On the truncated multilinear singular value decomposition

Nick Vannieuwenhoven *Raf Vandebril*
Karl Meerbergen

Report TW 589, March 2011



Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

On the truncated multilinear singular value decomposition

Nick Vannieuwenhoven *Raf Vandebril*
Karl Meerbergen

Report TW 589, March 2011

Department of Computer Science, K.U.Leuven

Abstract

In this report, we investigate the truncated multilinear singular value decomposition (MLSVD), proposed in De Lathauwer *et al.* (2000). Truncating the MLSVD results in an approximation, with a prescribed multilinear rank, to a tensor. We present a new error expression for an approximate Tucker decomposition with orthogonal factor matrices. From this expression, new insights are obtained which lead us to propose a novel non-iterative algorithm, the sequentially truncated multilinear singular value decomposition (ST-MLSVD). It has several characteristics which make it a suitable replacement for the standard T-MLSVD. In absence of truncation, the ST-MLSVD computes the MLSVD. The approximation error can be characterized exactly in terms of some discarded singular values. Furthermore, the error bound of the T-MLSVD also holds for the ST-MLSVD. The approximation error of the ST-MLSVD is provably better than the T-MLSVD for some tensors. Unfortunately, it is not always better than the T-MLSVD. Therefore, we demonstrate numerically that the ST-MLSVD approximation often outperforms the T-MLSVD. A convincing example from image denoising is presented. The ST-MLSVD algorithm is much more efficient in computation time than the T-MLSVD algorithm by De Lathauwer *et al.* It can attain speedups over the latter algorithm, proportional to the order of the tensor, in the case of a rank-1 approximation.

Keywords : tensor, sequentially truncated multilinear singular value decomposition, higher-order singular value decomposition, multilinear orthogonal projection
MSC : Primary : 15A69, 65F99, 65Y20 Secondary : 15A03, 15A72.

ON THE TRUNCATED MULTILINEAR SINGULAR VALUE DECOMPOSITION[†]

NICK VANNIEUWENHOVEN[‡§], RAF VANDEBRIL[§], AND KARL MEERBERGEN[§]

Abstract. In this report, we investigate the truncated multilinear singular value decomposition (MLSVD), as presented in [L. De Lathauwer, B. De Moor and J. Vandewalle, *a multilinear singular value decomposition*, 2000]. The MLSVD comprises a Tucker decomposition [L.R. Tucker, *some mathematical notes on three-mode factor analysis*, 1966] with orthogonal factor matrices. Truncating the MLSVD results in an approximation, with a prescribed multilinear rank, to a tensor. We present a new error expression for an approximate Tucker decomposition with orthogonal factor matrices. From this expression, new insights are obtained which lead us to propose a novel non-iterative algorithm, the sequentially truncated multilinear singular value decomposition (ST-MLSVD). It has several characteristics which make it a suitable replacement for the standard T-MLSVD. In absence of truncation, the ST-MLSVD computes the MLSVD. The approximation error can be characterized exactly in terms of some discarded singular values. Furthermore, the error bound of the T-MLSVD, proposed by De Lathauwer *et al.*, also holds for the ST-MLSVD. The approximation error of the ST-MLSVD is provably better than the T-MLSVD for some tensors. Unfortunately, it is not always better than the T-MLSVD. Therefore, we demonstrate numerically that the ST-MLSVD approximation often outperforms the T-MLSVD. A convincing example from image denoising is presented. The ST-MLSVD algorithm is much more efficient in computation time than the T-MLSVD algorithm by De Lathauwer *et al.* It can attain speedups over the latter algorithm, proportional to the order of the tensor, in the case of a rank-1 approximation.

Key words. tensor, sequentially truncated multilinear singular value decomposition, higher-order singular value decomposition, multilinear orthogonal projection

AMS subject classifications. 15A03, 15A69, 15A72, 65F99, 65Y20

1. Introduction. In this paper, we seek a *good* rank-$(r_1, r_2, \ldots, r_d)$ approximation to the order-d tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, using only tools from standard numerical linear algebra. Formally, we are interested in an approximate solution to

$$\min_{U_1, U_2, \ldots, U_d} \|\mathcal{A} - (U_1 U_1^T, U_2 U_2^T, \ldots, U_d U_d^T) \cdot \mathcal{A}\|_F^2, \tag{1.1}$$

wherein $U_k \in \mathbb{R}^{n_k \times r_k}$, $1 \leq k \leq d$, has orthonormal columns. U_k is called a *factor matrix*.

Traditionally, the problem of approximating the solution of Eq. (1.1) has been addressed by the (truncated) *higher-order singular value decomposition* (T-HOSVD), or *multilinear singular value decomposition* (T-MLSVD) [5, 16]. The T-MLSVD is one of the core algorithms in tensor approximation. It has been applied to several practical problems, such as handwritten digit classification [19] and analysis of image ensembles [24]. We refer to Kolda's survey article [16] and the references therein.

In this paper, we are interested in the approximation from a theoretical perspective. The T-MLSVD is a core building block algorithm. It can be used to approximate the best rank-1 approximation to a tensor [20]. Often, the T-MLSVD is employed to initialize iterative algorithms to compute the best rank-1 approximation [3, 25] or best approximation of a specified multilinear rank [6, 10, 14]. It is also a building block for other tensor decompositions. The hierarchical Tucker decomposition

[†]A shorter and modified version of this technical report has been submitted to SIAM Journal on Matrix Analysis and Applications.

[‡]Corresponding author (nick.vannieuwenhoven@cs.kuleuven.be).

[§]Numerical Approximation and Linear Algebra Group, Department of Computer Science, K.U.Leuven.

of Grasedyck [13], for instance, requires, in essence, the computation of several T-MLSVDs. The Tucker-Tree decomposition of Oseledets and Tyrtyshnikov [17] employs the T-MLSVD, essentially for dimensionality reduction, in the first step of its construction. Other applications include dimensionality reduction in order to compute other decompositions more efficiently [7], such as the canonical polyadic decomposition (CP)—also CANDECOMP/PARAFAC.

In this paper, we present a new Tucker decomposition [22] with orthogonal factor matrices, similar to the T-MLSVD. It has several advantages over the T-MLSVD, improving both the number of operations required to compute the approximation, as the error of the approximation. Our investigation was spurred by an interesting remark made by Andersson and Bro in [1]. In proposing techniques to improve the computational efficiency of the HOOI algorithm, they briefly point out a different initialization scheme. Instead of initializing the factor matrices by a truncated MLSVD, they, essentially, propose to initialize it with the sequentially truncated MLSVD—the decomposition we investigate in this paper. However, the decomposition is never formalized, nor were its interesting properties investigated in [1].

In the next section, we present some necessary preliminaries concerning tensor algebra, multilinear orthogonal projections and some basic properties and definitions, as the latter are not yet fully standardized. Thereafter, in Section 3, we briefly review the multilinear singular value decomposition [5]. In Section 4, we present a novel expression for the error of any approximate Tucker decomposition with orthogonal factor matrices. Based on this error expression, we propose a new non-iterative algorithm, the sequentially truncated multilinear singular value decomposition (ST-MLSVD), to compute a good approximate Tucker decomposition with orthogonal factor matrices, in Section 5. There, we prove several important characteristics of the ST-MLSVD. The relationship between the error of the T-MLSVD and ST-MLSVD is investigated in Section 6. In Section 7, we demonstrate numerically that the ST-MLSVD outperforms T-MLSVD in terms of approximation error and computation time. Finally, in Section 8, we summarize our main conclusions and argue that the ST-MLSVD may, indeed, be preferable over the T-MLSVD in many cases.

Concerning notation, the following conventions apply throughout. Tensors are typeset in an upper-case calligraphic font ($\mathcal{A}, \mathcal{S}, \mathcal{B}$), matrices as upper-case letters (A, U, V, Σ, P), vectors as bold-face lower-case letters ($\mathbf{u}, \mathbf{v}, \mathbf{e}, \mathbf{a}, \mathbf{m}$) and scalars as lower-case letters (a, b). Occasionally, $\mathcal{U}$ and $\mathcal{V}$ are used to denote vector spaces. I denotes an identity matrix of suitable dimensions. The scalar d is reserved for the order of the considered tensor. The scalars m and n are reserved to indicate the length of vectors over the real numbers; with a subscript they indicate the length of a certain mode of a tensor. The letters i, j, p and q are reserved for integer indices. A multilinear orthogonal projector that projects along mode-k is denoted as π_k, see also Section 2.2. Projectors, matrices and tensors which are typeset with an overhead bar ($\bar{\pi}_1$, $\bar{U}_1$, $\bar{\mathcal{A}}$) are related to the T-MLSVD, and with a hat ($\hat{\pi}_1$, $\hat{U}_1$, $\hat{\mathcal{A}}$) they are related to the ST-MLSVD.

2. Preliminaries. In this section, some preliminary material is presented. We review some relevant aspects from the algebra of tensors, discuss multilinear orthogonal projections and present some useful properties. The first part is based on the presentation of some important tensor-algebraic concepts from [4,8]. For more details, the reader is referred to these references.

2.1. Tensor algebra. This paper is concerned with *tensors*. A tensor may be considered as an element of the tensor product of a set of vector spaces. In this paper,

we are interested in tensor products of real vector spaces. That is,

$$\mathcal{A} \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \otimes \cdots \otimes \mathbb{R}^{n_d},$$

is a tensor of *order* d over the real numbers. More practically, such an object can be represented as a d-array of numbers with respect to a given tensor basis. That is,

$$\mathcal{A} = [\![a_{i_1,i_2,\ldots,i_d}]\!]_{i_1,i_2,\ldots,i_d=1}^{n_1,n_2,\ldots,n_d} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}.$$

In the above, $\mathcal{A}$ is specified with respect to the *standard tensor basis of order* d

$$\mathcal{E}_d = \{\mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \cdots \otimes \mathbf{e}_{i_d}\}_{i_1,i_2,\ldots,i_d=1}^{n_1,n_2,\ldots,n_d},$$

where $\mathbf{e}_i$ is the i^{th} standard basis vector of suitable dimension. We may thus write $\mathcal{A}$ as a multilinear combination of these basis vectors, as follows

$$\mathcal{A} = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} a_{i_1,i_2,\ldots,i_d} \mathbf{e}_{i_1} \otimes \mathbf{e}_{i_2} \otimes \cdots \otimes \mathbf{e}_{i_d}. \tag{2.1}$$

In this paper, we will identify the tensor with the d-array representing its coordinates with respect to a suitable basis. A tensor may be multiplied in each of its modes with a (different) matrix. Let

$$\mathcal{A} = [\![a_{i_1,i_2,\ldots,i_d}]\!] \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \qquad \text{and} \qquad H^{(k)} = [\![h_{p,q}^{(k)}]\!] \in \mathbb{R}^{m_k \times n_k}.$$

Then, $\mathcal{A}$ may be transformed into a tensor $\mathcal{B} = [\![b_{j_1,j_2,\ldots,j_d}]\!] \in \mathbb{R}^{m_1 \times m_2 \times \cdots \times m_d}$ via the *multilinear multiplication* of $\mathcal{A}$ by $\{H^{(k)}\}_k$,

$$b_{j_1,j_2,\ldots,j_d} = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} a_{i_1,i_2,\ldots,i_d} h_{j_1,i_1}^{(1)} h_{j_2,i_2}^{(2)} \cdots h_{j_d,i_d}^{(d)}.$$

We will write this more concisely, in standard mathematical notation [8], as

$$\mathcal{B} = (H^{(1)}, H^{(2)}, \ldots, H^{(d)}) \cdot \mathcal{A}. \tag{2.2}$$

In two special cases, the above multilinear multiplication has an algebraic interpretation. First, if the matrices $\{H^{(k)}\}_k$ are orthogonal, and thus square, then Eq. (2.2) can be interpreted as a *change of basis*. That is, $\mathcal{B}$ represents the same object of the vector space $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_d}$ as $\mathcal{A}$. However, $\mathcal{B}$ comprises the coordinates of that object with respect to the tensor basis represented by the columns of the matrices $\{H^{(k)}\}_k$. More formally, let $H^{(k)} = \begin{bmatrix} \mathbf{h}_1^{(k)} & \mathbf{h}_2^{(k)} & \ldots & \mathbf{h}_{n_k}^{(k)} \end{bmatrix}$ be the column representation, then $b_{i_1,i_2,\ldots,i_d}$ is the coordinate of $\mathcal{A}$ with respect to the tensor basis vector $\mathbf{h}_{i_1}^{(1)} \otimes \mathbf{h}_{i_2}^{(2)} \otimes \cdots \otimes \mathbf{h}_{i_d}^{(d)}$. That is, in analogy with Eq. (2.1), we can write

$$\mathcal{A} = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \cdots \sum_{i_d=1}^{n_d} b_{i_1,i_2,\ldots,i_d} \mathbf{h}_{i_1}^{(1)} \otimes \mathbf{h}_{i_2}^{(2)} \otimes \cdots \otimes \mathbf{h}_{i_d}^{(d)}.$$

Another useful algebraic interpretation of the multilinear multiplication arises when the matrices $\{H^{(k)}\}_k$ are orthogonal projectors. This operation is then referred to as a *multilinear orthogonal projection*.

2.2. Multilinear orthogonal projections. Key to this paper is the use of multilinear orthogonal projections. Recall from e.g. [18], that an *orthogonal projector* is a linear transformation P that projects a vector $\mathbf{u} \in \mathbb{R}^n$ onto a vector space $\mathcal{U} \subseteq \mathbb{R}^n$, such that the residual $\mathbf{u} - P\mathbf{u}$ is orthogonal to $\mathcal{U}$. Such a projector can always be represented in matrix form as $P = UU^T$, assuming that the columns of U form an orthonormal basis for the vector space $\mathcal{U}$. De Silva and Lim [8] state that if ϕ_k is an orthogonal projector from the vector space $\mathcal{V}_k \subseteq \mathbb{R}^{n_k}$ onto $\mathcal{U}_k \subseteq \mathcal{V}_k$ then $\Phi = (\phi_1, \phi_2, \ldots, \phi_d)$ is a multilinear orthogonal projection from the tensor space $\mathcal{V} := \mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \cdots \otimes \mathcal{V}_d$ onto the tensor subspace $\mathcal{U}_1 \otimes \mathcal{U}_2 \otimes \cdots \otimes \mathcal{U}_d \subseteq \mathcal{V}$. In this paper, we deal with orthogonal projectors from $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_{k-1}} \otimes \mathbb{R}^{n_k} \otimes \mathbb{R}^{n_{k+1}} \otimes \cdots \otimes \mathbb{R}^{n_d}$ onto the subspace $\mathbb{R}^{n_1} \otimes \cdots \otimes \mathbb{R}^{n_{k-1}} \otimes \mathcal{U}_k \otimes \mathbb{R}^{n_{k+1}} \otimes \cdots \otimes \mathbb{R}^{n_d}$ exclusively. This multilinear orthogonal projection is given by

$$\pi_k \mathcal{A} := (\underbrace{I, \ldots, I}_{\substack{k-1 \\ \text{matrices}}}, U_k U_k^T, I, \ldots, I) \cdot \mathcal{A} \qquad \text{with} \quad \mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}, \tag{2.3}$$

where we assume that the columns of U_k form an orthonormal basis of the vector space $\mathcal{U}_k$. The subscript of the projector π_k indicates that it projects orthogonally along mode k. The projector satisfies the following properties. These properties can be proved straightforwardly from the properties of the multilinear multiplication, which are presented in the next subsection. Foremost, two projectors that operate on different modes, commute. That is, if $i \neq j$,

$$\pi_i \pi_j \mathcal{A} = \pi_j \pi_i \mathcal{A},$$

The projector π_k is also idempotent; $\pi_k \pi_k \mathcal{A} = \pi_k \mathcal{A}$. Furthermore, its orthogonal complement can be characterized explicitly by

$$(1 - \pi_k)\mathcal{A} = (I, \ldots, I, I - U_k U_k^T, I, \ldots, I) \cdot \mathcal{A},$$

due to Property 2.5, which is presented in the next subsection. Finally, the projector is orthogonal with respect to the Frobenius norm [8], defined as the square root of the sum of all squared elements of the tensor,

$$\|\mathcal{A}\|_F^2 = \|\pi_k \mathcal{A}\|_F^2 + \|(1 - \pi_k)\mathcal{A}\|_F^2. \tag{2.4}$$

The above projections will be used extensively, in this paper.

2.3. Basic definitions and properties. Here, we give some basic definitions that will be used throughout. We bundle a number of well-known properties.

Unfolding. Given a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, we define a *mode-$k$ vector* $\mathbf{v}$ as the vector one obtains by fixing all indices of $\mathcal{A}$, and varying the mode k index: $\mathbf{v} = \mathcal{A}_{i_1, \ldots, i_{k-1}, :, i_{k+1}, \ldots, i_d}$ with i_j a fixed value. We refer to the set of all mode-k vectors of $\mathcal{A}$ as the *mode-k vector space*. The *mode-k unfolding*, or matricization [10], of $\mathcal{A}$, denoted by $\mathcal{A}_{(k)}$, is an $n_k \times \Pi_{i \neq k} n_i$ matrix whose columns are all possible mode-k vectors. The actual order of the mode-k vectors within this unfolding is usually not important, as long as it is consistent. We assume the canonical order, as presented in [10]. More importantly, it should be noted that the column space of $\mathcal{A}_{(k)}$ is the mode-k vector space, hence justifying its name.

