

# A tensor decomposition approach to data compression and approximation of $ND$ systems

F. van Belzen · S. Weiland

Received: 25 November 2009 / Revised: 9 November 2010 / Accepted: 3 December 2010 /  
Published online: 25 December 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** 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  $ND$  nature of a distributed model into account. This results in a novel procedure for model reduction of systems with multiple independent variables. Data in multiple independent variables is associated with the mathematical structure of a tensor. We show how orthonormal decompositions of this tensor can be used to derive suitable projection spaces. These projection spaces prove useful for determining reduced order models by performing Galerkin projections on equation residuals. 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 an application in data compression. The proposed model reduction procedure is illustrated on a heat diffusion problem.

**Keywords** Model reduction ·  $ND$  systems · Proper orthogonal decompositions · Multi-linear algebra · Tensors

## 1 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). Since analytical solutions to systems of PDEs can generally not be inferred, one typically resorts to numerical techniques to obtain insight in the evolutions and solutions of these systems. Traditionally, these numerical techniques rely on finite element (FE) or finite volume methods in which both space and time are

---

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

---

F. van Belzen (✉) · S. Weiland  
Control Systems, Department of Electrical Engineering, Eindhoven University of Technology,  
P.O. Box 513, 5600 MB Eindhoven, The Netherlands  
e-mail: f.v.belzen@tue.nl

gridded. A substantial body of research is dedicated to gridding techniques so as to guarantee the accuracy of approximate solutions. They all have in common that, with increasing requirements on accuracy, the number of equations that need to be resolved increases. Depending on the dimensionality and level of accuracy, the numerical models become large-scale discrete systems that require a substantial computational effort to solve. In particular, this makes FE models less suitable for purposes of on-line monitoring, real-time prediction, optimization, or model-based control.

Model reduction methods aim to reduce the complexity of large-scale dynamical systems. The goal of model reduction is to obtain models that are computationally efficient and that, at the same time, offer an accurate description of the system under consideration. Previous work on model reduction for infinite dimensional systems, [Fujimoto and Ono \(2008\)](#), [Glover \(1988\)](#), considers mainly the evolution of system variables in one independent variable, namely time. These papers focus on the accurate representation of the input/output behavior in the reduced order model. Our interest lies in obtaining reduced order models that offer an accurate description of the state evolution of an  $ND$  system. In this paper, we explicitly aim to take the structure of multiple independent variables into account.

The method of Proper Orthogonal Decompositions (POD) [Kirby \(2001\)](#), [Thomee \(1997\)](#), [Kunisch and Volkwein \(2002\)](#) is among the few model reduction techniques that can be readily applied to systems with multiple independent variables. POD is also known under the names of Principal Component Analysis (PCA) [Jolliffe \(1986\)](#) and Karhunen-Loève-Decomposition [Kirby \(2001\)](#). 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  $ND$  system by separating time and space. That is, the independent variables are assumed to reside in a Cartesian product of a temporal and spatial domain.

It is the aim of this paper to generalize the POD method to explicitly take the multi-dimensionality of independent variables into account. This is accomplished by assuming a more general Cartesian structure for the independent variables. In our point of view tensors and tensor decompositions provide a natural and suitable mathematical formalism for the representation of signals that evolve in multiple independent variables. When applied to  $ND$  systems, tensor decompositions can be used to define empirical projection spaces that resemble the projection spaces used in POD model reduction techniques. Moreover, when applied to  $ND$  systems, they offer distinct advantages over classical POD techniques in the sense that they allow more flexibility in approximating spatial evolutions.

Throughout this paper tensors are defined as multi-linear functionals. We believe that this level of abstraction has advantages over the alternative definition where tensors are viewed as multi-dimensional arrays. Firstly, this displays the algebraic nature of tensors in a more direct manner. Secondly, it allows to clearly distinguish tensors as an algebraic object from its representation. Thirdly, basis changes are introduced in a most natural algebraic setting. The *order* of the tensor is the number of arguments of the functional. *Tensor decompositions* refer to expanding a multi-linear functional in more elementary building blocks, such as low-rank tensors. Our aim will be to show that orthogonal rank-one decompositions of tensors prove useful in questions of low-rank tensor approximations. Unlike the matrix case (order-2 tensors), where optimal low-rank approximations are readily found via the Singular Value Decomposition (SVD), finding optimal low-rank approximations for tensors is not straightforward at all. See [de Silva and Lim \(2008\)](#) for a thorough discussion of the issues that may arise.

