathcal{X}^{bs}, j]\|_F &\leq \|\mathcal{F} - \mathcal{X}^{bs}\| \leq \epsilon, \\ \|b[\mathcal{F}, j, k] - b[\mathcal{X}^{bs}, j, k]\|_F &\leq \|\mathcal{F} - \mathcal{X}^{bs}\| \leq \epsilon. \end{aligned} \tag{55}$$

Hence, by the condition (ii), if $\varepsilon > 0$ is small enough, we have

$$\|Y^{j,k} - M^{j,k}[G^{bs}]\| = O(\varepsilon). \tag{56}$$

for all $(j, k) \in \Upsilon$. This follows from perturbation analysis for linear least squares (see [42, Theorem 3.4]).

By (48) and Theorem 3, for $s = 1, \dots, r$ and $(j, k) \in \Upsilon$, it holds that

$$M^{j,k}[G^{bs}](x^{s,1})_{1:r} = (x^{s,j})_k (x^{s,1})_{1:r}.$$

This means that each $(x^{s,1})_{1:r}$ is an eigenvector of $M^{j,k}[G^{bs}]$, associated with the eigenvalue $(x^{s,j})_k$, for each $s = 1, \dots, r$. The matrices $M^{j,k}[G^{bs}]$ are simultaneously diagonalizable, by the condition (i). So $M[\xi, G^{bs}]$ is also diagonalizable. Note the eigenvalues of $M[\xi, G^{bs}]$ are the sums in (52). They are distinct from each other, by the condition (iv). When $\varepsilon > 0$ is small enough, $M[\xi, G^{bs}]$ also has distinct eigenvalues. Write that

$$Q = [(x^{1,1})_{1:r} \cdots (x^{r,1})_{1:r}].$$

Note that $Q^{-1}M[\xi, G^{bs}]Q = D$ is an eigenvalue decomposition. Up to a scaling on \hat{P} in algorithm 3, it holds that

$$\|\hat{p}_s - x^{s,1}\|_2 = O(\varepsilon), \quad \|D - \Lambda\|_F = O(\varepsilon). \tag{57}$$

We refer to [43] for the perturbation bounds in (57). The constants in the above $O(\cdot)$ eventually only depend on \mathcal{F}, ξ .

Note that $(\hat{y}_s, \dots, \hat{y}_r)$ is the least squares solution to (43) and

$$\mathcal{X}_{1:r, :, \dots, :}^{bs} = \sum_{s=1}^r x^{s,1} \otimes x^{s,2} \otimes x^{s,3} \otimes \dots \otimes x^{s,m}. \tag{58}$$

Due to perturbation analysis of linear least squares, we also have

$$\|\hat{y}_s - x^{s,2}\|_2 = O(\varepsilon). \tag{59}$$

Note that the subvectors $(x^{s,1})_{r+1:n_1}$ satisfy the equation

$$\mathcal{X}_{r+1:n_1, :, \dots, :}^{bs} = \sum_{s=1}^r (x^{s,1})_{r+1:n_1} \otimes x^{s,2} \otimes \dots \otimes x^{s,m}. \tag{60}$$

Recall that $(\hat{z}_1, \dots, \hat{z}_r)$ is the least squares solution to (44). Due to perturbation analysis of linear least squares, we further have the error bound

$$\|(x^{s,1})_{r+1:n_1} - \hat{z}_s\|_2 = O(\varepsilon). \tag{61}$$

Summarizing the above, we eventually get $\|\mathcal{X}^{gp} - \mathcal{X}^{bs}\| = O(\varepsilon)$, so

$$\|\mathcal{F} - \mathcal{X}^{gp}\| \leq \|\mathcal{F} - \mathcal{X}^{bs}\| + \|\mathcal{X}^{bs} - \mathcal{X}^{gp}\| = O(\varepsilon).$$

The constant for the above $O(\cdot)$ eventually only depends on \mathcal{F} , ξ .

5.2 Reshaping for Low Rank Approximations

Similar to tensor decompositions, the reshaping trick as in Sect. 4.2 can also be used for computing low rank tensor approximations. For $m > 3$, a tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$ can be reshaped as a cubic tensor $\widehat{\mathcal{F}} \in \mathbb{C}^{p_1 \times p_2 \times p_3}$ as in (12). Similarly, Algorithm 3 can be used to compute low rank tensor approximations. Suppose the computed rank- r approximating tensor for $\widehat{\mathcal{F}}$ is

$$\widehat{\mathcal{X}}^{gp} := \sum_{s=1}^r \widehat{w}^{s,1} \otimes \widehat{w}^{s,2} \otimes \widehat{w}^{s,3}. \tag{62}$$

Typically, the decomposing vectors $\widehat{w}^{s,1}$, $\widehat{w}^{s,2}$, $\widehat{w}^{s,3}$ may not be reshaped to rank-1 tensors. Suppose the reshaping is such that $I_1 \cup I_2 \cup I_3 = \{1, 2, \dots, m\}$ is a union of disjoint label sets and the reshaped dimensions are

$$p_1 = \prod_{i \in I_1} n_i, \quad p_2 = \prod_{i \in I_2} n_i, \quad p_3 = \prod_{i \in I_3} n_i.$$

