21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

On the problem of low rank approximation of tensors

Hanumant Singh Shekhawat and Siep Weiland[1]

***Abstract*— This paper considers the problem of low-rank approximations of multi-linear functionals. Multi-linear functionals are generally referred to as tensors and provide a natural object of study in multi-dimensional signal and system analysis. We propose a decomposition of an arbitrary order N tensor in terms of mutually orthogonal rank 1 tensors. A novel numerically efficient algorithm is provided to compute such decompositions. It is shown how this decomposition can be used to approximate arbitrary tensors.**

I. Introduction

Many applications in signal processing naturally require a study of signals with multiple independent and multiple dependent variables. Examples include signals that evolve over spatial and temporal coordinates, and represent multiple physical quantities. In imaging and video processing, the independent variables represent pixel locations, an R-G-B intensity and possibly time. In the study of multi-rate dynamical systems, one typically distinguishes signals with multiple time-scales or signals that are parameter varying.

The data structures that are necessary to store and process these signals may become very large and pose significant challenges to questions on data transmission, data processing, storage, compression and analysis. Among the standard ways to compress large data objects is to remove irrelevant or insignificant information from the data, to reduce the data resolution, down sampling, or to characterize and detect specific data features. These techniques are known as data compression or data reduction techniques and have applications way beyond the image and video processing community. In particular, generic compression techniques are of quite some independent relevance to important areas such as model reduction, signal approximation, spectral analysis, controller synthesis and system identification [1], [3].

This paper addresses the problem to compress data objects in multiple independent variables. For this, we consider the mathematical framework to represent multidimensional data objects as tensors. A tensor or a multi-array is a multi-linear functional defined on the Cartesian product of Hilbert spaces. We believe that the abstract definition of tensors provide a natural and generic framework to study data compression problems. In fact, tensors can be viewed as the natural algebraic generalization of the concepts of scalars, vectors and matrices. In spite of this observation, it is remarkable that very few algebraic and numerical tools have been developed to use or analyze tensors in system theoretic applications.

The problem of compressing a given tensor by a low rank approximation is known as the *tensor approximation* of *low rank approximation problem*. This problem has been studied in several papers and theses [4], [7], [5], [11]. We refer to [8] for a detailed review. Contrary to the matrix case, the low rank approximation problem is a delicate problem in tensor analysis: it may be ill-posed and it may have non-unique solutions [5]. It is for this reason that the tensor approximation problem needs concise formalizations of concepts such as tensor ranks, orthogonality and decomposability of tensors.

Tensor approximation problems generally require tensor decompositions. One basically distinguishes two types of decompositions: the *CANDECOMP-PARAFAC (CP)* and Tucker decompositions. See [8] and the references therein. In 1997, Lathauwer proposed a popular and easily computable decomposition that is generally referred to as the higher order singular value decomposition (HOSVD) for a Tucker decomposition of tensors. Belzen and Weiland also proposed a number of algorithms for a Tucker decomposition with performance similar to the HOSVD [3], [10]. CP decompositions can be obtained by alternating least squares approximation methods and many other algorithms (see [8, §3.4] for a review). However, all of these algorithms do not provide optimal solutions to the tensor decomposition problem.

One approach to obtain a decomposition of a tensor, is to generalize specific properties of the singular value decomposition (SVD). Various approaches have been explored in [10] to obtain a Tucker decomposition of a tensor. One method in [10] is known as the tensor singular value decomposition (TSVD) which, in essence, defines singular values and vectors of a tensor in specific subspaces so as to obtain a Tucker decomposition. The dimension of the search space of singular values and vectors decreases with increasing numbers of singular values. In this paper, we propose a decomposition that uses a larger search space to allow more accurate Tucker approximations. We prove that the proposed method is equivalent to the standard SVD for order 2 tensors. We

[1]H S Shekhawat and Siep Weiland are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. h.s.shekhawat@tue.nl and s.weiland@tue.nl

ISBN: 978-90-367-6321-9

provide simulations that illustrate that the new approximation method outperforms the TSVD algorithm and that its performance is comparable to other widely used algorithms such as the HOSVD. In addition, we provide a numerically efficient scheme for the computation of such a decomposition.

This paper is organized as follows. In Section III, we briefly discuss the basics of tensors. The tensor approximation problems that we consider in this paper are formalized in Section IV. The Tucker decomposition proposed in [10] is briefly discussed in Section V. In Section VI, we introduce new tensor decomposition methods, and discuss advantages and the use of these decompositions in the tensor approximation problem. An algorithm to calculate singular vectors and singular values of a tensor is presented in Section VII. Finally, in Section VIII, we provide simulation results and discuss the performance of the new algorithms suggested in this paper when compared with existing algorithms.

II. Notation

Throughout $\mathbb{R}$ denotes the space of real numbers. For a given positive integer N, we define $\mathcal{N} := \{1, \cdots, N\}$. SVD stands for the singular value decomposition. $\mathfrak{D}(n)$ denotes the nth element of an (ordered or indexed) set $\mathfrak{D}$. The set $\mathfrak{D}(n:m) := \{\mathfrak{D}(n), \cdots, \mathfrak{D}(m)\}$.

III. Tensor basics

In this section, we define the concept of tensors and summarize their properties. For a detailed discussion, see [10], [4], [7], [8] and the references therein.

A tensor is a multi-linear functional. More specifically, a functional $T : \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_N \to \mathbb{R}$ that is linear in each of its N arguments is an order-N tensor defined on a Cartesian product of Hilbert spaces $\mathbb{X}_k$ over $\mathbb{R}$. We will assume that for all k, $\mathbb{X}_k$ has finite dimension $L_k := \dim(\mathbb{X}_k)$. Under this assumption, the tensor T can be identified with a multi-linear array in $\mathbb{R}^{L_1} \times \cdots \times \mathbb{R}^{L_N}$. To elaborate on this, first consider a rank-1 tensor that is defined by the functional

$$U(x_1, \cdots, x_N) = \langle x_1, u_1\rangle_{\mathbb{X}_1} \cdots \langle x_N, u_N\rangle_{\mathbb{X}_N}.$$