Multilinear rank. This paper is concerned with the problem of approximating a tensor by another one with low(er) multilinear rank. The multilinear rank of a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a d-tuple $(r_1, r_2, \ldots, r_d)$, wherein r_k is the dimension of the mode-k vector space. Alternatively, r_k is the column rank of $\mathcal{A}_{(k)}$.

Inner product and norm. The *Frobenius inner product* of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is defined by

$$\langle \mathcal{A}, \mathcal{B} \rangle_F = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} a_{i_1,\ldots,i_d} b_{i_1,\ldots,i_d}.$$

The corresponding *Frobenius norm* is defined by

$$\|\mathcal{A}\|_F^2 := \langle \mathcal{A}, \mathcal{A} \rangle_F = \|\mathcal{A}_{(k)}\|_F^2,$$

for any mode-k unfolding of $\mathcal{A}$. An expression of the Frobenius inner product in terms of the unfoldings of the tensors, is also possible. It is not hard to see that the inner product of $\mathcal{A}$ and $\mathcal{B}$ is simply the sum of the inner products of their mode-k vectors. We thus obtain the following property [10, Lemma 2.1].

PROPERTY 2.1 (Inner product [10]). *Let $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. Then, if $1 \leq k \leq d$,*

$$\langle \mathcal{A}, \mathcal{B} \rangle_F = \operatorname{trace}\left(\mathcal{B}_{(k)}^T \mathcal{A}_{(k)}\right) = \operatorname{trace}\left(\mathcal{A}_{(k)}^T \mathcal{B}_{(k)}\right) = \langle \mathcal{B}, \mathcal{A} \rangle_F,$$

where $\operatorname{trace}(A) := \sum_i a_{ii}$ *is the trace of* A.

Of course, the above expression should not be used to compute the inner product. However, it will yield an interesting insight into the orthogonality of orthogonally projected tensors, in Section 4.

The Frobenius norm for matrices is unitarily invariant.

PROPERTY 2.2. *(Orthonormality preserves Frobenius norm) If $A \in \mathbb{R}^{m \times n}$ and $U \in \mathbb{R}^{r \times m}$, with $r \geq m$, is a matrix with orthonormal columns, and $V \in \mathbb{R}^{s \times n}$, with $s \geq n$, is a matrix with orthonormal columns, then*

$$\|UAV^T\|_F^2 = \|A\|_F^2.$$

Proof. Let the columns of A and U be given respectively by

$$A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_n \end{bmatrix} \qquad \text{and} \qquad U = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_m \end{bmatrix},$$

then,

$$\|UA\|_F^2 = \sum_{i=1}^{n} \|U\mathbf{a}_i\|_2^2 = \sum_{i=1}^{n} \left\| \sum_{j=1}^{m} a_{ij}\mathbf{u}_j \right\|_2^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \|a_{ij}\mathbf{u}_j\|_2^2 = \|A\|_F^2.$$

Herein, the first and last equality are due to the definition of the Frobenius norm. The third equality follows from the orthonormality of the columns of U, $U^TU = I$, and the definition of the 2-norm. From $\|A\|_F = \|A^T\|_F$, it also follows that $\|AV^T\|_F^2 = \|A\|_F^2$. Then, $\|UAV^T\|_F^2 = \|BV^T\|_F^2 = \|B\|_F^2 = \|UA\|_F^2 = \|A\|_F^2$ concludes the proof. ∎

Multilinear multiplication. In this paragraph, we assume that the dimensions of the matrices involved in the presented multilinear multiplications are compatible with the latter. Multilinear multiplication in one mode, say $1 \leq k \leq d$, with a matrix M can be interpreted as multiplying the mode-k vectors by M. That is, if M is at position k in the tuple, then

$$[(I, \ldots, I, M, I, \ldots, I) \cdot \mathcal{A}]_{(k)} = M\mathcal{A}_{(k)},$$

is the most basic property, which we will use throughout. In general, the unfolding of a multilinear multiplication has a Kronecker structure [5, 10].

PROPERTY 2.3 (Unfolding of multilinear multiplication [10]). *Let a tensor* $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *be given. Then,*

$$[(M_1, M_2, \ldots, M_d) \cdot \mathcal{A}]_{(k)} = M_k \mathcal{A}_{(k)} \left(M_1 \otimes M_2 \otimes \cdots \otimes M_{k-1} \otimes M_{k+1} \otimes \cdots \otimes M_d\right)^T.$$

The following basic properties of multilinearity can be found in [8].

PROPERTY 2.4 (Component-wise multiplication [8]). *Two multilinear multiplications can be transformed into one multilinear multiplication, by component-wise multiplication of the matrices,*

$$(L_1, L_2, \ldots, L_d) \cdot [(M_1, M_2, \ldots, M_d) \cdot \mathcal{A}] = (L_1 M_1, L_2 M_2, \ldots, L_d M_d) \cdot \mathcal{A}$$

PROPERTY 2.5 (Multilinearity [8]). *The multilinear multiplication is multilinear in its arguments,*

$$\begin{aligned}(A_1, \ldots, A_{k-1}, A_k + B_k, A_{k+1}, \ldots, A_d) \cdot \mathcal{A} = \\ (A_1, \ldots, A_{k-1}, A_k, A_{k+1}, \ldots, A_d) \cdot \mathcal{A} + (A_1, \ldots, A_{k-1}, B_k, A_{k+1}, \ldots, A_d) \cdot \mathcal{A}.\end{aligned}$$

Kronecker product. Some of the basic properties of the Kronecker product can be found in [23]. A trivial property that we employ, is $(A \otimes B)^T = A^T \otimes B^T$. The Kronecker product also preserves orthogonality.

PROPERTY 2.6 (Preserves orthonormality [23]). *The kronecker product of matrices with orthonormal columns (rows) is a matrix with orthonormal columns (rows).*

3. Multilinear singular value decomposition. De Lathauwer, De Moor and Vandewalle proved the following theorem in [5].

THEOREM 3.1 (Multilinear singular value decomposition [5]). *Every tensor* $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *admits a* multilinear singular value decomposition:

$$\mathcal{A} = (U_1, U_2, \ldots, U_d) \cdot \mathcal{S}, \tag{3.1}$$

where the factor matrix U_k *is an orthogonal* $n_k \times n_k$ *matrix, and the* core tensor $\mathcal{S} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ *satisfies the following properties:*

$$\begin{aligned}&\textit{all-orthogonality:} && \langle \mathcal{S}_{i_k=\alpha}, \mathcal{S}_{i_k=\beta} \rangle_F = 0, \quad \forall \alpha \neq \beta, \textit{ and} && (3.2)\\ &\textit{ordering:} && \|\mathcal{S}_{i_k=1}\| \geq \|\mathcal{S}_{i_k=2}\| \geq \ldots \geq \|\mathcal{S}_{i_k=n_k}\| \geq 0, && (3.3)\end{aligned}$$

for all $1 \leq k \leq d$. *Herein,* $\mathcal{S}_{i_k=\alpha}$ *is the subtensor of* $\mathcal{S}$ *whose mode-k index is fixed to* α. *The MLSVD is unique up to some trivial indeterminacies.*

De Lathauwer *et al.* also demonstrated that the MLSVD can be computed by means of d singular value decompositions of the distinct unfoldings of $\mathcal{A}$.

THEOREM 3.2 (De Lathauwer *et al.*'s MLSVD algorithm [5]). *Let* $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. *If the factor matrix* U_k *is obtained from the SVD of the mode-k unfolding of* $\mathcal{A}$,

$$\mathcal{A}_{(k)} = U_k \Sigma_k V_k^T, \tag{3.4}$$

for all $1 \leq k \leq d$, *and if the core tensor is given by,* $\mathcal{S} = (U_1^T, U_2^T, \ldots, U_d^T) \cdot \mathcal{A}$, *then* $\mathcal{A}$ *can be decomposed as* $\mathcal{A} = (U_1, U_2, \ldots, U_d) \cdot \mathcal{S}$, *i.e. it is the MLSVD of* $\mathcal{A}$.

The MLSVD can be employed to construct a low multilinear rank approximation to a tensor. Suppose we want to approximate $\mathcal{A}$ by a rank-$(r_1, r_2, \ldots, r_d)$ tensor $\bar{\mathcal{A}}$,

with $r_k \leq n_k$ for all $1 \leq k \leq d$. A factor matrix $\bar{U}_k$ of the truncated MLSVD is obtained from a truncation of the SVD of the mode-k unfolding of the tensor [5, 16],

$$\mathcal{A}_{(k)} = U_k \Sigma_k V_k^T = \begin{bmatrix} \bar{U}_k & \tilde{U}_k \end{bmatrix} \begin{bmatrix} \bar{\Sigma}_k & \\ & \tilde{\Sigma}_k \end{bmatrix} \begin{bmatrix} \bar{V}_k^T \\ \tilde{V}_k^T \end{bmatrix}, \qquad \text{with} \quad \bar{U}_k \in \mathbb{R}^{n_k \times r_k}. \tag{3.5}$$

The approximation is obtained by an orthogonal projection onto the tensor basis, represented by these factor matrices. That is,

$$\bar{\mathcal{A}} := \bar{\pi}_1 \bar{\pi}_2 \cdots \bar{\pi}_d \mathcal{A} := (\bar{U}_1 \bar{U}_1^T, \bar{U}_2 \bar{U}_2^T, \ldots, \bar{U}_d \bar{U}_d^T) \cdot \mathcal{A} =: (\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_d) \cdot \bar{\mathcal{S}} \approx \mathcal{A},$$

wherein the *truncated core tensor* is defined as,

$$(\bar{U}_1^T, \bar{U}_2^T, \ldots, \bar{U}_d^T) \cdot \mathcal{A} =: \bar{\mathcal{S}} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d}.$$

In the remainder, we will always denote the T-MLSVD projector onto mode k by $\bar{\pi}_k \mathcal{A} := (I, \ldots, I, \bar{U}_k \bar{U}_k^T, I, \ldots, I) \cdot \mathcal{A}$. The error of such an approximation can be bounded from above by,

$$\|\mathcal{A} - \bar{\mathcal{A}}\|_F^2 \leq \|\tilde{\Sigma}_1\|_F^2 + \|\tilde{\Sigma}_2\|_F^2 + \cdots + \|\tilde{\Sigma}_d\|_F^2,$$

which was stated as property 10 in [5]. A new, elegant proof of this error bound, providing new insights into the source of the approximation error of a Tucker decomposition with orthogonal factor matrices, is presented in the next section.

4. Error of a truncated orthogonal Tucker decomposition. In this section, we derive a new explicit formula for the error of an approximate Tucker decomposition with orthogonal factor matrices. This formula provides insights into the structure of the optimization problem. In the next section, we propose a new greedy optimization algorithm, based on the error expansion that is presented in the next theorem.

THEOREM 4.1 (Error of a truncated orthogonal Tucker decomposition). *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. Let $\mathcal{A}$ be approximated by*

$$\breve{\mathcal{A}} := \breve{\pi}_1 \breve{\pi}_2 \cdots \breve{\pi}_d \mathcal{A} := (\breve{U}_1 \breve{U}_1^T, \breve{U}_2 \breve{U}_2^T, \ldots, \breve{U}_d \breve{U}_d^T) \cdot \mathcal{A},$$

in which the factor matrices $\{\breve{U}_k\}_k$ have orthonormal columns. Then, the approximation error is given, for any permutation $\mathbf{p}$ of $\{1, 2, \ldots, d\}$, by

$$\|\mathcal{A} - \breve{\mathcal{A}}\|_F^2 = \|\breve{\pi}_{p_1}^{\perp} \mathcal{A}\|_F^2 + \|\breve{\pi}_{p_2}^{\perp} \breve{\pi}_{p_1} \mathcal{A}\|_F^2 + \|\breve{\pi}_{p_3}^{\perp} \breve{\pi}_{p_1} \breve{\pi}_{p_2} \mathcal{A}\|_F^2 + \cdots + \|\breve{\pi}_{p_d}^{\perp} \breve{\pi}_{p_1} \cdots \breve{\pi}_{p_{d-1}} \mathcal{A}\|_F^2$$

and, it is bounded by

$$\|\mathcal{A} - \breve{\mathcal{A}}\|_F^2 \leq \sum_{k=1}^{d} \|\breve{\pi}_{p_k}^{\perp} \mathcal{A}\|_F^2.$$

wherein $\breve{\pi}_k^{\perp} := 1 - \breve{\pi}_k$.

Proof. If the factor matrices $\{\breve{U}_k\}_k$ are fixed, then the core tensor $\breve{\mathcal{S}}$ that minimizes the approximation error

$$\|\mathcal{A} - (\breve{U}_1, \breve{U}_2, \ldots, \breve{U}_d) \cdot \breve{\mathcal{S}}\|_F,$$

is given by $\breve{\mathcal{S}} = (\breve{U}_1^T, \breve{U}_2^T, \ldots, \breve{U}_d^T) \cdot \mathcal{A}$, as proved in [6,20]. Therefore, given fixed factor matrices, the optimal approximation to $\mathcal{A}$ is obtained by *a multilinear orthogonal*

projection onto the tensor basis spanned by the columns of the factor matrices. It is thus justified to *define* the error of an orthogonal Tucker decomposition, given a set of factor matrices, as the error of this optimal orthogonal multilinear projection of $\mathcal{A}$.

Let us assume, without loss of generality, that $\mathbf{p} = \langle 1, 2, \ldots, d \rangle$. We can introduce a telescoping sum, as follows

$$\begin{aligned}\mathcal{A} - \breve{\mathcal{A}} &= (\mathcal{A} - \breve{\pi}_1\mathcal{A}) + (\breve{\pi}_1\mathcal{A} - \breve{\pi}_2\breve{\pi}_1\mathcal{A}) + \cdots + (\breve{\pi}_{d-1}\cdots\breve{\pi}_1\mathcal{A} - \breve{\pi}_d\cdots\breve{\pi}_1\mathcal{A}), \\ &= \breve{\pi}_1^{\perp}\mathcal{A} + \breve{\pi}_2^{\perp}\breve{\pi}_1\mathcal{A} + \breve{\pi}_3^{\perp}\breve{\pi}_1\breve{\pi}_2\mathcal{A} + \cdots + \breve{\pi}_d^{\perp}\breve{\pi}_1\cdots\breve{\pi}_{d-1}\mathcal{A}.\end{aligned}$$

Let, for $i < j$,

$$\begin{aligned}\breve{\mathcal{B}}_{(i)} &:= (I - \breve{U}_i\breve{U}_i^T)\mathcal{A}_{(i)}(\breve{U}_1\breve{U}_1^T \otimes \cdots \otimes \breve{U}_{i-1}\breve{U}_{i-1}^T \otimes I \otimes \cdots \otimes I), \\ \breve{\mathcal{C}}_{(i)} &:= \breve{U}_i\breve{U}_i^T\mathcal{A}_{(i)}(\breve{U}_1\breve{U}_1^T \otimes \cdots \otimes \breve{U}_{i-1}\breve{U}_{i-1}^T \otimes \breve{U}_{i+1}\breve{U}_{i+1}^T \otimes \cdots \otimes \breve{U}_{j-1}\breve{U}_{j-1}^T \otimes \\ &\qquad I - \breve{U}_j\breve{U}_j^T \otimes I \otimes \cdots \otimes I),\end{aligned}$$

then we notice that $\breve{\mathcal{C}}_{(i)}^T\breve{\mathcal{B}}_{(i)} = 0$, because $\breve{U}_i\breve{U}_i^T$ is a projector. From Property 2.1, we derive that the trace of the above expression equals the inner product

$$\langle \breve{\pi}_i^{\perp}\breve{\pi}_1\cdots\breve{\pi}_{i-1}\mathcal{A},\ \breve{\pi}_j^{\perp}\breve{\pi}_1\cdots\breve{\pi}_i\cdots\breve{\pi}_{j-1}\mathcal{A}\rangle_F = \operatorname{trace}\left(\breve{\mathcal{C}}_{(i)}^T\breve{\mathcal{B}}_{(i)}\right) = 0.$$

This entails that $\breve{\pi}_i^{\perp}\breve{\pi}_1\cdots\breve{\pi}_{i-1}\mathcal{A} \perp \breve{\pi}_j^{\perp}\breve{\pi}_1\cdots\breve{\pi}_i\cdots\breve{\pi}_{j-1}\mathcal{A}$, i.e. these projected tensors are orthogonal with respect to the Frobenius norm. As the above holds for any $i < j$ and orthogonality is reflexive, all the terms are orthogonal with respect to one another, in the Frobenius norm. This results in

$$\|\mathcal{A} - \breve{\mathcal{A}}\|_F^2 = \|\breve{\pi}_1^{\perp}\mathcal{A}\|_F^2 + \|\breve{\pi}_2^{\perp}\breve{\pi}_1\mathcal{A}\|_F^2 + \|\breve{\pi}_3^{\perp}\breve{\pi}_1\breve{\pi}_2\mathcal{A}\|_F^2 + \cdots + \|\breve{\pi}_d^{\perp}\breve{\pi}_1\cdots\breve{\pi}_{d-1}\mathcal{A}\|_F^2,$$

completing the first part of the proof. The second part follows readily from the observation that an orthogonal projection onto a subspace can only decrease the Frobenius norm. That is clear from Eq. (2.4). ∎

Theorem 4.1 readily demonstrates that the error of the T-MLSVD can be bounded from above and below.

