Approximation of n D Systems using Tensor Decompositions

F. van Belzen and S. Weiland

Abstract—This paper considers reduced order modeling of systems with multiple independent variables. The method of Proper Orthogonal Decompositions (POD) is a data-based method that is suitable for the reduction of large-scale distributed systems. In this paper we propose a generalization of the POD method so as to take the n D nature of the distributed model into account. Suitable projection spaces can be computed by associating a tensor with the measurement data and computing a suitable decomposition of this tensor. We demonstrate how prior knowledge about the structure of the model reduction problem can be used to improve the quality of approximations. The tensor decomposition techniques are demonstrated on a data approximation example and then the model reduction process is illustrated using a heat diffusion problem.

Index Terms—Model Reduction, n D Systems, Proper Orthogonal Decompositions, Multilinear Algebra, Tensors

I. INTRODUCTION

Systems that have space and time as independent variables occur frequently in all fields of science and engineering. Typically, dynamic relations among variables of these systems are described by Partial Differential Equations (PDEs). In most cases, analytical solutions to the PDEs are not known, which makes numerical solutions necessary. Numerical solutions can be computed using Finite Element (FE) methods on discrete grids for both space and time. In a lot of applications accurate solutions are required on fine grids, leading to large-scale discrete systems and, consequently, to large simulation times. This makes FE models not very suitable for on-line monitoring, prediction or model-based control.

Model reduction methods can be used to reduce the complexity of large-scale dynamical systems. The goal of model reduction is to obtain models that are computationally efficient and at the same time offer an accurate description of the system under consideration. Previous work on model reduction for infinite dimensional systems, [2], [1], considers only one independent variable, namely time. Furthermore, these papers focus on the accurate representation of the input/output behavior in the reduced models. Our interest lies in obtaining reduced models that offer an accurate description of the state evolution of an n D system. In addition, the aim is to take multiple independent variables into account.

Proper Orthogonal Decompositions (POD) [3], [4], [5] is among the most popular model reduction methods that can be applied to systems with multiple independent variables. POD is a projection-based method that relies on the computation

of empirical projection spaces from a representative set of measurement or simulation data. In its classical formulation, the projection spaces are used in a spectral expansion that separates space and time. No further structure is assumed for the spatial variables. This makes POD basically a two-variable method since it deals with an n D system by separating time and space, where independent variables are assumed to belong to a Cartesian product of time and space.

In this paper we will show that it is possible to generalize the POD method to explicitly take the n D nature of the original problem into account. This can be accomplished by assuming a more general Cartesian structure for the independent variables. In this paper we will advocate that tensors provide an adequate representation of signals and a suitable mathematical approach for the reduction of n D systems. Then, suitable projection spaces can be computed using tensor decompositions. This approach makes it possible to reduce each spatial variable separately, instead of reducing the total spatial domain at once, as is usually done in POD. In the second part of the paper we will show how prior knowledge about the structure of the model reduction problem can be used to improve the quality of approximations.

This paper is organized as follows. In Section II we will give a summary of the POD method. In Section III we will generalize the POD method to arbitrary n D systems. Then, in Section IV the POD basis construction is changed based on prior knowledge, to improve accuracy. In Section V two examples are discussed and finally conclusions are given. Proofs to the theorems can be found in the appendix.

II. PROPER ORTHOGONAL DECOMPOSITIONS

Consider an arbitrary linear distributed system described by the Partial Differential Equation (PDE)

$$R\left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N}\right)w = 0. \quad (1)$$

Here, $R \in \mathbb{R}^{\times 1}[\xi_1, \dots, \xi_N]$ is a real matrix valued polynomial in N indeterminates and (1) is viewed as a PDE in the signal $w : \mathbb{X} \subset \mathbb{R}^N \rightarrow \mathbb{R}$ that evolves over N independent variables. The domain of the signal w , $\mathbb{X}$, is assumed to have a Cartesian structure $\mathbb{X} = \mathbb{X}' \times \mathbb{X}''$, which is typically the product of a spatial and a temporal domain. A Hilbert space $\mathcal{H}$ of square integrable functions on $\mathbb{X}'$ is introduced. $\mathcal{H}$ is assumed to be separable, which means that a countable orthonormal basis $\{\varphi_n, n = 1, 2, \dots\}$ for $\mathcal{H}$ exists. Solutions w of (1) are assumed to satisfy $w(\cdot, x'') \in \mathcal{H}$ for all $x'' \in \mathbb{X}''$ and can be represented by a spectral expansion

$$w(x', x'') = \sum_n a_n(x'')\varphi_n(x') \quad (2)$$

This work is supported by the Dutch Technologiestichting STW under project number EMR.7851.

Control Systems, Department of Electrical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. f.v.belzen@tue.nl, s.weiland@tue.nl

in which the modal coefficients a_n are uniquely determined by $a_n = \langle w, \varphi_n \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathcal{H}$. For $r > 0$ reduction of the signal space $\mathcal{H}$ can be defined by the truncation

$$w_r(x', x'') = \sum_{n=1}^r a_n(x'') \varphi_n(x') \quad (3)$$

For $r > 0$, the reduced order model is then defined by the collection of solutions $w_r(x', x'') = \sum_{n=1}^r a_n(x'') \varphi_n(x')$ that satisfy the Galerkin projection

$$\langle R \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N} \right) w_r, \varphi \rangle = 0 \quad \forall \varphi \in \mathcal{H}_r \quad (4)$$

where $\mathcal{H}_r$ is the finite dimensional projection space $\mathcal{H}_r = \text{span}\{\varphi_1, \dots, \varphi_r\}$. If the spectral expansion of w_r is substituted in (4) and $\mathbb{X}'' \subseteq \mathbb{R}$, then (4) becomes a system of r ordinary differential equations in the modal coefficients a_n , $n = 1, \dots, r$.