Traditionally, two types of tensor decompositions are found in the literature, the *Tucker* decomposition [Tucker \(1966\)](#) and the *PARAFAC* or *CANDECOMP* decomposition [Carroll and Chang \(1970\)](#). In the Tucker decomposition an order- $N$  tensor is represented as a product of a core tensor of the same size as the original tensor with  $N$  nonsingular matrices. The columns of these matrices span the domain of each of the arguments of the original tensor. The well-known Higher-Order Singular Value Decomposition (HOSVD) [De Lathauwer et al. \(2000a\)](#) is a special case of the Tucker decomposition. Recently, in [Weiland and van Belzen \(2010\)](#) the authors of this paper proposed a method for tensor decompositions that generalizes the classical singular value decomposition of matrices. It proves useful for finding low-rank approximations of tensors. The method presented is of the Tucker form, with additional orthogonality constraints. The second class of tensor decompositions represents a tensor as a linear combination of normalized rank-one tensors. This decomposition is found under the names of Parallel Factor (PARAFAC) Decomposition or Canonical Decomposition (CANDECOMP). More recently, in [Oseledets and Tyrtshnikov \(2009\)](#), [Oseledets and Tyrtshnikov \(2010\)](#), an alternative to these two classes of tensor decompositions was proposed. In the so-called *Tree-Tucker* (TT) representation, a tensor is decomposed as a tree of order-3 tensors.

In addition to the work on tensor decompositions, a lot of effort is geared towards using tensors to reduce computation time of multidimensional functions on discretized grids. In [Beylkin and Mohlenkamp \(2005\)](#) the authors introduce the concepts of separated rank and separated representation to accelerate computations of multidimensional functionals on discretized grids. It is their aim to arrive at function approximations, rather than tensor decompositions. Therefore, in the construction of the separated representation, the authors do not require minimality of rank, nor orthonormality of the decomposition in whatever sense, nor optimality of the approximation. Since it is the aim of this paper to construct empirical projection spaces spanned by orthonormal bases, the work of [Beylkin and Mohlenkamp \(2005\)](#) is not considered further in this paper. In [Khoromskij and Khoromskaia \(2009\)](#) the authors attempt to combine the strengths of the Tucker and CANDECOMP decomposition to decrease the time involved in computations with function related multidimensional arrays. Numerical algorithms for tensor decompositions and approximations of Tucker type are presented, with the additional constraint that the core array is to be presented in a low-rank canonical format. As an application of tensors to accelerate computation of multidimensional functions on discretized grids, [Hackbusch \(2008\)](#) uses tensors to solve elliptic eigenvalue problems.

This paper uses tensor decompositions to derive empirical projection spaces from measured or simulated data. These spaces can be used to define reduced order models through Galerkin projections. Since these projection spaces are computed once, and usually off-line, the computational efficiency of computing these projection spaces is not a severe restriction. Therefore, the focus in this paper will be on deriving projection spaces using tensor decompositions of the Tucker form. It will be shown that these decompositions can be robustly computed and offer the required orthonormality.

The contribution of this paper is twofold. First, we review the method presented in [Weiland and van Belzen \(2010\)](#) and show how more flexibility in reduction orders can be exploited to improve the quality of reduced-order models. An algorithm to compute the decompositions is introduced and a numerical example is presented. Second, the techniques of tensor decompositions and Proper Orthogonal Decompositions are combined to obtain a model reduction framework for  $ND$  systems. We show that this model reduction technique has decisive advantages in that the  $ND$  structure of the original problem is left invariant. As a result of this, it allows to reduce spatially dependent variables independently in each spatial direction.

This paper is organized as follows. In Sect. 2, a compact introduction to the method of Proper Orthonormal Decompositions is provided. Then, in Sect. 3 tensors and tensor decompositions are introduced and discussed. Section 4 proposes some changes to the POD method that may be useful when considering  $ND$  systems. The numerical computation of a tensor decomposition is discussed in Sect. 5. We end with our conclusions. Proofs to the theorems can be found in the appendix.

## 2 Introduction to proper orthogonal decompositions

This section offers a brief introduction to the method of Proper Orthogonal Decompositions. More information on this topic can be found in Kirby (2001), Thomee (1997), Kunisch and Volkwein (2002).

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

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

Here,  $R \in \mathbb{R}^{\times 1}[\xi_1, \dots, \xi_N]$  is a real matrix valued polynomial in  $N$  indeterminates. That is, with  $\xi$  defined as the multi-indexed indeterminate,  $\xi = (\xi_1, \dots, \xi_N)$  and  $\ell$  the multi-indexed object  $\ell = (\ell_1, \dots, \ell_N)$ ,  $R$  is the degree- $L$  polynomial

$$R(\xi_1, \dots, \xi_N) := R(\xi) = \sum_{0 \leq |\ell| \leq L} R_\ell \xi^\ell = \sum_{0 \leq |\ell| \leq L} R_{\ell_1 \dots \ell_N} \xi_1^{\ell_1} \dots \xi_N^{\ell_N}$$

where  $|\ell| = \sum \ell_n$  and  $R_\ell$  are real-valued coefficients. With  $\xi_n$  replaced by the partial derivative  $\xi_n = \frac{\partial}{\partial x_n}$ ,  $R$  defines a polynomial differential operator as in (1). That is, (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 functions on  $\mathbb{X}'$  is introduced. Solutions  $w$  of (1) are assumed to be sufficiently often differentiable functions that satisfy  $w(\cdot, x'') \in \mathcal{H}$  for all  $x'' \in \mathbb{X}''$ . In addition,  $\mathcal{H}$  is assumed to be a separable Hilbert space, which means that a countable orthonormal basis  $\{\varphi_n, n = 1, 2, \dots\}$  for  $\mathcal{H}$  exists and every solution  $w$  to (1) admits a spectral expansion