Let $m_i = |I_i|$ for $i = 1, 2, 3$. By the reshaping, the vectors $\widehat{w}^{s,i}$ can be reshaped back to a tensor $\widehat{W}^{s,i}$ of order m_i , for each $i = 1, 2, 3$. If $m_i = 1$, $\widehat{W}^{s,i}$ is a vector. If $m_i = 2$, we can find a best rank-1 matrix approximation for $\widehat{W}^{s,i}$. If $m_i \geq 3$, we can apply Algorithm 3 with $r = 1$ to get a rank-1 approximation for $\widehat{W}^{s,i}$. In application, we are mostly interested in reshaping such that all $m_i \leq 2$. Finally, this produces a rank- r approximation for \mathcal{F} .

The following is a low rank tensor approximation algorithm via reshaping tensors.

Algorithm 4 (low rank tensor approximations via reshaping.)

- Input A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$ and a rank r .
- Step 1 Reshape \mathcal{F} to a cubic order tensor $\widehat{\mathcal{F}} \in \mathbb{C}^{p_1 \times p_2 \times p_3}$.
- Step 2 Use Algorithm 3 to compute a rank- r approximating tensor $\widehat{\mathcal{X}}^{gp}$ as in (62) for $\widehat{\mathcal{F}}$.
- Step 3 For each $i = 1, 2, 3$, reshape each vector $\widehat{w}^{s,i}$ back to a tensor $\widehat{W}^{s,i}$ of order m_i as above.
- Step 4 For each $i = 1, 2, 3$, compute a rank-1 approximating tensor $\widehat{X}^{s,i}$ for $\widehat{W}^{s,i}$ of order m_i as above.
- Output Reshape the sum $\sum_{s=1}^r \widehat{X}^{s,1} \otimes \widehat{X}^{s,2} \otimes \widehat{X}^{s,3}$ to a tensor in $\mathbb{C}^{n_1 \times n_2 \times \dots \times n_m}$, which is a rank- r approximation for \mathcal{F} .

We can do a similar approximation analysis for Algorithm 4 as for Theorem 6. For cleanness of the paper, we do not repeat that.

6 Numerical Experiments

In this section, we apply Algorithms 1 and 3 to compute tensor decompositions and low rank tensor approximations. We implement these algorithms in MATLAB 2020b on a workstation with Ubuntu 20.04.2 LTS, Intel ®Xeon(R) Gold 6248R CPU @ 3.00GHz and memory 1TB. For computing low rank tensor approximations, we use the function *cpd_nls* provided in Tensorlab 3.0 [44] to solve nonlinear least squares optimization (46). The \mathcal{X}^{sp} denotes the approximating tensor returned by Algorithm 3, and \mathcal{X}^{opt} denotes the approximating tensor obtained by solving (46), with \mathcal{X}^{sp} as the initial point. In our numerical experiments, if the rank r is unknown, we use the most square flattening matrix to estimate r as in (4) and Lemma 1.

Example 2 Consider the tensor $\mathcal{F} \in \mathbb{C}^{4 \times 4 \times 3}$ whose slices $\mathcal{F}_{\dots,1}, \mathcal{F}_{\dots,2}, \mathcal{F}_{\dots,3}$ are, respectively

$$\left[\begin{array}{cccc|cccc|cccc} 27 & 25 & 35 & 42 & 44 & 32 & 52 & 56 & 42 & 26 & 48 & 45 \\ 48 & 68 & 80 & 80 & 68 & 76 & 100 & 96 & 64 & 60 & 88 & 76 \\ 26 & 24 & 34 & 40 & 42 & 30 & 50 & 52 & 47 & 27 & 53 & 47 \\ 33 & 41 & 49 & 66 & 46 & 46 & 62 & 76 & 45 & 37 & 57 & 60 \end{array} \right].$$

By Lemma 1, the estimated rank is $r = 4$.

Applying Algorithm 1 with $r = 4$, we get the rank-4 decomposition $\mathcal{F} = U^{(1)} \circ U^{(2)} \circ U^{(3)}$, with

$$U^{(1)} = \begin{bmatrix} 8 & 6 & 4 & 9 \\ 8 & 12 & 16 & 12 \\ 4 & 6 & 4 & 12 \\ 4 & 12 & 8 & 9 \end{bmatrix}, \quad U^{(2)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ \frac{1}{2} & 1 & 3 & \frac{1}{3} \\ 1 & 1 & 3 & 1 \\ 1 & 4 & 1 & \frac{2}{3} \end{bmatrix}, \quad U^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 1 & \frac{2}{3} & \frac{3}{4} & 3 \end{bmatrix}.$$