THEOREM 4.2 (T-MLSVD error bounds). *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and let $\bar{\mathcal{A}}$ be the rank-$(r_1, \ldots, r_d)$ truncated multilinear singular value decomposition of $\mathcal{A}$. Let*

$$\mathcal{A}_{(k)} = \begin{bmatrix}\bar{U}_k & \tilde{U}_k\end{bmatrix}\begin{bmatrix}\bar{\Sigma}_k & \\ & \tilde{\Sigma}_k\end{bmatrix}\begin{bmatrix}\bar{V}_k^T \\ \tilde{V}_k^T\end{bmatrix}, \qquad \text{with } \bar{U}_k \in \mathbb{R}^{n_k \times r_k},$$

be a singular value decomposition. The error of the truncated MLSVD approximation $\bar{\mathcal{A}}$ to $\mathcal{A}$ is then bounded by

$$\max_k \|\tilde{\Sigma}_k\|_F^2 \leq \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2 \leq \|\tilde{\Sigma}_1\|_F^2 + \cdots + \|\tilde{\Sigma}_d\|_F^2.$$

Proof. The proof of the lower bound follows from noticing that Theorem 4.1 holds for any permutation $\mathbf{p}$ of $\{1, 2, \ldots d\}$, the independence of the error on the permutation $\mathbf{p}$ and the positivity of the terms.

The proof of the upper bound follows from Theorem 4.1 and by the definition of the T-MLSVD factor matrices. We have

$$\begin{aligned}\|\mathcal{A} - \bar{\pi}_1\cdots\bar{\pi}_d\mathcal{A}\|_F^2 &\leq \|\bar{\pi}_1^{\perp}\mathcal{A}\|_F^2 + \|\bar{\pi}_2^{\perp}\mathcal{A}\|_F^2 + \|\bar{\pi}_3^{\perp}\mathcal{A}\|_F^2 + \cdots + \|\bar{\pi}_d^{\perp}\mathcal{A}\|_F^2, \\ &= \|\tilde{U}_1\tilde{\Sigma}_1\tilde{V}_1^T\|_F^2 + \cdots + \|\tilde{U}_d\tilde{\Sigma}_d\tilde{V}_d^T\|_F^2,\end{aligned}$$

which, combined with Property 2.2, yields the desired result. ∎

From Theorem 4.1, it is clear that the T-MLSVD can be interpreted as an algorithm that minimizes the above upper bound, providing a quite strong rationale for the T-MLSVD approximation. Indeed, minimizing the upper bound yields,

$$\min_{\pi_1,\ldots,\pi_d} \|\mathcal{A} - \pi_1 \cdots \pi_d \mathcal{A}\|_F^2 \leq \min_{\pi_1,\ldots,\pi_d} \sum_{k=1}^{d} \|\pi_k^{\perp}\mathcal{A}\|_F^2 = \sum_{k=1}^{d} \min_{\pi_k} \|\pi_k^{\perp}\mathcal{A}\|_F^2, \tag{4.1}$$

whose solution is given by choosing π_k to project onto the dominant subspace of the mode-k vector space, as in the T-MLSVD. Coincidentally, this also minimizes the lower bound in Theorem 4.1.

Theorem 4.1 provides an intuitive argument why the truncated MLSVD generally performs well. Suppose that $\bar{\pi}_k\mathcal{A}$ is a very good approximation to $\mathcal{A}$, for all modes k. One can then expect that $\bar{\pi}_1 \cdots \bar{\pi}_{k-1}\mathcal{A}$ is still a good approximation to $\mathcal{A}$. If this is the case, then $\bar{\pi}_k^{\perp}$, which projects onto the least dominant subspace in mode k, will still be a good approximation to the actual projector onto the least dominant subspace of the mode-k vector space of $\bar{\pi}_1 \cdots \bar{\pi}_{k-1}\mathcal{A}$. In other words, $\bar{\pi}_k^{\perp}$ would be a fairly good approximation to the optimal orthogonal projector, which would project exactly onto the aforementioned vector space. However, it is also clear that if at least one $\bar{\pi}_j\mathcal{A}$ is not good approximation to $\mathcal{A}$, then it is unlikely that $\bar{\pi}_1 \cdots \bar{\pi}_j \cdots \bar{\pi}_{k-1}\mathcal{A}$ approximates $\mathcal{A}$ well. Consequently, it is much less plausible that the least dominant subspace in mode-k of that projected tensor still resembles the least dominant subspace of the original tensor. The T-MLSVD approximation may be bad in that case. Alternatively, this can also be deduced from the error bound in Theorem 4.2. A bad approximation in at least one mode, results in a higher lower (and upper) bound. In our numerical experiments, we demonstrate that both the higher-order orthogonal iteration (HOOI) algorithm [6] as our new truncation algorithm, which is presented in the next section, can be much better than the truncated MLSVD approximation, precisely when the tensor is strongly truncated in one or more modes.

In the next section, we propose a new technique to truncated the multilinear singular value decomposition, inspired by Theorem 4.1.

5. Sequentially truncated multilinear singular value decomposition. In this section, we propose a novel truncated multilinear singular value decomposition. Contrary to the T-MLSVD, the order in which the modes are processed is relevant, and leads to different approximations. The order in which modes are processed, is denoted by a sequence $\mathbf{p}$. Throughout this section, we present our results only for the processing order $\mathbf{p} = \langle 1, 2, \ldots, d\rangle$, as this significantly simplifies the notation. It should be stressed, however, that many of the results actually *do* depend on the particular permutation $\mathbf{p}$ of $\{1, 2, \ldots, d\}$ that is chosen. For instance, the approximation error of our novel algorithm depends on the order in which the modes are processed.

5.1. Definition. Consider again Theorem 4.1. The solution of optimization problem (1.1) can be expressed as

$$\begin{aligned}
&\min_{\pi_1,\ldots,\pi_d} \|\mathcal{A} - \pi_1\pi_2 \cdots \pi_d\mathcal{A}\|_F^2 \qquad (5.1)\\
&\quad = \min_{\pi_1,\ldots,\pi_d} \left(\|\pi_1^{\perp}\mathcal{A}\|_F^2 + \|\pi_2^{\perp}\pi_1\mathcal{A}\|_F^2 + \cdots + \|\pi_d^{\perp}\pi_1 \cdots \pi_{d-1}\mathcal{A}\|_F^2\right),\\
&\quad = \min_{\pi_1} \left[\|\pi_1^{\perp}\mathcal{A}\|_F^2 + \min_{\pi_2} \left[\|\pi_2^{\perp}\pi_1\mathcal{A}\|_F^2 + \min_{\pi_3} \left[\cdots + \min_{\pi_d} \|\pi_d^{\perp}\pi_1 \cdots \pi_{d-1}\mathcal{A}\|_F^2\right]\right]\right].
\end{aligned}$$

Consider the minimization over π_1. It is not unreasonable to assume that the final error depends more strongly on the term $\|\pi_1^\perp \mathcal{A}\|_F^2$ than on the other terms. The subsequent terms will be minimized over (a number of) other projectors, hereby diminishing the importance of a single projector. Therefore, we are led to believe that the projector

$$\hat{\pi}_1 := \arg\min_{\pi_1} \|\pi_1^\perp \mathcal{A}\|_F^2,$$

might be a good approximation to the optimal projector. By repeating the above argument for the projector π_2, and then π_3, and so on, we arrive at

$$\begin{aligned}\min_{\pi_1,\ldots,\pi_d} \|\mathcal{A} - \pi_1\pi_2\cdots\pi_d\mathcal{A}\|_F^2 &\leq \|\hat{\pi}_1^\perp \mathcal{A}\|_F^2 + \|\hat{\pi}_2^\perp \hat{\pi}_1 \mathcal{A}\|_F^2 + \cdots + \|\hat{\pi}_d^\perp \hat{\pi}_1\cdots\hat{\pi}_{d-1}\mathcal{A}\|_F^2, \\ &= \|\mathcal{A} - \hat{\pi}_1\cdots\hat{\pi}_d\mathcal{A}\|_F^2. \end{aligned} \tag{5.2}$$

Herein, the hat-projectors are defined recursively by

$$\hat{\pi}_k := \arg\min_{\pi_k} \|\pi_k\hat{\pi}_1\cdots\hat{\pi}_{k-1}\mathcal{A}\|_F^2 = \arg\min_{U_k} \|U_k U_k^T \left[\hat{\pi}_1\cdots\hat{\pi}_{k-1}\mathcal{A}\right]_{(k)}\|_F^2. \tag{5.3}$$

wherein, as we recall, $U_k \in \mathbb{R}^{n_k \times r_k}$. Keep in mind that every permutation $\mathbf{p}$ of $\{1, 2, \ldots, d\}$ yields a different optimization problem, and solution, of the above form. The novel truncation algorithm we propose, consists of computing the solution to this recursively defined optimization problem.

By comparing Eq. (5.2) and Eq. (5.3) with Eq. (4.1), we arrive at an interesting relationship between the minimization problem solved by the T-MLSVD, and optimization problem (5.3). Clearly,

$$\min_{\pi_1,\ldots,\pi_d} \|\mathcal{A} - \pi_1\pi_2\cdots\pi_d\mathcal{A}\|_F^2 \leq \sum_{i=1}^{d} \min_{\pi_i} \|\pi_i\hat{\pi}_1\cdots\hat{\pi}_{i-1}\mathcal{A}\|_F^2 \leq \sum_{i=1}^{d} \min_{\pi_i} \|\pi_i\mathcal{A}\|_F^2,$$

because additional multilinear orthogonal projections can only decrease the Frobenius norm. The optimization problem solved by the truncated MLSVD can thus be considered as a simplification of optimization problem (5.3). This provides a strong rationale to investigate the solution of Eq. (5.3).

The solution to optimization problem (5.3), for a fixed $1 \leq k \leq d$, can be obtained from a truncated singular value decomposition of the mode-k unfolding of $\hat{\pi}_1\cdots\hat{\pi}_{k-1}\mathcal{A}$. It is a tensor of the same size as $\mathcal{A}$, namely $n_1 \times n_2 \times \cdots \times n_d$. However, if we let $\hat{\pi}_j\mathcal{A} := (I, \ldots, I, \hat{U}_j\hat{U}_j^T, I, \ldots, I) \cdot \mathcal{A}$, then

$$\begin{aligned}\hat{U}_k &= \arg\min_{U_k} \|U_k U_k^T \mathcal{A}_{(k)}(\hat{U}_1\hat{U}_1^T \otimes \hat{U}_2\hat{U}_2^T \otimes \cdots \otimes \hat{U}_{k-1}\hat{U}_{k-1}^T \otimes I \otimes \cdots \otimes I)^T\|_F^2, \\ &= \arg\min_{U_k} \|U_k U_k^T \mathcal{A}_{(k)}(\hat{U}_1 \otimes \hat{U}_2 \otimes \cdots \otimes \hat{U}_{k-1} \otimes I \otimes \cdots \otimes I)\|_F^2,\end{aligned}$$

due to Property 2.2. Consequently, it is also possible to compute the optimal projector by means of a truncated SVD of $(\hat{U}_1^T, \ldots, \hat{U}_{k-1}^T, I, \ldots, I) \cdot \mathcal{A}$, which is a tensor of size $r_1 \times r_2 \times \cdots \times r_{k-1} \times n_k \times \cdots \times n_d$. Whenever the tensor is strongly truncated, r_i is much smaller than n_i, which improves the computational performance. Based on this observation, we present Algorithm 1, which computes the solution of optimization problem (5.3). Notice that it expects the chosen permutation $\mathbf{p}$ as input. A MATLAB® implementation of this algorithm, using the Tensor Toolbox [2], is presented in the appendix.

Algorithm 1 Computing the sequentially truncated singular value decomposition

$\hat{\mathcal{S}} \leftarrow \mathcal{A}$
for $k = p_1, p_2, \ldots, p_d$ **do**
  { Compute the left singular vectors of $\hat{\mathcal{S}}_{(k)}$ }
  $\hat{\mathcal{S}}_{(k)} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$, with $U_1 \in \mathbb{R}^{n_k \times r_k}$
  $\hat{U}_k \leftarrow U_1$
  $\hat{\mathcal{S}}_{(k)} \leftarrow \Sigma_1 V_1^T$
end for

We can now formally define the new truncation strategy for the MLSVD.

DEFINITION 5.1 (Sequentially truncated singular value decomposition). *A rank-*$(r_1, \ldots, r_d)$ sequentially truncated singular value decomposition (ST-MLSVD) *of a tensor* $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$*, corresponding to the permutation* $\mathbf{p} = \langle 1, 2, \ldots, d \rangle$*, is an approximate decomposition of the form*

$$\mathcal{A} \approx (\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_d) \cdot \hat{\mathcal{S}} =: \hat{\mathcal{A}}_{\mathbf{p}} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d},$$

whose truncated core tensor *is defined as*

$$(\hat{U}_1^T, \hat{U}_2^T, \ldots, \hat{U}_d^T) \cdot \mathcal{A} =: \hat{\mathcal{S}} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d},$$

and wherein the factor matrix $\hat{U}_k \in \mathbb{R}^{n_k \times r_k}$*, has orthonormal columns for all* $1 \leq k \leq d$*. The* k^{th} partially truncated core tensor *is defined as*

$$(\hat{U}_1^T, \hat{U}_2^T, \ldots, \hat{U}_k^T, I, \ldots, I) \cdot \mathcal{A} =: \hat{\mathcal{S}}^k \in \mathbb{R}^{r_1 \times \cdots \times r_k \times n_{k+1} \times \cdots \times n_d}, \tag{5.4}$$

with $\hat{\mathcal{S}}^0 := \mathcal{A}$ *and* $\hat{\mathcal{S}}^d = \hat{\mathcal{S}}$*. The* rank-$(r_1, \ldots, r_k, n_{k+1}, \ldots, n_d)$ partial approximation to $\mathcal{A}$ *is defined as*

$$(\hat{U}_1, \hat{U}_2, \ldots, \hat{U}_k, I, \ldots, I) \cdot \hat{\mathcal{S}}^k =: \hat{\mathcal{A}}^k \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d},$$

with $\hat{\mathcal{A}}^0 := \mathcal{A}$ *and* $\hat{\mathcal{A}}^d = \hat{\mathcal{A}}$*.*

The factor matrix $\hat{U}_k$*,* $1 \leq k \leq d$*, is obtained from the (truncated) singular value decomposition of the* $(k-1)^{th}$ *partially truncated core tensor, as follows*

$$\hat{\mathcal{S}}_{(k)}^{k-1} = U_k \Sigma_k V_k^T = \begin{bmatrix} \hat{U}_k & \tilde{U}_k \end{bmatrix} \begin{bmatrix} \hat{\Sigma}_k & \\ & \tilde{\Sigma}_k \end{bmatrix} \begin{bmatrix} \hat{V}_k^T \\ \tilde{V}_k^T \end{bmatrix}, \tag{5.5}$$

wherein

$$\hat{\Sigma}_k = \operatorname{diag}(\hat{\sigma}_{k,1}, \hat{\sigma}_{k,2}, \ldots, \hat{\sigma}_{k,r_k}) \quad \textit{and} \quad \tilde{\Sigma}_k = \operatorname{diag}(\tilde{\sigma}_{k,r_k+1}, \tilde{\sigma}_{k,r_k+2}, \ldots, \tilde{\sigma}_{k,n_k}).$$

The factor matrix $\hat{U}_k$ *is the matrix of the* r_k *dominant left singular vectors of the mode-k vector space of* $\hat{\mathcal{S}}^{k-1}$*.*

Alternatively, we write

$$\hat{\mathcal{A}} := \hat{\pi}_1 \hat{\pi}_2 \cdots \hat{\pi}_d \mathcal{A} := (\hat{U}_1 \hat{U}_1^T, \hat{U}_2 \hat{U}_2^T, \ldots, \hat{U}_d \hat{U}_d^T) \cdot \mathcal{A}.$$

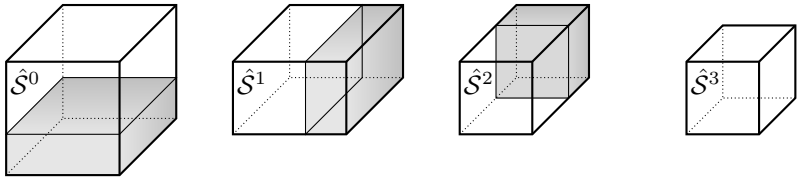


Figure 5.1: A graphical depiction of the sequential truncation of a third-order tensor.