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

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  $r > 0$  a lower rank approximation to  $w$  is defined by the truncation

$$w_r(x', x'') = \sum_{n=1}^r a_n(x'') \varphi_n(x'). \tag{2}$$

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

$$\left\langle R \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_N} \right) w_r, \varphi \right\rangle = 0, \quad \forall \varphi \in \mathcal{H}_r \tag{3}$$

where  $\mathcal{H}_r$  is the finite dimensional projection space  $\mathcal{H}_r = \text{span}\{\varphi_1, \dots, \varphi_r\}$ . The interest of (3) lies in the crucial observation that, whenever the spectral expansion of  $w_r$  is substituted in (3) and  $\mathbb{X}'' \subseteq \mathbb{R}$ , then (3) becomes a system of  $r$  ordinary differential equations in the modal coefficients  $a_n, n = 1, \dots, r$ . This reduces the PDE (1) to an approximate model of  $r$  ordinary differential equations.

Clearly, the quality of the reduced order model (3) 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  $w : \mathbb{X} \rightarrow \mathbb{R}$  simulated from (1). This set of measured or simulated data is assumed to contain a collection of trajectories that is representative of the system dynamics of interest. Specifically, for given data  $w : \mathbb{X} \rightarrow \mathbb{R}$  with  $w(\cdot, x'') \in \mathcal{H}$  and  $x'' \in \mathbb{X}''$ , the basis functions  $\varphi_n$  are chosen so as to minimize the criterion function

$$J(\varphi_1, \dots, \varphi_r) := \int_{\mathbb{X}''} \|w(\cdot, x'') - \sum_{n=1}^r \langle w(\cdot, x''), \varphi_n \rangle \varphi_n\|^2 dx'' \tag{4}$$

subject to the constraint that

$$\langle \varphi_n, \varphi_m \rangle = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases} \quad n, m = 1, \dots, r.$$

Here, the inner product is the inner product of the Hilbert space  $\mathcal{H}$  and the optimization is carried out for an arbitrary approximation degree  $r$ . A solution  $(\varphi_1, \dots, \varphi_r)$  to this constrained optimization problem is called a *POD basis of order  $r$* . Hence, a POD basis of order  $r$  minimizes the integrated error  $\int_{\mathbb{X}''} \|w - w_r\|^2 dx''$  over all rank  $r$  approximations  $w_r$  of  $w$  of the form (2). A complete orthonormal basis  $\{\varphi_n, n \in \mathbb{I}\}$  of  $\mathcal{H}$  is said to be a *POD basis* if for all  $r$  the collection  $(\varphi_1, \dots, \varphi_r)$  is a POD basis of order  $r$ . Here  $\mathbb{I}$  is a countable set of indices with cardinality equal to the dimension of  $\mathcal{H}$ .

The constrained optimization problem has an elegant solution in terms of the data correlation operator  $\Phi : \mathcal{H} \rightarrow \mathcal{H}$  that is implicitly defined as

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

Note that,  $\Phi$  is a well defined linear, bounded, self-adjoint and non-negative operator on  $\mathcal{H}$ .

**Theorem 1** *Suppose that  $\{\varphi_n, n \in \mathbb{I}\}$  is an orthonormal basis of  $\mathcal{H}$  and suppose that the eigenvalues of  $\Phi$  are absolute summable. Then  $\{\varphi_n, n \in \mathbb{I}\}$  is a POD basis if and only if  $\Phi \varphi_n = \lambda_n \varphi_n, n \in \mathbb{I}$  where the eigenvalues  $\lambda_n$  are ordered according to  $\lambda_1 \geq \lambda_2 \geq \dots$ . Moreover, in that case the error*

$$J(\varphi_1, \dots, \varphi_r) = \sum_{n>r} \lambda_n$$

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

Hence, the eigenfunctions of the data correlation operator determine the POD basis. We refer to the Appendix for a proof.

In applications, the PDE (1) can be discretized by sampling the domain  $\mathbb{X} = \mathbb{X}' \times \mathbb{X}''$  and by discretizing the PDE. This typically leads to difference equations of the form

$$D(\zeta_1, \dots, \zeta_N)w = 0 \tag{5}$$

where  $D \in \mathbb{R}^{\times 1}[\xi_1, \dots, \xi_N]$  is a real matrix-valued polynomial in  $N$  indeterminates and  $\zeta_n$  is the forward shift operator acting on the spatial discretization in the  $n$ th mode according to  $\zeta_n w(x_1^{(\ell_1)}, \dots, x_n^{(\ell_n)}, \dots, x_N^{(\ell_N)}) = w(x_1^{(\ell_1)}, \dots, x_n^{(\ell_{n+1})}, \dots, x_N^{(\ell_N)})$  where  $w : \mathbb{X} \rightarrow \mathbb{R}$  with  $\mathbb{X} = \prod_{n=1}^N \mathbb{X}_n$ . An example of a simple finite element discretization rule defines  $D$  from the PDE (1) according to

$$D(\zeta_1, \dots, \zeta_N) = R \left( \frac{\zeta_1 - 1}{h_1}, \dots, \frac{\zeta_N - 1}{h_N} \right)$$

where  $h_n, n = 1, \dots, N$  is the (uniform) sampling interval of  $\mathbb{X}_n$ .