Example 3 Consider the tensor in $\mathbb{C}^{5 \times 4 \times 3 \times 3}$

$$\mathcal{F} = V^{(1)} \circ V^{(2)} \circ V^{(3)} \circ V^{(4)},$$

where the matrices $V^{(i)}$ are

$$V^{(1)} = \begin{bmatrix} 10 & 5 & -9 & -5 & 7 \\ 8 & 6 & -3 & -9 & 7 \\ -9 & -1 & 7 & -3 & -1 \\ 9 & -7 & -8 & 8 & -5 \\ -1 & 10 & 7 & -3 & 10 \end{bmatrix}, \quad V^{(2)} = \begin{bmatrix} -1 & 9 & -8 & 8 & 2 \\ 0 & -1 & -4 & 6 & 8 \\ 7 & -7 & -2 & 2 & 10 \\ 2 & 10 & -3 & -1 & -3 \end{bmatrix},$$

$$V^{(3)} = \begin{bmatrix} 5 & 2 & -2 & -7 & 3 \\ 9 & -3 & -7 & 7 & -2 \\ 0 & -10 & 10 & 6 & 10 \end{bmatrix}, \quad V^{(4)} = \begin{bmatrix} 8 & 2 & -7 & 10 & -5 \\ 4 & -8 & 4 & -6 & -10 \\ 5 & 0 & 7 & -1 & -2 \end{bmatrix}.$$

By Lemma 1, the estimated rank $r = 5$.

Applying Algorithm 1 with $r = 5$, we get the rank-5 tensor decomposition $\mathcal{F} = U^{(1)} \circ U^{(2)} \circ U^{(3)} \circ U^{(4)}$, where the computed matrices $U^{(i)}$ are

$$\begin{aligned}
 U^{(1)} &= \begin{bmatrix} -400 & 180 & 1008 & 2800 & -210 \\ -320 & 216 & 336 & 5040 & -210 \\ 360 & -36 & -784 & 1680 & 30 \\ -360 & -252 & 896 & -4480 & 150 \\ 40 & 360 & -784 & 1680 & -300 \end{bmatrix}, & U^{(2)} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & -\frac{1}{9} & \frac{1}{2} & \frac{3}{4} & 4 \\ -7 & -\frac{7}{9} & \frac{1}{4} & \frac{1}{4} & 5 \\ -2 & \frac{10}{9} & \frac{3}{8} & -\frac{1}{8} & -\frac{3}{2} \end{bmatrix}, \\
 U^{(3)} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \frac{9}{5} & -\frac{3}{2} & \frac{7}{2} & -1 & -\frac{2}{3} \\ 0 & -5 & -5 & -\frac{6}{7} & \frac{10}{3} \end{bmatrix}, \\
 U^{(4)} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ \frac{1}{2} & -4 & -\frac{4}{7} & -\frac{3}{5} & 2 \\ \frac{5}{8} & 0 & -1 & -\frac{1}{10} & \frac{2}{5} \end{bmatrix}.
 \end{aligned}$$

Example 4 Consider the tensor $\mathcal{F} \in \mathbb{C}^{5 \times 5 \times 4}$ such that

$$\mathcal{F}_{i_1, i_2, i_3} = i_1 + \frac{i_2}{2} + \frac{i_3}{3} + \sqrt{i_1^2 + i_2^2 + i_3^2}$$

for all i_1, i_2, i_3 in the corresponding range. The 5 biggest singular values of the flattening matrix $\text{Flat}(\mathcal{F})$ are

$$109.7393, \quad 5.2500, \quad 0.1068, \quad 8.325 \times 10^{-3}, \quad 3.401 \times 10^{-4}.$$

Applying Algorithm 3 with rank $r = 2, 3, 4, 5$, we get the approximation errors

| r | 2 | 3 | 4 | 5 |
|---------------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| $\ \mathcal{F} - \mathcal{X}^{gp}\ $ | 5.1237×10^{-1} | 6.8647×10^{-2} | 1.0558×10^{-2} | 9.9449×10^{-3} |
| $\ \mathcal{F} - \mathcal{X}^{opt}\ $ | 1.5410×10^{-1} | 1.3754×10^{-2} | 2.6625×10^{-3} | 4.9002×10^{-4} |

For the case $r = 3$, the computed approximating tensor by Algorithm 3 and by solving (46) is $U^{(1)} \circ U^{(2)} \circ U^{(3)}$, with

$$U^{(1)} = \begin{bmatrix} -0.4973 & -7.6813 & 11.7465 \\ -0.2525 & -6.9651 & 12.4970 \\ -0.0872 & -6.0497 & 13.2858 \\ -0.0132 & -5.0521 & 14.1423 \\ -0.0010 & -4.0469 & 15.0771 \end{bmatrix},$$

$$U^{(2)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.5058 & 0.9211 & 1.0306 \\ 0.1713 & 0.8167 & 1.0649 \\ 0.0262 & 0.7003 & 1.1042 \\ 0.0136 & 0.5807 & 1.1490 \end{bmatrix},$$

$$U^{(3)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.5075 & 0.9289 & 1.0216 \\ 0.1756 & 0.8323 & 1.0469 \\ 0.0399 & 0.7231 & 1.0771 \end{bmatrix}.$$

