Received: 26 March 2019 | Revised: 25 October 2019 | Accepted: 4 February 2020

DOI: 10.1002/nla.2290

RESEARCH ARTICLE

WILEY

Decompositions of third-order tensors: HOSVD, T-SVD, and Beyond

Chao Zeng | **Michael K. Ng**

Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

Correspondence
Chao Zeng, Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong.
Email: zengchao@nankai.edu.cn

Funding information
Research Grants Council, University Grants Committee, Grant/Award Numbers: 12200317, 12300218, 12306616, 12300519; University Research Committee, University of Hong Kong, Grant/Award Number: 104005583

Summary

The higher order singular value decomposition, which is regarded as a generalization of the matrix singular value decomposition (SVD), has a long history and is well established, while the T-SVD is relatively new and lacks systematic analysis. Because of the unusual tensor-tensor product that the T-SVD is based on, the form of the T-SVD may be difficult to comprehend. The main aim of this article is to establish a connection between these two decompositions. By converting the form of the T-SVD into the sum of outer product terms, we compare the forms of the two decompositions. Moreover, from establishing the connection, a new decomposition which has a specific nonzero pattern, is proposed and developed. Numerical examples are given to demonstrate the useful ability of the new decomposition for approximation and data compression.

KEYWORDS

HOSVD, O-SVD, T-SVD, third-order tensor

1 | INTRODUCTION

Tensors are multidimensional extension of matrices, which are used to represent ubiquitous multidimensional data, such as color images, hyperspectral images, videos, and facial recognition datasets. In the past several decades, tensor analysis has attracted a lot of attention in chemometrics,[1] signal processing,[2] computer vision,[3] psychometrics,[4] data mining,[5] and machine learning.[6] The compression, sort, analysis, and many other processing of tensor data rely on tensor decomposition. Various decompositions such as CANDECOMP/PARAFAC,[4,7] higher order singular value decomposition (HOSVD),[8,9] hierarchical singular value decomposition (SVD),[10,11] T-SVD,[12–14] tensor-train[15] have been developed in the literature. The interested readers can refer to References 16, 17 for a thorough review.

Among all these decompositions, T-SVD is a special one: the T-SVD regards a third-order tensor as a matrix with each element being a tube, that is, tubal matrix. An algebraic framework of the T-SVD is constructed,[18] and the analogous definitions of matrices, such as tensor-tensor product, identity tensor, transpose, orthogonal tensor, diagonal tensor, and tubal rank, are defined in References 12,13,19. Although the tube-based strategy makes the T-SVD have a nice form similar to the matrix SVD, the so-called t-product between tensors results in some issues. For example, for most decompositions, such as CANDECOMP/PARAFAC, HOSVD, and tensor-train, each entry of the original tensor can be represented by the elements in the decompositions, while for the T-SVD, there is no such expression; we cannot express the T-SVD as the sum of outer product terms directly; and the sparsity of the core tensor corresponding to the T-SVD is measured by the tubal rank, which is not so clear compared with other definitions of tensor rank. In Reference 14, the T-SVD is generalized to arbitrary order, and the situation becomes more complicated.

Our motivation arises from establishing a connection between HOSVD and T-SVD. Since the T-SVD mainly aims at third-order tensors, we focus on this type of tensors. Because the form of the HOSVD is easy to understand and the multilinear rank is well studied, the establishment of their connection is necessary to understand the essence of the T-SVD. Here we would like to answer the following two questions:

1. What is the relationship between the multilinear rank and the tubal rank?
2. Can we express the T-SVD with the sum of outer product terms?

After we understand the relationship between HOSVD and T-SVD, we propose a new decomposition based on them. Like the T-SVD, the new decomposition also aims at tensors with fixed orientations. In applications, we usually confront this type of tensors, involving time series or other ordered data which have high correlation among frontal slices. Examples include, but not limited to, computed tomography,[20] multichannel images,[21] hyperspectral images,[19,22] facial recognition datasets,[23] mass spectrometry imaging,[24] and videos.[24–27] The new tensor decomposition is obtained by finding a basis of the space spanned by all frontal slices and computing SVD on each element of the basis. It is interesting to note that the new decomposition has a specific nonzero pattern in the new decomposition. By using this property, we can obtain a much better approximation than HOSVD and T-SVD. Numerical results will be reported to demonstrate the approximation ability of the new tensor decomposition.

The rest of this article is organized as follows. In Section 2, we introduce some basic definitions and preliminaries. The connection between the HOSVD and the T-SVD is established in Section 3. In Section 4, a new decomposition is proposed. In Section 5, we give numerical experiments. Some conclusion remarks are presented in Section 6.

2 | TENSOR PRELIMINARIES

Throughout this article, scalars are denoted by lowercase letters $(a, b, \dots)$, vectors are denoted by bold-face lowercase letters $(\mathbf{a}, \mathbf{b}, \dots)$, matrices are denoted by bold-face capitals $(\mathbf{A}, \mathbf{B}, \dots)$, and tensors are written as calligraphic letters $(\mathcal{A}, \mathcal{B}, \dots)$. The ith entry of a vector $\mathbf{a}$ is denoted by a_i, and the (i, j, k)th element of a third-order tensor $\mathcal{A}$ is denoted by a_{ijk}.

We adopt the Matlab indexing notation. A mode-n fiber is a column vector defined by fixing every index but the nth index, and a mode-(m, n) slice is a matrix defined by fixing every index but the mth index and the nth index. For example, for a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\mathcal{A}(:, j, k) = [a_{1jk}, a_{2jk}, \dots , a_{I_1jk}]^T$ is a mode-1 fiber. A mode-(1,2) slice is also called a frontal slice. The kth frontal slice has the following form:

$$\mathcal{A}(:, :, k) = \begin{bmatrix} a_{11k} & a_{12k} & \dots & a_{1I_2k} \\ a_{21k} & a_{22k} & \dots & a_{2I_2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I_11k} & a_{I_12k} & \dots & a_{I_1I_2k} \end{bmatrix}.$$

The Frobenius norm of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ is given by

$$\|\mathcal{A}\| = \sqrt{\sum_{i=1}^{I_1} \sum_{j=1}^{I_2} \sum_{k=1}^{I_3} |a_{ijk}|^2}.$$

The n-mode product of $\mathcal{A}$ by $\mathbf{M} \in \mathbb{C}^{J_n \times I_n}$, denoted by $\mathcal{A} \times_n \mathbf{M}$, is a tensor generated by multiplying each mode-n fiber of $\mathcal{A}$ by $\mathbf{M}$. For two tensors with some type of fibers having the same length, we can define a corresponding type of tensor-tensor product with the matrix-matrix product on each certain type of slice.

Definition 1. [tensor-tensor product] The three-mode product of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\mathcal{B} \in \mathbb{C}^{I_2 \times J \times I_3}$, denoted by $\mathcal{A} *_3 \mathcal{B}$, is of size $I_1 \times J \times I_3$, which is given by

$$(\mathcal{A} *_3 \mathcal{B})(:, :, k) = \mathcal{A}(:, :, k)\mathcal{B}(:, :, k), \quad k = 1, \dots , I_3.$$

We want to express the matrix-tensor product with the tensor-tensor product. The following definition is needed.

Definition 2. Let $\mathbf{M} \in \mathbb{C}^{I_1 \times I_2}$ and J be a positive integer. We define a tensor $\text{ten}(\mathbf{M}, J) \in \mathbb{C}^{I_1 \times I_2 \times J}$, given by

$$\text{ten}(\mathbf{M}, J)(:, :, k) = \mathbf{M},$$

for $k = 1, \dots, J$.

Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\mathbf{M}^{(1)} \in \mathbb{C}^{J_1 \times I_1}$, $\mathbf{M}^{(2)} \in \mathbb{C}^{J_2 \times I_2}$. Then we have the following relationship

$$\mathcal{A} \times_1 \mathbf{M}^{(1)} = \text{ten}\left(\mathbf{M}^{(1)}, I_3\right) *_3 \mathcal{A}, \quad \mathcal{A} \times_2 \mathbf{M}^{(2)} = \mathcal{A} *_3 \text{ten}\left(\mathbf{M}^{(2)^T}, I_3\right). \tag{1}$$

The mode-n unfolding of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ is denoted by $\mathbf{A}_{(n)}$ and arranges the mode-n fibers to be the columns of the resulting matrix. The tensor element (i_1, i_2, i_3) is mapped to the matrix element (i_n, j), where

$$j = 1 + \sum_{k=1, k \neq n}^{3} (i_k - 1) J_k \ \text{ with } \ J_k = \prod_{m=1, m \neq n}^{k-1} I_m.$$

2.1 | Tensor Rank

There are various ways to generalize the rank concept of matrices for high-order tensors. The multilinear rank is a generalization of the notion of column and row rank of matrices.

Definition 3. [see 9] The n-rank of $\mathcal{A}$, denoted by $\text{rank}_n(\mathcal{A})$, is the dimension of the vector space spanned by all n-mode fibers. The three-tuple $(\text{rank}_1(\mathcal{A}), \text{rank}_2(\mathcal{A}), \text{rank}_3(\mathcal{A}))$ is called the multilinear rank of $\mathcal{A}$.