We will assume that (5) is an accurate representation of (1) and refer to solutions of (5) as Finite Element solutions. Such solutions again assume the form  $w : \mathbb{X}' \times \mathbb{X}'' \rightarrow \mathbb{R}$  where both  $\mathbb{X}'$  and  $\mathbb{X}''$  are sets of finite cardinality. For the discretized model (5) the solution space  $\mathcal{H}$  then becomes finite, but large, dimensional. Also the data correlation operator  $\Phi$  becomes a symmetric non-negative definite matrix and the calculation of POD basis functions becomes a straightforward algebraic eigenvalue or singular value decomposition problem. The following example illustrates the method described in this section.

*Example 1* Consider the following Partial Differential Equation, describing the evolution of temperature over some two-dimensional spatial domain and time

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

Here,  $w(x, y, t)$  denotes the temperature at position  $(x, y) \in \mathbb{X}' := [0, L_X] \times [0, L_Y]$  and time  $t \in \mathbb{X}'' := [0, L_T]$ . Let  $\mathcal{H} = \mathcal{L}_2(\mathbb{X}', \mathbb{R})$ . Given a set of POD basis functions  $\{\varphi_n(x, y)\}_{n=1}^\infty$ ,  $w$  can be represented as

$$w(x, y, t) = \sum_n a_n(t) \varphi_n(x, y) = \sum_n a_n(x'') \varphi_n(x').$$

Truncation of this sum to the level  $r$  gives an approximation of  $w$  which will be denoted by  $w_r$ . A Galerkin projection of the form (3) is given by

$$\left\langle -\rho C_p \frac{\partial w_r}{\partial t} + \kappa_x \frac{\partial^2 w_r}{\partial x^2} + \kappa_y \frac{\partial^2 w_r}{\partial y^2}, \varphi_k \right\rangle = 0, \quad \text{for } k = 1, \dots, r.$$

Using the definition of  $w_r$ , this expression can be reworked into the following ODE

$$\rho C_p \dot{a}_k(t) = \kappa_x \sum_{n=1}^r a_n(t) \left\langle \frac{\partial^2}{\partial x^2} \varphi_n, \varphi_k \right\rangle + \kappa_y \sum_{n=1}^r a_n(t) \left\langle \frac{\partial^2}{\partial y^2} \varphi_n, \varphi_k \right\rangle \quad \text{for } k = 1, \dots, r$$

which defines a reduced model of order  $r$ .

### 3 Exploiting structure via tensor decompositions

In most model reduction applications,  $\mathbb{X}'$  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 lumped and this yields a large-dimensional data correlation operator  $\Phi$ . In this way, the ND nature of the original problem is replaced by an artificial 2-D structure with space and time as independent variables.

In this paper we will adapt the POD method in such a way that the  $ND$  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, i.e.  $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_{N-1}$ . Whenever this assumption holds, the measured/simulated data can be stored in a multi-dimensional array  $[[w]]$ . This array defines a tensor  $W$  and the projection spaces used in model reduction can be computed via a decomposition of this tensor.

In this section we will first give some definitions regarding tensors and their properties. Then, we will discuss tensor decompositions and introduce the tensor SVD method that will be used throughout this paper. We show how prior knowledge of the structure of the model reduction problem can be used to improve the accuracy of the approximation of the tensor. Finally, an approximation example with three-dimensional MRI data is presented.

### 3.1 Tensors and tensor decompositions

Assume that the domain  $\mathbb{X}$  of (1) has the Cartesian structure  $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_N$  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  $[[w]] \in \mathbb{R}^{L_1 \times \cdots \times L_N}$  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*  $W$  is a multi-linear functional  $W : X_1 \times \cdots \times X_N \rightarrow \mathbb{R}$  operating on  $N$  vector spaces  $X_1, \dots, X_N$ . The elements of  $W$ ,  $w_{\ell_1 \dots \ell_N}$ , are defined with respect to bases for  $X_1, \dots, X_N$  according to  $w_{\ell_1 \dots \ell_N} = W(e_1^{\ell_1}, \dots, e_N^{\ell_N})$ , where  $\{e_n^{\ell_n}, \ell_n = 1, \dots, L_n\}$  is a basis for  $X_n$ ,  $n = 1, \dots, N$ . For example,  $W(x_1, x_2) := \langle x_1, Ax_2 \rangle$  defines an order-2 tensor whose element  $w_{\ell_1 \ell_2}$  is the  $(\ell_1, \ell_2)$ th entry of the matrix  $A$ .