We refer to this (approximate) decomposition as a *sequentially truncated multilinear singular value decomposition*. The sequential truncation algorithm can be understood to compute *a sequence* of approximations $\langle \hat{\mathcal{A}}^0, \hat{\mathcal{A}}^1, \ldots, \hat{\mathcal{A}}^d \rangle$, such that the multilinear rank of $\hat{\mathcal{A}}^k$ corresponds in the first k modes to the desired dimension of the corresponding vector space. We term our approach "sequential" in that the mode-k projector depends on the previously computed projectors. In the remainder, we will always denote the ST-MLSVD projector onto mode k by $\hat{\pi}_k \mathcal{A} := (I, \ldots, I, \hat{U}_k \hat{U}_k^T, I, \ldots, I) \cdot \mathcal{A}$.

In Figure 5.1, we illustrate the operation of Algorithm 1. It represents the truncation of a third-order tensor $\mathcal{A} = \hat{\mathcal{S}}^0$ to the ST-MLSVD core tensor $\hat{\mathcal{S}} = \hat{\mathcal{S}}^3$. The algorithm can be understood as follows. First, the singular value decomposition of the mode-1 vector space is computed. By projecting onto the span of the matrix of left singular vectors $\hat{U}_1$, the "energy" in the tensors is reordered. More formally, as in the T-MLSVD, Eq. (3.3) holds for $\hat{\mathcal{S}}^0$, which we will prove later. The Frobenius norm of the area shaded in gray is $\|\tilde{\Sigma}_1\|_F^2$, whereas the white area has a Frobenius norm equal to $\|\hat{\Sigma}_1\|_F^2$. By projecting onto the dominant subspace of the mode-1 vector space, we retain only the non-shaded area of $\hat{\mathcal{S}}^0$, resulting in the approximation $\hat{\mathcal{S}}^1$. The shaded area is, essentially, set to zero. Its energy is lost. In the second mode we again compute the SVD, and by projecting onto the space spanned by the left singular vectors, the energy is again reordered according to Eq. (3.3). To obtain the next approximation, the shaded area of $\hat{\mathcal{S}}^1$ is set to zero. The procedure proceeds analogously in the last mode. In the end, the ST-MLSVD core tensor is obtained.

5.2. Basic properties. We continue to investigate some of the properties of this alternative truncation strategy for the multilinear singular value decomposition. First, we demonstrate that the ST-MLSVD reduces to the multilinear singular value decomposition [5], in absence of truncation.

THEOREM 5.2 (An alternative MLSVD algorithm). *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and set $\mathcal{S}^0 := \mathcal{A}$. In every step $1 \leq k \leq d$, $\mathcal{A}$ is decomposed as*

$$(U_1^T, U_2^T, \ldots, U_k^T, I, \ldots, I) \cdot \mathcal{A} =: \mathcal{S}^k \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}. \tag{5.6}$$

If $k \leq d-1$, the next $n_{k+1} \times n_{k+1}$ factor matrix *U_{k+1} is determined as the matrix of left singular vectors in the SVD of the mode-$(k+1)$ unfolding of $\mathcal{S}^k$. That is,*

$$\mathcal{S}^k_{(k+1)} := U_{k+1} \Sigma_{k+1} V_{k+1}^T. \tag{5.7}$$

If the core tensor *is defined as*

$$\mathcal{S} := (U_1^T, U_2^T, \ldots, U_d^T) \cdot \mathcal{A} = \mathcal{S}^d,$$

then the MLSVD decomposition of $\mathcal{A}$ is given by

$$\mathcal{A} = (U_1, U_2, \ldots, U_d) \cdot \mathcal{S}. \tag{5.8}$$

Proof. The proof is largely inspired by the proof of Theorem 3.1 in [5]. Consider any $1 \leq k \leq d$. By Eq. (5.8), we have

$$\mathcal{A}_{(k)} = U_k \mathcal{S}_{(k)} (U_1 \otimes \cdots \otimes U_{k-1} \otimes U_{k+1} \otimes \cdots \otimes U_d)^T. \tag{5.9}$$

From Eq. (5.6) and Eq. (5.7), we derive that

$$\mathcal{S}_{(k)}^{k-1} = I\mathcal{A}_{(k)} (U_1 \otimes \cdots \otimes U_{k-1} \otimes I \otimes \cdots \otimes I) = U_k \Sigma_k V_k^T,$$

and, because the Kronecker product preserves orthogonality (Property 2.6),

$$\mathcal{A}_{(k)} = U_k \Sigma_k V_k^T (U_1 \otimes \cdots \otimes U_{k-1} \otimes I \otimes \cdots \otimes I)^T. \tag{5.10}$$

Combining the above with Eq. (5.9), while eliminating U_k, yields

$$\mathcal{S}_{(k)} (U_1 \otimes \cdots \otimes U_{k-1} \otimes U_{k+1} \otimes \cdots \otimes U_d)^T = \Sigma_k V_k^T (U_1 \otimes \cdots \otimes U_{k-1} \otimes I \otimes \cdots \otimes I)^T.$$

Again exploiting Property 2.6, and using the orthogonality of $\{U_j\}_j$, results in

$$\mathcal{S}_{(k)} = \Sigma_k V_k^T (I \otimes \cdots \otimes I \otimes U_{k+1} \otimes \cdots \otimes U_d) := \Sigma_k V_k^T Z^T. \tag{5.11}$$

Let $V_k := \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_{n_k} \end{bmatrix}$. Notice that ZV_k has orthonormal columns.

We are now ready to prove the two properties stated in Theorem 3.1, demonstrating that the decomposition in Eq. (5.8) is the (unique) MLSVD decomposition. Consider first the all-orthogonality of $\mathcal{S}$. By its construction in Eq. (5.7),

$$\Sigma_k := \operatorname{diag}(\sigma_{k,1}, \sigma_{k,2}, \dots, \sigma_{k,n_k}), \quad \text{where} \quad \sigma_{k,1} \geq \sigma_{k,2} \geq \dots \geq \sigma_{k,n_k} \geq 0. \tag{5.12}$$

Notice that row α of $\mathcal{S}_{(k)}$ contains the values of the core tensor for which the index in mode k is fixed to α, i.e. it is the vectorization of $\mathcal{S}_{i_k=\alpha}$. Therefore, from Eq. (5.11),

$$\langle \mathcal{S}_{i_k=\alpha}, \mathcal{S}_{i_k=\beta} \rangle = \langle \sigma_{k,\alpha} Z\mathbf{v}_\alpha, \sigma_{k,\beta} Z\mathbf{v}_\beta \rangle = \begin{cases} \sigma_{k,\alpha}^2 & \text{if } \alpha = \beta \\ 0 & \text{otherwise} \end{cases},$$

for any $1 \leq \alpha, \beta \leq n_k$. That proves the all-orthogonality condition.

The proof of ordering follows readily from

$$\|\mathcal{S}_{i_k=\alpha}\| = \langle \mathcal{S}_{i_k=\alpha}, \mathcal{S}_{i_k=\alpha} \rangle = \sigma_{k,\alpha}^2$$

and the ordering of the mode-k singular values in Eq. (5.12). ∎

Consequently, the ST-MLSVD inherits the properties of the MLSVD in absence of truncation. In particular, the ST-MLSVD of an order-2 tensor reduces to the matrix SVD, even in presence of truncation, which is not hard to prove.

We note that the ST-MLSVD algorithm requires less floating point operations to compute the approximation.

PROPERTY 5.3. *Let $\mathcal{A}$ be an order-d tensor of size $n_1 \times n_2 \times \cdots \times n_d$. Assume that $\mathcal{A}$ is truncated to a rank-$(r_1, r_2, \dots, r_d)$ tensor by the ST-MLSVD. Assume an $O(m^2 n)$ algorithm to compute the singular value decomposition of an $m \times n$ matrix, $m \leq n$. Then, the ST-MLSVD requires*

$$O\left(\sum_{k=1}^{d} r_1 \cdots r_{k-1} \cdot n_k^2 \cdot n_{k+1} \cdots n_d + \sum_{k=1}^{d} r_1 \cdots r_k \cdot n_{k+1} \cdots n_d \right)$$

operations to compute the approximation by Algorithm 1.

Proof. In every mode, the singular value decomposition of an $n_k \times r_1 \cdots r_{k-1} \cdot n_{k+1} \cdots n_d$ matrix should be computed, leading to the first summation. The next partially truncated core tensor can be obtained, simply by scaling the right singular vectors with the corresponding singular values. This amount of operations conforms to the latter summation. ∎

The T-MLSVD, on the other hand, requires

$$O\left(\sum_{k=1}^{d} n_1 \cdots n_{k-1} \cdot n_k^2 \cdot n_{k+1} \cdots n_d + \sum_{k=1}^{d} r_1 \cdots r_{k-1} \cdot r_k \cdot n_k \cdot n_{k+1} \cdots n_d\right)$$

operations to compute the core tensor and the factor matrices. The first summation is due to the singular value decompositions, in every mode, of an $n_k \times \Pi_{j \neq k} n_j$ matrix, in order to compute the factor matrices. The second summation is required to compute the core tensor, by means of d matrix multiplications. The T-MLSVD is clearly much more expensive than a ST-MLSVD if $r_i \ll n_i$. Consider for instance a rank-1 approximation to an order-d cubic tensor of size $n \times n \times \cdots \times n$. Then the T-MLSVD requires $O(dn^{d+1})$ operations, versus $O(n^{d+1} + n^d + \cdots + n)$ for the ST-MLSVD. The speedup of ST-MLSVD over T-MLSVD can then be quite close to the order, d, of the tensor. We illustrate this numerically, in Section 7.3.

Notice that the permutation is very relevant in Property 5.3, and may lead to large differences in the total number of operations required to compute the decomposition. Nevertheless, the ST-MLSVD algorithm always requires less operations than the T-MLSVD, regardless of the chosen permutation $\mathbf{p}$ or the rank of the approximation. It is a good heuristic to sort the modes by decreasing compression ratio, i.e. by n_k/r_k, and compute the ST-MLSVD in that order. We have observed experimentally that this order often leads to a good approximation error as well.

5.3. Approximation error. Another important aspect of the ST-MLSVD concerns its approximation error. That will be a major topic in the remainder of this paper. First, it is shown that the approximation error of the ST-MLSVD can be expressed in terms of the singular values that are obtained readily from an execution of Algorithm 1.

THEOREM 5.4 (Error of the ST-MLSVD). *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ and let $\hat{\mathcal{A}}$ be the rank-$(r_1, \ldots, r_d)$ sequentially truncated singular value decomposition of $\mathcal{A}$, as defined in Definition 5.1. Then*

$$\|\mathcal{A} - \hat{\mathcal{A}}\|_F^2 = \sum_{k=1}^{d} \sum_{i_k=r_k+1}^{n_k} \tilde{\sigma}_{k,i_k}^2, \tag{5.13}$$

i.e. the error is the sum of the squares of the discarded singular values.

Proof. We derive an explicit formula for the error of two successive approximations $\hat{\mathcal{A}}^k$ and $\hat{\mathcal{A}}^{k+1}$, as they are defined in Definition 5.1, with $0 \leq k \leq d-1$. That is,

$$\begin{aligned}\epsilon_k^2 &:= \|\hat{\mathcal{A}}^k - \hat{\mathcal{A}}^{k+1}\|_F^2 \\ &= \|(\hat{U}_1, \ldots, \hat{U}_k, I, \ldots, I) \cdot \hat{\mathcal{S}}^k - (\hat{U}_1, \ldots, \hat{U}_{k+1}, I, \ldots, I) \cdot \hat{\mathcal{S}}^{k+1}\|_F^2.\end{aligned}$$

We notice that the factor matrices have orthonormal columns. Therefore,

$$\begin{aligned}\epsilon_k^2 &= \left\| \left(\hat{\mathcal{S}}^k_{(k+1)} - \hat{U}_{k+1} \hat{\mathcal{S}}^{k+1}_{(k+1)} \right) (\hat{U}_1 \otimes \cdots \otimes \hat{U}_k \otimes I \otimes \cdots \otimes I)^T \right\|_F^2, \\ &= \left\| \hat{\mathcal{S}}^k_{(k+1)} - \hat{U}_{k+1} \hat{\mathcal{S}}^{k+1}_{(k+1)} \right\|_F^2, \end{aligned} \tag{5.14}$$

From the definition of the partially truncated core tensor in Eq. (5.4), we note that

$$\begin{aligned}\hat{U}_{k+1}^T \hat{\mathcal{S}}_{(k+1)}^k &= \Big[(I,\ldots,I,\hat{U}_{k+1}^T,I,\ldots,I)\cdot\hat{\mathcal{S}}^k\Big]_{(k+1)} \\ &= \Big[(\hat{U}_1^T,\ldots,\hat{U}_{k+1}^T,I,\ldots,I)\cdot\mathcal{A}\Big]_{(k+1)} = \hat{\mathcal{S}}_{(k+1)}^{k+1}.\end{aligned}$$

However, from Eq. (5.5) it is also clear that

$$\hat{U}_{k+1}\hat{\mathcal{S}}_{(k+1)}^{k+1} = \hat{U}_{k+1}\hat{U}_{k+1}^T\hat{\mathcal{S}}_{(k+1)}^k = \hat{U}_{k+1}\hat{\Sigma}_{k+1}\hat{V}_{k+1}^T.$$

Substituting the above in Eq. (5.14) and using Eq. (5.5) again, we obtain

$$\begin{aligned}\epsilon_k^2 &= \left\|\hat{U}_{k+1}\hat{\Sigma}_{k+1}\hat{V}_{k+1}^T + \tilde{U}_{k+1}\tilde{\Sigma}_{k+1}\tilde{V}_{k+1}^T - \hat{U}_{k+1}\hat{\Sigma}_{k+1}\hat{V}_{k+1}^T\right\|_F^2 \\ &= \left\|\tilde{U}_{k+1}\tilde{\Sigma}_{k+1}\tilde{V}_{k+1}^T\right\|_F^2 = \sum_{i_k=r_k+1}^{n_k}\tilde{\sigma}_{k,i_k}^2. \end{aligned} \tag{5.15}$$

Next, we note that $\hat{\mathcal{A}}^k - \hat{\mathcal{A}}^{k+1} = \hat{\pi}_{k+1}^{\perp}\hat{\pi}_1\cdots\hat{\pi}_k\mathcal{A}$, so that we can immediately apply Theorem 4.1. That results in

$$\begin{aligned}\|\mathcal{A}-\hat{\mathcal{A}}^k\|_F^2 &= \|\hat{\pi}_1^{\perp}\mathcal{A}\|_F^2 + \|\hat{\pi}_2^{\perp}\hat{\pi}_1\mathcal{A}\|_F^2 + \cdots + \|\hat{\pi}_d^{\perp}\hat{\pi}_1\hat{\pi}_2\cdots\hat{\pi}_d\mathcal{A}\|_F^2, \\ &= \sum_{k=1}^{d}\epsilon_k^2 = \sum_{k=1}^{d}\sum_{i_k=r_k+1}^{n_k}\tilde{\sigma}_{k,i_k}^2,\end{aligned}$$

concluding the proof. ∎

It should be stressed that the approximation error of the ST-MLSVD depends on the permutation $\mathbf{p}$, as the singular values will in general be different.

Example. Consider, for instance, the third-order tensor $\mathcal{A}\in\mathbb{R}^{3\times3\times3}$:

$$\mathcal{A}_{:,:,1} = \begin{bmatrix}2&4&7\\5&6&3\\9&3&5\end{bmatrix}, \qquad \mathcal{A}_{:,:,2} = \begin{bmatrix}7&5&3\\9&2&8\\9&2&3\end{bmatrix}, \qquad \mathcal{A}_{:,:,3} = \begin{bmatrix}8&4&6\\3&2&5\\9&3&4\end{bmatrix}.$$

If we approximate this tensor by a rank-$(2,2,2)$ ST-MLSVD decomposition, the following errors are obtained for the different permutations of $\{1,2,3\}$:

$$\begin{aligned}&\|\mathcal{A}-\hat{\mathcal{A}}_{\langle1,2,3\rangle}\|_F = 8.1912, \quad \|\mathcal{A}-\hat{\mathcal{A}}_{\langle1,3,2\rangle}\|_F = 8.1932, \quad \|\mathcal{A}-\hat{\mathcal{A}}_{\langle2,1,3\rangle}\|_F = 7.4799,\\ &\|\mathcal{A}-\hat{\mathcal{A}}_{\langle2,3,1\rangle}\|_F = 7.4497, \quad \|\mathcal{A}-\hat{\mathcal{A}}_{\langle3,1,2\rangle}\|_F = 7.5001, \quad \|\mathcal{A}-\hat{\mathcal{A}}_{\langle1,2,3\rangle}\|_F = 7.4835.\end{aligned}$$

The error of the rank-$(2,2,2)$ T-MLSVD approximation is 8.8188, which is worse than every ST-MLSVD approximation. ♦

The error of the ST-MLSVD can also be bounded in terms of the "original" singular values of the different unfoldings of the tensor. This bound is presented next.