For a third-order tensor $\mathcal{A}$, one has

$$\text{rank}_n(\mathcal{A}) = \text{rank}\left(\mathbf{A}_{(n)}\right), \quad n = 1, 2, 3.$$

Also suppose $\mathcal{S} = \mathcal{A} \times_1 \mathbf{M}^{(1)} \times_2 \mathbf{M}^{(2)} \times_3 \mathbf{M}^{(3)}$. By Reference 28, (2.18), (2.19), we have

$$\text{rank}_n(\mathcal{S}) \leq \text{rank}_n(\mathcal{A}), \quad n = 1, 2, 3. \tag{2}$$

In particular, if $\mathbf{M}^{(1)}, \mathbf{M}^{(2)}, \mathbf{M}^{(3)}$ are invertible, then

$$\text{rank}_n(\mathcal{S}) = \text{rank}_n(\mathcal{A}), \quad n = 1, 2, 3. \tag{3}$$

The following definition is related to the T-SVD.

Definition 4. Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. The (1,2)-rank, denoted by $\text{rank}_{(1,2)}(\mathcal{A})$, is given by

$$\text{rank}_{(1,2)}(\mathcal{A}) = \max\left\{\text{rank}\left(\mathcal{A}(:, :, 1)\right), \dots, \text{rank}\left(\mathcal{A}(:, :, I_3)\right)\right\}.$$

Similarly,

$$\begin{aligned}
\text{rank}_{(2,3)}(\mathcal{A}) &= \max\left\{\text{rank}\left(\mathcal{A}(1, :, :)\right), \dots, \text{rank}\left(\mathcal{A}(I_1, :, :)\right)\right\}, \\
\text{rank}_{(1,3)}(\mathcal{A}) &= \max\left\{\text{rank}\left(\mathcal{A}(:, 1, :)\right), \dots, \text{rank}\left(\mathcal{A}(:, I_2, :)\right)\right\}.
\end{aligned}$$

The three-tuple $\left(\text{rank}_{(1,2)}(\mathcal{A}), \text{rank}_{(2,3)}(\mathcal{A}), \text{rank}_{(1,3)}(\mathcal{A})\right)$ is called the slice rank of $\mathcal{A}$.

Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\mathcal{M}^{(1)} \in \mathbb{C}^{I_1 \times I_1 \times I_3}$, $\mathcal{M}^{(2)} \in \mathbb{C}^{I_2 \times I_2 \times I_3}$. Suppose $\mathcal{M}^{(1)}(:, :, k)$ and $\mathcal{M}^{(2)}(:, :, k)$ are invertible for $k = 1, \dots, I_3$. Then the following equation is straightforward:

$$\text{rank}_{(1,2)}\left(\mathcal{M}^{(1)} *_3 \mathcal{A} *_3 \mathcal{M}^{(2)}\right) = \text{rank}_{(1,2)}(\mathcal{A}). \tag{4}$$

There is a subtle difference between the slice rank and the multilinear rank. The multilinear rank concerns all fibers of one type, while the slice rank concentrates on each slice, which consists of a part of fibers of the tensor.

Proposition 1. *Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. Then*

$$\operatorname{rank}_{(m,n)}(\mathcal{A}) \leq \min \{\operatorname{rank}_m(\mathcal{A}), \operatorname{rank}_n(\mathcal{A})\}, \tag{5}$$

where $1 \leq m < n \leq 3$.

Proof. We show the case for $(m, n) = (1, 2)$. The other cases can be shown similarly.

Suppose $\operatorname{rank}_{(1,2)}(\mathcal{A}) = \operatorname{rank}(\mathcal{A}(:,:,k_0))$, where $k_0 \in \{1, 2, \dots, I_3\}$. Since the columns of $\mathcal{A}(:,:,k_0)$ are mode-1 fibers and the rows of $\mathcal{A}(:,:,k_0)$ are mode-2 fibers, $\mathcal{A}(:,:,k_0)$ is a submatrix of $\mathbf{A}_{(1)}$ and $\mathcal{A}(:,:,k_0)^T$ is a submatrix of $\mathbf{A}_{(2)}$. It follows that

$$\begin{aligned}
\operatorname{rank}_{(1,2)}(\mathcal{A}) &= \operatorname{rank}(\mathcal{A}(:,:,k_0)) \leq \operatorname{rank}\left(\mathbf{A}_{(1)}\right) = \operatorname{rank}_1(\mathcal{A}), \\
\operatorname{rank}_{(1,2)}(\mathcal{A}) &= \operatorname{rank}(\mathcal{A}(:,:,k_0)) = \operatorname{rank}(\mathcal{A}(:,:,k_0)^T) \leq \operatorname{rank}\left(\mathbf{A}_{(2)}\right) = \operatorname{rank}_2(\mathcal{A}).
\end{aligned}$$

■

Corollary 1. *Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}, \mathbf{M}^{(1)} \in \mathbb{C}^{J_1 \times I_1}, \mathbf{M}^{(2)} \in \mathbb{C}^{J_2 \times I_2}, \mathbf{M}^{(3)} \in \mathbb{C}^{J_3 \times I_3}$. Suppose $\mathcal{S} = \mathcal{A} \times_1 \mathbf{M}^{(1)} \times_2 \mathbf{M}^{(2)} \times_3 \mathbf{M}^{(3)}$. Then*

$$\operatorname{rank}_{(m,n)}(\mathcal{S}) \leq \min \{\operatorname{rank}_m(\mathcal{A}), \operatorname{rank}_n(\mathcal{A})\}, \quad 1 \leq m < n \leq 3.$$

Proof. It follows from Equations (2) and (5) that

$$\operatorname{rank}_{(m,n)}(\mathcal{S}) \leq \min \{\operatorname{rank}_m(\mathcal{S}), \operatorname{rank}_n(\mathcal{S})\} \leq \min \{\operatorname{rank}_m(\mathcal{A}), \operatorname{rank}_n(\mathcal{A})\}.$$

■

3 | CONNECTIONS BETWEEN HOSVD AND T-SVD

3.1 | Higher order singular value decomposition

The HOSVD is introduced in Reference 9, where the authors showed that it is a convincing generalization of the matrix SVD. For third-order tensors, the HOSVD has the following form.

Theorem 1. *[HOSVD 9] Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. There exist three unitary matrices $\mathbf{U}^{(1)} \in \mathbb{C}^{I_1 \times I_1}, \mathbf{U}^{(2)} \in \mathbb{C}^{I_2 \times I_2}, \mathbf{U}^{(3)} \in \mathbb{C}^{I_3 \times I_3}$, such that*

$$\mathcal{A} = \mathcal{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \tag{6}$$

where $\mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ satisfies

$$\mathcal{S}(i,j,k) = 0$$

if $i > \operatorname{rank}_1(\mathcal{A})$ or $j > \operatorname{rank}_2(\mathcal{A})$ or $k > \operatorname{rank}_3(\mathcal{A})$.

The unitary matrix $\mathbf{U}^{(n)}$ is called the n-mode singular matrix and can directly be found as the left singular matrix of $\mathbf{A}_{(n)}$, that is,

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} \mathbf{\Sigma}^{(n)} \mathbf{V}^{(n)^H}$$

is the SVD of $\mathbf{A}_{(n)}$. The tensor $\mathcal{S}$ is called the *core tensor* of the HOSVD.

3.2 | Singular value decomposition

The T-SVD of third-order tensors was first proposed in Reference 29, and then developed in References 12,13,18,19. The authors consider the tensor decomposition from a different perspective - by defining a novel tensor-tensor product. We introduce the updated definition presented in Reference 19, which covers the earliest definition in Reference 29 as a special case.

Let $\mathbf{a} \in \mathbb{C}^I, \mathbf{b} \in \mathbb{C}^I$. Denote by $\boldsymbol{a} \odot \boldsymbol{b}$ the Hadamard product of $\mathbf{a}, \mathbf{b}$, which is an element of $\mathbb{C}^I$ given by

$$(\boldsymbol{a} \odot \boldsymbol{b})(i) = a_i b_i, \quad i = 1, \dots, I.$$

Definition 5. [t-product of two vectors] Given an invertible matrix $\mathbf{L} \in \mathbb{C}^{I \times I}$, the t-product $\star_{\mathbf{L}} : \mathbb{C}^{I} \times \mathbb{C}^{I} \to \mathbb{C}^{I}$ is defined as

$$\boldsymbol{a} \star_{\mathbf{L}} \boldsymbol{b} = \mathbf{L}^{-1} \left((\mathbf{L}\boldsymbol{a}) \odot (\mathbf{L}\boldsymbol{b}) \right),$$

where $(\mathbf{L}\boldsymbol{a}) \odot (\mathbf{L}\boldsymbol{b})$ is the Hadamard product of $\mathbf{L}\boldsymbol{a}$ and $\mathbf{L}\boldsymbol{b}$.

When $\mathbf{L}$ is the identity matrix, $\star_{\mathbf{L}}$ is just the Hadamard product. When $\mathbf{L}$ is the discrete Fourier transform matrix, $\star_{\mathbf{L}}$ is the convolution.[29] In Reference 19, $\mathbf{L}$ is the discrete cosine transform (DCT) matrix. In Reference 22, $\mathbf{L}$ is wavelet transform matrix.