Clearly, the quality of the reduced order model (4) entirely depends on the choice of basis functions $\{\varphi_n\}$. In the POD method, the orthonormal basis functions $\{\varphi_n\}$ of $\mathcal{H}$ are determined empirically, either from measurements or data simulated from (1). Specifically, for given data w with $w(\cdot, x'') \in \mathcal{H}$ and $x'' \in \mathbb{X}''$, the basis functions φ_n are the ordered normalized eigenfunctions of the data correlation operator $\Phi : \mathcal{H} \rightarrow \mathcal{H}$ that is defined as

$$\langle \psi_1, \Phi \psi_2 \rangle := \int_{\mathbb{X}''} \langle \psi_1, w(\cdot, x'') \rangle \cdot \langle w(\cdot, x''), \psi_2 \rangle dx''$$

$\psi_1, \psi_2 \in \mathcal{H}.$

The data correlation operator Φ is a well defined linear, bounded, self-adjoint and non-negative operator on $\mathcal{H}$. That is, the basis functions φ_n satisfy $\Phi \varphi_n = \lambda_n \varphi_n$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and achieve that the error

$$\int_{\mathbb{X}''} \|w(\cdot, x'') - w_r(\cdot, x'')\|^2 dx'' = \sum_{n>r} \lambda_n$$

is minimal for all truncation levels $r > 0$.

In applications, the PDE (1) is discretized by sampling the domain $\mathbb{X} = \mathbb{X}' \times \mathbb{X}''$ and by discretizing the PDE according to

$$D(\varsigma_1, \dots, \varsigma_N) w = 0 \quad (5)$$

where $D \in \mathbb{R}^{\times 1}[\xi_1, \dots, \xi_N]$ is a real matrix-valued polynomial in N indeterminates and ς_n is the forward shift operator acting on the spatial discretization in the n th mode according to $\varsigma_n s(x_1^{(\ell_1)}, \dots, x_n^{(\ell_n)}, \dots, x_N^{(\ell_N)}) = s(x_1^{(\ell_1)}, \dots, x_n^{(\ell_{n+1})}, \dots, x_N^{(\ell_N)})$. We will assume that (5) is an accurate representation of (1) and refer to solutions of (5) as Finite Element solutions. For the discretized model (5) the solution space $\mathcal{H}$ becomes finite, but large, dimensional, provided that $\mathbb{X}'$ is finite. Then Φ becomes a symmetric non-negative definite matrix and the calculation of POD basis functions becomes an algebraic eigenvalue or singular value decomposition problem.

This section forms a very brief introduction to the method of Proper Orthogonal Decompositions only, more information can be found in [3], [4], [5].

III. EXPLOITING STRUCTURE

Usually in POD $\mathbb{X}'$ is identified with space and $\mathbb{X}''$ is identified with time. No further structure is assumed for the spatial domain. In particular, for larger dimensional Euclidean geometries, all spatial variables are stacked and this yields a large-dimensional data correlation operator Φ . In this way, the N-D nature of the original problem is replaced by an artificial 2-D structure. Furthermore, in high-dimensional applications stacking of spatial variables leads to large dimensional basis vectors, which are difficult to compute or to process.

In this paper we will adapt the POD method in such a way that the N-D nature of the original system is explicitly taken into account. The key to make this possible is to assume structure for the spatial variables instead of stacking them. We will assume that the spatial domain has a Cartesian structure. It is then possible to associate a tensor with the snapshot data. POD basis functions can be determined by computing a suitable decomposition of this tensor.

In this section we will first give some definitions regarding tensors and their properties. Then, we will introduce a Singular Value Decomposition (SVD) for tensors and show how this can be used to compute POD basis functions. Finally, the construction of reduced models is discussed.

A. Tensors

Assume that the domain $\mathbb{X}$ of (1) has the Cartesian structure $\mathbb{X} = \mathbb{X}_1 \times \dots \times \mathbb{X}_N$ and that, for $n = 1, \dots, N$, the domain $\mathbb{X}_n$ is gridded into a finite set of L_n elements and let $X_n := \mathbb{R}^{L_n}$ be equipped with its standard Euclidean inner product. Suppose that w is a *Finite Element* (FE) solution of (5) that is defined on the $L = \prod_{n=1}^N L_n$ grid elements. Then w defines a multidimensional array $[[w]] \in \mathbb{R}^{L_1 \times \dots \times L_N}$ in which the $(\ell_1, \dots, \ell_N)$ th entry is the sample $w_{\ell_1 \dots \ell_N}$ on the Cartesian grid.

At a more abstract level $[[w]]$ defines a tensor. An *order- N tensor* T is a multilinear functional $T : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$ operating on N vector spaces $X_1, \dots, X_N$. The elements of T , $t_{\ell_1 \dots \ell_N}$, are defined with respect to bases for $X_1, \dots, X_N$ according to $t_{\ell_1 \dots \ell_N} = T(e_1^{\ell_1}, \dots, e_N^{\ell_N})$, where $\{e_n^{\ell_n}, \ell_n = 1, \dots, L_n\}$ is a basis for X_n , $n = 1, \dots, N$. For example, $T(x_1, x_2) := \langle x_1, A x_2 \rangle$ defines an order-2 tensor whose element t_{ℓ_1, ℓ_2} is the (ℓ_1, ℓ_2) th entry of the matrix A .

A FE solution w , or its associated multidimensional array $[[w]]$, therefore defines the tensor

$$W = \sum_{\ell_1=1}^{L_1} \dots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} e_1^{\ell_1} \otimes \dots \otimes e_N^{\ell_N} \quad (6)$$

where $e_1 \otimes \dots \otimes e_N$ is shorthand for the *rank-1 tensor* $E : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$, defined by $E(x_1, \dots, x_N) := \prod_{n=1}^N e_n^T x_n$ and where $w_{\ell_1 \dots \ell_N}$ is the data element on the sample point with index $(\ell_1, \dots, \ell_N)$.

The *inner product* of two tensors $S, T \in \mathcal{T}_N$ with elements $s_{k_1, \dots, k_N}$ and $t_{\ell_1, \dots, \ell_N}$, both defined with respect to the same

bases for $X_1, \dots, X_N$, is given by

$$\langle S, T \rangle := \sum_{k_1=1}^{L_1} \cdots \sum_{k_N=1}^{L_N} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} s_{k_1, \dots, k_N} t_{\ell_1, \dots, \ell_N} \langle e_1^{k_1}, e_1^{\ell_1} \rangle \cdots \langle e_N^{k_N}, e_N^{\ell_N} \rangle.$$

It is immediate that the right-hand side of this expression is invariant under unitary basis transformations (i.e., transformations $Q_n : X_n \rightarrow X_n$ for which $\|Q_n x\|_n = \|x\|_n$ for all $x \in X_n$) and so $\mathcal{T}_N$ becomes a well defined inner product space. The Frobenius norm of $T \in \mathcal{T}_N$ is defined as