THEOREM 5.5 (Error bound of the ST-MLSVD). *Let $\mathcal{A}\in\mathbb{R}^{n_1\times\cdots\times n_d}$ and let $\hat{\mathcal{A}}$ be the rank-$(r_1,\ldots,r_d)$ sequentially truncated singular value decomposition of $\mathcal{A}$, as defined in Definition 5.1. Let*

$$\mathcal{A}_{(k)} = \begin{bmatrix}\bar{U}_k & \tilde{U}_k\end{bmatrix}\begin{bmatrix}\bar{\Sigma}_k & \\ & \tilde{\Sigma}_k\end{bmatrix}\begin{bmatrix}\bar{V}_k^T\\ \tilde{V}_k^T\end{bmatrix}, \qquad \text{with } \hat{U}_k\in\mathbb{R}^{n_k\times r_k},$$

be the singular value decomposition in mode k. Then

$$\min_k \|\tilde{\Sigma}_k\|_F^2 \leq \|\mathcal{A} - \hat{\mathcal{A}}\|_F^2 \leq \sum_{k=1}^{d} \|\tilde{\Sigma}_k\|_F^2$$

are bounds on the error of the ST-MLSVD.

Proof. The proof of the upper bound follows from

$$\|\hat{\pi}_k^{\perp} \hat{\pi}_1 \cdots \hat{\pi}_{k-1} \mathcal{A}\|_F^2 \leq \|\bar{\pi}_k^{\perp} \hat{\pi}_1 \cdots \hat{\pi}_{k-1} \mathcal{A}\|_F^2 \leq \|\bar{\pi}_k^{\perp} \mathcal{A}\|_F^2 = \|\tilde{\Sigma}_k\|_F^2,$$

for all $1 \leq k \leq d$. Herein, $\bar{\pi}_k$ is the T-MLSVD projector in mode k. The first inequality in the above formula is due to the fact that the $\hat{\pi}_k$ projector is optimal, as it is derived from a truncated SVD. The second inequality is due to Eq. (2.4).

The lower bound is a consequence of Theorem 4.1 and the fact that the approximation error of the ST-MLSVD depends on the order in which it is computed. ∎

The sequentially truncated MLSVD thus has the same upper bound on the error as the T-MLSVD, but its lower bound is better.

5.4. Invariance to choice of tensor basis. We show that choice of tensor basis does not affect the outcome of the T-MLSVD or ST-MLSVD algorithm. This property does not automatically extend to other algorithms.

THEOREM 5.6 (Invariance of tensor basis). *Let $\mathcal{A}, \mathcal{L} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be order-d tensors. Let $\mathcal{A}$ and $\mathcal{L}$ be related via a multilinear change of basis, that is,*

$$\mathcal{A} = (Z_1, Z_2, \ldots, Z_d) \cdot \mathcal{L}, \quad \text{wherein } Z_k \in \mathbb{R}^{n_k \times n_k} \text{ is orthogonal for all } k. \tag{5.16}$$

Let, furthermore,

$$\breve{\mathcal{L}} := (\breve{U}_1 \breve{U}_1^T, \breve{U}_2 \breve{U}_2^T, \ldots, \breve{U}_d \breve{U}_d^T) \cdot \mathcal{L}$$

be the rank-$(r_1, r_2, \ldots, r_d)$ T-MLSVD (ST-MLSVD) approximation to $\mathcal{L}$, as in Theorem 3.2 (Definition 5.1). Then, the T-MLSVD (ST-MLSVD) approximation of the same multilinear rank of $\mathcal{A}$ is given by

$$\breve{\mathcal{A}} = (Z_1, Z_2, \ldots, Z_d) \cdot \breve{\mathcal{L}},$$

i.e., the approximation error of the T-MLSVD (ST-MLSVD) algorithm is independent of the tensor basis in which $\mathcal{A}$ is expressed.

Proof. We begin with the proof for the T-MLSVD. By Eq. (5.16),

$$\begin{aligned}
\mathcal{A}_{(k)} &= Z_k \mathcal{L}_{(k)} (Z_1 \otimes \cdots \otimes Z_{k-1} \otimes Z_{k+1} \otimes \cdots \otimes Z_d)^T \\
&= Z_k \begin{bmatrix} \breve{U}_k & \tilde{U}_k \end{bmatrix} \begin{bmatrix} \breve{\Sigma}_k & \\ & \tilde{\Sigma}_k \end{bmatrix} \begin{bmatrix} \breve{V}_k^T \\ \tilde{V}_k^T \end{bmatrix} (Z_1 \otimes \cdots \otimes Z_{k-1} \otimes Z_{k+1} \otimes \cdots \otimes Z_d)^T,
\end{aligned}$$

where in the second equality, we have replaced $\mathcal{L}_{(k)}$ with its singular value decomposition, wherein $\breve{U}_k \in \mathbb{R}^{n_k \times r_k}$. Clearly, the first two factors form an orthogonal matrix and equally so for the last two factors. Hence, the above represents the unique SVD of $\mathcal{A}_{(k)}$. Consequently, truncating it to the first r_k singular values yields the orthogonal projector $\breve{\pi}_k$ in mode k. That is,

$$\breve{\pi}_k \mathcal{A} = (I, \ldots, I, Z_k \breve{U}_k \breve{U}_k^T Z_k^T, I, \ldots, I) \cdot \mathcal{A}. \tag{5.17}$$

As the above comments hold for all $1 \leq k \leq d$, we soon arrive at

$$\begin{aligned}\breve{\mathcal{A}} = \breve{\pi}_1\breve{\pi}_2\cdots\breve{\pi}_d\mathcal{A} &= (Z_1\breve{U}_1\breve{U}_1^T Z_1^T, Z_2\breve{U}_2\breve{U}_2^T Z_2^T, \ldots, Z_d\breve{U}_d\breve{U}_d^T Z_d^T)\cdot\mathcal{A},\\ &= (Z_1\breve{U}_1\breve{U}_1^T, Z_2\breve{U}_2\breve{U}_2^T, \ldots, Z_d\breve{U}_d\breve{U}_d^T)\cdot\mathcal{L},\\ &= (Z_1, Z_2, \ldots, Z_d)\cdot\breve{\mathcal{L}},\end{aligned}$$

which completes the proof for the T-MLSVD.

To prove the theorem for the ST-MLSVD, we proceed along the same line. In the first unfolding, the ST-MLSVD projector is the same as the T-MLSVD projector. It is thus given by $\breve{\pi}_1\mathcal{A} = (Z_1\breve{U}_1\breve{U}_1^T Z_1^T, I, \ldots, I)\cdot\mathcal{A}$. In the second mode, the ST-MLSVD algorithm computes the SVD of

$$\begin{aligned}\mathcal{A}_{(2)}(Z_1\breve{U}_1\breve{U}_1^T Z_1^T \otimes I) &= Z_2\mathcal{L}_{(2)}(Z_1\otimes Z_3\otimes\cdots\otimes Z_d)^T(Z_1\breve{U}_1\breve{U}_1^T Z_1^T\otimes I),\\ &= Z_2\mathcal{L}_{(2)}(\breve{U}_1\breve{U}_1^T\otimes I)(Z_1\otimes Z_3\otimes\cdots\otimes Z_d)^T,\end{aligned}$$

where the first equality is due to the change of basis in Eq. (5.16). Next, we note that

$$\mathcal{A}_{(2)}(Z_1\breve{U}_1\breve{U}_1^T Z_1^T\otimes I) = Z_2\begin{bmatrix}\breve{U}_2 & \tilde{U}_2\end{bmatrix}\begin{bmatrix}\breve{\Sigma}_2 & \\ & \tilde{\Sigma}_2\end{bmatrix}\begin{bmatrix}\breve{V}_2^T\\ \tilde{V}_2^T\end{bmatrix}(Z_1\otimes Z_3\otimes\cdots\otimes Z_d)^T,$$

from the definition of the ST-MLSVD algorithm. Now, the first two factors again form an orthogonal matrix, as do the last two factors. Consequently, the right-hand side is the SVD of the left-hand side. The second ST-MLSVD projector is thus

$$\breve{\pi}_2\mathcal{A} = (I, Z_2\breve{U}_2\breve{U}_2^T Z_2^T, I, \ldots, I)\cdot\mathcal{A}.$$

It is not hard to show, by induction in the above manner, that, in general, Eq. (5.17) holds for the ST-MLSVD projectors as well (with different matrices, of course). Consequently, by repeating the argument employed in the previous part of the proof, we arrive at the desired conclusion. □

From an algebraic viewpoint, the above theorem reveals that both the T-MLSVD as the ST-MLSVD operate on *tensors as elements of a vector space*, not just on the d-array representing their coordinates with respect to some tensor basis. In fact, this theorem proves that yet another property of the SVD generalizes to the (S)T-MLSVD.

6. Error with respect to the truncated MLSVD. In this rather technical section, we investigate the relationship between the approximation error of the classic T-MLSVD truncation strategy and the ST-MLSVD truncation strategy. Throughout this section, we will denote the rank-$(r_1, r_2, \ldots, r_d)$ ST-MLSVD approximation to $\mathcal{A}\in\mathbb{R}^{n_1\times\cdots\times n_d}$, corresponding to permutation $\mathbf{p}$, by

$$\hat{\pi}_{p_1}\hat{\pi}_{p_2}\cdots\hat{\pi}_{p_d}\mathcal{A} =: \hat{\mathcal{A}}_{\mathbf{p}}\in\mathbb{R}^{n_1\times\cdots\times n_d}.$$

Whenever the particular permutation $\mathbf{p}$ is of no importance, the subscript in $\hat{\mathcal{A}}_{\mathbf{p}}$ will be dropped. The T-MLSVD approximation of the same multilinear rank to $\mathcal{A}$ will be denoted by

$$\bar{\pi}_1\bar{\pi}_2\cdots\bar{\pi}_d\mathcal{A} =: \bar{\mathcal{A}}\in\mathbb{R}^{n_1\times\cdots\times n_d}.$$

Recall, from Section 5.1, that the optimization problem solved by the T-MLSVD can be considered an approximation to the optimization problem that ST-MLSVD

solves. Both approximate the solution of the actual problem (1.1). Given this observation, we wonder whether the ST-MLSVD approximation is better than the T-MLSVD.

CONJECTURE 6.1 (Error of the ST-MLSVD w.r.t. the T-MLSVD). *Let a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be given. Let $\hat{\mathcal{A}}_{\mathbf{p}*} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be the best rank-$(r_1, \ldots, r_d)$ sequentially truncated multilinear singular value decomposition of $\mathcal{A}$. Let $\bar{\mathcal{A}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be the rank-$(r_1, \ldots, r_d)$ truncated multilinear singular value decomposition of $\mathcal{A}$. Then,*

$$\|\mathcal{A} - \hat{\mathcal{A}}_{\mathbf{p}^*}\|_F^2 \overset{?}{\leq} \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2.$$

Unfortunately, the above conjecture does *not* hold, in general, as the next counterexample shows.

6.1. A counterexample. Let us consider the following fourth order tensor

$$\mathcal{A}_{:,:,1,1} = \begin{bmatrix} 0.5 & -1.7 \\ -1.3 & -0.6 \end{bmatrix}, \qquad \mathcal{A}_{:,:,2,1} = \begin{bmatrix} -2.4 & -0.1 \\ -0.7 & 1.4 \end{bmatrix},$$
$$\mathcal{A}_{:,:,1,2} = \begin{bmatrix} 0.1 & 0.1 \\ 2.2 & -0.8 \end{bmatrix}, \qquad \mathcal{A}_{:,:,2,2} = \begin{bmatrix} -0.3 & -2.5 \\ 0.0 & 0.3 \end{bmatrix}.$$

The rank-$(1,1,1)$ T-MLSVD approximation is given by

$$\bar{\mathcal{A}} = \left(\begin{bmatrix} -0.97325 \\ -0.22975 \end{bmatrix}, \begin{bmatrix} -0.78940 \\ 0.61388 \end{bmatrix}, \begin{bmatrix} -0.31546 \\ 0.94894 \end{bmatrix}, \begin{bmatrix} -0.88167 \\ 0.47186 \end{bmatrix} \right) \cdot \begin{bmatrix} 2.57934 \end{bmatrix}.$$

The best ST-MLSVD approximation corresponds to the permutation $\mathbf{p}^* = \langle 1, 3, 2, 4 \rangle$. It is given by

$$\hat{\mathcal{A}}_{\mathbf{p}^*} = \left(\begin{bmatrix} -0.97325 \\ -0.22975 \end{bmatrix}, \begin{bmatrix} -0.97310 \\ 0.23037 \end{bmatrix}, \begin{bmatrix} -0.09956 \\ 0.99503 \end{bmatrix}, \begin{bmatrix} -0.99692 \\ 0.07841 \end{bmatrix} \right) \cdot \begin{bmatrix} 2.53595 \end{bmatrix}.$$

The approximation errors are given respectively by

$$\|\bar{\mathcal{A}} - \mathcal{A}\|_F^2 = 18.68700 \qquad \text{and} \qquad \|\hat{\mathcal{A}}_{\mathbf{p}^*} - \mathcal{A}\|_F^2 = 18.90896,$$

which demonstrates that the T-MLSVD is better than the ST-MLSVD approximation. The ST-MLSVD approximation is nearly 0.6% worse than T-MLSVD. This counterexample was found by randomly sampling the fourth-order tensors of size $2 \times 2 \times 2 \times 2$, whose entries were drawn from a Gamma distribution with mean 1, and truncating them to rank $(1,1,1,1)$. Our experiments yielded 7 counterexamples in ten million samples. The above example was the best counterexample.

Our experiments indicate that the counterexamples to Conjecture 6.1 are extremely thinly spread for tensors whose entries are drawn from a normal, Gamma or uniform distribution, at least for third-order tensors. Furthermore, the probability of encountering a counterexample decreases as the multilinear rank to approximate with is decreased. Somewhat related, the ST-MLSVD approximation appears to be more competitive as the tensor is further truncated, a result which is supported by the numerical experiments in Section 7.

6.2. Sufficient conditions. Here, we investigate two sufficient conditions under which Conjecture 6.1 does hold. We demonstrate that a non-trivial and non-empty set of tensors exists that satisfy one of these conditions.

First, we prove that a third-order tensor that is fully truncated in one mode, will be approximated better by the ST-MLSVD.

THEOREM 6.2. *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Let $\hat{\mathcal{A}}_{\mathbf{p}}$ be the rank-$(1, r_1, r_2)$ ST-MLSVD of $\mathcal{A}$ corresponding to the permutation $\mathbf{p}$, and let $\bar{\mathcal{A}}$ be the T-MLSVD of $\mathcal{A}$ of the same multilinear rank. Then,*

$$\|\mathcal{A} - \hat{\mathcal{A}}_{\mathbf{p}}\|_F^2 \leq \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2. \tag{6.1}$$

at least for $\mathbf{p} = \langle 1, 2, 3 \rangle$.

Proof. Let $\hat{\pi}_1$ be the ST-MLSVD projector along the first mode and $\bar{\pi}_1$ the T-MLSVD projector. Both are equal and reduce the third order tensor to a matrix. That is, $A = \bar{\pi}_1 \mathcal{A}$. From the Eckart-Young theorem, we know that the truncated singular value decomposition of A yields the best low-rank approximation [9,11]. Clearly, the ST-MLSVD will truncate the SVD of A to rank $\min\{r_1, r_2\}$. That is because it will select the r_1 left singular vectors, which define the projector $\hat{\pi}_2$, project onto them and then compute the r_2 left singular vectors of the projected matrix, which comprise the projector $\hat{\pi}_3$. However, these r_2 left singular vectors correspond to the r_2 dominant right singular vectors of A. Thus, the combined projector $\hat{\pi}_2 \hat{\pi}_3 A$ simply yields the best rank $\min\{r_1, r_2\}$ approximation of A. The T-MLSVD, on the other hand, does not compute the projectors in this manner. Hence, $\bar{\pi}_2$ and $\bar{\pi}_3$, nor their combined projector, are optimal in general. That entails

$$\|A - \hat{\pi}_2 \hat{\pi}_3 A\|_F^2 \leq \|A - \bar{\pi}_2 \bar{\pi}_3 A\|_F^2,$$
$$\|\hat{\pi}_1^{\perp} \mathcal{A}\|_F^2 + \|\hat{\pi}_1 \mathcal{A} - \hat{\pi}_2 \hat{\pi}_3 \hat{\pi}_1 \mathcal{A}\|_F^2 \leq \|\hat{\pi}_1^{\perp} \mathcal{A}\|_F^2 + \|\hat{\pi}_1 \mathcal{A} - \bar{\pi}_2 \bar{\pi}_3 \bar{\pi}_1 \mathcal{A}\|_F^2.$$

The first term can be brought into the norm, on both sides of the inequality, as in the proof of Theorem 4.2. Theorem 4.1 completes the proof. ∎

The above theorem holds for any third order tensor that is approximated by a tensor of lower multilinear rank in which at least one mode is truncated to rank one. The specific order of the rank that was assumed in Theorem 6.2, namely $(1, r_1, r_2)$, does not limit the generality because we can renumber the modes in such a manner that the theorem applies. As an immediate corollary, it is clear that third-order tensors of multilinear rank $(2, 2, 2)$, and lower, are approximated better by the ST-MLSVD than the T-MLSVD. In particular, we note that the rank-$(1, 1, 1)$ ST-MLSVD approximation is better than the T-MLSVD approximation. The ST-MLSVD is thus likely to yield a more interesting starting point for iterative algorithms that approximate the optimal rank-1 approximation to a tensor, e.g. [3,6,16,25]. Our preliminary numerical experiments confirm this suspicion, and reveal that the number of iterations required for convergence may decrease. However, more rigorous numerical experiments are necessary to confirm these results.