Definition 6. [t-product of two tensors] Let $\mathbf{L} \in \mathbb{C}^{I_3 \times I_3}$ be invertible. The t-product $\star_{\mathbf{L}} : \mathbb{C}^{I_1 \times J \times I_3} \times \mathbb{C}^{J \times I_2 \times I_3} \to \mathbb{C}^{I_1 \times I_2 \times I_3}$ between two tensors $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$(\mathcal{A} \star_{\mathbf{L}} \mathcal{B})(i, j, :) = \sum_{l=1}^{J} \mathcal{A}(i, l, :) \star_{\mathbf{L}} \mathcal{B}(l, j, :).$$

We can see the definition of the t-product of two tensors is just like the normal product of two matrices, except that the product of two scalars is replaced by the $\star_{\mathbf{L}}$ product of two tubes. There is a relationship between the t-product and the three-mode tensor-tensor product (Definition 1) given in Reference 19 as follows.

Proposition 2. *Let $\mathcal{A} \in \mathbb{C}^{I_1 \times J \times I_3}, \mathcal{B} \in \mathbb{C}^{J \times I_2 \times I_3}$. Then*

$$\mathcal{A} \star_{\mathbf{L}} \mathcal{B} = \left((\mathcal{A} \times_3 \mathbf{L}) *_3 (\mathcal{B} \times_3 \mathbf{L}) \right) \times_3 \mathbf{L}^{-1}.$$

By this proposition, when $\mathbf{L}$ is the identity matrix, $\star_{\mathbf{L}}$ is just $*_3$.

Theorem 2. *[T-SVD 19] Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ and $\mathbf{L} \in \mathbb{C}^{I_3 \times I_3}$ be invertible. There exist three tensors $\mathcal{U} \in \mathbb{C}^{I_1 \times I_1 \times I_3}$, $\mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times I_3}, \mathcal{V} \in \mathbb{C}^{I_2 \times I_2 \times I_3}$, such that*

$$\mathcal{A} = \mathcal{U} \star_{\mathbf{L}} \mathcal{S} \star_{\mathbf{L}} \mathcal{V}, \tag{7}$$

where $\mathcal{S}(:, :, k) \in \mathbb{C}^{I_1 \times I_2}$ is a diagonal matrix for $k = 1, \dots, I_3$.

The tensor $\mathcal{S}$ is called the *core tensor* of the T-SVD. The algorithm for the T-SVD[19] is given as Algorithm 1.

Algorithm 1. T-SVD induced by $\star_{\mathbf{L}}$[19]

Input: $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$.
Output: $\mathcal{U} \in \mathbb{C}^{I_1 \times I_1 \times I_3}, \mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times I_3}, \mathcal{V} \in \mathbb{C}^{I_2 \times I_2 \times I_3}$.

$\hat{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{L}$;
for $k = 1, \dots, I_3$ **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}^H] = \mathtt{svd}[\hat{\mathcal{A}}(:, :, k)]$;
  $\hat{\mathcal{U}}(:, :, k) = \mathbf{U}, \hat{\mathcal{S}}(:, :, k) = \mathbf{S}, \hat{\mathcal{V}}(:, :, k) = \mathbf{V}$;
end for
$\mathcal{U} = \hat{\mathcal{U}} \times_3 \mathbf{L}^{-1}, \mathcal{S} = \hat{\mathcal{S}} \times_3 \mathbf{L}^{-1}, \mathcal{V} = \hat{\mathcal{V}} \times_3 \mathbf{L}^{-1}$.

The motivation of introducing $\mathbf{L}$ is described as constructing an interaction among slices in References 19,29. The explanation is that the main purpose of introducing $\mathbf{L}$ is to obtain a decomposition with sparse core tensor. A question proposed in Reference 19 is how to choose a suitable transform matrix $\mathbf{L}$ for the T-SVD. We can answer this question after comparing HOSVD and T-SVD.

3.3 | The Relation

If we express HOSVD and T-SVD with tensor-tensor product (Definition 1), we can discover that these two decompositions are very similar. First, by Equation (1), we can rewrite HOSVD in Equation (6) as follows:

$$\mathcal{A} = \left(\operatorname{ten}\left(\mathbf{U}^{(1)}, I_3 \right) *_3 \mathcal{S} *_3 \operatorname{ten}\left({\mathbf{U}^{(2)}}^{T}, I_3 \right) \right) \times_3 \mathbf{U}^{(3)}. \tag{8}$$

Now we consider the T-SVD. From Algorithm 1, we can find the procedure of the T-SVD is very simple: We just need to apply a linear transform on every tube of $\mathcal{A}$ and solve the SVD of each frontal slice of the transformed tensor $\hat{\mathcal{A}}$. To express the T-SVD with a close form concerning $\star_{\mathbf{L}}$, all articles[12,13,19,29] apply the inverse transform on every tube of the decomposition tensors $\hat{\mathcal{U}}, \hat{\mathcal{S}}, \hat{\mathcal{V}}$, that is, the last step of Algorithm 1. Actually, this step is unnecessary. We can use the following form of T-SVD directly as follows:

$$\mathcal{A} = \left(\hat{\mathcal{U}} *_3 \hat{\mathcal{S}} *_3 \hat{\mathcal{V}}\right) \times_3 \mathbf{L}^{-1}. \tag{9}$$

By comparing Equations (8) and (9), we can discover the forms of the two decompositions are very similar and find the following differences:

1. In Equation (8), $\mathbf{U}^{(3)}$ is a special unitary matrix related to $\mathcal{A}$, while in Equation (9), $\mathbf{L}^{-1}$ can be any invertible matrix and $\mathbf{L}^{-1}$ is not necessary to be dependent on $\mathcal{A}$;
2. In Equation (8), ten $\left(\mathbf{U}^{(1)}, I_3\right)$ is an $(I_1 \times I_1 \times I_3)$-tensor with all frontal slices being the same unitary matrix, while in Equation (9), $\hat{\mathcal{U}}$ is an $(I_1 \times I_1 \times I_3)$-tensor with all frontal slices being different unitary matrices. The same case occurs for ten $\left(\mathbf{U}^{(2)}, I_3\right)$ and $\hat{\mathcal{V}}$;
3. In Equation (8), the entries of $\mathcal{S}$ except those in a $(\mathrm{rank}_1(\mathcal{A}) \times \mathrm{rank}_2(\mathcal{A}) \times \mathrm{rank}_3(\mathcal{A}))$-tensor are all zero, while in Equation (9), each frontal slice of $\hat{\mathcal{S}}$ is diagonal. In addition, $\mathcal{S}$ is not unique for the HOSVD, while $\hat{\mathcal{S}}$ is fixed for the T-SVD.

For the HOSVD in Equation (8), we use a common matrix $\mathbf{U}^{(1)}$ and a common matrix $\mathbf{U}^{(2)T}$ to change each frontal slice of $\mathcal{A} \times_3 \mathbf{U}^{(3)^H}$ into a corresponding frontal slice of $\mathcal{S}$. For the T-SVD in Equation (9), we use different matrices $\hat{\mathcal{U}}(:, :, k)$ and different matrices $\hat{\mathcal{V}}(:, :, k)$ to change each frontal slice of $\mathcal{A} \times_3 \mathbf{L}^{-1}$ into a corresponding frontal slice of $\hat{\mathcal{S}}$. This difference may result in $\hat{\mathcal{S}}$ being sparser than $\mathcal{S}$ in general.

Each frontal slice of $\hat{\mathcal{S}}$ being diagonal means that, for all tubes, only $\hat{\mathcal{S}}(i, i, :)$ can be nonzero. Actually, $\hat{\mathcal{S}}(i, i, :)$ can also be zero. Next, we give an estimate for the nonzero elements of $\hat{\mathcal{S}}$ in the T-SVD, which can be described by $\mathrm{rank}_{(1,2)}\left(\hat{\mathcal{S}}\right)$.

Theorem 3. *For the T-SVD Equation (9), we have*

$$\mathrm{rank}_{(1,2)}\left(\hat{\mathcal{S}}\right) = \mathrm{rank}_{(1,2)}(\mathcal{A} \times_3 \mathbf{L}) \le \min\left\{\mathrm{rank}_1(\mathcal{A}), \mathrm{rank}_2(\mathcal{A})\right\}.$$

Proof. It follows from Equation (4) and Corollary 1 that

$$\begin{aligned}\mathrm{rank}_{(1,2)}\left(\hat{\mathcal{S}}\right) &= \mathrm{rank}_{(1,2)}\left(\hat{\mathcal{U}} *_3 \hat{\mathcal{S}} *_3 \hat{\mathcal{V}}\right)\\ &= \mathrm{rank}_{(1,2)}(\mathcal{A} \times_3 \mathbf{L}) \le \min\left\{\mathrm{rank}_1(\mathcal{A}), \mathrm{rank}_2(\mathcal{A})\right\}.\end{aligned}$$

■

We remark that the value of $\mathrm{rank}_{(1,2)}\left(\hat{\mathcal{S}}\right)$ is called the *tubal rank* of $\mathcal{A}$ in many applications[21,24–26,30] with the T-SVD. By Theorem 3, we have, when $i > \min\left\{\mathrm{rank}_1(\mathcal{A}), \mathrm{rank}_2(\mathcal{A})\right\}$,

$$\hat{\mathcal{S}}(i, i, :) = 0.$$

As an illustration, we visualize the HOSVD with the multilinear rank $R_1 = \mathrm{rank}_1(\mathcal{A})$, $R_2 = \mathrm{rank}_2(\mathcal{A})$, and $R_3 = \mathrm{rank}_3(\mathcal{A})$ in Figure 1 and the T-SVD with tubal rank R in Figure 2, where the entries in the blank regions are zero; for $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}, \hat{\mathcal{U}}$, and $\hat{\mathcal{V}}$, the entries in the dark refer to singular vectors and the entries in the gray-value refer to vectors appending to singular vectors for the construction of a basis of the whole vector space.