Example 5 Consider the tensor $\mathcal{F} \in \mathbb{C}^{6 \times 6 \times 6 \times 5 \times 4}$ such that

$$\mathcal{F}_{i_1, i_2, i_3, i_4, i_5} = \arctan(i_1 + 2i_2 + 3i_3 + 4i_4 + 5i_5),$$

for all i_1, i_2, i_3, i_4, i_5 in the corresponding range. The 5 biggest singular values of the flattening matrix $\text{Flat}(\mathcal{F})$ are

$$101.71, \quad 7.7529 \times 10^{-2}, \quad 2.2870 \times 10^{-3}, \quad 7.2294 \times 10^{-5}, \quad 2.0633 \times 10^{-6}.$$

Applying Algorithm 3 with rank $r = 2, 3, 4, 5$, we get the approximation errors as follows:

| r | 2 | 3 | 4 | 5 |
|---------------------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| $\ \mathcal{F} - \mathcal{X}^{gp}\ $ | 9.8148×10^{-3} | 3.1987×10^{-3} | 5.7945×10^{-3} | 1.0121×10^{-5} |
| $\ \mathcal{F} - \mathcal{X}^{opt}\ $ | 5.3111×10^{-3} | 2.2623×10^{-4} | 3.0889×10^{-5} | 1.7523×10^{-6} |

For the case $r = 3$, the computed approximating tensor by Algorithm 3 and by solving (46) is $U^{(1)} \circ U^{(2)} \circ U^{(3)} \circ U^{(4)} \circ U^{(5)}$, with

$$U^{(1)} = \begin{bmatrix} -0.0134 & -0.0347 & 1.5524 \\ -0.0112 & -0.0329 & 1.5525 \\ -0.0094 & -0.0312 & 1.5526 \\ -0.0079 & -0.0295 & 1.5527 \\ -0.0066 & -0.0280 & 1.5528 \\ -0.0056 & -0.0265 & 1.5529 \end{bmatrix}, U^{(2)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.7011 & 0.8992 & 1.0001 \\ 0.4939 & 0.8080 & 1.0003 \\ 0.3485 & 0.7260 & 1.0004 \\ 0.2459 & 0.6523 & 1.0006 \\ 0.1734 & 0.5861 & 1.0007 \end{bmatrix},$$

$$U^{(3)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.5886 & 0.8523 & 1.0002 \\ 0.3490 & 0.7258 & 1.0004 \\ 0.2064 & 0.6183 & 1.0006 \\ 0.1214 & 0.5269 & 1.0008 \\ 0.0715 & 0.4489 & 1.0011 \end{bmatrix}, U^{(4)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 \\ 0.4949 & 0.8078 & 1.0003 \\ 0.2463 & 0.6521 & 1.0006 \\ 0.1211 & 0.5269 & 1.0008 \\ 0.0596 & 0.4256 & 1.0011 \end{bmatrix},$$

$$U^{(5)} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 0 \\ 0.4161 & 0.7656 & 1.0003 & \\ 0.1730 & 0.5862 & 1.0007 & \\ 0.0711 & 0.4489 & 1.0011 & \end{bmatrix}.$$

Example 6 As in Theorem 6, we have shown that if the tensor to be approximated is sufficiently close to a rank- r tensor, then the computed rank- r approximation \mathcal{X}^{gp} is quasi-optimal. It can be further improved to a better approximation \mathcal{X}^{opt} by solving nonlinear optimization (46). In this example, we explore the numerical performance of Algorithms 3 and 4 for computing low rank tensor approximations. For the given dimensions n_1, \dots, n_m , we generate the tensor

$$\mathcal{R} = \sum_{s=1}^r u^{s,1} \otimes u^{s,2} \otimes \dots \otimes u^{s,m},$$

where each $u^{s,j} \in \mathbb{C}^{n_j}$ is a complex vector whose real and imaginary parts are generated randomly, obeying the Gaussian distribution. We perturb \mathcal{R} by another tensor \mathcal{E} , whose entries are also generated with the Gaussian distribution. We scale the perturbing tensor \mathcal{E} to have a desired norm ϵ . The tensor \mathcal{F} is then generated as

$$\mathcal{F} = \mathcal{R} + \mathcal{E}.$$

We choose ϵ to be one of $10^{-2}, 10^{-4}, 10^{-6}$, and use the relative errors

$$\rho_{-gp} = \frac{\|\mathcal{F} - \mathcal{X}^{gp}\|}{\|\mathcal{E}\|}, \quad \rho_{-opt} = \frac{\|\mathcal{F} - \mathcal{X}^{opt}\|}{\|\mathcal{E}\|}$$

to measure the approximation quality of $\mathcal{X}^{gp}, \mathcal{X}^{opt}$, respectively. For each case of $(n_1, \dots, n_m), r$ and ϵ , we generate 10 random instances of $\mathcal{R}, \mathcal{F}, \mathcal{E}$. For the case $(n_1, \dots, n_m) = (20, 20, 20, 20, 10)$, Algorithm 4 is used to compute \mathcal{X}^{gp} . All other cases are solved by Algorithm 3. The computational results are reported in Tables 1. For each case of (n_1, \dots, n_m) and r , we also list the median of above relative errors and the average CPU time (in seconds). The t_{gp} and t_{opt} denote the average CPU time (in seconds) for Algorithms 3/4 and for solving (46), respectively.