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

The concept of *tensor rank* is a highly non-trivial extension of the same concept for linear mappings and has been discussed in considerable detail in, for example, [13], [8], [9], [10], [11], [12]. To define the *modal rank* of a tensor $W \in \mathcal{T}_N$, we first introduce the *n-mode kernel* of W to be the set

$$\ker_n(W) := \{x_n \in X_n \mid W(x_1, \dots, x_N) = 0, \text{ for all } x_k \in X_k, k \neq n\}. \quad (7)$$

The multi-linearity of W implies that $\ker_n(W)$ is a linear subspace of X_n . The *n-mode rank* of W , is defined by

$$R_n = \text{rank}_n(W) := L_n - \dim(\ker_n(W)), \quad n = 1, \dots, N.$$

Note that $\text{rank}_n(W)$ coincides with the dimension of the space spanned by stringing out all elements $w_{\ell_1, \dots, 1, \dots, \ell_N}$ till $w_{\ell_1, \dots, N, \dots, \ell_N}$ (where the indices $1, \dots, N$ are at the n th spot). Finally, the *modal rank* of W , denoted $\text{modrank}(W)$, is the vector of all *n-mode ranks*, i.e., $\text{modrank}(W) = (R_1, \dots, R_N)$, $R_n = \text{rank}_n(W)$. Throughout this section it is assumed that W is a given tensor in $\mathcal{T}_N$ of modal rank $\text{modrank}(W) = (R_1, \dots, R_N)$.

The tensor (6) associated with the FE solution defines suitable projection spaces by decomposing the tensor W in rank-1 tensors as follows. For each of the vector spaces X_n , $n = 1, \dots, N$ we propose the construction of an orthonormal basis $\{\varphi_n^{\ell_n}, \ell_n = 1, \dots, L_n\}$ such that a coordinate change of W with respect to these bases achieves that the truncated tensor

$$W_r := \sum_{\ell_1=1}^{r_1} \cdots \sum_{\ell_N=1}^{r_N} \hat{w}_{\ell_1, \dots, \ell_N} \varphi_1^{\ell_1} \otimes \cdots \otimes \varphi_N^{\ell_N} \quad (8)$$

with $r = (r_1, \dots, r_N)$ and $r_n \leq L_n$, $n = 1, \dots, N$, will minimize the error $\|W - W_r\|$, in a suitable tensor norm, [8], [6].

For order-2 tensors (matrices) this problem is solved by the singular value decomposition. For higher-order tensors, it is not straightforward how to construct proper sets of orthonormal bases with this property. Different methods exist, including the Higher-Order Singular Value Decomposition (HOSVD) [10] and the Tensor SVD [6]. Since the HOSVD is basically a matrix-method in which the multilinear structure of a tensor is replaced by multiple bilinear structures, we will

not consider this method in this paper. Instead we will focus on the Tensor SVD. Definitions of this SVD are summarized in the next section.

B. Computation of POD basis using a Singular Value Decomposition for Tensors

Let $W \in \mathcal{T}_N$ be an order- N tensor defined on the finite dimensional vector spaces $X_1, \dots, X_N$ where we suppose that $\dim(X_n) = L_n$. The *singular values* of W , denoted $\sigma_k(W)$, with $k = 1, \dots, K$ and $K = \min_n \text{modrank}(W)$ are defined as follows.

For $n = 1, \dots, N$ let

$$\mathcal{S}_n^{(1)} := \{x \in X_n \mid \|x\|_n = 1\}$$

denote the unit sphere in X_n . Define the first singular value of W by

$$\sigma_1(W) := \sup_{\substack{x_n \in \mathcal{S}_n^{(1)}, \\ 1 \leq n \leq N}} |W(x_1, \dots, x_N)|. \quad (9)$$

Since W is continuous and the Cartesian product $\mathcal{S}^{(1)} = \mathcal{S}_1^{(1)} \times \cdots \times \mathcal{S}_N^{(1)}$ of unit spheres is a compact set, an extremal solution of (9) exists (i.e., the supremum in (9) is a maximum) and is attained by an N -tuple

$$(\varphi_1^{(1)}, \dots, \varphi_N^{(1)}) \in \mathcal{S}^{(1)}.$$

Subsequent singular values of W are defined in an inductive manner by setting

$$\mathcal{S}_n^{(k)} := \{x \in X_n \mid \|x\|_n = 1, \langle x, \varphi_n^{(j)} \rangle_n = 0 \text{ for } j = 1, \dots, (k-1)\}$$

for $n = 1, \dots, N$, and by defining

$$\sigma_k(W) = \sup_{\substack{x_n \in \mathcal{S}_n^{(k)}, \\ 1 \leq n \leq N}} |W(x_1, \dots, x_N)|, \quad k \leq K. \quad (10)$$

Again, since the Cartesian product

$$\mathcal{S}^{(k)} = \mathcal{S}_1^{(k)} \times \cdots \times \mathcal{S}_N^{(k)}$$

is compact, the supremum in (10) is a maximum that is attained by an N -tuple

$$(\varphi_1^{(k)}, \dots, \varphi_N^{(k)}) \in \mathcal{S}^{(k)}.$$

It follows that the vectors $\varphi_n^{(1)}, \dots, \varphi_n^{(K)}$ are mutually orthonormal in X_n . If $K < L_n$ for any n , then we extend the collection of orthogonal elements $\varphi_n^{(1)}, \dots, \varphi_n^{(K)}$ to a complete orthonormal basis of X_n . This construction leads to a collection of orthonormal bases

$$\{\varphi_1^{(\ell_1)}, \ell_1 = 1, \dots, L_1\}, \dots, \{\varphi_N^{(\ell_N)}, \ell_N = 1, \dots, L_N\} \quad (11)$$

for the vector spaces $X_1, \dots, X_N$, respectively.

Definition III.1 The *singular values* of a tensor $W \in \mathcal{T}_N$ are the numbers $\sigma_1, \dots, \sigma_K$, $K = \min_n \text{modrank}_n(W)$ defined by (9) and (10). The *singular vectors of order k* are the extremal solutions $(\varphi_1^{(k)}, \dots, \varphi_N^{(k)}) \in \mathcal{S}^{(k)}$ that attain the

maximum in (10). A *singular value decomposition* (SVD) of the tensor W is a representation of W with respect to the basis (11), i.e.,

$$W = \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{\ell_1, \dots, \ell_N} \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_N^{(\ell_N)}. \quad (12)$$

The N -way array $[[w_{\ell_1, \dots, \ell_N}]] \in \mathbb{R}^{L_1 \times \cdots \times L_N}$ in (12) is called the *singular value core* of W .