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

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

where  $e_1 \otimes \cdots \otimes e_N$  is shorthand for the *rank-1 tensor*  $U : X_1 \times \cdots \times X_N \rightarrow \mathbb{R}$ , defined by  $U(x_1, \dots, x_N) := \prod_{n=1}^N \langle e_n, x_n \rangle$  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  $V, W \in \mathcal{T}_N$  with elements  $v_{k_1, \dots, k_N}$  and  $w_{\ell_1, \dots, \ell_N}$ , both defined with respect to the same bases for  $X_1, \dots, X_N$ , is given by

$$\langle V, W \rangle := \sum_{k_1=1}^{L_1} \cdots \sum_{k_N=1}^{L_N} \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} v_{k_1 \dots k_N} w_{\ell_1 \dots \ell_N} \langle e_1^{k_1}, e_1^{\ell_1} \rangle \cdots \langle e_N^{k_N}, e_N^{\ell_N} \rangle. \tag{7}$$

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  $W \in \mathcal{T}_N$  is defined as

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

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, [de Silva and Lim \(2008\)](#), [Kolda \(2001, 2003\)](#), [De Lathauwer et al. \(2000a\)](#), [De Lathauwer et al. \(2000b\)](#). To define the *modal rank*, also referred to as the *multi-linear rank*, of a tensor  $W \in \mathcal{T}_N$ , we first introduce the *n-mode kernel* of  $W$  as the set

$$\ker_n(W) := \{x_n \in X_n \mid W(x_1, \dots, x_N) = 0, \text{ for all } x_k \in X_k, k \neq n\}. \tag{8}$$

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  $w_{\ell_1 \dots \ell_N}$  till  $w_{\ell_1 \dots \ell_N}$  (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 can be used to define 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

$$W_r := \sum_{\ell_1=1}^{r_1} \dots \sum_{\ell_N=1}^{r_N} \hat{w}_{\ell_1 \dots \ell_N} \varphi_1^{\ell_1} \otimes \dots \otimes \varphi_N^{\ell_N} \tag{9}$$

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, [Kolda \(2001\)](#), [van Belzen et al. \(2007\)](#).

For order-2 tensors (matrices) this problem is solved by the singular value decomposition. For higher-order tensors, it is not straightforward to construct proper sets of orthonormal bases with this property. As mentioned in the introduction, there is a vast literature on tensor decompositions. The most appropriate in the current setting are decompositions of the Tucker type. Consider a tensor  $W \in \mathcal{T}_N$ , a Tucker decomposition is of the form

$$W = \sum_{\ell_1=1}^{L_1} \dots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} \varphi_1^{(\ell_1)} \otimes \dots \otimes \varphi_N^{(\ell_N)} \tag{10}$$

where  $\{\varphi_n^{(\ell_n)}\}_{\ell_n=1}^{L_n}$  forms a basis for  $\mathbb{R}^{L_n}$ . The Higher-Order Singular Value Decomposition (HOSVD) is the best known example of a Tucker decomposition with some additional orthonormality constraints on the basis functions. Since the HOSVD is basically a matrix-method in which the multi-linear 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 [van Belzen et al. \(2007\)](#), [Weiland and van Belzen \(2010\)](#). This is also a Tucker decomposition with orthonormality constraints on the basis vectors, but unlike the HOSVD it leaves the multi-linearity of the original tensor intact in the computation of the decomposition.

Unlike the matrix case, truncation of the Tucker decomposition of a tensor of order larger than two does not give the best low multi-linear rank approximation of a tensor. Instead, the truncated Tucker decomposition can be used as an initial guess in any of the numerical algorithms that compute best low multi-linear rank approximations, as is discussed in [Ishteva et al. \(2010\)](#). Examples of algorithms that can be used to compute optimal lower modal rank

approximations of tensors can be found in for example [De Lathauwer et al. \(2000b\)](#), [Eldén and Savas \(2009\)](#). In this paper, optimal lower modal rank approximations are not considered as we are interested in tensor *decompositions*. Details of the tensor decomposition are given in the next section.

### 3.2 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

$$\sigma_1(W) := \sup_{\substack{x_n \in \mathcal{S}_n^{(1)}, \\ 1 \leq n \leq N}} |W(x_1, \dots, x_N)|. \tag{11}$$

Since  $W$  is continuous and the Cartesian product  $\mathcal{S}^{(1)} = \mathcal{S}_1^{(1)} \times \dots \times \mathcal{S}_N^{(1)}$  of unit spheres is a compact set, an extremal solution of (11) exists (i.e., the supremum in (11) 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

$$\sigma_k(W) = \sup_{\substack{x_n \in \mathcal{S}_n^{(k)}, \\ 1 \leq n \leq N}} |W(x_1, \dots, x_N)|, \quad k \leq K. \tag{12}$$

Again, since the Cartesian product

$$\mathcal{S}^{(k)} = \mathcal{S}_1^{(k)} \times \dots \times \mathcal{S}_N^{(k)} \tag{13}$$

is compact, the supremum in (12) 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

$$\{\varphi_1^{(\ell_1)}, \ell_1 = 1, \dots, L_1\}, \dots, \{\varphi_N^{(\ell_N)}, \ell_N = 1, \dots, L_N\} \tag{14}$$

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

**Definition 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 (11) and (12). 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 (12). A *singular value*

decomposition (SVD) of the tensor  $W$  is a representation of  $W$  with respect to the basis (14), i.e.,

$$W = \sum_{\ell_1=1}^{L_1} \cdots \sum_{\ell_N=1}^{L_N} w_{\ell_1 \dots \ell_N} \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_N^{(\ell_N)}. \tag{15}$$

The  $N$ -way array  $[[w_{\ell_1, \dots, \ell_N}]] \in \mathbb{R}^{L_1 \times \dots \times L_N}$  in 15 is called the *singular value core* of  $W$ .