In the following, we give a comparison with the generalized eigenvalue decomposition (GEVD) method, which is a classical one for computing tensor decompositions when the rank $r \leq n_2$. We refer to [45, 46] for the work about the GEVD method. Consider a cubic order tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with $n_1 \geq n_2 \geq n_3$. Suppose $\mathcal{F} = U^{(1)} \circ U^{(2)} \circ U^{(3)}$ is a rank- r decomposition and $r \leq n_2$. Assume its first and second decomposing matrices $U^{(1)}, U^{(2)}$ have full column ranks and the third decomposing matrix $U^{(3)}$ does not have colinear columns. Denote the slice matrices

$$F_1 := \mathcal{F}_{1:r,1:r,1}, \quad F_2 := \mathcal{F}_{1:r,1:r,2}. \tag{63}$$

Table 1 Computational performance of Algorithms 3 and 4 and of nonlinear optimization (46)

| r | ϵ | ρ_{gp} | t_{gp} | ρ_{opt} | t_{opt} | r | ϵ | ρ_{gp} | t_{gp} | ρ_{opt} | t_{opt} |
|---|------------|-------------|----------|--------------|-----------|--------------------------------------|------------|-------------|----------|--------------|-----------|
| $(n_1, n_2, n_3)=(50,50,50)$ | | | | | | $(n_1, n_2, n_3)=(60, 50, 40)$ | | | | | |
| 10 | 10^{-2} | 1.63 | 0.08 | 0.99 | 1.57 | 15 | 10^{-2} | 17.49 | 0.19 | 0.99 | 2.17 |
| | 10^{-4} | 6.32 | 0.10 | 0.99 | 1.16 | | 10^{-4} | 10.80 | 0.15 | 0.99 | 1.36 |
| | 10^{-6} | 3.84 | 0.09 | 0.99 | 0.83 | | 10^{-6} | 5.16 | 0.20 | 0.99 | 1.10 |
| 20 | 10^{-2} | 25.83 | 0.29 | 0.99 | 2.99 | 30 | 10^{-2} | 28.70 | 0.40 | 0.98 | 6.95 |
| | 10^{-4} | 5.41 | 0.28 | 0.99 | 1.99 | | 10^{-4} | 15.77 | 0.37 | 0.98 | 3.61 |
| | 10^{-6} | 30.41 | 0.29 | 0.99 | 1.49 | | 10^{-6} | 50.96 | 0.37 | 0.98 | 2.27 |
| 30 | 10^{-2} | 27.91 | 0.50 | 0.98 | 7.08 | 45 | 10^{-2} | 35.48 | 0.61 | 0.97 | 25.73 |
| | 10^{-4} | 213.82 | 0.43 | 0.98 | 3.73 | | 10^{-4} | 35.03 | 0.63 | 0.97 | 8.08 |
| | 10^{-6} | 17.97 | 0.47 | 0.98 | 2.20 | | 10^{-6} | 34.67 | 0.61 | 0.97 | 5.69 |
| $(n_1, n_2, n_3)=(100,100,100)$ | | | | | | $(n_1, n_2, n_3)=(150,150,150)$ | | | | | |
| 20 | 10^{-2} | 11.21 | 0.86 | 1.00 | 6.36 | 30 | 10^{-2} | 8.59 | 2.92 | 1.00 | 17.17 |
| | 10^{-4} | 3.48 | 0.85 | 1.00 | 4.24 | | 10^{-4} | 3.18 | 3.05 | 1.00 | 11.20 |
| | 10^{-6} | 3.88 | 0.83 | 1.00 | 3.20 | | 10^{-6} | 4.24 | 3.42 | 1.00 | 11.75 |
| 40 | 10^{-2} | 24.17 | 1.76 | 0.99 | 17.80 | 60 | 10^{-2} | 49.80 | 6.04 | 1.00 | 87.31 |
| | 10^{-4} | 11.60 | 1.65 | 0.99 | 11.02 | | 10^{-4} | 13.77 | 5.89 | 1.00 | 24.96 |
| | 10^{-6} | 11.09 | 1.61 | 0.99 | 7.97 | | 10^{-6} | 17.49 | 6.07 | 1.00 | 18.81 |
| 60 | 10^{-2} | 18.71 | 3.40 | 0.99 | 28.16 | 90 | 10^{-2} | 29.44 | 10.64 | 0.99 | 98.78 |
| | 10^{-4} | 26.28 | 3.41 | 0.99 | 17.25 | | 10^{-4} | 152.49 | 10.53 | 0.99 | 43.58 |
| | 10^{-6} | 19.12 | 3.49 | 0.99 | 13.14 | | 10^{-6} | 17.01 | 10.06 | 0.99 | 26.98 |
| $(n_1, n_2, n_3, n_4)=(20,20,20,20,10)$ | | | | | | $(n_1, n_2, n_3, n_4)=(60,50,40,30)$ | | | | | |
| 24 | 10^{-2} | 37.93 | 0.88 | 1.00 | 45.56 | 20 | 10^{-2} | 31.42 | 2.78 | 1.00 | 31.16 |
| | 10^{-4} | 9.10 | 0.92 | 1.00 | 15.86 | | 10^{-4} | 1.17 | 2.76 | 1.00 | 9.39 |
| | 10^{-6} | 715.16 | 0.91 | 1.00 | 15.63 | | 10^{-6} | 4.14 | 2.79 | 1.00 | 9.48 |
| 48 | 10^{-2} | 166.00 | 1.95 | 1.00 | 270.56 | 40 | 10^{-2} | 6.99 | 7.52 | 1.00 | 31.81 |
| | 10^{-4} | 161.62 | 1.93 | 1.00 | 40.63 | | 10^{-4} | 2.58 | 7.32 | 1.00 | 20.07 |
| | 10^{-6} | 52.01 | 1.93 | 1.00 | 21.71 | | 10^{-6} | 2.49 | 7.22 | 1.00 | 20.22 |
| 72 | 10^{-2} | 73.70 | 3.10 | 1.00 | 102.90 | 60 | 10^{-2} | 11.48 | 9.83 | 1.00 | 48.08 |
| | 10^{-4} | 113.13 | 3.06 | 1.00 | 70.13 | | 10^{-4} | 6.38 | 9.80 | 1.00 | 38.97 |
| | 10^{-6} | 34.28 | 3.03 | 1.00 | 36.72 | | 10^{-6} | 16.35 | 9.76 | 1.00 | 30.38 |