Here, $u_k \in \mathbb{X}_k$ are given vectors for $k \in \mathcal{N} := \{1, \cdots, N\}$. If $\|u_k\|_{\mathbb{X}_k} = 1$ for all k, then U is called a *normalized or unit tensor of rank*-1. In more compact form, the tensor U is written as $u_1 \otimes \cdots \otimes u_n$. Now suppose that, for $k \in \mathcal{N}$, the unit vectors $\{e_k^{(l)}\}_{l \in \{1, \cdots, L_k\}}$ represent a basis of $\mathbb{X}_k$. Then the tensor T can be represented as

$$T = \sum_{l_1=1}^{L_1} \cdots \sum_{l_N=1}^{L_N} t_{l_1 \ldots l_N}\, e_1^{(l_1)} \otimes \cdots \otimes e_N^{(l_N)}. \qquad (1)$$

where $t_{l_1 \cdots l_n} \in \mathbb{R}$ is called an *element* of the tensor T (obviously with respect to the given basis). The tensor T can be viewed as a mapping from the vector space spanned by $\{e_1^{(l_1)} \otimes \cdots \otimes e_N^{(l_N)} | l_i \in \{1, \cdots, L_i\}\}$ to $\mathbb{R}$. This mapping can be represented as a vector of size $L_1 L_2 \cdots L_N$ or, equivalently, as an N-way array $\mathrm{T} \in \mathbb{R}^{L_1 \times \cdots \times L_N}$ where elements of T are defined as

$$t_{l_1 \ldots l_N} := T(e_1^{(l_1)}, \cdots, e_N^{(l_N)}).$$

In this way, an order 1 tensor is a vector, and an order 2 tensor is a matrix.

From this discussion it follows that any tensor of order-N can be represented as (1) with respect to a given basis $\{e_k^{(l)}\}_{l \in \{1, \cdots, L_k\}}$ of $\mathbb{X}_k$, $k \in \mathcal{N}$. With the standard notions of addition and scalar multiplication, it is immediate that the set of all order-N tensors becomes a vector space over $\mathbb{R}$. We denote this space by $\mathcal{T}_N$.

A. Tensor norms

For the approximation of a tensor $T \in \mathcal{T}_N$, we equip the space $\mathcal{T}_N$ with a norm. First, we define the *induced norm* of a tensor in $\mathcal{T}_N$ as a direct extension of the induced norm of a linear operator. It is defined as

$$\|T\| := \sup_{\|x_k\|_{\mathbb{X}_k} = 1, k \in \mathcal{N}} |T(x_1, \cdots, x_N)|.$$

Next, we define the Frobenius norm of a tensor in $\mathcal{T}_N$. This norm is a direct extension of the Frobenius or Hilbert-Schmidt norm of a linear operator. As in the linear case, this norm requires the definition of an inner-product of two tensors. Suppose that unit vectors $\{e_k^{(l)}\}_{l \in \{1, \cdots, L_k\}}$ define a basis of $\mathbb{X}_k$ for all $k \in \mathcal{N}$. Then the inner product of two tensors $S, T \in \mathcal{T}_N$ with elements $s_{k_1, \ldots, k_N}$ and $t_{l_1, \ldots, l_N}$ is defined as

$$\langle S, T\rangle := \sum_{k_1=1}^{L_1} \cdots \sum_{k_N=1}^{L_N} \sum_{l_1=1}^{L_1} \cdots \sum_{l_N=1}^{L_N} s_{k_1, \ldots, k_N} \\ t_{l_1, \ldots, l_N} \left\langle e_1^{(l_1)}, e_1^{(k_1)} \right\rangle \cdot \ldots \cdot \left\langle e_N^{(l_N)}, e_N^{(k_N)} \right\rangle.$$

If the unit vectors $\{e_k^{(l)}\}_{l \in \{1, \cdots, L_k\}}$ are mutually orthonormal then

$$\langle S, T\rangle := \sum_{l_1=1}^{L_1} \cdots \sum_{l_N=1}^{L_N} s_{l_1, \ldots, l_N} t_{l_1, \ldots, l_N}.$$

Now, the Frobenius norm of a tensor $T \in \mathcal{T}_N$ is defined as

$$\|T\|_F := \sqrt{\langle T, T\rangle}$$

Moreover, $\mathcal{T}_N$ becomes an inner product space.

Note that the induced and Frobenius norm of a normalized rank-1 tensor is one.

B. Orthogonality of rank-1 tensors

With the inner products defined on $\mathbb{X}_k$ and $\mathcal{T}_N$ we obtain different notions of orthogonality of rank-1 tensors.

Definition III.1. *Let $U = u_1 \otimes \cdots \otimes u_N$ and $V = v_1 \otimes \cdots \otimes v_N$ be two rank-1 tensors.*

1) *If $\langle U, V\rangle = \prod_{n=1}^{N} \langle v_n, u_n\rangle_{\mathbb{X}_n} = 0$ then U and V are said to be* orthogonal. *This is denoted as $U \perp V$.*
2) *If $\langle v_n, u_n\rangle_{\mathbb{X}_n} = 0 \forall n \in \mathcal{N}$ then U and V are said to be* completely orthogonal. *This is denoted $U \perp_c V$.*

C. Tensor decomposition

Any tensor $T \in \mathcal{T}_N$ can be decomposed as

$$T = \sum_{i=1}^{R} \sigma_i U_i \tag{2}$$

where $R \leq L_1 \cdots L_N$ and U_i are normalized rank-1 tensors. This type of decomposition is known as a *CANDECOMP-PARAFAC* or *CP* decomposition. In the CP decomposition, we do not assume orthogonality of U_i. However, if $\langle U_i, U_j\rangle = \delta_{ij}$ for all $(i, j) = 1, \ldots, L$ then (2) is called a *(CP) orthogonal decomposition* of the tensor T. Similarly, if we assume that U_i and U_j are complete orthogonal for all $(i, j) = 1, \ldots, L$ then (2) is said to be a *(CP) completely orthogonal decomposition* of the tensor T. In either of these cases, $\|T\|_F^2 := \sum_{i=1}^{R} \sigma_i^2$. It is shown in [7] that neither orthogonal nor completely orthogonal decomposition of a tensor are unique in general. Using (1), it is trivial to show that an orthogonal decomposition of a tensor actually exists. However, a completely orthogonal decomposition of a tensor does not necessarily exist [7, corollary 3.9].

The *Tucker decomposition* is an alternative tensor decomposition. For a given tensor T, a Tucker decomposition is an expression of the form

$$T = \sum_{l_1=1}^{R_1} \cdots \sum_{l_N=1}^{R_N} \sigma_{l_1 \cdots l_N} u_1^{(l_1)} \otimes \cdots \otimes u_N^{(l_N)} \tag{3}$$

One of the Tucker decompositions of the tensor T is given by (1). For each $n \in \mathbb{N}$, if the vectors in the set $\{u_n^{l_n}\}_{1 \leq l_n \leq R_n}$ are mutually orthonormal, then the Tucker decomposition is known as a *modal-rank decomposition*. By re-arrangement of a Tucker decomposition, we can show that a CP-decomposition is a special case of a Tucker decomposition.