We refer to [6], [7], for more details on this decomposition.

The tensor SVD can be used to find a suitable POD basis. This is done by using low-rank approximations of the tensor. Given the collection (11) of bases of singular vectors of W , we define the subspaces

$$\mathcal{M}_n^{(k)} = \text{span}\{\varphi_n^{(1)}, \dots, \varphi_n^{(k)}\}, \quad n = 1, \dots, N.$$

Definition III.2 For a vector of integers $r = (r_1, \dots, r_N)$ with $r_n \leq R_n$, the *modal truncation* W_r^* is defined by the restriction $W_r^* := W|_{\mathcal{M}_1^{r_1} \times \cdots \times \mathcal{M}_N^{r_N}}$ and is represented by the expansion

$$W_r^* = \sum_{\ell_1=1}^{r_1} \cdots \sum_{\ell_N=1}^{r_N} w_{\ell_1, \dots, \ell_N} \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_N^{(\ell_N)} \quad (13)$$

where $[[w_{\ell_1, \dots, \ell_N}]]$ is the singular value core tensor of W .

An important result on the approximation properties of this decomposition is the following theorem, the proof can be found in the appendix.

Theorem III.3 The tensor $W_1^* := \sigma_1 \varphi_1^{(1)} \otimes \cdots \otimes \varphi_N^{(1)}$ is the *optimal rank-1 approximation of W in the sense that $\|W - W_1^*\|_F$ is minimal among all rank 1 approximations of W .*

Now, a POD basis can be defined.

Definition III.4 The vector of integers $r = (r_1, \dots, r_N)$ with $r_n \leq R_n$ is said to achieve a relative approximation error $\epsilon > 0$ if

$$\frac{\|W - W_r^*\|_F}{\|W\|_F} \leq \epsilon \quad (14)$$

In that case, we say that the basis functions $\{\varphi_n^{(1)}, \dots, \varphi_n^{(r_n)}\}$ for $n = 1, \dots, N$ constitute a POD basis for the model (5) derived from the tensor W .

C. Model Reduction

Now that a POD basis is defined, the model reduction framework can be completed by defining a corresponding spectral expansion and Galerkin projection. We assume that $\mathbb{X} \subset \mathbb{R}^N$ is a Cartesian product $\mathbb{X} = \mathbb{X}' \times \mathbb{X}''$ in which

$$\begin{aligned} \mathbb{X}' &= \mathbb{X}_1 \times \cdots \times \mathbb{X}_{N-1} \\ \mathbb{X}'' &= \mathbb{X}_N. \end{aligned}$$

The reason for this choice is that in practical situations, $\mathbb{X}_1, \dots, \mathbb{X}_{N-1}$ are associated with the spatial variables and

$\mathbb{X}_N$ with time. Using this structure, a spectral expansion is defined as follows

$$w(x_1, \dots, x_N) = \sum_{\ell_1} \cdots \sum_{\ell_{N-1}} a_{\ell_1, \dots, \ell_{N-1}}(x_N) \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_{N-1}^{(\ell_{N-1})}. \quad (15)$$

Reduction of the signal space is now defined by the truncation

$$w_r(x_1, \dots, x_N) = \sum_{\ell_1=1}^{r_1} \cdots \sum_{\ell_{N-1}=1}^{r_{N-1}} a_{\ell_1, \dots, \ell_{N-1}}(x_N) \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_{N-1}^{(\ell_{N-1})} \quad (16)$$

for a vector of integers $r = (r_1, \dots, r_{N-1})$ with $r_n \leq R_n$. Using a Galerkin projection, the reduced order model is defined by the collection of solutions w_r that satisfy

$$\begin{aligned} \langle \varphi_n^{(\ell_n)}, D(\varsigma_1, \dots, \varsigma_{N-1}) w_r \rangle_n &= 0 \\ \text{for } n = 1, \dots, N-1, \ell_n &= 1, \dots, r_n \end{aligned} \quad (17)$$

IV. IMPROVED ACCURACY

In this section some changes will be made to the Singular Value Decomposition for tensors described in Section III-B. The tensor SVD constructs a complete decomposition of the tensor, whereas we are only interested in obtaining orthonormal bases for the vector space defined on $\mathbb{X}'$. In this section it will be shown that it may be advantageous to use this prior knowledge for the construction of a decomposition that is more dedicated to the problem. To achieve this the SVD described in Section III-B will be adapted in such a way that only the vector spaces of interest will be orthonormalized. This ensures that all unnecessary constraints are removed from the optimization. The additional freedom that is created in this way will ensure that more information is captured in the POD basis functions. This will then lead to more accurate reduced models.

The SVD for tensors is constructed inductively. In each step a singular value and a collection of singular vectors are computed (10). To every consecutive step additional constraints are added to ensure that the singular vectors are orthogonal to the singular vectors computed in previous steps. When the tensor SVD is used for model reduction purposes, the only interest is in obtaining orthonormal bases for the vector spaces defined on $\mathbb{X}'$. Yet, the SVD orthonormalizes the vector spaces defined on $\mathbb{X}''$ as well. Since we do not reduce the bases found for vector spaces defined on $\mathbb{X}''$ it makes sense to remove the constraints that enforce the orthonormalization of these vector spaces. The additional freedom that is thus created will give the optimization more flexibility in determining the POD basis.

The new construction for the decomposing a tensor is as follows. Let $W \in \mathcal{T}_N$ be an order- N tensor defined on the finite dimensional vector spaces $X_1, \dots, X_N$ where we suppose that $\dim(X_n) = L_n$. Furthermore, let $X' := X_1 \times \cdots \times X_i$ and $X'' := X_{i+1} \times \cdots \times X_N$, where $0 < i < N$. The *dedicated singular values* of W , denoted $\hat{\sigma}_k(W)$, with

$k = 1, \dots, K$ and $K = \min_{n=1, \dots, i} \text{modrank}(W)$ are defined as follows.