One can show that

$$F_1 = U_{1:r,:}^{(1)} \cdot \text{diag}(U_{1,:}^{(3)}) \cdot (U_{1:r,:}^{(2)})^\top, \quad F_2 = U_{1:r,:}^{(1)} \cdot \text{diag}(U_{2,:}^{(3)}) \cdot (U_{1:r,:}^{(2)})^\top. \quad (64)$$

This implies that the columns of $(U_{1:r,r}^{(1)})^{-\top}$ are generalized eigenvectors of the matrix pair (F_1^\top, F_2^\top) . Consider the transformed tensor

$$\hat{\mathcal{F}} = (U_{1:r,r}^{(1)})^{-1} \times_1 \mathcal{F}_{1:r, :, :}, \tag{65}$$

For each $s = 1, \dots, r$, the slice $\hat{\mathcal{F}}_{s, :, :} = U_{:,s}^{(2)} \cdot (U_{:,s}^{(3)})^\top$ is a rank-1 matrix. The matrices $U^{(2)}, U^{(3)}$ can be obtained by computing rank-1 decompositions for the slices $\hat{\mathcal{F}}_{s, :, :}$. After this is done, we can solve the linear system

$$U^{(1)} \circ U^{(2)} \circ U^{(3)} = \mathcal{F} \tag{66}$$

to get the matrix $U^{(1)}$. The following is the GEVD method for computing cubic order tensor decompositions when the rank $r \leq n_2$.

Algorithm 5 (The GEVD method.)

Input A tensor $\mathcal{F} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with the rank $r \leq n_2$.

1. Formulate the tensor $\hat{\mathcal{F}}$ as in (65).
2. For $s = 1, \dots, r$, compute $U_{:,s}^{(2)}, U_{:,s}^{(3)}$ from the rank-1 decomposition of the matrix $\hat{\mathcal{F}}_{s, :, :}$.
3. Solve linear system (66) to get $U^{(1)}$.

Output The decomposing matrices $U^{(1)}, U^{(2)}, U^{(3)}$.

We compare the performance of Algorithm 1 and Algorithm 5 for randomly generated tensors with the rank $r \leq n_2$. We generate $\mathcal{F} = U^{(1)} \circ U^{(2)} \circ U^{(3)}$ such that each $U^{(i)} \in \mathbb{C}^{n_i \times r}$. The entries of $U^{(i)}$ are randomly generated complex numbers. Their real and imaginary parts are randomly generated, obeying the Gaussian distribution. For each case of (n_1, \dots, n_m) and r , we generate 20 random instances of \mathcal{F} . Algorithm 5 is implemented by the function `cpd_gevd` in the software `Tensorlab`. All the tensor decompositions are computed correctly by both methods. The average CPU time (in seconds) for Algorithm 1 is denoted as `time-gp`, while the average CPU time for the GEVD method is denoted as `time-gevd`. The computational results