D. Tensor rank

For a given tensor $T \in \mathcal{T}_N$, the rank $\mathrm{rank}(T)$ is the minimal non-negative integer R such that T can be represented as (2). Similarly, for a given tensor $T \in \mathcal{T}_N$, the (completely) orthogonal rank is the minimal non-negative integer R such that T can be represented as in (2) with mutually (completely) orthogonal rank-1 tensors U_i. It is shown in [7] that $\mathrm{rank}(T) \leq \mathrm{rank}_\perp(T) \leq \mathrm{rank}_{\perp_c}(T)$ where $\mathrm{rank}_\perp(T)$ is the orthogonal rank and $\mathrm{rank}_{\perp_c}(T)$ is the completely orthogonal rank of a given tensor T. Only for order-2 tensors these rank concepts coincide.

Keeping in mind that the Tucker decomposition is defined as in (3), it is interesting to bound or minimize the values R_n. Clearly $R_n \leq \dim \mathbb{X}_n$. To obtain a lower bound on R_n, we define the *n-mode rank* $\mathrm{rank}_n(T)$ as

$$\mathrm{rank}_n(T) = \dim \mathbb{X}_n - \dim(\mathrm{Ker}_n\, T)$$

for all $n \in \mathcal{N}$. Here, the *n-mode kernel* of the tensor T is defined as

$$\begin{aligned}\mathrm{Ker}_n(T) :=& \{x_n \in \mathbb{X}_n \mid T(x_1, \cdots, x_N) = 0,\\ & \forall x_k \in \mathbb{X}_k, k \neq n\}\end{aligned}$$

Clearly, $\mathrm{rank}_n(T) = \min R_n$. The *modal-rank* of a tensor $T \in \mathcal{T}_N$ is the N-tuple $(\mathrm{rank}_1(T), \cdots, \mathrm{rank}_N(T))$. For details we refer to [10].

IV. Problem formulation

We can now provide concise problem formulations as follows.

Problem $\mathcal{P}_1$: *Let $T \in \mathcal{T}_N$ be a given N-order tensor. Given a non-negative integer $r \leq \mathrm{rank}(T)$, determine a tensor*

$$T_r^* \in \operatorname*{arg\,inf}_{\mathrm{rank}(T_r) \leq r} \|T - T_r\|_F.$$

Here, $\|\cdot\|_F$ is the Frobenius norm of a tensor.

Similarly, we consider the modal rank approximation problem.

Problem $\mathcal{P}_2$: *Let $T \in \mathcal{T}_N$ be a given N-order tensor. Given an N-tuple of non-negative integers $r = (r_1, \cdots, r_N)$, determine a tensor*

$$T_r^* \in \operatorname*{arg\,inf}_{\mathrm{rank}_n(T_r) \leq r_n \forall n \in \mathbb{N}} \|T - T_r\|_F.$$

Here, $\|\cdot\|_F$ is the Frobenius norm of a tensor.

For $N \leq 2$, the solution of these problems is well-known. For $r = 1$ and $N > 2$, optimal solutions exist for both problems and are described, e.g. in [3, theorem 3.6.1]. For $r > 1$ and $N > 2$, it is known that Problem $\mathcal{P}_1$ is ill-posed [5], [8]. On the other hand, Problem $\mathcal{P}_2$ is well defined but a closed form solution is not known [6], [5].

Various alternative formulations have been suggested that avoid the ill-posed nature of problems $\mathcal{P}_1$ and $\mathcal{P}_2$. One way to obtain a tensor approximation is to generalize the property of the SVD as an orthogonal or completely orthogonal decomposition of a matrix. For example, singular vectors and singular values of a matrix M can be defined recursively as

$$\{x_i, y_i\} = \underset{\substack{\|x\|=1,\|y\|=1 \\ x\perp\{x_0,\cdots,x_{i-1}\} \\ y\perp\{y_0,\cdots,y_{i-1}\}}}{\arg\sup} x^\top My,\ \sigma_i := x_i^\top My_i \tag{4}$$

for $i > 1$. Here $x_0 = 0$ and $y_0 = 0$. This property of the SVD is generalized in [10] to tensors and leads to a completely orthogonal decomposition of a tensor whenever it exists. As an alternative, singular vectors and singular values of a matrix M can also be defined as

$$\{x_i, y_i\} = \underset{\substack{\|x\|=1,\|y\|=1 \\ x\perp\{x_0,\cdots,x_{i-1}\}}}{\arg\sup} x^\top My,\ \sigma_i := x_i^\top My_i \tag{5}$$

for $i > 1$. As before, $x_0 = 0$ and $y_0 = 0$. Here, we relax the orthogonality constraint on y. It is shown in the Appendix that this definition also leads to an SVD of a matrix.

In this paper, we generalize singular values and singular vectors as defined in (5) to tensors. This leads to an *orthogonal decomposition* of a tensor. When compared to a completely orthogonal decomposition, this provides significant advantages in approximation accuracy, as we will see later.

A. Tensor approximation

A common approach in tensor approximation is to truncate a CP or Tucker decomposition of a given tensor (see Section III-C). This is known as *tensor truncation*. Specifically, for a given CP decomposition, a truncation of (2) at level $r < R$ is defined as

$$T_r = \sum_{i=1}^{r} \sigma_i U_i. \tag{6}$$

Clearly $\mathrm{rank}(T_r) \leq r$. If the U_i's are mutually orthogonal normalized rank-1 tensors then

$$\|T - T_r\|_F^2 = \|T\|_F^2 - \sum_{i=1}^{r} \sigma_i^r.$$

Similarly, for Tucker decompositions, we define truncations of the decomposition (3) as

$$T_{r_1:r_n} = \sum_{l_1=1}^{r_1} \cdots \sum_{l_N=1}^{r_N} \sigma_{l_1\cdots l_N} u_1^{(l_1)} \otimes \cdots \otimes u_N^{(l_N)} \tag{7}$$

where $r_n \leq R_n$. We call this a *Tucker truncation* at the multi-level $(r_1, \cdots, r_N)$. For each $n \in \mathbb{N}$, if the vectors in the set $\{u_n^{l_n}\}_{1\leq l_n \leq R_n}$ are mutually orthonormal then the Tucker truncation is known as a *modal truncation* at multi-level $(r_1, \cdots, r_N)$. Obviously, the modal rank $\mathrm{rank}_n(T_{r_1:r_n}) < \mathrm{rank}_n(T)$. Due to the orthogonality, it is straightforward to calculate the approximation error in a modal truncation. Indeed, the modal truncation error is given by

$$\|T - T_{r_1:r_n}\|_F^2 = \|T\|_F^2 - \sum_{l_1=1}^{r_1} \cdots \sum_{l_N=1}^{r_N} \sigma_{l_1\cdots l_N}^2$$