Let

$$\begin{aligned}\mathcal{S}_n^{(1)} &:= \{x \in X_n \mid \|x\|_n = 1\} \quad \text{for } n = 1, \dots, i \\ \mathcal{S}_n &:= \{x \in X_n \mid \|x\|_n = 1\} \quad \text{for } n = i+1, \dots, N\end{aligned}$$

denote the unit sphere in X_n . Define the first dedicated singular value of W by

$$\hat{\sigma}_1(W) := \sup_{\substack{x_n \in \mathcal{S}_n^{(1)}, 1 \leq n \leq i \\ x_n \in \mathcal{S}_n, (i+1) \leq n \leq N}} |W(x_1, \dots, x_N)|. \quad (18)$$

Since W is continuous and the Cartesian product $\mathcal{S}^{(1)} = \mathcal{S}_1^{(1)} \times \dots \times \mathcal{S}_i^{(1)} \times \mathcal{S}_{i+1} \times \dots \times \mathcal{S}_N$ of unit spheres is a compact set, an extremal solution of (18) exists (i.e., the supremum in (9) is a maximum) and is attained by an N -tuple

$$(\psi_1^{(1)}, \dots, \psi_N^{(1)}) \in \mathcal{S}^{(1)}.$$

Subsequent dedicated singular values of W are defined in an inductive manner by setting

$$\begin{aligned}\mathcal{S}_n^{(k)} &:= \{x \in X_n \mid \|x\|_n = 1, \\ &\quad \langle x, \psi_n^{(j)} \rangle_n = 0 \text{ for } j = 1, \dots, (k-1)\}\end{aligned}$$

for $n = 1, \dots, i$, and by defining

$$\hat{\sigma}_k(W) = \sup_{\substack{x_n \in \mathcal{S}_n^{(k)}, 1 \leq n \leq i \\ x_n \in \mathcal{S}_n, (i+1) \leq n \leq N}} |W(x_1, \dots, x_N)|, \quad k \leq K. \quad (19)$$

Again, since the Cartesian product

$$\mathcal{S}^{(k)} = \mathcal{S}_1^{(k)} \times \dots \times \mathcal{S}_i^{(k)} \times \mathcal{S}_{i+1} \times \dots \times \mathcal{S}_N$$

is compact, the supremum in (19) is a maximum that is attained by an N -tuple

$$(\psi_1^{(k)}, \dots, \psi_N^{(k)}) \in \mathcal{S}^{(k)}.$$

It follows that the vectors $\psi_n^{(1)}, \dots, \psi_n^{(K)}$ are mutually orthonormal in X_n , for $n = 1, \dots, i$. If $K < L_n$ for any $1 \leq n \leq i$, then we extend the collection of orthogonal elements $\psi_n^{(1)}, \dots, \psi_n^{(K)}$ to a complete orthonormal basis of X_n . This construction leads to a collection of orthonormal bases

$$\{\psi_1^{(\ell_1)}, \ell_1 = 1, \dots, L_1\}, \dots, \{\psi_i^{(\ell_i)}, \ell_i = 1, \dots, L_i\} \quad (20)$$

for the vector spaces $X_1, \dots, X_i$, respectively. We will call elements of these orthonormal bases *dedicated singular vectors* of the tensor W .

Since there is no construction of orthonormal bases for the vector spaces $X_{i+1}, \dots, X_N$, it is not possible nor appropriate to define a singular-value-like decomposition of the tensor using dedicated singular vectors. Instead, we define a *dedicated representation* of the tensor, which can be used to define a *dedicated modal truncation*.

Definition IV.1 Given an order- N tensor $W \in \mathcal{T}_N$, with $W : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$. Assume $X' = X_1 \times \dots \times X_i$ and

$X'' = X_{i+1} \times \dots \times X_N$. Then, a *dedicated representation* of W can be defined as a representation of W with respect to the bases (20) for X' , where the original bases for X'' are kept intact, i.e.

$$W^d = \sum_{\ell_1=1}^{R_1} \dots \sum_{\ell_N=1}^{R_N} w_{\ell_1 \dots \ell_N} \psi_1^{(\ell_1)} \otimes \dots \otimes \psi_i^{(\ell_i)} \otimes e_{i+1}^{(\ell_{i+1})} \otimes \dots \otimes e_N^{(\ell_N)}. \quad (21)$$

Using this representation of W , a dedicated modal truncation can be defined which will then lead to a definition for the dedicated POD basis.

Definition IV.2 Given an order- N tensor $W \in \mathcal{T}_N$, with dedicated representation W^d and a vector of integers $r = (r_1, \dots, r_i)$ with $r_n \leq R_n$ for $n = 1, \dots, i$. Let

$$\mathcal{M}_n^{(k)} = \text{span}\{\psi_n^{(1)}, \dots, \psi_n^{(k)}\}, \quad n = 1, \dots, i.$$

with $k \leq R_n$. A *dedicated modal truncation* is then defined by the restriction $W_r^d := W^d|_{\mathcal{M}_1^{(r_1)} \times \dots \times \mathcal{M}_i^{(r_i)}}$ and is represented by the expansion

$$W_r^d = \sum_{\ell_1=1}^{r_1} \dots \sum_{\ell_i=1}^{r_i} \sum_{\ell_{i+1}=1}^{R_{i+1}} \dots \sum_{\ell_N=1}^{R_N} w_{\ell_1 \dots \ell_N} \psi_1^{(\ell_1)} \otimes \dots \otimes \psi_i^{(\ell_i)} \otimes e_{i+1}^{(\ell_{i+1})} \otimes \dots \otimes e_N^{(\ell_N)}. \quad (22)$$

Theorem IV.3 Consider $W \in \mathcal{T}_N$.

1) Both the original and dedicated singular values of a tensor are ordered

$$\sigma_1 \geq \dots \geq \sigma_K \geq 0 \quad (23)$$

$$\hat{\sigma}_1 \geq \dots \geq \hat{\sigma}_K \geq 0 \quad (24)$$

2) $\sigma_1 = \hat{\sigma}_1$ and

$$\sigma \cdot \varphi_1^{(1)} \otimes \dots \otimes \varphi_N^{(1)} = \hat{\sigma}_1 \cdot \psi_1^{(1)} \otimes \dots \otimes \psi_N^{(1)}$$

3) $\hat{\sigma}_2 \geq \sigma_2$

Theorem IV.4 Given $W \in \mathcal{T}_N$. Define for $r = (k, R_2, \dots, R_N)$ the approximation error $E_k = W - W_r^*$. Then we have $\|E_{k+1}\|_F \leq \|E_k\|_F$.