Next, we consider arbitrary low-rank approximations of third-order tensors. To this end, we shall rely on the next theorem.

THEOREM 6.3. *Let a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be given. Let $\hat{\mathcal{A}}_{\mathbf{p}}$ be the rank-$(r_1, \ldots, r_d)$ ST-MLSVD of $\mathcal{A}$ corresponding to the permutation $\mathbf{p}$, and let $\bar{\mathcal{A}}$ be the T-MLSVD of $\mathcal{A}$ of the same multilinear rank. If there is a permutation $\mathbf{p}$ that satisfies*

$$\hat{\pi}_{p_j} = \bar{\pi}_{p_j} \qquad \text{for all} \qquad 1 \leq j \leq d-1$$

then,

$$\|\mathcal{A} - \hat{\mathcal{A}}_{\mathbf{p}}\|_F^2 \leq \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2.$$

Proof. From Theorem 4.1, it is clear that

$$\|\hat{\pi}_{p_j}^{\perp} \hat{\pi}_{p_1} \cdots \hat{\pi}_{p_{j-1}} \mathcal{A}\|_F^2 = \|\bar{\pi}_{p_j}^{\perp} \bar{\pi}_{p_1} \cdots \bar{\pi}_{p_{j-1}} \mathcal{A}\|_F^2 \qquad \text{for all} \qquad 1 \leq j \leq d-1.$$

From Definition 5.1, it is clear that $\hat{\pi}_{p_d}$ is optimal, where $\bar{\pi}_{p_d}$ is not, as the former results from the (optimal) truncated SVD of $[\hat{\pi}_{p_1} \cdots \hat{\pi}_{p_{d-1}} \mathcal{A}]_{(d)}$. Consequently,

$$\|\hat{\pi}_{p_d}^{\perp} \hat{\pi}_{p_1} \cdots \hat{\pi}_{p_{d-1}} \mathcal{A}\|_F^2 \leq \|\bar{\pi}_{p_d}^{\perp} \bar{\pi}_{p_1} \cdots \bar{\pi}_{p_{d-1}} \mathcal{A}\|_F^2$$

which, combined with Theorem 4.1, completes the proof. ∎

The above theorem is quite simple. However, it is not clear whether tensors exist that satisfy the above conditions. From Theorem 5.6, we already know that both the T-MLSVD as the ST-MLSVD are equivalent under orthogonal transformations. De Silva and Lim [8] refer to this as O-equivalence; equivalence under the action of the orthogonal group $O_{n_1,n_2,\ldots,n_d}(\mathbb{R}) := O_{n_1}(\mathbb{R}) \times O_{n_2}(\mathbb{R}) \times \cdots \times O_{n_d}(\mathbb{R})$, wherein $O_n(\mathbb{R})$ denotes the group of orthogonal $n \times n$ matrices over $\mathbb{R}$. In [8], De Silva and Lim consider the problem of classifying third-order tensors into a finite set of "canonical tensors", so that every tensor $\mathcal{A} \in \mathbb{R}^{2\times 2\times 2}$ is equivalent to a canonical tensor under the action of the general linear group $\mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R}) \times \mathrm{GL}_2(\mathbb{R})$, wherein $\mathrm{GL}_2(\mathbb{R})$ denotes the set of non-singular 2×2 matrices over $\mathbb{R}$. This is called a classification of the GL-orbits of $\mathbb{R}^{2\times 2\times 2}$ [8]. They come up with a list of eight such orbits. In this paper, we are interested in the O-orbits of $\mathbb{R}^{n_1\times n_2\times \cdots \times n_d}$, as this would allow us to precisely characterize in which O-orbits the ST-MLSVD algorithm outperforms the T-MLSVD algorithm. However, De Silva and Lim point out that such a task may be Herculean, as, in general, "an explicit classification [of GL-orbits] must necessarily include continuous parameters" [8]. Their argument is based on counting the degrees of freedom, and applies to the O-orbits as well. Notice that a classification of order-2 tensors is possible. The singular value decomposition reveals that there is one parameterized O-orbit for the matrices of rank r. They are all equivalent to a diagonal matrix with exactly r (strictly) positive diagonal entries.

We already know from our counterexample in Section 6.1 that ST-MLSVD does not outperform T-MLSVD everywhere. We wonder whether there exist (parameterized) O-orbits of $\mathbb{R}^{n_1\times n_2\times n_3}$ such that the best ST-MLSVD approximation is no worse than the T-MLSVD approximation. This question can be answered in the affirmative, at least if $n_1 = n_2$. It is addressed by our next theorem.

THEOREM 6.4. *Let $\mathcal{A} \in \mathbb{R}^{n\times n\times m}$ be a tensor whose mode-1 unfolding is given by*

$$\mathcal{A}_{(1)} = \begin{bmatrix} A_1 & A_2 & \ldots & A_m \end{bmatrix}.$$

If there exist orthogonal matrices $Z_1, Z_2 \in \mathbb{R}^{n\times n}$ and $Z_3 \in \mathbb{R}^{m\times m}$ such that

$$Z_1^T \mathcal{A}_{(1)} (Z_2 \otimes Z_3) = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \ldots & \Lambda_m \end{bmatrix},$$

where Λ_i is a real diagonal matrix for all i, then

$$\|\mathcal{A} - \hat{\mathcal{A}}_{\mathbf{p}}\|_F^2 \leq \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2.$$

for $\mathbf{p} = \langle 1, 2, 3\rangle$.

Proof. Let us introduce an order-3 tensor $\mathcal{L} \in \mathbb{R}^{n\times n\times m}$, which is defined by

$$\mathcal{L}_{(1)} = \mathcal{L}_{(2)} = \begin{bmatrix} \Lambda_1 & \Lambda_2 & \ldots & \Lambda_m \end{bmatrix}. \tag{6.2}$$

Then, $\mathcal{A} = (Z_1, Z_2, Z_3) \cdot \mathcal{L}$. From Theorem 5.6, we know that the ST-MLSVD and T-MLSVD approximations of $\mathcal{A}$ can be obtained through a change of basis from the ST-MLSVD, respectively T-MLSVD, approximation of $\mathcal{L}$. We thus restrict our

attention to the latter tensor. We will assume, without loss of generality, that $r_1 \geq r_2$, in the remainder.

First, we consider the T-MLSVD approximation in the first two modes. Consider the singular value decomposition of $\mathcal{L}$ in the first mode. From the specific structure in Eq. (6.2), it is clear that $\mathcal{L}_{(1)} = U\Sigma V$ wherein U is a permutation matrix and

$$\Sigma = \begin{bmatrix} \Sigma' & 0 \end{bmatrix} \quad \text{with} \quad {\Sigma'}^2 = \operatorname{diag}\left(\sum_{i=1}^{m} (\Lambda_i)^2_{q_1,q_1},\ \sum_{i=1}^{m} (\Lambda_i)^2_{q_2,q_2},\ \ldots,\ \sum_{i=1}^{m} (\Lambda_i)^2_{q_n,q_n} \right), \tag{6.3}$$

where $\mathbf{q}$ is a permutation of $\{1, 2, \ldots, n\}$, such that the diagonal entries of Σ' are sorted in decreasing magnitude. Notice that the SVD is not unique whenever some of the diagonal values are equal. In that case, U is not necessarily a permutation matrix. However, the basis spanning the subspace of those equal singular values can always be chosen to consist of standard basis vectors, yielding a U that is a permutation matrix. We assume that, in the remainder, this additional constraint is imposed. The T-MLSVD truncates Σ to the first r_1 singular values. Denote by $\bar{U}_1$, the corresponding left singular vectors. Notice that the diagonal entries of Σ' contain the norms of the rows of $\mathcal{L}_{(1)}$, sorted in decreasing magnitude. Since U is a permutation matrix, the projector

$$\bar{\pi}_1 \mathcal{L} := (\bar{U}_1 \bar{U}_1^T, I, I) \cdot \mathcal{L}$$

simply "selects" the r_1 rows of $\mathcal{L}_{(1)}$ that have the largest norm. The other rows are zeroed out. More formally, let $S_{r_i} := \{q_1, q_2, \ldots, q_{r_i}\}$ denote the set of the first r_i values in $\mathbf{q}$. Then,

$$\bar{U}_1 \bar{U}_1^T = \operatorname{diag}\left([1 \in S_{r_i}], [2 \in S_{r_i}], \ldots, [n \in S_{r_i}] \right), \tag{6.4}$$

wherein we used Iverson's convenient notation [12, p 24]. That is, $[j \in S]$ is one if j is in the set S and zero otherwise.

In the second mode, the T-MLSVD algorithm does exactly the same as in the first mode. However, now it selects the r_2 largest singular values in Eq. (6.3). Let $\bar{U}_2$ denote the matrix of corresponding left singular vectors. Then, the second projector

$$\bar{\pi}_2 \mathcal{L} := (I, \bar{U}_2 \bar{U}_2^T, I) \cdot \mathcal{L},$$

selects the r_2 rows of $\mathcal{L}_{(2)}$ with the largest norm. The combined action of $\bar{\pi}_1$ and $\bar{\pi}_2$ is given by

$$\begin{aligned} [\bar{\pi}_2 \bar{\pi}_1 \mathcal{L}]_{(2)} &= \bar{U}_2 \bar{U}_2^T \mathcal{L}_{(2)} (\bar{U}_1 \bar{U}_1^T \otimes I), \\ &= \begin{bmatrix} \bar{U}_2 \bar{U}_2^T \Lambda_1 \bar{U}_1 \bar{U}_1^T & \bar{U}_2 \bar{U}_2^T \Lambda_2 \bar{U}_1 \bar{U}_1^T & \ldots & \bar{U}_2 \bar{U}_2^T \Lambda_m \bar{U}_1 \bar{U}_1^T \end{bmatrix}, \\ &= \begin{bmatrix} \bar{\Lambda}_1 & \bar{\Lambda}_2 & \ldots & \bar{\Lambda}_m \end{bmatrix}, \end{aligned}$$

with $\bar{\Lambda}_i := \bar{U}_2 \bar{U}_2^T \Lambda_i \bar{U}_1 \bar{U}_1^T$. We can simplify the latter to

$$\begin{aligned} \bar{\Lambda}_i &= \operatorname{diag}([1 \in S_{r_2}], \ldots, [n \in S_{r_2}]) \Lambda_i \operatorname{diag}([1 \in S_{r_1}], \ldots, [n \in S_{r_1}]), \\ &= \operatorname{diag}([1 \in S_{r_2}], \ldots, [n \in S_{r_2}]) \Lambda_i, \end{aligned}$$

where the last equation follows from the observation that $[j \in S_{r_2}] \cdot [j \in S_{r_1}] = [j \in S_{r_2} \cap S_{r_1}]$ and $r_2 \leq r_1$ by assumption.

Second, we consider the ST-MLSVD. Clearly, the mode-1 ST-MLSVD projector $\hat{\pi}_1$ equals the T-MLSVD projector $\bar{\pi}_1$. In the second mode, ST-MLSVD computes the singular value decomposition of

$$\begin{aligned}\mathcal{L}_{(2)}(\hat{U}_1\hat{U}_1^T \otimes I) &= \begin{bmatrix}\Lambda_1\hat{U}_1\hat{U}_1^T & \Lambda_2\hat{U}_1\hat{U}_1^T & \dots & \Lambda_m\hat{U}_1\hat{U}_1^T\end{bmatrix}, \\ &= \begin{bmatrix}\hat{U}_1\hat{U}_1^T\Lambda_1 & \hat{U}_1\hat{U}_1^T\Lambda_2 & \dots & \hat{U}_1\hat{U}_1^T\Lambda_m\end{bmatrix} = \hat{U}_1\hat{U}_1^T\mathcal{L}_{(1)}.\end{aligned}$$

The second equality is due the fact that $\hat{U}_1\hat{U}_1^T$ is diagonal. The SVD is given by $\mathcal{L}_{(2)}(\hat{U}_1\hat{U}_1^T \otimes I) = U\Sigma V$ wherein $\Sigma = \begin{bmatrix}\Sigma' & 0\end{bmatrix}$ and

$$\Sigma'^2 = \operatorname{diag}\left(\sum_{i=1}^m (\Lambda_i)^2_{q_1,q_1},\ \sum_{i=1}^m (\Lambda_i)^2_{q_2,q_2},\ \dots,\ \sum_{i=1}^m (\Lambda_i)^2_{q_{r_1},q_{r_1}}, 0, \dots, 0\right),$$

and $U := \begin{bmatrix}\hat{U}_1 & \hat{U}_1^\perp\end{bmatrix}$ is an orthogonal matrix. Truncating the SVD to the first $r_2 \leq r_1$ columns, it is clear that the corresponding left singular vectors are given by $\bar{U}_2$, as in the T-MLSVD. Therefore, the ST-MLSVD projector in the second mode equals the T-MLSVD projector. That is, $\hat{\pi}_2 = \bar{\pi}_2$.

The proof can now be completed by applying Theorem 6.3. ∎

One particular instance of this theorem results if $Z_3 = I$. The theorem then states that if the family of matrices $\{A_i\}_i$ is simultaneously diagonalizable by an orthogonal congruence transformation, then ST-MLSVD outperforms the T-MLSVD.

Example. To conclude this section, we present a numerical example of a tensor satisfying the conditions of Theorem 6.4. Consider the third order tensor $\mathcal{A} \in \mathbb{R}^{7\times7\times5}$,

$$\begin{aligned}\mathcal{A}_{:,:,1} &= \operatorname{diag}(\ \ 1.1,\ \ 1.1, -0.9,\ \ 0.1, -1.2, -1.1,\ \ 0.0),\\ \mathcal{A}_{:,:,2} &= \operatorname{diag}(\ \ 1.5, -0.8,\ \ 0.4, -0.2,\ \ 1.1, -1.1,\ \ 0.0),\\ \mathcal{A}_{:,:,3} &= \operatorname{diag}(\ \ 0.6,\ \ 1.1,\ \ 1.5,\ \ 0.1, -1.5, -0.7, -1.1),\\ \mathcal{A}_{:,:,4} &= \operatorname{diag}(\ \ 2.4, -0.6,\ \ 0.7, -0.2,\ \ 0.9, -0.8, -1.4),\\ \mathcal{A}_{:,:,5} &= \operatorname{diag}(-1.4,\ \ 0.5, -0.2, -0.2,\ \ 1.4,\ \ 0.3,\ \ 0.2).\end{aligned}$$

The rank-$(3,3,2)$ T-MLSVD approximation is given by $\bar{\mathcal{A}} = (\bar{U}_1\bar{U}_1^T, \bar{U}_2\bar{U}_2^T, \bar{U}_3\bar{U}_3^T)\cdot\mathcal{A}$, where

$$\bar{U}_1 = \bar{U}_2 = \begin{bmatrix}1&0&0\\0&0&0\\0&0&1\\0&0&0\\0&1&0\\0&0&0\\0&0&0\end{bmatrix}, \qquad \text{and} \qquad \bar{U}_3 = \begin{bmatrix}0.2751 & 0.4734\\ 0.4088 & -0.4555\\ 0.3858 & 0.5975\\ 0.6917 & -0.3840\\ -0.3604 & -0.2529\end{bmatrix}.$$

Note the particular form of the projectors in modes one and two. They indeed satisfy Eq. (6.4). The relative approximation error $\|\mathcal{A} - \bar{\mathcal{A}}\|_F / \|\mathcal{A}\|_F = 0.649552$.