V. Completely orthogonal decomposition

In [10], Weiland and Belzen introduced an algorithm for a completely orthogonal (CP) decomposition of a tensor (if exists). They also explained a method for a modal rank decomposition. In this section, we briefly describe the algorithm proposed in [10]. It is based on a generalization of an SVD property given in (4).

Consider an order-N tensor $T : \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_N \to \mathbb{R}$ where, for each $k = 1, \ldots, N$, $\mathbb{X}_k$ is a Hilbert space over $\mathbb{R}$ of finite dimension $L_k := \dim(\mathbb{X}_k)$. Define the *completely orthogonal singular values* or, in short, the *CO-singular values* of the tensor T as follows. The first singular value of the tensor T is defined as

$$\hat{\sigma}_1(T) := \sup_{\|x_n\|_{\mathbb{X}_n}=1, n\in\mathcal{N}} |T(x_1, \cdots, x_N)|. \tag{8}$$

Since all $\mathbb{X}_n$ are finite dimensional, T is a bounded mapping on the Cartesian domain of unit vectors. Therefore $\sigma_1(T)$ exists and is attained by an N-tuple, say $\{\hat{w}_1^{(1)}, \cdots, \hat{w}_N^{(1)}\}$.

In the next step, one determines a normalized rank-1 tensor which is completely orthogonal to tensor $U_1 := \hat{w}_1^{(1)} \otimes \hat{w}_2^{(1)} \otimes \cdots \otimes \hat{w}_N^{(1)}$. This can be achieved by maximizing the gain $|T(x_1, \cdots, x_N)|$ over unit vectors $x_n \in \mathbb{X}_n$ perpendicular to $\hat{w}_n^{(1)}$ for all $n \in \mathcal{N}$ (see Definition III.1.2). Proceeding in this way, subsequent singular values of T are defined in a recursive manner as

$$\hat{\sigma}_k(T) = \sup_{n\in\mathbb{N}, x_n\in\hat{S}_n^k} |T(x_1, \cdots, x_N)| \tag{9}$$

where $\hat{S}_n^k := \{x_n \in \mathbb{X}_n \mid \|x_n\|_{\mathbb{X}_n=1}, \left\langle x_n, \hat{w}_n^{(i)} \right\rangle_{\mathbb{X}_n} = 0\ \forall i \in \{1, \cdots, k-1\}\}$. Assume that the maxima are attained by an N-tuple, say $\{\hat{w}_1^{(k)}, \cdots, \hat{w}_N^{(k)}\}$. This algorithm proceeds until one reaches a zero singular value. It is shown in [10] that this results in $K := \min_{n\in\mathcal{N}} \mathrm{rank}_n(T)$ completely orthogonal singular values. Now define the decomposition

$$\hat{T}_c := \sum_{k=1}^{K} \hat{\sigma}_k(T) \hat{w}_1^{(k)} \otimes \cdots \otimes \hat{w}_N^{(k)}$$

Example V.1. *It is shown in [3, e.g. 3.5.5] that if* $T := 2\, e_1 \otimes e_1 \otimes e_1 + \frac{1}{\sqrt{2}}\, (e_1 + e_2) \otimes e_2 \otimes e_2$ *where* $e_1 := \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ *and* $e_2 := \begin{bmatrix} 0 & 1 \end{bmatrix}^T$*, then* $\hat{T}_c = 2\, e_1 \otimes e_1 \otimes e_1 + \frac{1}{\sqrt{2}}\, e_2 \otimes e_2 \otimes e_2$*. Clearly,* $\|T - \hat{T}_c\| = \frac{1}{\sqrt{2}}$*.*

Conclude from Example V.1 that one cannot guarantee that a completely orthogonal CP-decomposition of a tensor exists. To complete the decomposition of a given tensor, [10, §IV.A] proposed a method for complementing the basis. In this method, if $K < \mathrm{rank}_n(T)$ then we can get an orthonormal basis of $\mathbb{X}_n$, by obtaining an orthonormal basis of the orthogonal complement of $\overline{\mathrm{span}}\{\hat{w}_n^{(1)}, \cdots, \hat{w}_n^{(K)}\}$. We denote this basis by $\{\hat{w}_n^{(K+1)}, \cdots, \hat{w}_n^{(L_n)}\}$. The tensor T then has a modal-rank decomposition (hence a CP orthogonal decomposition) given as

$$T = \sum_{l_1=1}^{L_1} \cdots \sum_{l_N=1}^{L_N} t_{l_1 \cdots l_N} \hat{w}_1^{(l_1)} \otimes \cdots \otimes \hat{w}_N^{(l_N)}$$

Any modal truncation of the above decomposition is called the TSVD approximation in [10].

The main issues with the decompositions discussed in this section are listed below.

1) Singular values at level $k > K$ are zero. This leads to arbitrary completion of an orthogonal basis without using any information about the tensor.
2) Searching and optimizing in a space perpendicular to the space of all previous singular vectors is very restrictive in general.

VI. Orthogonal tensor approximation

In this section, we present a novel method for a tensor decomposition. Our aim is to find a decomposition method that avoids the main issues with the decompositions discussed in Section V. The new method is based on a generalization of the SVD property given in (5) and has its roots in the algorithms proposed by Belzen and Weiland in [3], [2], [10].

One issue with the decompositions discussed in Section V is that the search for a completely orthogonal decomposition is too restrictive. Since any completely orthogonal decomposition is also a orthogonal decomposition, one can consider orthogonal decomposition as an alternative to improve approximation error. If it exists, the completely orthogonal decomposition of a tensor (of order at least 3) is unique [11, theorem 3.2]. On the other hand, there exists a large number of possibilities to define orthogonal decompositions. Hence, the search for an orthogonal decomposition is more exhaustive and beneficial then a completely orthogonal decompositions.