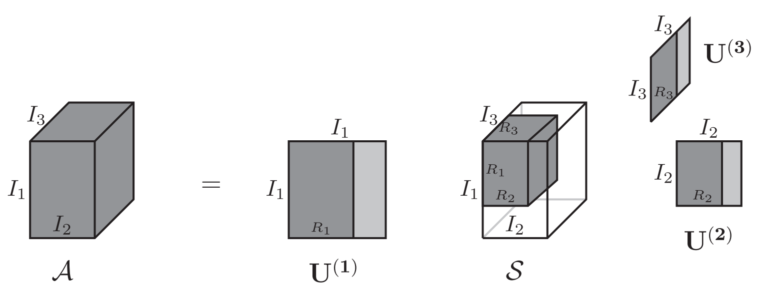

FIGURE 1 The HOSVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $R_1 = \mathrm{rank}_1(\mathcal{A}), R_2 = \mathrm{rank}_2(\mathcal{A}), R_3 = \mathrm{rank}_3(\mathcal{A})$. HSOVD, higher order singular value decomposition

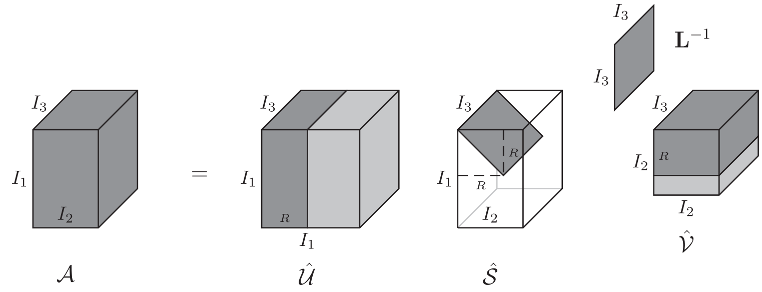

FIGURE 2 The T-SVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $R = \text{rank}_{(1,2)}(\mathcal{A} \times_3 \mathbf{L}) \leq \min\{\text{rank}_1(\mathcal{A}), \text{rank}_2(\mathcal{A})\}$. SVD, singular value decomposition

4 | ORIENTED SINGULAR VALUE DECOMPOSITION

In general, the core tensor $\mathcal{S}$ in Equation (6) is not very sparse, not necessarily nonnegative, and can even be complex when $\mathcal{A}$ is complex. These defects limit the applications of the HOSVD. In addition, the HOSVD is not orientation dependent. In applications, as already mentioned in the introduction, we usually confront orientation-dependent data which have high correlation among frontal slices. By contrast, the T-SVD is orientation dependent. However, $\mathbf{L}$ is independent of $\mathcal{A}$. Hence, T-SVD cannot embody the data feature and is not suitable for all orientation-dependent data.

We consider a new decomposition for third-order tensors, combining the ideas of HOSVD and T-SVD. Specifically, we have the following result.

Theorem 4. *Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. There exist a unitary matrix $\mathbf{U}^{(3)} \in \mathbb{C}^{I_3 \times I_3}$, three tensors $\mathcal{U}^{(3)} \in \mathbb{C}^{I_1 \times I_1 \times I_3}$, $\mathcal{S}^{(3)} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, $\mathcal{V}^{(3)} \in \mathbb{C}^{I_2 \times I_2 \times I_3}$, such that*

$$\mathcal{A} = \left(\mathcal{U}^{(3)} *_3 \mathcal{S}^{(3)} *_3 \mathcal{V}^{(3)}\right) \times_3 \mathbf{U}^{(3)}, \tag{10}$$

where

1. *$\mathcal{U}^{(3)}(:,:,k), \mathcal{V}^{(3)}(:,:,k)$ are unitary for $k = 1, \ldots, \text{rank}_3(\mathcal{A})$, and $\mathcal{U}^{(3)}(:,:,k) = \mathcal{V}^{(3)}(:,:,k) = 0$ for $k = \text{rank}_3(\mathcal{A}) + 1, \ldots, I_3$;*
2. *$\mathcal{S}^{(3)}(:,:,k) \in \mathbb{R}^{I_1 \times I_2}$ is a nonnegative diagonal matrix for $k = 1, \ldots, \text{rank}_3(\mathcal{A})$, $\mathcal{S}^{(3)}(:,:,k) = 0$ for $k = \text{rank}_3(\mathcal{A}) + 1, \ldots, I_3$, and have the following properties:*
   (a) *ordering:*
   $$\|\mathcal{S}^{(3)}(:,:,1)\|_F \geq \|\mathcal{S}^{(3)}(:,:,2)\|_F \geq \cdots \geq \|\mathcal{S}^{(3)}(:,:,I_3)\|_F \geq 0,$$
   where $\|\mathcal{S}^{(3)}(:,:,k)\|_F = \sigma_k^{(3)}$, which is the kth singular value of $\mathbf{A}_{(3)}$; and
   $$\sigma_k^{(3)} \geq \mathcal{S}^{(3)}(1,1,k) \geq \mathcal{S}^{(3)}(2,2,k) \geq \cdots \geq \mathcal{S}^{(3)}(I,I,k) \geq 0$$
   for $k = 1, \ldots, I_3$, where $I = \min\{I_1, I_2\}$.
   (b) *rank property:*
   $$\text{rank}_{(1,2)}\left(\mathcal{S}^{(3)}\right) = \text{rank}_{(1,2)}\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right) \leq \min\{\text{rank}_1(\mathcal{A}), \text{rank}_2(\mathcal{A})\}. \tag{11}$$

Proof. Suppose the SVD of $\mathbf{A}_{(3)} \in \mathbb{C}^{I_3 \times I_1 I_2}$ is

$$\mathbf{A}_{(3)} = \mathbf{U}^{(3)} \mathbf{\Sigma}^{(3)} \mathbf{V}^{(3)},$$

where $\mathbf{U}^{(3)} \in \mathbb{C}^{I_3 \times I_3}, \mathbf{V}^{(3)} \in \mathbb{C}^{I_1 I_2 \times I_1 I_2}$ are unitary, and $\mathbf{\Sigma}^{(3)} = \text{diag}(\sigma_1^{(3)}, \ldots, \sigma_{R_3}^{(3)}, 0, \ldots, 0)$, with $R_3 = \text{rank}_3(\mathcal{A}) = \text{rank}(\mathbf{A}_{(3)})$. Then $\mathbf{U}^{(3)^H} \mathbf{A}_{(3)} = \mathbf{\Sigma}^{(3)} \mathbf{V}^{(3)}$ satisfies

$$\left\|\left(\mathbf{U}^{(3)^H} \mathbf{A}_{(3)}\right)(k,:)\right\| = \begin{cases} \sigma_k^{(3)}, & k = 1, \ldots, R_3, \\ 0, & k = R_3 + 1, \ldots, I_3. \end{cases}$$

Since $\left\|\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k)\right\| = \left\|\left(\mathbf{U}^{(3)^H} \mathbf{A}_{(3)}\right)(k,:)\right\|$ for $k = 1, \ldots, I_3$, applying the SVD on $\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,1), \ldots, \left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,R_3)$ leads to the desired form Equation (10).

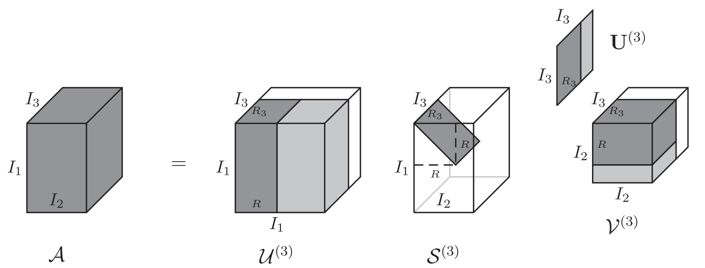

FIGURE 3 The O-SVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, where $R_3 = \text{rank}_3(\mathcal{A})$, $R = \text{rank}_{(1,2)}\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right) \leq \min\{\text{rank}_1(\mathcal{A}), \text{rank}_2(\mathcal{A})\}$. SVD, singular value decomposition

The verifying of the property of ordering in 2.(a) is direct. For the rank property in 2.(b), by the definition, we also have $\text{rank}_{(1,2)}\left(\mathcal{S}^{(3)}\right) = \text{rank}_{(1,2)}\left(\mathcal{U}^{(3)} *_3 \mathcal{S}^{(3)} *_3 \mathcal{V}^{(3)}\right)$. Therefore, like Theorem 3, Equation (11) is obtained. ■

We call the form Equation (10) the *oriented singular value decomposition* (O-SVD) of $\mathcal{A}$, which is visualized in Figure 3. In the figure, the entries in the blank regions are zero; for $\mathbf{U}^{(3)}$, $\mathcal{U}^{(3)}$, and $\mathcal{V}^{(3)}$, the entries in the dark refer to singular vectors and the entries in the gray-value refer to vectors appending to singular vectors for the construction of a basis of the whole vector space. Based on the proof of the theorem, the algorithm for the O-SVD is given in Algorithm 2.