The rank-$(3,3,2)$ ST-MLSVD approximation corresponding to $\mathbf{p} = \langle 1,2,3\rangle$ is $\hat{\mathcal{A}} = (\hat{U}_1\hat{U}_1^T, \hat{U}_2\hat{U}_2^T, \hat{U}_3\hat{U}_3^T)\cdot\mathcal{A}$, where

$$\hat{U}_1 = \hat{U}_2 = \bar{U}_1 = \bar{U}_2, \qquad \text{and} \qquad \hat{U}_3 = \begin{bmatrix}0.2615 & 0.3785\\ 0.4092 & -0.4518\\ 0.3256 & 0.5295\\ 0.6838 & -0.4132\\ -0.4367 & -0.4489\end{bmatrix}.$$

In this case, the relative approximation error $\|\mathcal{A} - \hat{\mathcal{A}}\|_F / \|\mathcal{A}\|_F = 0.637842$, which is indeed better than the T-MLSVD. In fact, the relative approximation error is lowest for the permutation $\mathbf{p} = \langle 3, 1, 2 \rangle$. The error is 0.607711 in that case. ♦

6.3. Diagonalizable tensors. A tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ is diagonalizable iff there exist orthogonal $Z_k \in \mathbb{R}^{n_k \times n_k}$, $1 \leq k \leq d$, such that

$$(Z_1, Z_2, \ldots, Z_d) \cdot \mathcal{A} = \mathcal{D}, \qquad \text{with } \mathcal{D} \text{ diagonal.}$$

A tensor $\mathcal{D} = [\![d_{i_1 i_2 \ldots i_d}]\!]$ is called *diagonal* if it is zero everywhere, except (possibly) $d_{i,i,\ldots,i}$, with $1 \leq i \leq \min\{n_1, n_2, \ldots, n_d\}$. In the matrix case, this problem is well-studied. It is known that any matrix can be represented as a diagonal matrix if a suitable orthonormal basis is chosen for its domain and range [9, Section 3.2.3]. Furthermore, the singular value decomposition [9,11] of the matrix completely resolves the question which bases should be chosen. Unfortunately, in the tensor case it has been shown that not every tensor is diagonalizable by a change of basis [3, 5, 15, 25]. That is, not every tensor can be represented by a diagonal tensor in a suitable tensor basis. However, for tensors that are diagonalizable, we can prove that T-MLSVD and ST-MLSVD approximate the original tensor equally well.

COROLLARY 6.5. *Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ be a diagonalizable by a change of basis. That is, there exist orthogonal $Z_k \in \mathbb{R}^{n_k \times n_k}$, $1 \leq k \leq d$, such that*

$$(Z_1, Z_2, \ldots, Z_d) \cdot \mathcal{A} = \mathcal{D}, \qquad \text{where } d_{i_1, i_2, \ldots, i_d} = 0 \text{ if } \exists j, k : i_j \neq i_k.$$

Let $\hat{\mathcal{A}}$ be any ST-MLSVD approximation to $\mathcal{A}$ and let $\bar{\mathcal{A}}$ be the T-MLSVD approximation. Then,

$$\|\mathcal{A} - \hat{\mathcal{A}}\|_F^2 = \|\mathcal{A} - \bar{\mathcal{A}}\|_F^2,$$

i.e., ST-MLSVD and T-MLSVD approximate a diagonalizable tensor equally well.

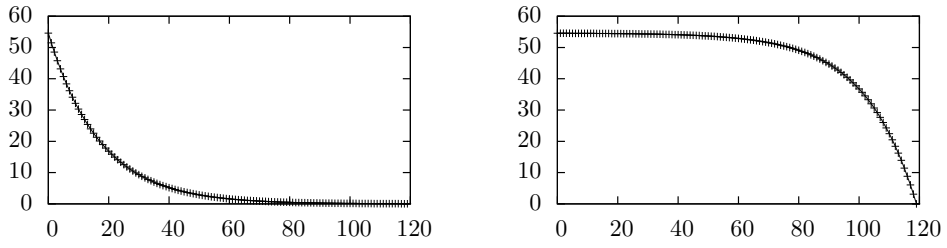
Proof. By Theorem 5.6, we can restrict our proof to the diagonal tensor $\mathcal{D}$. The theorem can then be proved in essentially the same way as Theorem 6.4, by applying the arguments for mode 1 and 2 to the other modes as well. In the end, the core tensor $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times \cdots \times r_d}$ is given, in either algorithm, by

$$s_{i,i,\ldots,i} = d_{p_i, p_i, \ldots, p_i}, \qquad \text{with } 1 \leq i \leq \min\{r_1, r_2, \ldots, r_d\},$$

and zero everywhere else. Herein, p_i denotes the position of the i^{th} largest element on the diagonal of $\mathcal{D}$. The factor matrices of the T-MLSVD and ST-MLSVD decomposition of $\mathcal{D}$, in corresponding modes, are equal. □

7. Numerical experiments. In this section, we analyze the performance of the ST-MLSVD and T-MLSVD through a series of numerical experiments. Both algorithms were implemented[1] in MATLAB® 7.9.0, using the MultiLinear Toolbox [21]. The experiments were conducted on a laptop comprising an Intel Core2Duo P7350 processing unit (2.0GHz) and 4GB main memory. It should be noted that our computational experiments are only indicative of the relative performance of T-MLSVD and ST-MLSVD.

[1] The source code can be obtained from:
`http://people.cs.kuleuven.be/~nick.vannieuwenhoven/`



(a) The weights $\mathbf{s}$. (b) The weights $\mathbf{t}$.

Figure 7.1: The weights of model problems one and two.

7.1. Model problems. Here, three model problems are presented. In the first two model problems, we consider an order-d cubic tensor $\mathcal{A} \in \mathbb{R}^{n \times \cdots \times n}$ that is approximately decomposable as a sum of r rank-1 terms. Let

$$\mathcal{C}(\mathbf{s}) = (F^{(1)}, F^{(2)}, \ldots, F^{(d)}) \cdot \mathcal{D} = \sum_{i=1}^{r} s_i \mathbf{f}_i^{(1)} \otimes \mathbf{f}_i^{(2)} \otimes \cdots \otimes \mathbf{f}_i^{(d)}, \tag{7.1}$$

be an order-d tensor of size $n \times \cdots \times n$ and every $F^{(i)} \in \mathbb{R}^{n \times r}$ a random Gaussian matrix[2] that has been scaled to Frobenius norm unity. The diagonal tensor $\mathcal{D}$ is defined as $\operatorname{diag}(s_1, s_2, \ldots, s_r)$. The two model problems, employ a different set of weights $\mathbf{s}$ in Eq. (7.1). Let α be the vector of r uniformly spread numbers in $[-3, 4]$, sorted from largest to smallest. We define two sets of weights

$$s_i = e^{\alpha_i}, \qquad \text{and} \qquad t_i = e^4 - s_{r+1-i}.$$

The vector $\mathbf{s} = [s_i]_i$ is employed in model problem one, and problem two uses $\mathbf{t} = [t_i]_i$. These numbers effectively determine the importance of the rank-1 terms in Eq. (7.1). The decay of the weights is visualized in Figure 7.1. The tensor considered in the first, respectively second, model problem is then given by

$$\mathcal{A}_1(d, n, r) := \frac{\mathcal{C}(\mathbf{s})}{\|\mathcal{C}(\mathbf{s})\|_F} + 0.05 \frac{\mathcal{E}}{\|\mathcal{E}\|_F} \quad \text{and} \quad \mathcal{A}_2(d, n, r) := \frac{\mathcal{C}(\mathbf{t})}{\|\mathcal{C}(\mathbf{t})\|_F} + 0.05 \frac{\mathcal{E}}{\|\mathcal{E}\|_F},$$

wherein $\mathcal{E}$ is a Gaussian tensor. Both model problems are thus random in the factor matrices and in the "noise term" $\mathcal{E}$. As third model problem, we consider an orthogonal Tucker decomposition of multilinear rank $(r, r, \ldots, r)$:

$$\mathcal{T} = (U_1, U_2, \ldots, U_d) \cdot \mathcal{S},$$

wherein every $U_i \in \mathbb{R}^{n \times r}$ has orthonormal columns, and is constructed by taking the Q-factor in the QR decomposition of a random Gaussian matrix. The core tensor $\mathcal{S}$ is a random Gaussian tensor. The third model problem, then reads

$$\mathcal{A}_3(d, n, r) := \frac{\mathcal{T}}{\|\mathcal{T}\|_F} + 0.05 \frac{\mathcal{E}}{\|\mathcal{E}\|_F},$$

wherein $\mathcal{E}$ represents white noise, i.e. it is a random Gaussian tensor.

[2]We call a tensor (matrix) whose entries are randomly sampled from the standard normal distribution, a Gaussian tensor (matrix).

7.2. Approximation error. Here, we investigate the error of T-MLSVD, ST-MLSVD and HOOI on the model problems from the previous subsection. The higher-order orthogonal iteration (HOOI) algorithm, introduced in [6], is an iterative algorithm to determine a (local) optimizer of optimization problem (1.1). In general, it is not guaranteed to converge to the global optimum, but when initialized with the factor matrices of the T-MLSVD, it is known to result in a good approximation to the optimum [6]. It is thus natural to compare the T-MLSVD and ST-MLSVD with respect to the HOOI solution, as this is probably close to the optimum. Unless indicated otherwise, we initialize the HOOI algorithm with the factor matrices of the T-MLSVD. The iterative process was halted if the improvement of the approximation error dropped below 10^{-4}.

We have selected $\mathcal{A}_1(3, 40, 120)$ as our first model problem, $\mathcal{A}_2(3, 40, 120)$ as second problem and $\mathcal{A}_3(3, 40, 30)$ as final problem. We approximated these model problem tensors by the three algorithms with a rank-(r, r, r) tensor, for every $1 \leq r \leq 25$, and repeated the experiment on 500 different instances of each model problem. For every pair of model problem and algorithm, the average of the relative approximation error over these 500 experiments was determined. The results are shown in Table 7.1. The relative approximation error is a quantity in $[0, 1]$, and can be interpreted as the fraction of the data that is not explained by our low rank orthogonal Tucker model. Recall that the ST-MLSVD approximation error depends on the order in which the modes are processed. By worst ST-MLSVD (STw), we thus imply the ST-MLSVD approximation corresponding to the permutation that caused the highest approximation error. By best ST-MLSVD (STb), we imply the ST-MLSVD approximation corresponding to the permutation that yielded the lowest approximation error. These worst and best permutations were determined by brute force.

As a first observation, we note that the model problems result in different approximation errors, when considering an approximation of the same multilinear rank. In particular, we note that on model 2, the error of a rank-$(25, 25, 25)$ approximation is still quite high, whereas in model problem 1 the error is much lower. This can be attributed to the difference in the distribution of the singular values in every mode. In model problem 1, the exponential decay of the singular values in every mode also results in an approximation error that behaves in a similar fashion. On the other hand, the slow decay of the singular values in every mode, in model problem 2, results in a slow decay of the error of the approximation.

The data in Table 7.1 indicates that on every model problem, for every multilinear rank, and for every permutation of the modes, the ST-MLSVD approximation yielded a better approximation error than the T-MLSVD.

We note that the difference between T-MLSVD and the other algorithms is largest if the tensor is approximated by one of (very) low multilinear rank, as we suspected at the end of Section 4. In the first model problem, the difference is largest at around rank $(3, 3, 3)$ and $(4, 4, 4)$. In model problem 2, the difference peaks at around rank $(6, 6, 6)$ and $(7, 7, 7)$, and in model problem 3 the maximum is attained approximately at rank $(10, 10, 10)$.

In these model problems, the difference in approximation error between ST-MLSVD and T-MLSVD was approximately half of the difference between HOOI and T-MLSVD. The difference between HOOI and T-MLSVD is rather limited. We have not observed more than 15% difference on these model problems. However, in terms of execution time, the HOOI algorithm is much more expensive than T-MLSVD. Consider, for instance, the speedup of ST-MLSVD over T-MLSVD and HOOI on the third

r	**Model Problem 1**				**Model Problem 2**				**Model Problem 3**			
	T	STw	STb	H	T	STw	STb	H	T	STw	STb	H
1	92.9	89.9	88.0	86.2	99.8	98.8	98.4	97.6	100.0	99.3	99.1	98.9
2	85.3	81.2	78.8	75.5	99.3	97.7	97.1	95.4	99.8	98.7	98.4	97.8
3	77.6	73.2	70.7	66.7	98.4	96.4	95.7	93.3	99.4	97.9	97.6	96.5
4	69.8	65.6	63.4	59.0	97.3	94.9	94.3	91.2	98.8	96.8	96.5	95.0
5	62.6	58.7	56.7	52.3	95.8	93.2	92.6	89.2	97.8	95.5	95.1	93.3
6	55.7	52.3	50.7	46.4	94.0	91.4	90.8	87.1	96.5	93.9	93.5	91.5
7	49.3	46.5	45.1	41.2	92.0	89.3	88.7	85.1	94.8	92.0	91.6	89.3
8	43.6	41.1	40.0	36.5	89.7	87.1	86.5	83.0	92.8	89.9	89.5	87.0
9	38.3	36.3	35.4	32.3	87.2	84.7	84.2	80.8	90.3	87.4	86.9	84.4
10	33.6	31.9	31.2	28.6	84.5	82.1	81.6	78.5	87.5	84.6	84.1	81.5
11	29.4	28.0	27.4	25.2	81.6	79.4	79.0	76.0	84.2	81.4	80.4	78.3
12	25.6	24.5	24.0	22.2	78.6	76.5	76.2	73.5	80.6	77.9	77.4	74.9
13	22.3	21.4	21.0	19.6	75.4	73.6	73.2	70.8	76.5	74.0	73.6	71.2
14	19.4	18.6	18.3	17.2	72.2	70.5	70.2	68.1	72.1	69.8	69.4	67.1
15	16.8	16.2	16.0	15.1	68.9	67.4	67.1	65.2	67.2	65.2	64.8	62.7
16	14.6	14.1	13.9	13.2	65.5	64.2	63.9	62.2	61.9	60.2	59.8	58.0
17	12.6	12.2	12.0	11.5	62.1	60.9	60.7	59.2	56.3	54.9	54.5	52.9
18	10.8	10.5	10.4	10.0	58.6	57.6	57.4	56.2	50.3	49.1	48.9	47.5
19	9.3	9.1	9.0	8.7	55.1	54.3	54.1	53.0	44.0	43.1	42.8	41.8
20	7.9	7.8	7.7	7.5	51.6	50.9	50.8	49.9	37.3	36.7	36.5	35.7
21	6.8	6.6	6.6	6.4	48.2	47.6	47.4	46.7	30.3	29.9	29.7	29.2
22	5.8	5.7	5.6	5.5	44.8	44.2	44.1	43.5	23.1	22.8	22.7	22.4
23	4.9	4.8	4.8	4.7	41.4	40.9	40.9	40.4	15.6	15.5	15.4	15.3
24	4.1	4.1	4.0	4.0	38.0	37.7	37.6	37.2	7.9	7.9	7.9	7.8
25	3.5	3.4	3.4	3.4	34.8	34.5	34.4	34.1	0.2	0.2	0.2	0.2

Table 7.1: The relative approximation error ($100 \cdot \|\mathcal{A} - \breve{\mathcal{A}}\|_F / \|\mathcal{A}\|_F$) of T-MLSVD (T), the worst ST-MLSVD approximation (STw), the best ST-MLSVD approximation (STb) and HOOI (H) in function of the multilinear rank (r, r, r) of the approximation tensor.

model problem, which is plotted in Figure 7.2. The ST-MLSVD approximation can be computed about twice as fast as the T-MLSVD, while reaching about half of the possible improvement in terms of approximation error (where we assume that HOOI reaches the maximum). The HOOI approximation, on the other hand, is extremely expensive to compute, but is able to reduce the approximation error further.

From the above observations, we are convinced that the ST-MLSVD approximation is, in general, a viable candidate to replace the T-MLSVD, as it improves both the approximation error and the execution time.

7.3. Computation time. We briefly investigate the execution times of the T-MLSVD and ST-MLSVD algorithms, in function of the approximation rank and the order of the tensor. In this experiment, we selected three tensors from model problem three; $\mathcal{A}_3(3, 30, 25)$, $\mathcal{A}_3(4, 30, 25)$ and $\mathcal{A}_3(5, 30, 25)$. We computed the (S)T-MLSVD approximation of multilinear rank $(r, \ldots, r)$, for every $1 \leq r \leq 30$, and repeated this five times. Of these five runs, the minimum execution time was selected. The speedup of ST-MLSVD with respect to T-MLSVD is displayed in Figure 7.3

Our numerical experiments verify that the speedup of ST-MLSVD can be as high as the order of the tensor, if the latter is approximated by a very low multilinear rank. The speedup also appears to decrease linearly with the rank.