V. EXAMPLES

In this section two examples will be given to illustrate the theory discussed above. We first present a data compression problem in 3-D imaging. The example shows the advantages of the dedicated tensor SVD over the generic tensor SVD. The second example considers model reduction of a 2-D heat transfer problem.

Fig. 1. 10th slice of the original data.

A. Application to 3D MRI compression

To illustrate the advantages of the dedicated tensor SVD over the generic SVD, we consider a data compression problem in 3-D imaging*. The data consists of pixel intensities of an MRI scan in which each of the L_3 slices is an image of $L_1 \times L_2$ pixels. For the original scan the dimensions are $L_1 = 262$, $L_2 = 262$ and $L_3 = 29$. The original data has modal rank $(243, 199, 29)$ and we are interested in obtaining rank approximations of the form $(r_1, r_2, 29)$, i.e. leaving the third dimension of the tensor intact. In Fig. 1 the 10th slice of the data is shown.

For this data both generic and dedicated singular values were computed, these can be found in Table I. From this table it is clear that the generic singular values decay much faster than the dedicated singular values. This would imply that the generic singular vectors give better results in approximation. However, examination of Table II, which lists the approximation errors, shows that exactly the opposite is the case. Using the dedicated singular vectors for approximation gives approximation errors that are much smaller than those obtained when using the generic singular vectors. Fig. 2 shows this clearly. On the left, a slice of the rank-(10, 10, 29) approximant is shown that was computed using a Tensor SVD, on the right the equivalent slice is shown of the rank -(10, 10, 29) approximant that was computed using the dedicated construction.

TABLE I
GENERIC AND DEDICATED SINGULAR VALUES.

σ_1	102773.20	$\tilde{\sigma}_1$	102773.20
σ_2	5815.49	$\tilde{\sigma}_2$	49916.03
σ_3	3265.52	$\tilde{\sigma}_3$	19275.82
σ_4	1948.73	$\tilde{\sigma}_4$	11779.07
σ_5	1489.41	$\tilde{\sigma}_5$	9920.99

*The data was obtained from TU/e-BME, Biomedical Image Analysis, in collaboration with Prof. Dr. med. Berthold Wein, Aachen, Germany

Fig. 2. 10th slice of rank-(10, 10, 29) approximant, computed using Tensor SVD (left) and dedicated construction (right).

TABLE II
APPROXIMATION ERROR RESULTS, FOR GENERIC TENSOR SVD (LEFT)
AND DEDICATED TENSOR SVD (RIGHT).

r	$\frac{\ W - W_r\ _F}{\ W\ _F}$	$\frac{\ W - W_r^d\ _F}{\ W\ _F}$
(1, 1)	0.5181	0.5181
(3, 3)	0.5126	0.2646
(5, 5)	0.5108	0.2305
(7, 7)	0.428	0.207
(10, 10)	0.4265	0.1872

B. Model reduction of a heat transfer process

Consider the following model of a heat transfer process on a rectangular plate of size $L_x \times L_y$:

$$0 = -\rho c_p \frac{\partial w}{\partial t} + \kappa_x \frac{\partial^2 w}{\partial x^2} + \kappa_y \frac{\partial^2 w}{\partial y^2}. \quad (25)$$

Here, $w(x, y, t)$ denotes temperature on position (x, y) and time $t \in \mathbb{T} := [0, T_f]$ and the rectangular spatial geometry defines the Cartesian product $\mathbb{X} \times \mathbb{Y} := [0, L_x] \times [0, L_y]$. The plate is assumed to be insulated from its environment. Let $\mathcal{H} = \mathcal{L}_2(\mathbb{X} \times \mathbb{Y})$ be the Hilbert space of square integrable functions on $\mathbb{X} \times \mathbb{Y}$ and let $\mathcal{H}_r = \mathcal{X}_{r_1} \times \mathcal{Y}_{r_2}$ with $\mathcal{X}_{r_1} \subseteq \mathcal{X} = \mathcal{L}_2(\mathbb{X})$ and $\mathcal{Y}_{r_2} \subseteq \mathcal{Y} = \mathcal{L}_2(\mathbb{Y})$ be finite dimensional subspaces spanned by r_1 and r_2 orthonormal bases functions $\{\varphi_1^{(\ell_1)}\}$ and $\{\varphi_2^{(\ell_2)}\}$, respectively.

Solutions of the reduced model are then given by $w_r(x, y, t) = \sum_{\ell_1=1}^{r_1} \sum_{\ell_2=1}^{r_2} a_{\ell_1 \ell_2}(t) \varphi_1^{(\ell_1)}(x) \otimes \varphi_2^{(\ell_2)}(y)$ with $a_{\ell_1 \ell_2}(t) = [A(t)]_{\ell_1 \ell_2}$ a solution of the matrix differential equation

$$0 = -\rho c_p \dot{A} + \kappa_x F A + \kappa_y A P. \quad (26)$$

Fig. 3. First and final time slices of the FE solution of (25).

Fig. 4. $\varphi_1^{(1)}$ (left), $\varphi_2^{(1)}$ (middle) and $\varphi_1^{(1)} \otimes \varphi_2^{(1)}$ (right). These basis functions were computed using the Tensor Singular Value Decomposition.

Here, F and P are defined as:

$$F = \begin{bmatrix} \langle \varphi_1^{(1)}, \ddot{\varphi}_1^{(1)} \rangle & \dots & \langle \varphi_1^{(1)}, \ddot{\varphi}_1^{(r_1)} \rangle \\ \vdots & & \vdots \\ \langle \varphi_1^{(r_1)}, \ddot{\varphi}_1^{(1)} \rangle & \dots & \langle \varphi_1^{(r_1)}, \ddot{\varphi}_1^{(r_1)} \rangle \end{bmatrix}; \quad P = \begin{bmatrix} \langle \varphi_2^{(1)}, \ddot{\varphi}_2^{(1)} \rangle & \dots & \langle \varphi_2^{(1)}, \ddot{\varphi}_2^{(r_2)} \rangle \\ \vdots & & \vdots \\ \langle \varphi_2^{(r_2)}, \ddot{\varphi}_2^{(1)} \rangle & \dots & \langle \varphi_2^{(r_2)}, \ddot{\varphi}_2^{(r_2)} \rangle \end{bmatrix}$$

Alternatively, $a_{\ell_1 \ell_2}(t)$ is the solution of the ordinary differential equation

$$\rho c_p \dot{a}_{\ell_1 \ell_2}(t) = \kappa_x \sum_{k_1=1}^{r_1} a_{k_1 \ell_2}(t) \langle \ddot{\varphi}_1^{(k_1)}(x), \varphi_1^{(\ell_1)}(x) \rangle + \kappa_y \sum_{k_2=1}^{r_2} a_{\ell_1 k_2}(t) \langle \ddot{\varphi}_2^{(k_2)}(y), \varphi_2^{(\ell_2)}(y) \rangle \quad (27)$$

for $1 \leq \ell_1 \leq r_1$ and $1 \leq \ell_2 \leq r_2$. A FE solution of (25) has been computed with physical and discretization parameters as given in Table III. Time slices, including the initial condition, of the simulation data can be seen in Fig. 3. The boundary conditions are chosen so as to represent that the plate is insulated from its environment.