Algorithm 2. O-SVD

Input: $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$.
Output: $\mathbf{U}^{(3)} \in \mathbb{C}^{I_3 \times I_3}$, $\mathcal{U}^{(3)} \in \mathbb{C}^{I_1 \times I_1 \times I_3}$, $\mathcal{S}^{(3)} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$, $\mathcal{V}^{(3)} \in \mathbb{C}^{I_2 \times I_2 \times I_3}$.

$[\mathbf{U}^{(3)}, \mathbf{\Sigma}^{(3)}, \mathbf{V}^{(3)}] = \texttt{svd}[\mathbf{A}_{(3)}]$;
$R_3 = \text{rank}(\mathbf{\Sigma}^{(3)})$;
$\hat{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{U}^{(3)^H}$;
for $k = 1, \dots, R_3$ **do**
  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \texttt{svd}[\hat{\mathcal{A}}(:, :, k)]$;
  $\mathcal{U}^{(3)}(:, :, k) = \mathbf{U}, \mathcal{S}^{(3)}(:, :, k) = \mathbf{S}, \mathcal{V}^{(3)}(:, :, k) = \mathbf{V}^H$;
end for
for $k = R_3 + 1, \dots, I_3$ **do**
  $\mathcal{U}^{(3)}(:, :, k) = 0, \mathcal{S}^{(3)}(:, :, k) = 0, \mathcal{V}^{(3)}(:, :, k) = 0$;
end for

In practical application, we can change $\mathbf{U}^{(3)}$ into $\mathbf{M}^{(3)}$, which consists of the $\text{rank}_3(\mathcal{A})$ leading singular vectors of $\mathbf{A}_{(3)}$, and obtain an economy-size decomposition:

$$\mathcal{A} = \left(\widetilde{\mathcal{U}}^{(3)} *_3 \widetilde{\mathcal{S}}^{(3)} *_3 \widetilde{\mathcal{V}}^{(3)}\right) \times_3 \mathbf{M}^{(3)},$$

where $\mathbf{M}^{(3)} \in \mathbb{C}^{I_3 \times R_3}$, $\widetilde{\mathcal{U}}^{(3)} \in \mathbb{C}^{I_1 \times I_1 \times R_3}$, $\widetilde{\mathcal{S}}^{(3)} \in \mathbb{R}^{I_1 \times I_2 \times R_3}$, $\widetilde{\mathcal{V}}^{(3)} \in \mathbb{C}^{I_2 \times I_2 \times R_3}$ with $R_3 = \text{rank}_3(\mathcal{A})$.

For an orientation-dependent tensor $\mathcal{A}$, usually the numerical rank of $\mathbf{A}_{(3)}$ is much smaller than I_3. We can utilize the CUR decomposition[31,32] or randomized methods[33] to find $\mathbf{U}^{(3)}$ efficiently.

4.1 | Properties

Now we study some properties of the O-SVD. The following lemma is needed.

Lemma 1. *Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. Then*

$$\text{span}\{\mathcal{A}(i,j,:) : 1 \leq i \leq I_1, 1 \leq j \leq I_2\} \simeq \text{span}\{\mathcal{A}(:,:,k) : 1 \leq k \leq I_3\};$$
$$\text{span}\{\mathcal{A}(i,:,k) : 1 \leq i \leq I_1, 1 \leq k \leq I_3\} \simeq \text{span}\{\mathcal{A}(:,j,:) : 1 \leq j \leq I_2\};$$
$$\text{span}\{\mathcal{A}(:,j,k) : 1 \leq j \leq I_2, 1 \leq k \leq I_3\} \simeq \text{span}\{\mathcal{A}(i,:,:) : 1 \leq i \leq I_1\}.$$

Proof. We only prove the first relationship, and the others can be proved similarly.

The space span $\{\mathcal{A}(i,j,:) : 1 \le i \le I_1, 1 \le j \le I_2\}$ is a subspace of $\mathbb{R}^{I_1 \times I_2}$. Let $E_{ij} \in \mathbb{R}^{I_1 \times I_2}$ be the matrix with element (i,j) being 1 and the other elements being zero. Then $\{E_{ij} : 1 \le i \le I_1, 1 \le j \le I_2\}$ is a basis of $\mathbb{R}^{I_1 \times I_2}$. It follows that

$$
\begin{aligned}
&(\mathcal{A}(:,:,1), \dots, \mathcal{A}(:,:,I_3)) \\
&\quad = \left(E_{11}, E_{21}, \dots, E_{I_1 1}, E_{12}, \dots, E_{I_1 2}, \dots, E_{1 I_2}, \dots, E_{I_1 I_2}\right) \mathbf{P},
\end{aligned}
$$

where $\mathbf{P} \in \mathbb{R}^{I_1 I_2 \times I_3}$. We can check that the columns of $\mathbf{P}^T$ are just all tube fibers of $\mathcal{A}$, which completes the proof. ∎

Property 1. Let $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$. Suppose $\mathbf{U}^{(3)}$ is the three-mode singular matrix of $\mathcal{A}$. Then, for any invertible matrix $\mathbf{L}$,

$$
\#\left\{k : \left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k) \neq 0\right\} \le \#\{k : (\mathcal{A} \times_3 \mathbf{L})(:,:,k) \neq 0\},
$$

and $\left\{\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k) : 1 \le k \le \mathrm{rank}_3(\mathcal{A})\right\}$ is a basis of $\mathrm{span}\{\mathcal{A}(:,:,k) : 1 \le k \le I_3\}$.

Proof. For any invertible matrix $\mathbf{L}$, it follows from Lemmas 1 and (3) that

$$
\dim\left(\mathrm{span}\{(\mathcal{A} \times_3 \mathbf{L})(:,:,k) : 1 \le k \le I_3\}\right) = \mathrm{rank}_3(\mathcal{A} \times_3 \mathbf{L}) = \mathrm{rank}_3(\mathcal{A}). \tag{12}
$$

Therefore,

$$
\#\{k : (\mathcal{A} \times_3 \mathbf{L})(:,:,k) \neq 0\} \ge \mathrm{rank}_3(\mathcal{A}) = \#\left\{k : \left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k) \neq 0\right\} \tag{13}
$$

for any invertible matrix $\mathbf{L}$.

For any matrix $\mathbf{M} \in \mathbb{R}^{J \times I_3}$, a frontal slice of $\mathcal{A} \times_3 \mathbf{M}$ is just a linear combination of some frontal slices of $\mathcal{A}$. Thus,

$$
\mathrm{span}\{(\mathcal{A} \times_3 \mathbf{M})(:,:,k) : 1 \le k \le J\} \subseteq \mathrm{span}\{\mathcal{A}(:,:,k) : 1 \le k \le I_3\}.
$$

It follows from Equations (12) and (13) that

$$
\mathrm{span}\{(\mathcal{A} \times_3 \mathbf{L})(:,:,k) : 1 \le k \le I_3\} = \mathrm{span}\{\mathcal{A}(:,:,k) : 1 \le k \le I_3\}
$$

and $\left\{\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k) : 1 \le k \le \mathrm{rank}_3(\mathcal{A})\right\}$ is a basis of $\mathrm{span}\{\mathcal{A}(:,:,k) : 1 \le k \le I_3\}$. ∎

Property 1 shows that there are two steps in the computational procedure of the O-SVD. The first step is to find a basis of the space spanned by all frontal slices. The second step is to apply the SVD on each element of the basis. This procedure benefits the sparsity of the core tensor, which can be discovered by the visualizations of the three decompositions in Figures 1 to 3. A sparser core tensor means we can express the original tensor with the sum of fewer outer product terms. Denote $R_1 = \mathrm{rank}_1(\mathcal{A}), R_2 = \mathrm{rank}_2(\mathcal{A}), R_3 = \mathrm{rank}_3(\mathcal{A}), \hat{R} = \mathrm{rank}_{(1,2)}\left(\hat{\mathcal{S}}\right), R = \mathrm{rank}_{(1,2)}\left(\mathcal{S}^{(3)}\right)$. We list the outer product forms of the three decompositions here:

$$
\text{HOSVD} \qquad \mathcal{A} = \sum_{i=1}^{R_1} \sum_{j=1}^{R_2} \sum_{k=1}^{R_3} s_{ijk}\, \mathbf{U}^{(1)}(:,i) \circ \mathbf{U}^{(2)}(:,j) \circ \mathbf{U}^{(3)}(:,k); \tag{14}
$$

$$
\text{T-SVD} \qquad \mathcal{A} = \sum_{i=1}^{\hat{R}} \sum_{k=1}^{I_3} \hat{s}_{iik}\, \hat{\mathcal{U}}(:,i,k) \circ \hat{\mathcal{V}}(i,:,k) \circ \mathbf{L}^{-1}(:,k); \tag{15}
$$

$$
\text{O-SVD} \qquad \mathcal{A} = \sum_{i=1}^{R} \sum_{k=1}^{R_3} s_{iik}^{(3)}\, \mathcal{U}^{(3)}(:,i,k) \circ \mathcal{V}^{(3)}(i,:,k) \circ \mathbf{U}^{(3)}(:,k). \tag{16}
$$