An orthogonal decomposition can be found by an exhaustive search of singular values and singular vectors as explained next. For an order-N tensor T defined on finite dimensional Hilbert spaces $\mathbb{X}_1, \cdots, \mathbb{X}_N$ the first singular value is

$$\sigma_{\langle 1 \rangle}(T) = \hat{\sigma}_1(T)$$

where $\hat{\sigma}_1(T)$ is defined in (8). For ease of notations, *singular vectors at level* 1 are denoted by $\{w_1^{\langle 1 \rangle}, \cdots, w_N^{\langle 1 \rangle}\}$ where $w_i^{\langle 1 \rangle} = \hat{w}_i^{(1)}$ (see (8)).

In the next step we need a normalized rank-1 tensor which is orthogonal to $U_1 := w_1^{\langle 1 \rangle} \otimes w_2^{\langle 1 \rangle} \otimes \cdots \otimes w_N^{\langle 1 \rangle}$. This can be achieved by proceeding in N different directions i.e. perpendicular to $w_k^{\langle 1 \rangle}$ for some $k \in \mathcal{N}$. Let us denote the orthogonal complement of the space $\overline{\mathrm{span}}\{w_k^{\langle 1 \rangle}\}$, by $\mathcal{S}_{\langle 1,k \rangle}$. Now, corresponding to each of the N orthogonal complements, we calculate N singular values at level 2. For all $k \in \{1, \cdots, N\}$, we calculate N singular values at the second level as

$$\sigma_{\langle 1,k \rangle}(T) := \sup_{\substack{\|x_n\|=1, n \in \mathcal{N}, \\ \left\langle x_k, w_k^{\langle 1 \rangle} \right\rangle_{\mathbb{X}_k} = 0}} |T(x_1, \cdots, x_N)|.$$

Again, assume that this maximum is attained by an N-tuple, say $\{w_1^{\langle 1,k \rangle}, \cdots, w_N^{\langle 1,k \rangle}\}$. We call this N-tuple the *singular vectors in the space $\mathcal{S}_{\langle 1,k \rangle}$ at level* 2. Clearly, we can identify the level from the cardinality of the set $\langle 1, k \rangle$. Since, the search spaces $\mathcal{S}_{\langle 1,k \rangle} \forall k \in \mathcal{N}$ at the second level are supersets of the search space required in $\hat{\sigma}_2(T)$, it follows that $\sigma_{\langle 1,k \rangle}(T) \geq \hat{\sigma}_2(T)$ for all $k \in \mathcal{N}$.

Continuing in this way, at the third level we consider a maximum of N^3 different search spaces. These spaces are orthogonal complements of the spaces $\overline{\mathrm{span}}\{w_i^{\langle 1 \rangle}\} \times \overline{\mathrm{span}}\{w_j^{\langle 1,k \rangle}\}$ for all $i \neq j$ and $i, j, k \in \mathcal{N}$ and $\overline{\mathrm{span}}\{w_i^{\langle 1 \rangle}, w_j^{\langle 1,k \rangle}\}$ for all $i = j$. The number of search spaces increases exponentially with an increase of the level index. Specifically, at the mth level, we define the following set.

Definition VI.1. *A direction set at level m is a set*

$$\mathfrak{D}_m := \langle 1, k_2, k_3, \cdots, k_m \rangle.$$

where index k_2 refers to singular vector $w_{k_2}^{\langle 1 \rangle}$ at level 2, index k_3 refers to $w_{k_3}^{\langle 1,k2 \rangle}$ at level 3 and so on till index k_m refers to singular vector $w_{k_m}^{\langle 1,k_2,k_3,\cdots,k_{m-1} \rangle}$ at level $m-1$. Also, $\mathfrak{D}_1 := \langle 1 \rangle$.

The direction set indicates the choice of singular vectors that are selected at each level. At the mth level we find the singular value in a space orthogonal to the space of these singular vectors. Hence, we define the following search space.

Definition VI.2. *Given a direction set $\mathfrak{D}_m = \langle 1, k_2, k_3, \cdots, k_m \rangle$ at level m, the search space at level m is defined as*

$$\mathcal{S}^{\mathfrak{D}_m} := \mathcal{S}_1^{\mathfrak{D}_m} \times \cdots \times \mathcal{S}_N^{\mathfrak{D}_m}$$

where $\mathcal{S}_n^{\mathfrak{D}_m} := \left(\overline{\mathrm{span}}\{w_{k_i}^{\langle k_1,\cdots,k_i \rangle}\}_{\substack{1 \leq i < m \\ k_i = n}} \right)^{\perp}$. *Here $k_1 = 1$ and $\perp$ denotes the orthogonal complement of the space.*

Now, we define singular values and vectors at a given level in a given direction precisely.

Definition VI.3 (Singular value and vectors at a given level in a given direction)**.** *Given a direction set* $\mathfrak{D}_m = \langle 1, k_2, k_3, \cdots, k_m \rangle$ *at the level* m*, the* singular value in the space $\mathcal{S}^{\mathfrak{D}_m}$ *is defined as*

$$\sigma_{\mathfrak{D}_m}(T) := \sup_{\substack{\|x_k\|_{\mathbb{X}_k}=1, k\in\mathcal{N} \\ x_k \in \mathcal{S}_k^{\mathfrak{D}_m}}} |T(x_1, \cdots, x_N)|. \tag{10}$$

This singular value is attained by an N*-tuple, say* $\{w_1^{\mathfrak{D}_m}, \cdots, w_N^{\mathfrak{D}_m}\}$*. Every such* N*-tuple is referred as the* singular vectors in $\mathcal{S}^{\mathfrak{D}_m}$ at level m*.*

Note that there exists no definite relationship between CO-singular values (defined in Section V) and singular values after level-2.

To identify non-zero singular values in a search space $\mathcal{S}^{\mathfrak{D}_m}$ at the mth level, we denote the orthogonal projections into the space $(\mathcal{S}_n^{\mathfrak{D}_m})^\perp$ by $P_n^{\mathfrak{D}_m}$. With these notions it is trivial to show that $\sigma_{\mathfrak{D}_m}(T)$ of the tensor T exists and is non-zero if and only if $T((I - P_1^{\mathfrak{D}_m})x_1, \cdots, (I - P_N^{\mathfrak{D}_m})x_N) \neq 0$.