TABLE III
PDE PARAMETER VALUES

Parameter	Value	Unit
ρC_p	5	$\frac{J}{m^3 \cdot K}$
κ_x	0.5	$\frac{W}{m \cdot K}$
κ_y	0.5	$\frac{W}{m \cdot K}$
L_x	3	m
L_y	4	m
T_f	3.6	s
Δ_x	0.05	m
Δ_y	0.05	m
Δ_t	0.05	s

In this example the original orders $(L_1, L_2, L_3) = (61, 81, 72)$ are reduced to (r_1, r_2, L_3) where we take $r_1 = r_2$. The orthonormal bases $\{\varphi_1^{(\ell_1)}\}$ and $\{\varphi_2^{(\ell_2)}\}$ have been computed using Tensor SVD and dedicated Tensor SVD construction, where in the latter time was not orthonormalized, since these basis functions will not be used in the reduced model. The first basis functions for $\mathcal{X}$ and $\mathcal{Y}$ computed using Tensor SVD described in Section III-B are shown in Fig. 4.

The simulation time of the FE implementation is 17.22s, the reduced models have a simulation time of approximately 0.35s. Table IV gives the simulation error of the reduced model for different model orders. The reduced models were given the same initial condition as the was used to collect the snapshot data. Simulation errors are given for models that use

basis functions computed using TSVD and basis functions computed using the dedicated construction. As can be seen in Table IV using a dedicated construction to compute basis functions does not give a more accurate reduced model for this example.

TABLE IV
REDUCED MODEL SIMULATION ERROR RESULTS, BASIS FUNCTIONS WERE COMPUTED USING TSVD (LEFT) AND DEDICATED CONSTRUCTION (RIGHT).

r	$\ W - W_r\ _F$	$\ W - W_r^d\ _F$
(2, 2)	0.366	0.366
(3, 3)	0.347	0.336
(5, 5)	0.239	0.205
(7, 7)	0.174	0.162
(10, 10)	0.137	0.079

VI. CONCLUSIONS AND FUTURE WORK

In this paper we discussed approximation of systems that have both space and time as independent variables. The aim was to find reduced models that are computationally efficient and at the same time give an accurate description of the state evolution of an nD system. To this end, we have generalized the data-based method of Proper Orthogonal Decompositions. The key of the adaptation lies in recognizing the nD nature of the class of systems under study. In the first part of the paper it was shown how POD basis functions can be computed by associating a tensor with the measured or simulated data and then determining a lower rank decomposition of the tensor. In the second part of the paper we showed how prior knowledge about the model reduction problem can be used to obtain better approximation results. Although we did not obtain a formal proof that the dedicated construction outperforms the generic tensor SVD the examples show that prior knowledge on the reduction order improves the approximation results. The theoretical results of this paper were illustrated by two examples.

In the future, we plan to extend the framework presented here to systems with multiple *dependent* variables. Furthermore, we would like to extend the tensor SVD to be able to take constraints into account. This may allow the explicit inclusion of physical conservation laws, such as conservation of mass in fluid dynamics. With respect to applications, we plan to test the methods discussed in this paper on an industrial benchmark problem.

APPENDIX I PROOFS

For the first proof we need the following lemma

Lemma I.1 Let $W \in \mathcal{T}_N$ and $x_n \in X_n$ for $n = 1, \dots, N$. Then

$$W(x_1, \dots, x_N) = \langle W, x_1 \otimes \dots \otimes x_N \rangle.$$

Proof: Proof of Lemma .1 Using the standard bases, the tensor evaluation can be written as $W(x_1, \dots, x_N) =$

$\sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} \langle x_1, e_{\ell_1}^{(1)} \rangle \cdots \langle x_N, e_{\ell_N}^{(N)} \rangle$. Let $U := x_1 \otimes \cdots \otimes x_N$. The elements of U then satisfy $u_{\ell_1 \dots \ell_N} = \prod_{n=1}^N \langle e_{\ell_n}^{(n)}, x_n \rangle$. Consequently,

$$\begin{aligned} \langle W, U \rangle &= \sum_{k_1=1}^{L_1} \cdots \sum_{k_N=1}^{L_N} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{k_1 \dots k_N} u_{\ell_1 \dots \ell_N} \\ &\quad \cdot \underbrace{\langle e_{\ell_1}^{(1)}, e_{k_1}^{(1)} \rangle \cdots \langle e_{\ell_N}^{(N)}, e_{k_N}^{(N)} \rangle}_{0 \text{ unless } k_1 = \ell_1} \\ &= \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} u_{\ell_1 \dots \ell_N} \\ &= \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} \langle e_{\ell_1}^{(1)}, x_1 \rangle \cdots \langle e_{\ell_N}^{(N)}, x_N \rangle \end{aligned}$$

which is the tensor evaluation. ■

Proof: Proof of Theorem III.3 Let $W_1 \in \mathcal{T}_N$ be an arbitrary rank-1 tensor. Then W_1 can be written as $W_1 = \lambda U$ where $0 \neq \lambda \in \mathbb{R}$ and $U = u_1 \otimes \cdots \otimes u_N$ is a normalized rank-1 tensor in that $\|U\|_F = 1$. Using the definition of the Frobenius norm, we have

$$\begin{aligned} \|W - \lambda U\|_F^2 &= \langle W - \lambda U, W - \lambda U \rangle \\ &= \langle W, W \rangle - 2\lambda \langle W, U \rangle + \lambda^2. \end{aligned}$$