We refer to van Belzen et al. (2007), Weiland and van Belzen (2010), for more details on this decomposition. We will refer to this decomposition as *Tensor SVD* for convenience. As discussed in Weiland and van Belzen (2010), this method generalizes some, but not all, properties of the matrix SVD to tensors.

The tensor SVD can be used to find a suitable projection basis to infer low-order models. This is done by using low-rank approximations of the tensor. Given the collection (14) 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 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  $W_r^* := W|_{\mathcal{M}_1^{r_1} \times \dots \times \mathcal{M}_N^{r_N}}$  and is represented by the expansion

$$W_r^* = \sum_{\ell_1=1}^{r_1} \cdots \sum_{\ell_N=1}^{r_N} w_{\ell_1 \dots \ell_N} \varphi_1^{(\ell_1)} \otimes \cdots \otimes \varphi_N^{(\ell_N)} \tag{16}$$

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 2** 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  $W$ .

This result can be applied to define an algorithm of successive rank-one approximations to tensors, see Weiland and van Belzen (2010) for more details. If only one modal direction, say the  $p$ -th, is approximated, the modal truncations defined in (16) yield non-increasing approximation error for increasing approximation order, as shown in the following result.

**Theorem 3** Given  $W \in \mathcal{T}_N$  of modal rank  $R = (R_1, \dots, R_N)$ . Define  $r = (r_1, \dots, r_N)$  with  $r_n = R_n$ ,  $n \neq p$  and  $r_p = k < R_p$  and consider the modal truncation  $W_r^*$ . Then the approximation error  $E_k = W - W_r^*$  satisfies  $\|E_{k+1}\|_F \leq \|E_k\|_F$ .

Therefore, when only one of the arguments of the tensor is approximated, the approximation error is non-increasing for increasing approximation order.

### 3.3 Improved accuracy

In this section some changes will be made to the Singular Value Decomposition for tensors described in Sect. 3.2. The tensor SVD constructs a complete decomposition of the tensor, whereas we may only be interested in obtaining orthonormal bases for the vector space of functions defined on  $\mathbb{X}'$ . In this section it will be shown that it may be advantageous to use this fact for the construction of a decomposition that is more dedicated to the problem. To achieve

this the SVD described in Sect. 3.2 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 (12). For 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 of functions defined on  $\mathbb{X}'$ . Yet, the SVD orthonormalizes the vector spaces of functions defined on  $\mathbb{X}''$  as well. Since we do not use 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 a basis. Note that in this section, the vector space  $\mathbb{X}''$  is not necessarily restricted to time only.

The new construction for 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  $X' := X_1 \times \dots \times X_i$  and  $X'' := X_{i+1} \times \dots \times X_N$ , 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

$$\hat{\sigma}_1(W) := \sup_{\substack{x_n \in \mathcal{S}_n^{(1)}, 1 \leq n \leq i \\ x_n \in \mathcal{S}_n, (i+1) \leq n \leq N}} |W(x_1, \dots, x_N)|. \tag{17}$$

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 (17) exists (i.e., the supremum in (11) 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

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

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

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

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 (18) is a maximum that is attained by an  $N$ -tuple

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

Note that the set over which the optimization takes place,  $\mathcal{S}^{(k)}$ , is in general a larger subset of the Cartesian product of unit balls than the set  $\mathcal{S}^{(k)}$  as originally defined in (13). 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

$$\{\psi_1^{(\ell_1)}, \ell_1 = 1, \dots, L_1\}, \dots, \{\psi_i^{(\ell_i)}, \ell_i = 1, \dots, L_i\} \tag{19}$$

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 3** 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 (19) for  $X'$ , where the original bases for  $X''$  are kept intact, i.e.

$$W^d = \sum_{\ell_1=1}^{L_1} \dots \sum_{\ell_N=1}^{L_N} \tilde{w}_{\ell_1 \dots \ell_N} \psi_1^{(\ell_1)} \otimes \dots \otimes \psi_i^{(\ell_i)} \otimes e_{i+1}^{(\ell_{i+1})} \otimes \dots \otimes e_N^{(\ell_N)} \tag{20}$$

$$= \sum_{\ell_1=1}^{L_1} \dots \sum_{\ell_N=1}^{L_N} \tilde{w}_{\ell_1 \dots \ell_N} U_{\ell_1 \dots \ell_N}. \tag{21}$$

Using this representation of  $W$ , a dedicated modal truncation can be defined.

**Definition 4** 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

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

The following theorem states some properties of the dedicated representation of a tensor.