Now, assume that for a some M, $P_n^{\mathfrak{D}_M} = I$ for all $n \in \mathcal{N}$. Then we can write

$$\acute{T}_o = \sum_{m=1}^{M} \sigma_{\mathfrak{D}_m}(T) w_1^{\mathfrak{D}_m} \otimes \cdots \otimes w_N^{\mathfrak{D}_m} \tag{11}$$

To illustrate these ideas, let us consider the following example.

Example VI.4. *Let* T *be as in Example V.1. Proceed sequentially by searching singular values and singular vectors in the space* $\mathcal{S}^{\mathfrak{D}_m}$ *at the* m*th level. Here,* $\mathfrak{D}_m = \mathfrak{D}(1:m)$ *and* $\mathfrak{D} := \{1, 2, 3, 1\}$*. The set* $\mathfrak{D}(1:m)$ *denotes the set that consists of the first* m *elements of* $\mathfrak{D}$*. Hence,* $\acute{T}_o = 2\, e_1 \otimes e_1 \otimes e_1 + \frac{1}{\sqrt{2}}\,(e_1 + e_2) \otimes e_2 \otimes e_2$ *Clearly,* $T = \acute{T}_o$*.*

Comparing the above example with Example V.1, it follows that $\acute{T}_o$ is a better decomposition than $\hat{T}_c$. The above example shows that T admits an orthogonal CP decomposition. However, this is not true in general as the following example illustrates.

Example VI.5. *Consider the tensor* $T : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ *defined by* $T = e_1 \otimes e_1 \otimes e_1 + e_2 \otimes e_1 \otimes e_1 + e_2 \otimes e_2 \otimes e_2$ *where* $e_1 := \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ *and* $e_2 := \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ *are an orthonormal basis of* $\mathbb{R}^2$*. Now,* $T = (e_1 + e_2) \otimes e_1 \otimes e_1 + e_2 \otimes e_2 \otimes e_2$*. Proceed sequentially by searching singular values/vectors in the space* $\mathcal{S}^{\mathfrak{D}_m}$ *at the* m*th level. Here* $\mathfrak{D}_m = \mathfrak{D}(1:m)$ *and* $\mathfrak{D} := \{1, 1, 2, 3\}$*. Hence,* $\acute{T}_o = (e_1 + e_2) \otimes e_1 \otimes e_1 + \frac{1}{2}(-e_1 + e_2) \otimes e_2 \otimes e_2$*. Clearly,* $T - \acute{T}_o = \frac{1}{2}(e_1 + e_2) \otimes e_2 \otimes e_2$ *is not zero.*

Example VI.5 shows that $\acute{T}_o$ may not be equal to T. Therefore, $\acute{T}_o$ is not an orthogonal decomposition of the tensor T. However, we can easily obtain a modal-rank decomposition of the form

$$T = \sum_{l_1=1}^{L_1} \cdots \sum_{l_N=1}^{L_N} t_{l_1 \cdots l_N} u_1^{(l_1)} \otimes \cdots \otimes u_N^{(l_N)} \tag{12}$$

where $t_{l_1 \cdots l_N} = T(u_1^{(l_1)}, \cdots, u_N^{(l_N)})$. Few ways of obtaining $u_n^{(i)}$ are explained below.

1) *Directional modal rank decomposition:* Assume that the projection matrices $P_n^{\mathfrak{D}_M} = I$ at some Mth level for all $n \in \mathcal{N}$. Then we can have $\begin{bmatrix} u_n^{(1)} & \cdots & u_n^{(L_n)} \end{bmatrix} := P_n^{\mathfrak{D}_M}$ for all n.
2) *Modal rank decomposition via orthogonalization:* Assume that set of singular vector $\{w_n^{\mathfrak{D}_m}\}_{1\leq m\leq M}$ for some M forms a basis of the space $\mathbb{X}_n$. Then $u_n^{(i)}$ are the vectors obtained from Gram-Schmidt orthogonalization of the set $\{w_n^{\mathfrak{D}_m}\}_{1\leq m\leq M}$ for some M.
3) *Single directional modal-rank decomposition (SDM):* We can obtain n different $\acute{T}_o$ given in (11), if we proceed in n different directions given by set $\mathfrak{D}_m^n := \{\mathfrak{D}_m^n(k) = n \forall 1 < k \leq m \; and \; \mathfrak{D}_m^n(1) = 1\}$ for all levels $1 \leq m \leq \dim \mathbb{X}_n$ and $n \in \mathcal{N}$. Now, set $u_n^{(m)} = w_n^{\mathfrak{D}_m^n}$.

Note that $u_n^{(i)}$ always depends upon the tensor. This solves the issue 1 in Section V.

A. A short note on selection of the direction set

The above setting defines a class of different decompositions depending on the direction set we choose at each level. Now, we can get the direction set based on the maximum singular value at each level or the singular values which provide a best approximation (in Frobenius sense) of the tensor T at a given level. This is computationally intensive because of an exponential increase in the number of singular values at each level.

To reduce calculation, we use the constraint $\mathfrak{D}_i = \mathfrak{D}_{i+1}(1:i)$. Now, the singular values are ordered and never increase with levels as $\mathcal{S}^{\mathfrak{D}_i} \supseteq \mathcal{S}^{\mathfrak{D}_{i+1}}$. Also, the number of singular value we need to compute now is N at each level $m > 1$. We used this simple method to obtain $\acute{T}_o$ given in (11) and to obtain a modal rank decomposition via orthogonalization. We refer to this modal-rank decomposition as *maximum singular value modal-rank (MSVM)* decomposition.

We claim that all modal-rank decompositions discussed in this section are equivalent to the classical SVD for order-2 tensors (see Appendix for the proof).

Because of space limitations, we omit a discussion on the relation between the singular values and the upper bound of the errors inferred from the modal truncation the we considered in this section.

VII. Calculation of the singular value and vectors

In this section, we propose a numerical scheme to determine the singular value and singular vectors of an arbitrary direction set. Using Lagrangian multipliers, we obtain the following theorem (proof is skipped due to length constraints).