This is a convex function in λ that attains its minimum at $\lambda^* = \langle W, U \rangle$. But then

$$\begin{aligned} \|W - \lambda^* U\|_F^2 &= \langle W, W \rangle - 2\lambda^* \langle W, U \rangle + (\lambda^*)^2 = \\ &= \langle W, W \rangle - 2\langle W, U \rangle^2 + \langle W, U \rangle^2 \\ &= \langle W, W \rangle - \langle W, U \rangle^2 = \\ &= \langle W, W \rangle - |W(u_1, \dots, u_N)|^2 \end{aligned}$$

where the last equality follows from Lemma .1. The latter expression shows that minimizing $\|W - \lambda^* U\|_F$ over all rank-1 tensors U with $\|U\|_F = 1$ is equivalent to maximizing $|W(u_1, \dots, u_N)|$ over all unit vectors u_n , $\|u_n\|_n = 1$, $n = 1, \dots, N$. But this problem is (9) and has $U^* = \varphi_1^{(1)} \otimes \cdots \otimes \varphi_N^{(1)}$ as its optimal solution. Consequently, $\lambda^* = \langle W, U^* \rangle = \sigma_1$ and it follows that $W_1^* := \sigma_1 \varphi_1^{(1)} \otimes \cdots \otimes \varphi_N^{(1)}$ is the optimal rank-1 approximation of W . The error $\|W - W_1^*\|_F^2 = \|W\|_F^2 - \sigma_1^2$. ■

Proof: Proof of Theorem IV.3

- 1) This is by construction.
- 2) Since these optimization problems are identical, the first singular value and the singular vectors that are found will also be identical.
- 3) This uses the previous part of this theorem. Since the results from the first optimization are identical, the optimization domains $\mathcal{S}_n^{(2)}$, $n = 1, \dots, N$ will be the same for both optimization problems. As the dedicated SVD construction will incorporate less constraints for the second step, it only takes $\mathcal{S}_n^{(2)}$, $n = 1, \dots, i$ into account and uses the unit sphere for the rest of the vector spaces, $\hat{\sigma}_2 \geq \sigma_2$. ■

Proof: Proof of Theorem IV.4 $\hat{W} = W|_{\mathcal{M}_1^{(k)} \times X_2 \times \cdots \times X_N}$ with $\mathcal{M}_1^{(k)} = \text{span}\{\varphi_1^{(1)}, \dots, \varphi_1^{(k)}\}$. $W = W|_{\mathcal{M}_1^{(k)} \times X_2 \times \cdots \times X_N} + W|_{\mathcal{M}_1^{(k)\perp} \times X_2 \times \cdots \times X_N}$. Since $W|_{(\mathcal{M}_1^{(k)} \times X_2 \times \cdots \times X_N) \cap (\mathcal{M}_1^{(k)\perp} \times X_2 \times \cdots \times X_N)} = 0$, $(\mathcal{M}_1^{(k)} \times X_2 \times \cdots \times X_N) \cap (\mathcal{M}_1^{(k)\perp} \times X_2 \times \cdots \times X_N)$ is equal to $(\mathcal{M}_1^{(k)} \cap \mathcal{M}_1^{(k)\perp}) \times (X_2 \cap X_2) \times \cdots \times (X_N \cap X_N)$ and $W|_{\emptyset \times X_2 \times \cdots \times X_N} = 0$, the following expression for E_k can be obtained

$$E_k = W|_{\mathcal{M}_1^{(k)\perp} \times X_2 \times \cdots \times X_N}.$$

Since $\mathcal{M}_1^{(k)} \subseteq \mathcal{M}_1^{(k+1)}$ and, consequently, $\mathcal{M}_1^{(k)\perp} \supseteq \mathcal{M}_1^{(k+1)\perp}$, we conclude $\|E_{k+1}\| \leq \|E_k\|$. ■

REFERENCES

- [1] Glover, K. et al., Realisation and approximation of linear infinite-dimensional systems with error bounds, *SIAM J. Control and Optimization*, vol. 26, No.4, 1988, pp 863-898.
- [2] Fujimoto, K. and Ono, S., *On balanced realization and finite-dimensional approximation for infinite-dimensional nonlinear systems*, Proceedings of the 47th IEEE CDC, Cancun, Mexico, 2008.
- [3] M. Kirby, "Geometric data analysis, an empirical approach to dimensionality reduction and the study of patterns", John Wiley, New York, 2001.
- [4] V. Thomee, "Galerkin finite element methods for parabolic problems", Springer, Berlin, Computational Mathematics, Vol. 25, 1997.
- [5] K. Kunisch and S. Volkwein, "Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics," *SIAM J. Numer. Anal.*, 40:492–515, 2002.
- [6] Belzen, F. van, Weiland, S. and Graaf, J. de. *Singular value decompositions and low rank approximations of multi-linear functionals*, Proc. 46th IEEE Conf. on Decision and Control, 2007.
- [7] Belzen, F. van, Weiland, S. *Diagonalization and Low-Rank Approximation of Tensors: a Singular Value Decomposition Approach*, Proc. 18th Int. MTNS, 2008.
- [8] T.G. Kolda, "Orthogonal tensor decompositions," *SIAM J. Matrix Analysis and Applications*, Vol. 23, No. 1, pp. 243-255, 2001.
- [9] T.G. Kolda, "A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theorem for the orthogonal rank tensor decomposition," *SIAM J. Matrix Analysis and Applications*, Vol. 24, No. 3, pp. 762-767, 2003.
- [10] L. De Lathauwer, B. De Moor, J. Vandewalle, "A multi-linear singular value decomposition," *SIAM J. Matrix Analysis and Applications*, Vol. 21, No. 4, pp. 1253-1278, 2000.
- [11] L. De Lathauwer, B. De Moor, J. Vandewalle, "On the best rank-1 and rank $(r_1, r_2, \dots, r_n)$ approximation of higher order tensors," *SIAM J. Matrix Analysis and Applications*, Vol. 21, No. 4, pp. 1324-1342, 2000.
- [12] Lek-Heng Lim, "What's possible and what's not possible in tensor decompositions – a freshman's view," Workshop on Tensor Decompositions, American Institute of Mathematics, July, 2004.
- [13] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem", *SIAM J. Matrix Analysis and Applications: Special Issue on Tensor Decompositions and Applications*, 30(3), 2008, pp.1084-1127.