**Theorem 4** Consider  $W \in \mathcal{T}_N$ .

1. For all  $1 \leq i \leq N$  the dedicated representation of  $W$  exists.
2. The dedicated representation is an orthogonal decomposition of  $W$  in the sense that the rank-one tensors  $U_{\ell_1 \dots \ell_N}$  in (21) are mutually orthogonal

$$\langle U_{\ell_1 \dots \ell_N}, U_{\ell'_1 \dots \ell'_N} \rangle = 0, \quad \text{unless } \ell_n = \ell'_n, \quad \forall n = 1, \dots, N.$$

3.

$$\hat{\sigma}_1 \geq \dots \geq \hat{\sigma}_K \geq 0 \tag{23}$$

4. There exists an orthonormal basis  $\{\psi_n^{(1)}, \dots, \psi_n^{(L_n)}\}$  of  $X_n$  with  $n > i$  such that

$$W(\psi_1^{(k)}, \dots, \xi, \dots, \psi_i^{(k)}, \psi_{i+1}^{(\ell_{i+1})}, \dots, \psi_N^{(\ell_N)}) = 0 \tag{24}$$

for all  $\xi \perp \text{span}\{\psi_n^{(1)}, \dots, \psi_n^{(k)}\}$ , where  $\xi$  is at the  $n$ th spot, with  $1 \leq n \leq i$ .

*Remark 1* Consider a tensor  $W \in \mathcal{T}_N$ . Item 4 of Theorem 4 immediately yields the following results regarding the zero structure of the dedicated representation (20) of  $W$

1. The core  $[[\tilde{w}_{\ell_1 \dots \ell_N}]]$  of the dedicated representation of  $W$  satisfies

$$\tilde{w}_{\ell_1 \dots \ell_N} = \begin{cases} 0 & \text{if } \ell_1 = \dots = \ell_N > K \\ 0 & \text{if } \ell_n > \ell_1 = \dots = \ell_{n-1} = \ell_{n+1} = \dots = \ell_i \end{cases} \tag{25}$$

2. Consider the case that  $L_1 = L_2 = \dots = L_N = L$ . Then the number of zeros in the core of the dedicated representation of  $W$  is at least

$$i \binom{L(L-1)}{2}.$$

The following theorem establishes a relationship between the original and dedicated singular values.

**Theorem 5** Consider  $W \in \mathcal{T}_N$ .

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

$$\sigma_1 \geq \dots \geq \sigma_K \geq 0 \tag{26}$$

$$\hat{\sigma}_1 \geq \dots \geq \hat{\sigma}_K \geq 0 \tag{27}$$

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

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

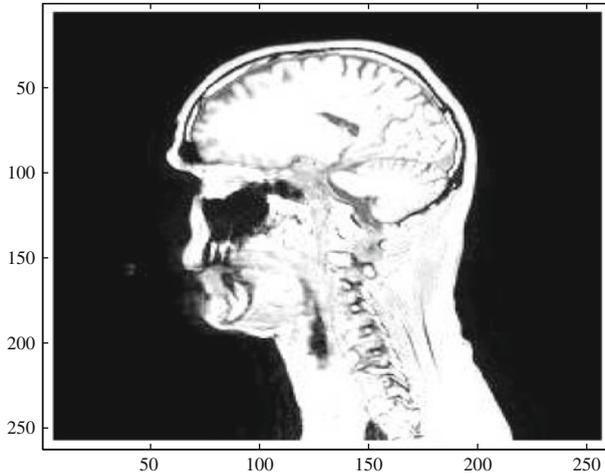
3.  $\hat{\sigma}_2 \geq \sigma_2$

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

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 scan consists of 1990676 pixels which corresponds to 2MB of storage. 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 original data is shown.

For this data both generic and dedicated singular values were computed. For computation of the singular values, the number of iterations of (45) needed are given in Table 3. The time to compute the first five singular values and sets of singular vectors for this example was 74.05 seconds on a 1.83 GHz Intel Dual Processor T2400. The numerical results of the computation of generic and dedicated singular values can be found in Table 1. From this table it is clear that the generic singular values decay much faster than the dedicated singular values. This



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

**Table 1** Generic and dedicated singular values

$\sigma_1$	102773.20	$\hat{\sigma}_1$	102773.20
$\sigma_2$	5815.49	$\hat{\sigma}_2$	49916.03
$\sigma_3$	3265.52	$\hat{\sigma}_3$	19275.82
$\sigma_4$	1948.73	$\hat{\sigma}_4$	11779.07
$\sigma_5$	1489.41	$\hat{\sigma}_5$	9920.99

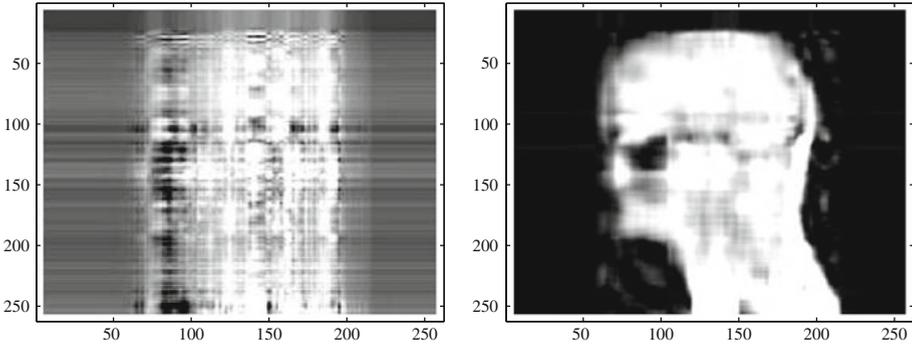
**Table 2** Approximation Error Results, for generic tensor SVD (left) and dedicated tensor SVD (right)