Theorem VII.1. *Given an order-N multi-linear functional $T : \mathbb{X}_1 \times \cdots \times \mathbb{X}_N \to \mathbb{R}$ defined on finite dimensional vector spaces $\mathbb{X}_n$ of dimension L_n. Let $\{y_n^{(i)}\}$ be a set of independent vectors in $\mathbb{X}_n$ where $n \in \mathcal{N}$, $i \in \{1, \cdots, I_n\}$ and $I_n \leq L_n$. Define*

$$\{\tilde{w}_1, \cdots, \tilde{w}_N\} := \underset{\substack{\|x_n\|=1, n\in\mathcal{N}, \\ \langle x_n, y_n^{(i)}\rangle_{\mathbb{X}_n}=0\, \forall i \leq I_n}}{\arg\sup} |T(x_1, \cdots, x_N)|.$$

and $\tilde{\sigma} := T(\tilde{w}_1, \cdots, \tilde{w}_N)$. Then for all $a_n \in \mathbb{X}_n$ such that $\left\langle a_n, y_n^{(i)} \right\rangle_{\mathbb{X}_n} = 0$, $n \in \mathcal{N}$ and $i \in \{1, \cdots, I_n\}$ with $I_n \leq L_n$ we have that

$$T(\tilde{w}_1, \cdots, a_n, \cdots, \tilde{w}_N) = \tilde{\sigma} \langle a_n, \tilde{w}_n \rangle_{\mathbb{X}_n},$$
$$\|\tilde{w}_n\|_{\mathbb{X}_n} = 1, \quad \left\langle \tilde{w}_n, y_n^{(i)} \right\rangle_{\mathbb{X}_n} = 0.$$

A. Algorithm

Theorem VII.1 provides a set of equations that have singular values and singular vectors as solutions. A numerical scheme for the computation of singular values and vectors is based on the operator

$$G_T(x) := \frac{1}{T(\frac{x_1}{\|x_1\|}, \cdots, \frac{x_N}{\|x_N\|})} \nabla T(\frac{x_1}{\|x_1\|}, \cdots, \frac{x_N}{\|x_N\|})$$

where $x_k \in \mathbb{X}_k$, $x = \begin{bmatrix} x_1^\top & \cdots & x_N^\top \end{bmatrix}^\top$, and ∇T denotes the gradient of T with respect to x. It can be shown that the fixed points of G_T coincide with the solutions of the set of equations given in Theorem VII.1. The algorithm iterates on G_T in the sense that we compute

$$x^{(n+1)} = G_T(x^{(n)}) \quad \text{for } n = 0, 1, \ldots.$$

Unfortunately, convergence of this algorithm is not guaranteed (even locally). Because of this, we propose to use the iterative map $H_T := \theta G_T + (1-\theta) I$ instead of G_T. With this map, (local) convergence of iterations to fixed points is actually guaranteed for specific values of θ. Specifically, we have the following theorem (proof is skipped due to length constraints).

Theorem VII.2 (Local convergence theorem)**.** *Define $H_T := \theta G_T + (1-\theta)I$ where $\theta \in (0, 1]$. If the fixed point of H_T is a strict maximum and $\theta \in (0, \frac{2}{N})$, then the iteration $x^{(n+1)} = H_T(x^{(n)})$ converges to the fixed point from every initial condition $x^{(0)}$ in an open neighborhood around the fixed point.*

Fig. 1. Starting from left: cross section of original MRI scan, $(15, 15, 15)$ approximation of the MRI scan computed using TSVD, HOSVD, MSVM and SDM.

Note that the fixed points of H_T and G_T are same. This result leads to the following power-type algorithm to compute the singular values and corresponding vectors for a given direction set $\mathfrak{D}_m$.

INPUT Tensor $T \in \mathcal{T}_N$

DESIRED Singular value/vectors of T at level m for a given direction set $\mathfrak{D}_m$.

Step 0 Set tolerance level $\varepsilon_{\text{tol}} > 0$. Define $\tilde{T}(x_1, \cdots, x_N) := T((I - P_1^{\mathfrak{D}_m})x_1, \cdots, (I - P_N^{\mathfrak{D}_m})x_N)$.

Step 1 Select random $x^{(0)} \in \mathbb{R}^{\sum_i^N L_i}$ such that $\|x_n^{(0)}\|_{\mathbb{X}_n} = 1$.

Step 2 Select $\theta \in (0, \frac{2}{N})$ and iterate the map $x^{(n+1)} = H_{\tilde{T}}(x^{(n)})$, for all $i = 0, 1, \ldots, i^*$ where i^* is such that $\|x^{(i^*)} - x^{(i^*-1)}\|/\|x^{(i^*)}\| < \varepsilon_{\text{tol}}$.

Step 3 Write $\tilde{w}_n^{\mathfrak{D}_m} = x_n^{(i^*)}$

Step 4 If $\tilde{T}(\tilde{w}_1^{\mathfrak{D}_m}, \cdots, \tilde{w}_N^{\mathfrak{D}_m}) < 0$ then set $\tilde{w}_1^{\mathfrak{D}_m} = -\tilde{w}_1^{\mathfrak{D}_m}$.

In the last step, we used multi-linearity to have a non-negative singular value. Initialization with the singular vectors obtained by the HOSVD sometimes helps to improve convergence in the above algorithm [4].

VIII. Simulation

In this section, we illustrate the power of the algorithms described in this paper. We consider a modal truncation of the MSVM and SDM decompositions and we compare the performance of these methods with the TSVD and the HOSVD algorithms [10], [4].

The data consists of a 3D MRI scan of a human head. The same data is used in [10]. The data is a 3-dimensional object of gray-scale levels defined on a grid of $262 \times 262 \times 29$ pixels. The data is stored as an order-3 tensor T of size $262 \times 262 \times 29$. The modal rank of the tensor T is $(255, 219, 29)$ (calculated by MATLAB [9]). One planar cross section of the data is shown in the left image of Figure 1. All simulations are performed with tolerance $\varepsilon_{\text{tol}} = 10^{-8}$ and $\theta = 0.5$.

We performed modal truncations at level (r_1, r_2, r_3) where $r_1 = r_2 = r_3 = r$ and $1 \leq r \leq 15$. The relative approximation errors $\|T - T_{r_1:r_3}\|_F / \|T\|_F$ for the various algorithms are shown in Table I. Here $T_{r_1:r_3}$ is the modal truncation at level (r_1, r_2, r_3) of T. It follows that both MSVM and SDM truncations outperform the TSVD and that accuracy of SDM is comparable to the

(r_1, r_2, r_3)	TSVD	HOSVD	MSVM	SDM
$(1,1,1)$	0.51814	0.51815	0.51814	0.51814
$(3,3,3)$	0.51269	0.26496	0.26480	0.26487
$(5,5,5)$	0.51089	0.23352	0.23073	0.23097
$(7,7,7)$	0.43250	0.21127	0.20730	0.20718
$(9,9,9)$	0.43104	0.19269	0.19255	0.19227
$(11,11,11)$	0.42965	0.18092	0.18598	0.18217
$(13,13,13)$	0.42752	0.17311	0.17890	0.17316
$(15,15,15)$	0.42573	0.16668	0.17524	0.16639

TABLE I

RELATIVE APPROXIMATION ERROR FOR VARIOUS ALGORITHMS

truncations of the HOSVD (see also Figure 1).The slow decay rate of the TSVD truncation error is attributed to the restrictions of the search space to define a completely orthogonal decomposition. Note that for level $(1,1,1)$ approximations, SDM, TSVD, MSVM are always better than HOSVD (see [10, theorem V.2]).