Table 2 A comparison for the performance of Algorithms 1 and 5

| (n_1, n_2, n_3) | r | time-gevd | time-gp |
|-------------------|-----|-----------|---------|
| (40,30,30) | 30 | 0.91 | 0.29 |
| (50,50,50) | 50 | 4.77 | 0.85 |
| (100,100,100) | 80 | 12.17 | 5.54 |
| (150,150,150) | 100 | 79.85 | 13.30 |
| (200,200,200) | 120 | 161.83 | 25.71 |
| (250,250,250) | 140 | 285.03 | 55.71 |
| (300,300,300) | 100 | 306.64 | 61.38 |
| (400,400,400) | 180 | 934.15 | 271.21 |
| (500,500,500) | 200 | 1 688.98 | 539.75 |

are reported in Table 2. The numerical experiments show that Algorithm 1 is more computationally efficient than Algorithm 5.

7 Conclusions

This paper gives computational methods for computing low rank tensor decompositions and approximations. The proposed methods are based on generating polynomials. For a generic tensor of rank $r \leq \min(n_1, N_3)$, its tensor decomposition can be obtained by Algorithm 1. Under some general assumptions, we show that if a tensor is sufficiently close to a low rank one, then the low rank approximating tensor produced by Algorithm 3 is quasi-optimal. Numerical experiments are presented to show the efficiency of the proposed methods.

Author Contributions The authors have done analysis and computational results.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- [1] Landsberg, J.: Tensors: Geometry and Applications, Grad. Stud. Math., 128, AMS, Providence, RI, (2012)
- [2] Lim, L.H.: Tensors and hypermatrices. In: Hogben, L. (ed.) Handbook of Linear Algebra, 2nd edn. CRC Press, Boca Raton (2013)
- [3] Breiding, P., Vannieuwenhoven, N.: A Riemannian trust region method for the canonical tensor rank approximation problem. *SIAM J. Optim.* **28**(3), 2435–2465 (2018)
- [4] Lathauwer, L., de Moor, B., Vandewalle, J.: Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition. *SIAM J. Matrix Anal. Appl.* **26**(2), 295–327 (2004)
- [5] Lathauwer, L.: A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM J. Matrix Anal. Appl.* **28**(3), 642–666 (2006)
- [6] Domanov, I., de Lathauwer, L.: Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition. *SIAM J. Matrix Anal. Appl.* **35**(2), 636–660 (2014)
- [7] Nie, J.: Generating polynomials and symmetric tensor decompositions. *Found. Comput. Math.* **17**(2), 423–465 (2017)
- [8] Sorber, L., van Barel, M., de Lathauwer, L.: Optimization-based algorithms for tensor decompositions: canonical polyadic decomposition, decomposition in rank- $(L_r, L_r, 1)$ terms, and a new generalization. *SIAM J. Optim.* **23**(2), 695–720 (2013)
- [9] Telen, S., Vannieuwenhoven, N.: Normal forms for tensor rank decomposition. (2021) [arXiv:2103.07411](https://arxiv.org/abs/2103.07411)
- [10] Cui, C.F., Dai, Y.H., Nie, J.: All real eigenvalues of symmetric tensors. *SIAM J. Matrix Anal. Appl.* **35**(4), 1582–1601 (2014)