Definition 7. Let a core tensor corresponding to the O-SVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ be $\mathcal{S}^{(3)}$ and denote $I = \min\{I_1, I_2\}$. Define the singular values of the pair $\left(\mathcal{A}, \mathcal{S}^{(3)}\right)$ as

$$
s_{iik}^{(3)}, \quad i = 1, \dots, I, k = 1, \dots, I_3.
$$

We write the singular values as $\sigma_j\left(\mathcal{A}, \mathcal{S}^{(3)}\right)$ in a nonincreasing ordering:

$$\sigma_1\left(\mathcal{A}, \mathcal{S}^{(3)}\right) \geq \sigma_2\left(\mathcal{A}, \mathcal{S}^{(3)}\right) \geq \cdots \geq \sigma_{II_3}\left(\mathcal{A}, \mathcal{S}^{(3)}\right).$$

Property 2. [Norm] Let a core tensor corresponding to the O-SVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ be $\mathcal{S}^{(3)}$ and denote $I = \min\{I_1, I_2\}$. Then the following property holds:

$$\|\mathcal{A}\|^2 = \|\mathcal{S}^{(3)}\|^2 = \sum_{i=1}^{II_3} \sigma_i^2\left(\mathcal{A}, \mathcal{S}^{(3)}\right).$$

Proof. Since the Frobenius norm is unitarily invariant, by Equation (10), we have

$$\begin{aligned}
\|\mathcal{A}\|^2 &= \left\|\mathcal{A} \times_3 \mathbf{U}^{(3)H}\right\|^2 = \left\|\mathcal{U}^{(3)} *_3 \mathcal{S}^{(3)} *_3 \mathcal{V}^{(3)}\right\|^2 \\
&= \sum_{k=1}^{\operatorname{rank}_3(\mathcal{A})} \left\|\mathcal{U}^{(3)}(:,:,k)\mathcal{S}^{(3)}(:,:,k)\mathcal{V}^{(3)}(:,:,k)\right\|^2 \\
&= \sum_{k=1}^{\operatorname{rank}_3(\mathcal{A})} \left\|\mathcal{S}^{(3)}(:,:,k)\right\|^2 = \|\mathcal{S}^{(3)}\|^2 = \sum_{i=1}^{II_3} \sigma_i^2\left(\mathcal{A}, \mathcal{S}^{(3)}\right).
\end{aligned}$$

■

Next we discuss the approximation property of O-SVD by considering a tensor obtained by keeping r terms for which s_{ijk} are largest in magnitude and discarding the other terms. We call this the *r terms approximation*.

Property 3. [Approximation] Let a core tensor corresponding to the O-SVD of $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ be $\mathcal{S}^{(3)}$ and denote $I = \min\{I_1, I_2\}$. Denote by $\mathcal{A}_r$ the r terms approximation by the O-SVD. Then we have

$$\|\mathcal{A} - \mathcal{A}_r\|^2 = \sum_{i=r+1}^{II_3} \sigma_i^2\left(\mathcal{A}, \mathcal{S}^{(3)}\right).$$

Proof. For $\mathcal{A}_r$, denote by $\mathcal{S}_r^{(3)}$ the tensor obtained by setting the corresponding entries of $\mathcal{S}^{(3)}$ equal to zero. We have

$$\|\mathcal{A} - \mathcal{A}_r\|^2 = \|\mathcal{S}^{(3)} - \mathcal{S}_r^{(3)}\|^2 = \sum_{i=r+1}^{II_3} \sigma_i^2\left(\mathcal{A}, \mathcal{S}^{(3)}\right).$$

■

5 | NUMERICAL EXAMPLES

We consider data compression based on the HOSVD, T-SVD, and O-SVD decompositions in Equations (14) to (16), respectively. The compression strategies are based on truncated decompositions of tensors. Compression based on HOSVD has been discussed in detail in Reference 29, section 2.2. There are two strategies. The first way is to employ an alternating least squares algorithm to obtain the truncated HOSVD[16] for approximation. Because of the use of alternating least squares algorithm, the computational cost is involved and the algorithm can be sensitive to initial guess. The second way is based on the strategy in Reference 34 by keeping the terms for which s_{ijk} are largest in magnitudes and discarding the other terms. We employ this strategy by using the r terms approximation as a substitute of the original tensor to compress data. More precisely, we define

$$r \text{ terms approximation of HOSVD: } \mathcal{A} = \sum_{\substack{(i,j,k) \text{ for the } r \\ \text{largest } |s_{ijk}|}} s_{ijk}\, \mathbf{U}^{(1)}(:,i) \circ \mathbf{U}^{(2)}(:,j) \circ \mathbf{U}^{(3)}(:,k);$$

$$r \text{ terms approximation of T-SVD: } \mathcal{A} = \sum_{\substack{(i,i,k) \text{ for the } r \\ \text{largest } |\hat{s}_{iik}|}} \hat{s}_{iik}\, \hat{\mathcal{U}}(:,i,k) \circ \hat{\mathcal{V}}(i,:,k) \circ \mathbf{L}^{-1}(:,k);$$

$$r \text{ terms approximation of O-SVD: } \mathcal{A} = \sum_{\substack{(i,i,k) \text{ for the } r \\ \text{largest } |s_{iik}^{(3)}|}} s_{iik}^{(3)} \, \mathcal{U}^{(3)}(:,i,k) \circ \mathcal{V}^{(3)}(i,:,k) \circ \mathbf{U}^{(3)}(:,k).$$

For the r terms approximation of HOSVD, we denote by $(i_1, j_1, k_1), \dots, (i_r, j_r, k_r)$ the coordinates of s_{ijk} that will be kept. Suppose $i_{\max} = \max\{i_1, \dots, i_r\}$, $j_{\max} = \max\{j_1, \dots, j_r\}$, $k_{\max} = \max\{k_1, \dots, k_r\}$. We need to store the values of $s_{i_1 j_1 k_1}, \dots, s_{i_r j_r k_r}$, the coordinates $(i_1, j_1, k_1), \dots, (i_r, j_r, k_r)$, the first $i_{\max}$ (respectively, $j_{\max}$ and $k_{\max}$) columns of $\mathbf{U}^{(1)}$ (respectively, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$). For convenience, we regard the storage size of (i, j, k) as that of s_{ijk}. Thus, the whole storage size is $2r + i_{\max} I_1 + j_{\max} I_2 + k_{\max} I_3$, and the compression ratio is

$$\text{Ratio(H)} = \frac{I_1 I_2 I_3}{2r + i_{\max} I_1 + j_{\max} I_2 + k_{\max} I_3}. \tag{17}$$

For O-SVD, we note that the r terms approximation of O-SVD yields the approximation of the basis of the frontal slices space, see Property 1. To be specific, for a fixed k, we denote by $(1, 1, k), \dots, (i_k, i_k, k)$ the coordinates of $s_{iik}^{(3)}$ that will be kept for the r terms approximation. Then $\sum_{i=1}^{i_k} s_{iik}^{(3)} \, \mathcal{U}^{(3)}(:,i,k) \circ \mathcal{V}^{(3)}(i,:,k)$ is also an approximation of the slice $\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k)$. Therefore, like the matrix compression using the SVD, if $i_k > \frac{I_1 I_2}{I_1 + I_2 + 1}$, we just store $\left(\mathcal{A} \times_3 \mathbf{U}^{(3)^H}\right)(:,:,k)$. For all $s_{iik}^{(3)}$ that will be kept, denote by $k_{\max}$ the largest k that occurs. Let $K = \#\{k : i_k > \frac{I_1 I_2}{I_1 + I_2 + 1}\}$ and $\hat{r} = \sum_{i_k \le \frac{I_1 I_2}{I_1 + I_2 + 1}} i_k$. The whole storage size is $\hat{r}(I_1 + I_2 + 1) + K I_1 I_2 + k_{\max} I_3$ and the compression ratio is

$$\text{Ratio(O)} = \frac{I_1 I_2 I_3}{\hat{r}(I_1 + I_2 + 1) + K I_1 I_2 + k_{\max} I_3}. \tag{18}$$

The case for T-SVD Equation (15) is very similar to that for O-SVD. The only difference is that we do not need to store $\mathbf{L}$. We use the same notation for the O-SVD, and the compression ratio is

$$\text{Ratio(T)} = \frac{I_1 I_2 I_3}{\hat{r}(I_1 + I_2 + 1) + K I_1 I_2}. \tag{19}$$