The compression achieved by truncation at level $(15,15,15)$ is $1 - \frac{15\times15\times15}{255\times219\times29} = 0.99792$ i.e. 99.792%.

IX. CONCLUSIONS

This paper addresses the approximation problem of multi-linear functionals defined on finite dimensional vector spaces. We proposed a number of decompositions of tensors as dyadic expansions of rank-1 elements and considered their properties as candidate lower rank approximations of a given tensor. This approximation problem is motivated by questions on signal and data compression for large-scale data elements that consist of multiple independent variables. This paper contributes with novel approximation methods of tensors by defining different classes of singular values and singular vectors of tensors. For general order-N tensors, the proposed decompositions depend on search directions at every decomposition level. This leads to a plethora of algorithms for tensor approximation. We focused on two easily computable algorithms (MSVM and SDM) for tensor approximations and compared these algorithms with the existing TSVD and HOSVD algorithm. Convergence of the numerical algorithms has been established and we have shown that these algorithms outperform the TSVD algorithm. The approximation accuracy of SDM is comparable with the HOSVD truncations.

APPENDIX

We now show that the singular values and vectors obtained in (5), forms an SVD of a given matrix M. It is shown in [10, e.g. IV.3] that (4) is equivalent to SVD of a matrix. Therefore, we just show that for a given order-2 tensor, the maximization problem defined in (5) is equivalent to the problem in (4).

Consider an order two tensor $T : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. The first singular value/vectors in (5) and (4) are same. Therefore, we have that $T(x_1, y) = \sigma_1 \langle y, y_1 \rangle_{\mathbb{R}_m}$,

$$T(x, y_1) = \sigma_1 \langle x, x_1 \rangle_{\mathbb{R}_n} \tag{13}$$

and $\|x_1\|_{\mathbb{R}_n} = 1$ and $\|y_1\|_{\mathbb{R}_m} = 1$. Therefore, we start with the second singular value/vectors. Using Theorem VII.1, we can show that the singular value σ_2 and singular vectors $\{x_2, y_2\}$ defined in (5) satisfy $\|x_2\|_{\mathbb{R}_n} = 1$, $\|y_2\|_{\mathbb{R}_m} = 1$, $\langle x_2, x_1 \rangle = 0$,

$$T(x_2, y) = \sigma_2 \langle y, y_2 \rangle_{\mathbb{R}_m} \tag{14}$$

and $T(x, y_2) = \sigma_2 \langle x, x_2 \rangle_{\mathbb{R}_n}$, for all $x \in \mathbb{R}_n$ and $y \in \mathbb{R}_m$. Now, setting $y = y_1$ in (14) and setting $x = x_2$ in (13), we have $T(x_2, y_1) = \sigma_2 \langle y_1, y_2 \rangle_{\mathbb{R}_m} = \sigma_1 \langle x_2, x_1 \rangle_{\mathbb{R}_n} = 0$. This implies $\langle y_1, y_2 \rangle_{\mathbb{R}_m} = 0$.

Now, we use induction to prove the rest. Assume that the $\{y_{n-1}, \cdots, y_1\}$ of right singular vectors is a mutually orthonormal set, we prove that y_n is orthogonal to this set. Using Theorem VII.1, we can obtain singular value σ_n and singular vectors $\{x_n, y_n\}$ defined in (5), satisfy $\|x_n\|_{\mathbb{R}_n} = 1$, $\|y_n\|_{\mathbb{R}_m} = 1$, $x_n \perp \{x_{n-1}, \cdots, x_1\}$ and $T(x_n, y) = \sigma_n \langle y, y_n \rangle_{\mathbb{R}_m}$, $T(x, y_n) = \sigma_n \langle x, x_n \rangle_{\mathbb{R}_n}$ for all $x \in \mathbb{R}_n$ and $y \in \mathbb{R}_m$. Setting $y = y_k$ $(1 \leq k < n)$, we have $T(x_n, y_k) = \sigma_n \langle y_k, y_n \rangle_{\mathbb{R}_m} = \sigma_k \langle x_n, x_k \rangle_{\mathbb{R}_n} = 0$. This implies $\langle y_k, y_n \rangle_{\mathbb{R}_m} = 0$ for all $1 \leq k < n$. We get the above result even if exchange x and y, therefore given any direction set we will have same decomposition in order-2 tensor case.

REFERENCES

[1] F. Belzen and S. Weiland. Efficient simulation of large-scale dynamical systems using tensor decompositions. In *Scientific Computing in Electrical Engineering 2008*, Mathematics in Industry, pages 413–420. Springer Berlin Heidelberg, 2010.

[2] F Belzen and S Weiland. A tensor decomposition approach to data compression and approximation of nd systems. *Multidimensional Systems and Signal Processing*, 23(1-2):209–236, 2012.

[3] F.V. Belzen. *Approximation of Multi-variable Signals and Systems: a Tensor Decomposition Approach.* PhD thesis, Technische Universiteit Eindhoven, 2011.

[4] L. De Lathauwer. *Signal Processing Based on Multilinear Algebra*. PhD thesis, Katholieke Universiteit Leuven, 1997.

[5] 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(3):1084–1127, September 2008.

[6] L. Eldán 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. on Matrix Analysis and Applications*, 31(2):248–271, 2009.

[7] T.G. Kolda. Orthogonal tensor decompositions. *SIAM J. on Matrix Analysis and Applications*, 23(1):243–255, 2001.

[8] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, September 2009.

[9] MATLAB. *version 7.12.0 (R2011a)*. The MathWorks Inc., Natick, Massachusetts, 2011.

[10] S. Weiland and F.V. Belzen. Singular value decompositions and low rank approximations of tensors. *IEEE Trans. Signal Processing*, 58(3):1171–1182, 2010.

[11] T. Zhang and G. H. Golub. Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.*, 23(2):534–550, February 2001.