- [11] Fan, J., Nie, J., Zhou, A.: Tensor eigenvalue complementarity problems. *Math. Progr.* **170**(2), 507–539 (2018)
- [12] Nie, J., Wang, L.: Semidefinite relaxations for best rank-1 tensor approximations. *SIAM J. Matrix Anal. Appl.* **35**(3), 1155–1179 (2014)
- [13] Nie, J., Zhang, X.: Real eigenvalues of nonsymmetric tensors. *Comput. Opt. Appl.* **70**(1), 1–32 (2018)
- [14] Nie, J., Yang, Z., Zhang, X.: A complete semidefinite algorithm for detecting copositive matrices and tensors. *SIAM J. Optim.* **28**(4), 2902–2921 (2018)
- [15] Xiong, L., Chen, X., Huang, T.-K., Schneider, J., Carbonell, J. G.: Temporal collaborative filtering with bayesian probabilistic tensor factorization. In: *Proceedings of the 2010 SIAM International Conference on Data Mining*, 211–222, (2010)
- [16] Dunlavy, D.M., Kolda, T.G., Acar, E.: Temporal link prediction using matrix and tensor factorizations. *ACM Trans. Knowl. Discov. Data* **5**(2), 1–27 (2011)
- [17] Araujo, M., Papadimitriou, S., Gunnemann, S., Faloutsos, C., Basu, P., Swami, A., Papalexakis, E. E., Koutra, D.: Com2: fast automatic discovery of temporal (‘comet’) communities. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 271–283, (2014)
- [18] Nickel, M., Tresp, V., Kriegel, H.-P.: A three-way model for collective learning on multi-relational data. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1258–1266, (2012)
- [19] Jenatton, R., Roux, N., Bordes, A., Obozinski, G.R.: A latent factor model for highly multi-relational data. *Adv. Neural. Inf. Process. Syst.* **25**, 3167–3175 (2012)
- [20] Boden, B., Gunnemann, S., Hoffmann, H., Seidl, T.: Mining coherent subgraphs in multi-layer graphs with edge labels. In: *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1258–1266, (2012)
- [21] Kofidis, E., Regalia, P.A.: Tensor approximation and signal processing applications. *Contemp. Math.* **280**, 103–134 (2001)
- [22] Pajarola, R., Suter, S. K., Ballester-Ripoll, R., Yang, H.: Tensor approximation for multidimensional and multivariate data. In: *Anisotropy Across Fields and Scales*, 73–98, (2021)
- [23] Guo, B., Nie, J., Yang, Z.: Learning diagonal gaussian mixture models and incomplete tensor decompositions. *Vietnam J. Math.* (2021). <https://doi.org/10.1007/s10013-021-00534-3>
- [24] Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009)
- [25] Nie, J., Yang, Z.: Hermitian tensor decompositions. *SIAM J. Matrix Anal. Appl.* **41**(3), 1115–1144 (2020)
- [26] Friedland, S., Wang, L.: Spectral norm of a symmetric tensor and its computation. *Math. Comput.* **89**, 2175–2215 (2020)
- [27] Qi, L., Hu, S.: Spectral norm and nuclear norm of a third order tensor (2019). [arXiv:1909.01529](https://arxiv.org/abs/1909.01529)
- [28] de Silva, V., Lim, L.-H.: Tensor rank and the ill-posedness of the best low rank approximation problem. *SIAM J. Matrix Anal. Appl.* **30**(3), 1084–1127 (2008)
- [29] Comon, P., Luciani, X., de Almeida, A.L.F.: Tensor decompositions, alternating least squares and other tales. *J. Chemom.* **23**, 393–405 (2009)
- [30] Mao, X., Yuan, G., Yang, Y.: A self-adaptive regularized alternating least squares method for tensor decomposition problems. *Anal. Appl.* **18**(01), 129–147 (2020)
- [31] Yang, Y.: The epsilon-alternating least squares for orthogonal low-rank tensor approximation and its global convergence. *SIAM J. Matrix Anal. Appl.* **41**(4), 1797–1825 (2020)
- [32] Lathauwer, L., de Moor, B., Vandewalle, J.: On the best rank-1 and rank- (R_1, R_2, \dots, R_n) approximation of higher-order tensors. *SIAM J. Matrix Anal. Appl.* **21**(4), 1324–1342 (2000)
- [33] Guan, Y., Chu, M.T., Chu, D.: Convergence analysis of an svd-based algorithm for the best rank-1 tensor approximation. *Linear Algebra Appl.* **555**, 53–69 (2018)
- [34] Friedland, S., Tammali, V.: Low-rank approximation of tensors. In: *Numerical Algebra, Matrix Theory, Differential-Algebraic Equations and Control Theory*. 377–411, Springer, (2015)
- [35] Nie, J.: Low rank symmetric tensor approximations. *SIAM J. Matrix Anal. Appl.* **38**(4), 1517–1540 (2017)
- [36] Cox, D., Little, J., O’Shea, D.: *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, Berlin (2013)
- [37] Schafarevich, I.: *Basic Algebraic Geometry I: Varieties in Projective Space*. Springer-Verlag, Berlin (1988)
- [38] Kruskal, J.B.: Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra Appl.* **18**(2), 95–138 (1977)

- [39] Chiantini, L., Ottaviani, G., Vannieuwenhoven, N.: Effective criteria for specific identifiability of tensors and forms. *SIAM J. Matrix Anal. Appl.* **38**(2), 656–681 (2017)
- [40] Nie, J.: Nearly low rank tensors and their approximations (2014). [arXiv:1412.7270](https://arxiv.org/abs/1412.7270)
- [41] Nie, J., Wang, L., Zheng, Z.: Higher Order Correlation Analysis for Multi-View Learning, *Pacific Journal of Optimization*, to appear, (2022)
- [42] Demmel, J.: *Applied Numerical Linear Algebra*. SIAM, New Delhi (1997)
- [43] Chatelin, F.: *Eigenvalues of Matrices*. SIAM, New Delhi (2012)
- [44] Vervliet, N., Debals, O., de Lathauwer, L.: Tensorlab 3.0 - numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization, In 2016 50th Asilomar Conference on Signals, Systems and Computers, 1733–1738, (2016)
- [45] Leurgans, S.E., Ross, R.T., Abel, R.B.: A decomposition for three-way arrays. *SIAM J. Matrix Anal. Appl.* **14**, 1064–1083 (1993)
- [46] Sanchez, E., Kowalski, B.R.: Tensorial resolution: a direct trilinear decomposition. *J. Chemometrics* **4**, 29–45 (1990)