Next we will test the approximation ability of the r terms approximation and data compression based on the three decompositions, and demonstrate the effectiveness of O-SVD. The relative error of an approximation tensor $\widetilde{\mathcal{A}}$ is defined as

$$\text{Rel} = \frac{\|\mathcal{A} - \widetilde{\mathcal{A}}\|}{\|\mathcal{A}\|}.$$

In all numerical examples, the DCT is employed for the T-SVD. The experiments for the HOSVD and the programs for the T-SVD and the O-SVD are based on Matlab Tensor Toolbox, version 3.0.[35]

5.1 | Hyperspectral Image

In this subsection, we test a hyperspectral image—the Samson.[36] The dataset is a third-order tensor (length × width × channels). In each image of the dataset, a region of 95 × 95 pixels is utilized, where each pixel is recorded at 156 frequency channels covering the wavelengths from 401 to 889 nm. Then the spectral resolution is highly up to 3.13 nm. Hence, the size of the resulting tensor is 95 × 95 × 156. We show 20th, 80th, and 140th bands of the image in the last column of Figure 5. The values on the same location of the bands change gradually as frequency channels change. Denote by $\mathcal{A}$ the tensor of the testing data. Under the relative error tolerance 0.005, (Matlab command: `hosvd(`$\mathcal{A}$`,0.005)`), the size of the nonzero part of the core tensor is 90 × 79 × 31. Hence, this is a well-oriented tensor. The relative errors of the r terms approximation are shown in Table 1. The advantage of the O-SVD over the other two decompositions is very significant, and the performance of the T-SVD is much better than that of the HOSVD when $r \ge 100$.

The values $10\log_{10}(\text{Rel})$ versus compression ratio are shown in Figure 4. The performance of the O-SVD is the best and the HOSVD outperforms the T-SVD. Figure 5 shows the compression results under compression ratio 70 for the three decompositions. From the visual comparison, especially the 20th and the 80th bands, we can see the performance of the O-SVD is the best, and the HOSVD outperforms the T-SVD.

TABLE 1 Relative errors of the r terms approximation for the hyperspectral image

r	HOSVD	T-SVD	O-SVD
1	0.3607	0.6471	0.3605
10	0.2126	0.2210	0.1316
100	0.1045	0.0537	0.0216
200	0.0820	0.0324	0.0110
500	0.0554	0.0144	0.0060
1000	0.0406	0.0086	0.0042

Abbreviations: HSOVD, higher order singular value decomposition; SVD, singular value decomposition.

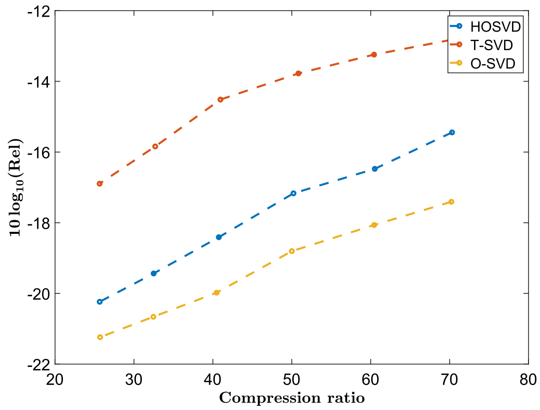

FIGURE 4 The values $10\log_{10}(\text{Rel})$ versus compression ratio for the three decompositions on the hyperspectral image

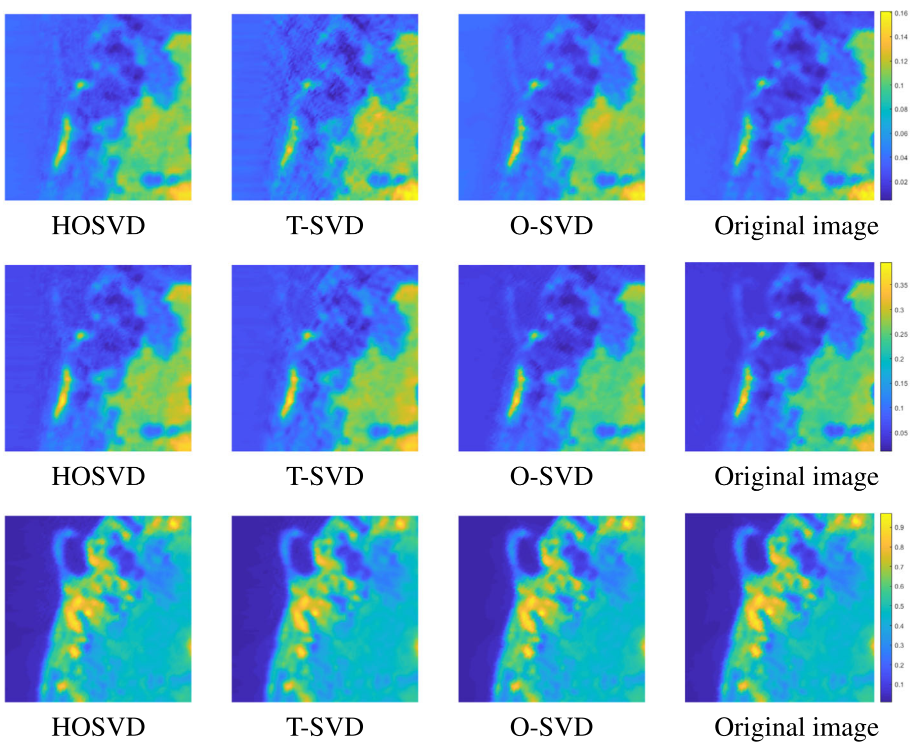

FIGURE 5 Compression results under compression ratio 70. From top to bottom: 20th band, 80th band, and 140th band. From left to right: compression by the HOSVD, Rel= 0.0285; compression by the T-SVD, Rel= 0.0522; compression by the O-SVD, Rel= 0.0182; original images. HSOVD, higher order singular value decomposition; SVD, singular value decomposition

TABLE 2 Relative errors of the r terms approximation for the video.

r	HOSVD	T-SVD	O-SVD
1	0.3805	0.3805	0.3805
10	0.2671	0.2508	0.2507
100	0.1647	0.0876	0.0856
200	0.1416	0.0504	0.0488
500	0.1156	0.0375	0.0353
1000	0.0984	0.0287	0.0255
2000	0.0844	0.0204	0.0167

Abbreviations: HSOVD, higher order singular value decomposition; SVD, singular value decomposition.

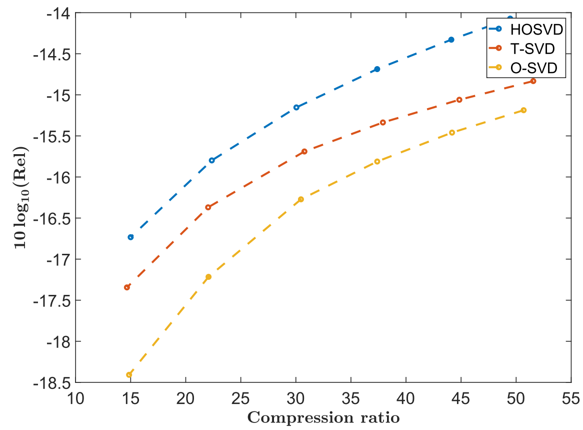

FIGURE 6 The values $10\log_{10}$(Rel) versus compression ratio for the three decompositions on the video

5.2 | Video

The second example is a video. The dataset is a surveillance video from YouTube, with size $250 \times 490 \times 200$ (height $\times$ width $\times$ frames). See Figure 7A. Most regions of the frames are stable, while the car in the blue box is moving. Under the relative error tolerance 0.01, the size of the nonzero part of the core tensor is $236 \times 422 \times 48$, which is a well-oriented tensor.

The relative errors of the r terms approximation are shown in Table 2. The O-SVD outperforms the T-SVD, and the performance of the T-SVD is much better than that of the HOSVD when $r \geq 100$. The values $10\log_{10}$(Rel) versus compression ratio are shown in Figure 6. The performance of the O-SVD is the best, and the T-SVD has a better performance over the HOSVD. Figure 7 shows the compression results with zoomed regions under compression ratio 15 for the three decompositions. All three decompositions perform well on most regions of the frame. However, for the zoomed region, the result of the O-SVD is much cleaner than the other two results. The result of the T-SVD has obvious trailing smear.

5.3 | Summary of experiment results

As we know, the DCT is very successful in two-dimensional (2D) image compression.[37] The core technique in the JPEG compression algorithm is using the DCT. This transform can convert a continuous signal into a sparse vector in the frequency domain, whose nonzero elements are mostly located on low frequency domain. The hyperspectral image and the surveillance video are continuous on the third direction and many frontal slices of $\mathcal{A} \times_3 \mathbf{L}$ are zero under a given tolerance. However, the DCT cannot be better than $\mathbf{U}^{(3)}$ in sparsity. See Property 1. Usually, we do not use the SVD for 2D image compression because we need to store the singular matrix. However, compared with the huge size of a third-order tensor, the storage of one matrix can be ignored. Compare Equations (18) and (19). Therefore, the O-SVD is better than the T-SVD in oriented tensor compression.