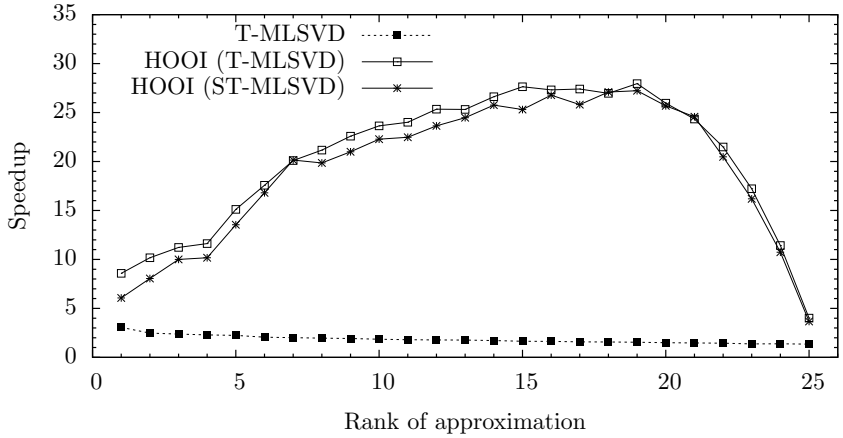


Figure 7.2: Speedup of ST-MLSVD over T-MLSVD, the HOOI algorithm initialized with the T-MLSVD factor matrices (HOOI (T-MLSVD)), and the HOOI algorithm initialized with the ST-MLSVD factor matrices (HOOI (ST-MLSVD)) on model problem 3.

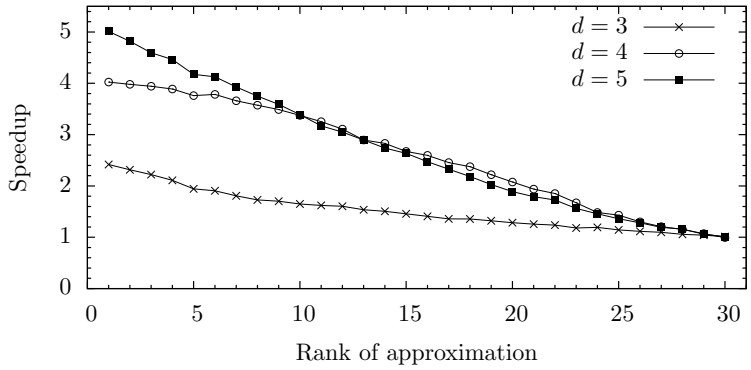


Figure 7.3: The speedup of ST-MLSVD over T-MLSVD on the third model problem $\mathcal{A}_3(d, 30, 25)$, in function of the approximation rank.

7.4. Case study: approximating a multidimensional function. In this subsection, we demonstrate that both the ST-MLSVD and HOOI can sometimes outperform T-MLSVD by a larger margin than in the previous examples. We consider here the problem of approximating the tensor

$$\mathcal{S} = \left[\frac{1}{i \sin(i + j + k)} \right]_{i,j,k=1}^{40,40,40}. \tag{7.2}$$

In Figure 7.4, the singular values of the mode-k unfoldings are displayed. The tensor is symmetric in modes 2 and 3, and hence the singular values are equal in those unfoldings. The shape of the decay of the singular values is strikingly different in the modes. In mode-1 the decay is very rapid, but in the other modes the decay is very

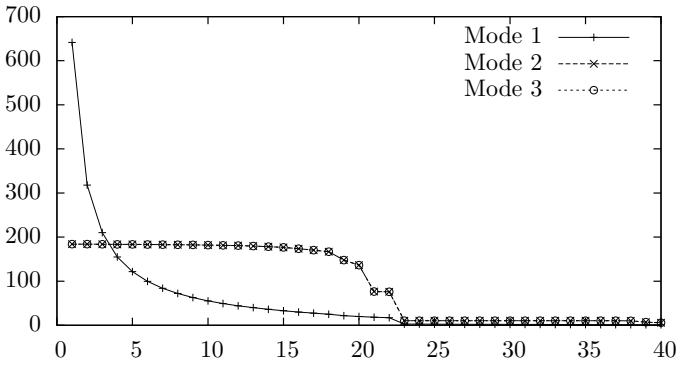


Figure 7.4: The singular values of the mode-k unfoldings of the tensor in Eq. (7.2).

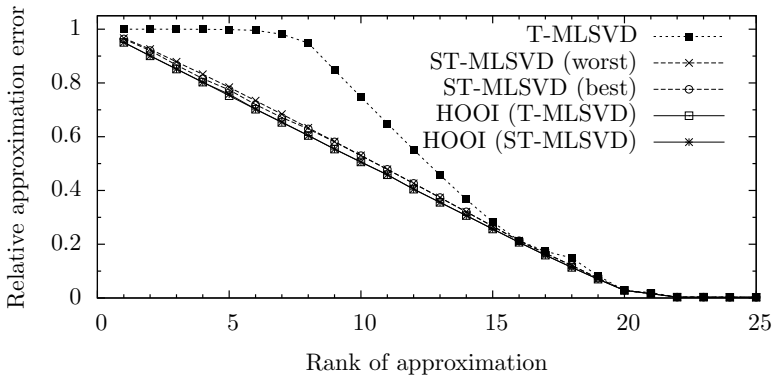


Figure 7.5: The relative approximation error $\|\mathcal{S} - \hat{\mathcal{S}}\|_F / \|\mathcal{S}\|_F$ of the approximation $\hat{\mathcal{S}}$, computed by various methods in function of the approximation rank (r, r, r).

slow. In fact, the difference between the first and tenth singular value, in mode 2, is less than 1.2%.

We approximate Eq. (7.2) by a rank-(r, r, r) tensor, for $1 \leq r \leq 25$. The results of our experiments are displayed in Figure 7.5. In this example, the ST-MLSVD approximation is very close to the solution found by the HOOI algorithm. The T-MLSVD performs poorly. For instance, a rank-$(5, 5, 5)$ T-MLSVD approximation still admits an error of 99.9%, whereas the worst ST-MLSVD has 78.4% error and HOOI only 74.2%. In terms of execution time, the average[3] speedup of ST-MLSVD over T-MLSVD was 2.1. An average speedup of 14.5 was obtained over the ST-MLSVD initialized HOOI algorithm (which includes the time to compute the ST-MLSVD approximation), and a speedup of 17.5 was obtained over HOOI initialized with a T-MLSVD. This discrepancy is mostly explained by the (small) difference in the number of iterations required for convergence. The ST-MLSVD initialized HOOI algorithm required, on average, 28.75 iterations, whereas the T-MLSVD initialization results in

[3]The average over the approximations with a different rank.

33.625 iterations, on average. In this example, the HOOI iteration steps were the predominant cost, not the computation of the (S)T-MLSVD.

7.5. Case study: image reconstruction. In this subsection, the truncated multilinear singular value decomposition is applied to a problem in image processing. The example in this section is only illustrative, however, as more advanced and appropriate tools may exist for the problem at hand. Nevertheless, we hope to demonstrate the versatility of these tensor decomposition approaches. This example illustrates that sometimes large compression rates, i.e. strongly truncated modes, do occur in practical applications. In such applications, large benefits in execution time may be expected from the ST-MLSVD. Another example of an application wherein large compression rates are achieved, may be found in the classification of handwritten digits [19]. Therein, the number of training images, say a few thousands, is reduced to a set of a few "canonical" images.

Here, we consider the problem of reconstructing an image from a set of seriously distorted source images. Let us call the matrix obtained by fixing the third mode index only, a mode-3 slice of the tensor. In Figure 7.6a, the source image[4], the lighthouse at the Cap Béar near Port-Vendres in France, is displayed. The source image can be represented as an order-3 tensor $\mathcal{I}$ of size $500 \times 667 \times 3$, where each of the mode-3 slices represents one of the RGB components of the image. We assume that, through some process, we are only able to obtain seriously distorted observations of this source image. In Figure 7.6b, we display one such input image. In our experiments, the input images $\mathcal{J}$ are generated from the source images, as follows. First a noise tensor is generated,

$$\mathcal{E}_{ijk} = X_g \exp(X_n - 1)^2$$

wherein $X_g \sim G(1)$ is a random sample from the Gamma distribution with mean 1 and $X_n \sim N(0, 1)$ represents a random sample from the standard normal distribution. The exponential factor ensures that there can be a large difference in the disturbance of different pixels. In fact, on average, about 15% of the entries of $\mathcal{E}$ will magnify X_g, while the remainder quickly diminishes it. The entries of the distorted image are given by

$$\mathcal{J}_{ijk} = \mathcal{I}_{ijk} + 2^{-9}\|\mathcal{I}\|_F \exp(X_u)\mathcal{E}_{ijk} \tag{7.3}$$

wherein $X_u \sim U(-0.5, 0.5)$ represents a random sample from the uniform distribution on $[-0.5; 0.5]$. The third factor represents a random scaling of the noise tensor, such that there can be a large difference in the total distortion of an input image.

We attempted to reconstruct the source image from the input image in Figure 7.6b by means of a truncated (S)T-MLSVD. Some of the obtained images are displayed in Figure 7.7. The results are quite poor, as the distortion of each individual input image is too large to properly denoise (using tensor techniques). As can be seen, the approximation is quite crude. In all three images, vertical and horizontal lines are discernible. By reducing the amount of noise in the image, we also quickly loose detail and the pictures appear blurry.

Next, we consider the situation in which we are allowed to sample the source images a number of times. We have sampled it 50 times, and collected the images in a

[4]The source image and input images may be obtained from:
`http://people.cs.kuleuven.be/~nick.vannieuwenhoven/`



(a) Original image



(b) Sample 11



(c) T-MLSVD approximation



(d) ST-MLSVD approximation

Figure 7.6: Image reconstruction from 50 significantly distorted sample images using tensor decompositions.



(a) Rank $(100, 100, 3)$ (b) Rank $(80, 80, 3)$ (c) Rank $(50, 50, 3)$

Figure 7.7: Image denoising of sample 11 using the ST-MLSVD.

fourth order tensor $\mathcal{L}$ of size $500 \times 667 \times 3 \times 50$. We would like to reconstruct, accurately, the source image from this tensor. Clearly, we would like to reduce the set of input images to a single image, a third-order tensor. Consider the rank $(500, 667, 3, 1)$ approximation to $\mathcal{L}$. Such an approximation must necessarily be a weighted linear combination of the input images. The amount of noise in this linear combination image, can then be reduced by computing an approximation of a yet lower multilinear rank. From the above observations, we might believe that computing, for example, a rank-$(250, 250, 3, 1)$ approximation to $\mathcal{L}$ yields an accurate reconstruction of the source image. We have done precisely this, in our experiment. The result is depicted in Figure 7.6c, where we used the T-MLSVD, and in Figure 7.6d, where the ST-MLSVD was employed. In case of the ST-MLSVD, we employed the ordering $\mathbf{p} = \langle 4, 1, 2, 3 \rangle$. That is, first we compress the input images to one image and then compress it further in the other modes.

The proposed ST-MLSVD algorithm outperforms the T-MLSVD algorithm in terms of execution time. ST-MLSVD computed the denoised image in 36.81 seconds, whereas T-MLSVD required 178.02 seconds. This represents a speedup of 4.84, illustrating that in some instances the speedup can exceed the order of the tensor. We attribute this effect to a more efficient memory and cache usage. That may be a consequence of the fact that ST-MLSVD reduces the size of the tensor in every step.

Comparing Figure 7.6c and Figure 7.6d, it is clear that ST-MLSVD yields a better approximation to the source image. The image produced by T-MLSVD has some clear visual artifacts. For instance, the lighthouse and the stones at the bottom of the picture, are much less sharp than in Figure 7.6d. Also, some noise is still visible left of the base of the lighthouse, and in the air at the top of the picture. It is not entirely a surprise that the ST-MLSVD algorithm performs better. In this case, ST-MLSVD first computes a weighted average of the input images. This already results in a good approximation to the source image. That is because the input images are distorted variants of the same source image. By simply averaging over all input images, a large part of the noise will be canceled out. Thereafter, it compresses that average image, which has the effect of removing the noise. As the average image is already strongly denoised, its compression yields a sharp denoised image, contrary to directly

denoising only one of the input images, as in Figure 7.7. The T-MLSVD, on the other hand, simultaneously computes the orthogonal projectors. It will take exactly the same linear combination of input images to compute the averaged image. However, the other two projectors are computed based on all of the input images. Its effect will thus be to simultaneously denoise all input images. However, we are only interested in denoising the averaged image, at this point. It is unlikely that the projectors computed from the entire set of input images results in a better approximation than the projectors computed from the averaged image, which is the one we would like to denoise. The effect is visible in Figure 7.6.

8. Conclusions. A novel error expression for truncated orthogonal Tucker decompositions, which yields insights into the structure of the optimization problem, was presented. From this, an improved truncation strategy was proposed for the multilinear singular value decomposition. Our numerical experiments indicate that this sequentially truncated MLSVD may be a suitable replacement for the T-MLSVD, as it significantly reduces the number of floating point operations to construct the approximation, and, in many cases, reduces the approximation error. We observed a maximum speedup proportional to the order of the tensor, if the tensor is approximated by another of very low multilinear rank. However, the ST-MLSVD also introduces a number of difficulties which should be addressed by future research. For instance, the sequential approach destroys most of the structure present in the original tensor. Therefore, straightforwardly applying the sequential truncation to sparse or otherwise structured tensors may be inadvisable. Techniques should be designed to overcome this difficulty. The ST-MLSVD is also sequential in nature, whereas the T-MLSVD is parallel in nature. In the latter, the SVDs of the different unfoldings can be computed entirely independent from one another. Future research should investigate insofar the ST-MLSVD algorithm is parallelizable.

Appendix. MATLAB implementation.

```
function [A] = stmlsvd(T,mlr,p)
% STMLSVD  Sequentially truncated multilinear singular value
% decomposition.
%
% A = stmlsvd(T,mlr,p) computes the ttensor A, as rank-mlr
% ST-MLSVD approximation corresponding to the order p of the
% tensor T.
%
% A = stmlsvd(T,mlr) computes the ttensor A, as rank-mlr ST-MLSVD
% approximation of T, corresponding to the order of the strongest
% reduction of the size of the partially truncated core tensor.
%
% NOTE: only supports dense unstructured tensors efficiently.
% Structured tensors may loose their structure after the first
% projection step.

% Sanity check.
mlr = min( size(T), mlr );

% Determine the order of the strongest reduction.
if nargin < 3
    [~,p] = sort( size(T) ./ mlr, 'descend' );
end
```

```
% Factor matrices and core tensor.
d = ndims(T);
U = cell(1,d);
S = T;

% ST-MLSVD algorithm.
for k = 1 : d
    m = p(k);

    % Compute dominant subspace.
    U{m} = nvecs(S,m,mlr(m));

    % Project onto dominant subspace.
    S = ttm(S, U{m}, m, 't');
end

A = ttensor(S,U);
end
```

Acknowledgements. Hendrik Speleers is kindly thanked for pointing out Property 2.2 and its proof to us. Laurent Sorber and Marc Van Barel are thanked for interesting and fundamental remarks. They are also thanked for their stimulating enthousiasm. Laurent Sorber is thanked for assistance with the tensor decomposition images and the code from the MultiLinear toolbox.

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State Science Policy Office. The scientific responsibility rests with its author(s).

REFERENCES

[1] C.A. Andersson and R. Bro, *Improving the speed of multi-way algorithms:: Part I. Tucker3*, Chemometrics and Intelligent Laboratory Systems, 42 (1998), pp. 93–103.
[2] B.W. Bader and T.G. Kolda, *MATLAB Tensor Toolbox Version 2.4*, Mar. 2010.
[3] J. Chen and Y. Saad, *On the tensor SVD and the optimal low rank orthogonal approximation of tensors*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1709–1734.
[4] P. Comon, G.H. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1254–1279.
[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[6] ———, *On the best rank-1 and rank-(r_1, r_2, ..., r_n) approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
[7] ———, *Computation of the canonical decomposition by means of a simultaneous generalized schur decomposition*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 295–327.
[8] V. de Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[9] J.W. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.
[10] L. Eldén and B. Savas, *A Newton–Grassmann method for computing the best multilinear rank-(r_1, r_2, r_3) approximation of a tensor*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 248–271.
[11] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, USA, 3rd ed., 1996.
[12] R.L. Graham, D.E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation For Computer Science*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[13] L. Grasedyck, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.
[14] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel, *Differential-geometric newton method for the best rank-(r_1, r_2, r_3) approximation of tensors*, Numer. Algorithms, 51 (2009), pp. 179–194.
[15] T.G. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255.
[16] T.G. Kolda and B.W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[17] I.V. Oseledets and E.E. Tyrtyshnikov, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759.
[18] Y. Saad, *Numerical Methods For Large Eigenvalue Problems*, 2nd ed., 2011. In preparation. Available from: http://www-users.cs.umn.edu/~saad/books.html.
[19] B. Savas and L. Eldén, *Handwritten digit classification using higher order singular value decomposition*, Pattern Recognition, 40 (2007), pp. 993–1003.
[20] ———, *Krylov-type methods for tensor computations*, Submitted to Linear Algebra Appl., (2010).
[21] L. Sorber, *Matlab multilinear toolbox v0.2*. In development. Private communication, Feb. 2011.
[22] L.R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.
[23] C.F. Van Loan, *The ubiquitous Kronecker product*, J. Comput. Appl. Math., 123 (2000), pp. 85–100.
[24] M.A.O. Vasilescu and D. Terzopoulos, *Multilinear subspace analysis of image ensembles*, Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, 2 (2003), pp. 93–99.
[25] T. Zhang and G.H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.