Comparing Equations (17) and (18) shows that, when the number of the terms used for the approximation increases, Ratio(O) decreases much faster than Ratio(H). The advantage of O-SVD over HOSVD in compression benefits from the

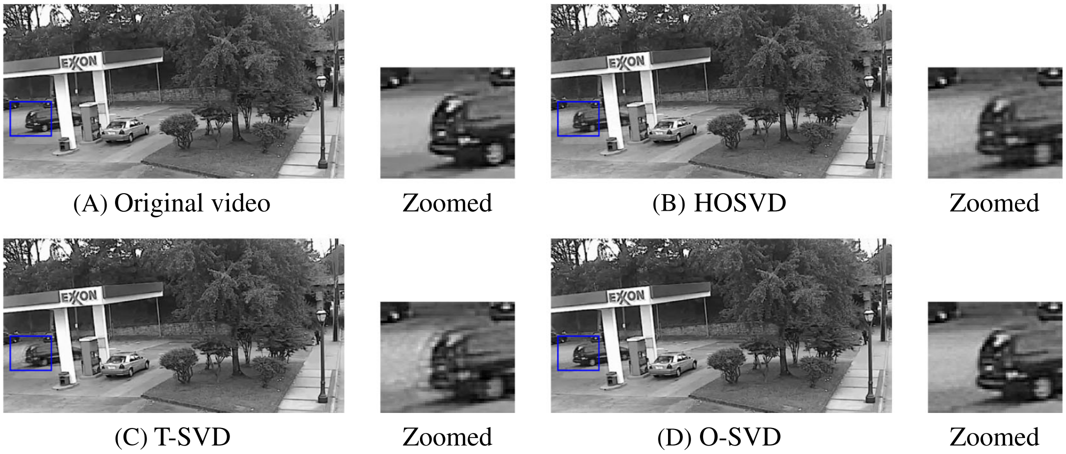

FIGURE 7 Compression results under compression ratio 15. (A) Original video; (B) compression by the HOSVD: Rel= 0.0212; (C) compression by the T-SVD: Rel= 0.0184; (D) compression by the O-SVD: Rel= 0.0144. HSOVD, higher order singular value decomposition; SVD, singular value decomposition

much better approximation ability of the O-SVD. See Tables 1 and 2. For oriented tensors, we can use the sum of few outer terms to achieve a good approximation. Hence, the performance of the O-SVD is better than that of the HOSVD in oriented tensor compression.

6 | CONCLUDING REMARKS

By defining some novel definitions such as slice rank, tensor-tensor product, we establish a connection between the HOSVD and the T-SVD for third-order tensors. After being converted into a new form and expressed as a linear combination of outer product terms, the T-SVD can be comprehended easily. The sparsity of the core tensors are compared for these two decompositions with the help of slice rank. Inspired by HOSVD and T-SVD, we propose O-SVD. This decomposition provides a new idea to decompose a tensor: Find a basis of the space spanned by all frontal slices and then apply SVD on each element of the basis. The new decomposition has some nice properties, including sparsity of the core tensor and norm invariance. For application, we consider data compression. The experiments illustrate the powerful ability for approximation and data compression of the new decomposition.

In the future research work, we would like to extend the idea to higher order tensors and study the tensor completion problem by using the slice rank minimization in the formulation. Moreover, we can consider and study the CUR decomposition[31,32] or randomized methods[33] to find unitary matrices efficiently in computing O-SVD.

ACKNOWLEDGEMENTS
We are very grateful to the two anonymous referees for their invaluable suggestions, which helped to improve this article from its original version. This work was supported in part by HKRGC GRF 12306616, 12200317, 12300218, 12300519, and HKU 104005583.

CONFLICTS OF INTEREST
This work does not have any conflict of interest.

ORCID
Chao Zeng https://orcid.org/0000-0002-0260-5682

REFERENCES
1. Smilde A, Bro R, Geladi P. Multi-way analysis: Applications in the chemical sciences. Hoboken, NJ: John Wiley & Sons, 2005.
2. Cichocki A, Mandic D, De Lathauwer L, et al. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. IEEE Signal Process Mag. 2015;32(2):145–163.
3. Liu J, Musialski P, Wonka P, Ye J. Tensor completion for estimating missing values in visual data. IEEE Trans Pattern Anal Mach Intell. 2013;35(1):208–220.

4. Carroll JD, Chang J-J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika. 1970;35(3):283–319.
5. Mørup M. Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscip Rev Data Min Knowl Discov. 2011;1(1):24–40.
6. Signoretto M, Dinh QT, De Lathauwer L, Suykens JA. Learning with tensors: A framework based on convex optimization and spectral regularization. Mach Learn. 2014;94(3):303–351.
7. Harshman RA. Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-mode factor analysis. *UCLA Working Papers in phonetics*. 1970;16:1-84
8. Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;31(3):279–311.
9. De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl. 2000;21(4):1253–1278.
10. Hackbusch W, Kühn S. A new scheme for the tensor representation. J Fourier Anal Appl. 2009;15(5):706–722.
11. Grasedyck L. Hierarchical singular value decomposition of tensors. SIAM J Matrix Anal Appl. 2010;31(4):2029–2054.
12. Kilmer ME, Martin CD. Factorization strategies for third-order tensors. Linear Algebra Appl. 2011;435(3):641–658.
13. Kilmer ME, Braman K, Hao N, Hoover RC. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM J Matrix Anal Appl. 2013;34(1):148–172.
14. Martin CD, Shafer R, LaRue B. An order-*p* tensor factorization with applications in imaging. SIAM J Scientific Comput. 2013;35(1):A474–A490.
15. Oseledets IV. Tensor-train decomposition. SIAM J Scientific Comput. 2011;33(5):2295–2317.
16. Kolda TG, Bader BW. Tensor decompositions and applications. SIAM Rev. 2009;51(3):455–500.
17. Cichocki A, Lee N, Oseledets I, et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. Found Trends Mach Learn. 2016;9(4-5):249–429.
18. Braman K. Third-order tensors as linear operators on a space of matrices. Linear Algebra Appl. 2010;433(7):1241–1253.
19. Kernfeld E, Kilmer M, Aeron S. Tensor–tensor products with invertible linear transforms. Linear Algebra Appl. 2015;485:545–570.
20. Semerci O, Hao N, Kilmer ME, Miller EL. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. IEEE Trans Image Process. 2014;23(4):1678–1693.
21. Lu C, Feng J, Chen Y, Liu W, Lin Z, Yan S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016:5249-5257.
22. Song GJ, Ng MK, Zhang X. Robust tensor completion using transformed tensor SVD. Numer Linear Algebra Appl (to appear).
23. Hao N, Kilmer ME, Braman K, Hoover RC. Facial recognition using tensor-tensor decompositions. SIAM J Imag Sci. 2013;6(1):437–463.
24. Tarzanagh DA, Michailidis G. Fast randomized algorithms for t-product based tensor operations and decompositions with applications to imaging data. SIAM J Imag Sci. 2018;11(4):2629–2664.
25. Zhang Z, Ely G, Aeron S, Hao N, Kilmer M. Novel methods for multilinear data completion and de-noising based on tensor-SVD. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014:3842-3849.
26. Zhang Z, Aeron S. Exact tensor completion using t-SVD. IEEE Trans Signal Process. 2017;65(6):1511–1526.
27. Jiang Q, Ng M. Robust low-tubal-rank tensor completion via convex optimization. Proceedings of the 28th International Joint Conference on Artificial Intelligence; 2019:2649-2655.AAAI Press.
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. 2008;30(3):1084–1127.
29. Kilmer ME, Martin CD, Perrone L. A third-order generalization of the matrix SVD as a product of third-order tensors. Tufts University, Department of Computer Science, Technical Report TR-2008-4; 2008.
30. Lu C, Feng J, Liu W, Lin Z, Yan S. Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Trans Pattern Anal Mach Intell. 2019. https://doi.org/10.1109/TPAMI.2019.2891760.
31. Goreinov SA, Tyrtyshnikov EE, Zamarashkin NL. A theory of pseudoskeleton approximations. Linear Algebra Appl. 1997;261(1-3):1–21.
32. Oseledets IV, Savostianov D, Tyrtyshnikov EE. Tucker dimensionality reduction of three-dimensional arrays in linear time. SIAM J Matrix Anal Appl. 2008;30(3):939–956.
33. Halko N, Martinsson P-G, Tropp JA. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 2011;53(2):217–288.
34. Nagy JG, Kilmer ME. Kronecker product approximation for preconditioning in three-dimensional imaging applications. IEEE Trans Image Process. 2006;15(3):604–613.
35. Bader BW, Kolda TG, et al. MATLAB Tensor Toolbox, Version 3.0. Available from: https://www.tensortoolbox.org, 2017.
36. Zhu F, Wang Y, Fan B, Xiang S, Meng G, Pan C. Spectral unmixing via data-guided sparsity. IEEE Trans Image Process. 2014;23(12):5412–5427.
37. Strang G. The discrete cosine transform. SIAM Rev. 1999;41(1):135–147.

How to cite this article: Zeng C, Ng MK. Decompositions of third-order tensors: HOSVD, T-SVD, and Beyond. *Numer Linear Algebra Appl*. 2020;e2290. https://doi.org/10.1002/nla.2290