$(r_1, r_2)$	$\frac{\ W - W_r\ _F}{\ W\ _F}$	$\frac{\ W - W_r^d\ _F}{\ W\ _F}$
(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

would imply that the generic singular vectors give better results in approximation. However, examination of Table 2, 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. Figure 2 shows this clearly. A possible explanation for this is that since the dedicated singular value decomposition uses less constraints, more information of the original data is captured in the decomposition. Hence the larger dedicated singular values and the better approximations. On the left of Fig. 2, 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. Both approximations involve a core tensor with less than 0.15% of the storage capacity of the original tensor. This achieves a data compression from 2 MB to about 3 KB.

**Table 3** Number of iterations for computation of singular values

$\sigma_1$	20
$\sigma_2$	56
$\sigma_3$	80
$\sigma_4$	130
$\sigma_5$	262



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

### 4 Approximation of ND systems

The previous section gave an introduction to tensors and tensor decompositions. In the current section these concepts will be combined to arrive at Galerkin projections of ND systems. The model reduction approach is demonstrated on a heat transfer example.

#### 4.1 Model reduction

As in Sect. 2, first empirical projection spaces are defined from measured/simulated data. Then, the reduced model is defined using a Galerkin projection. Assume a FE solution  $w$  of (5) is available on a Cartesian domain  $\mathbb{X} = \mathbb{X}_1 \times \dots \times \mathbb{X}_N$ . Here, for  $n = 1, \dots, N$ ,  $\mathbb{X}_n$  is gridded in  $L_n$  distinct points. Furthermore, let  $X_n := \mathbb{R}^{L_n}$  be equipped with the standard Euclidean inner product. Then,  $w$  defines a multidimensional array  $[[w]] \in \mathbb{R}^{L_1 \times \dots \times L_N}$  and thus a tensor  $W : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$ .  $W$  is represented with respect to the standard bases as

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

where  $w_{\ell_1 \dots \ell_N}$  takes the value of the data element on the sample point with index  $(\ell_1, \dots, \ell_N)$ .

Our aim here is to generalize the idea of a POD basis for spatial domains to a higher dimensional Euclidean product space. For this, following Definition 1, the data dependent tensor  $W$  is assumed to be decomposed in SVD from according to (15). Let

$$\mathcal{M}_n^{(r_n)} = \text{span}\{\varphi_n^{(1)}, \dots, \varphi_n^{(r_n)}\}$$

for  $n = 1, \dots, N$  and  $r_n \leq L_n$  define a  $r_n$ -dimensional projection space in  $X_n$ . Then,  $\varphi_n^{(1)}, \dots, \varphi_n^{(r_n)}$  form an orthonormal basis of  $\mathcal{A}_n^{(r_n)}$ . Let  $W_r^*$  be the modal truncation of  $W$  as defined in Definition 2. Then, a generalized projection basis can be defined as follows.

**Definition 5** 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

$$\frac{\|W - W_r^*\|_F}{\|W\|_F} \leq \epsilon \tag{28}$$

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

Let  $W_r^d$  be the *dedicated* modal truncation of  $W$  as defined in Definition 4. Then, a *dedicated* projection basis is defined as follows.

**Definition 6** The vector of integers  $r = (r_1, \dots, r_i)$  with  $r_n \leq R_n$  is said to achieve a relative approximation error  $\epsilon > 0$  if

$$\frac{\|W - W_r^d\|_F}{\|W\|_F} \leq \epsilon \tag{29}$$

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

Now that two alternative projection bases have been defined, the model reduction framework can be completed by defining a corresponding spectral expansion and Galerkin projection. We assume that  $\mathbb{X}$  is a Cartesian product  $\mathbb{X} = \mathbb{X}' \times \mathbb{X}''$  in which

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

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

$$w(x_1, \dots, x_N) = \sum_{\ell_1} \dots \sum_{\ell_{N-1}} a_{\ell_1 \dots \ell_{N-1}}(x_N) \varphi_1^{(\ell_1)}(x_1) \otimes \dots \otimes \varphi_{N-1}^{(\ell_{N-1})}(x_{N-1}). \tag{30}$$

Reduction of the signal space is now defined by the truncation

$$w_r(x_1, \dots, x_N) = \sum_{\ell_1=1}^{r_1} \dots \sum_{\ell_{N-1}=1}^{r_{N-1}} a_{\ell_1 \dots \ell_{N-1}}(x_N) \varphi_1^{(\ell_1)}(x_1) \otimes \dots \otimes \varphi_{N-1}^{(\ell_{N-1})}(x_{N-1}) \tag{31}$$

where  $r = (r_1, \dots, r_{N-1})$  with  $r_n \leq R_n$ .

To define the reduced order system we go back to the difference equation

$$D(\zeta_1, \dots, \zeta_N)w = 0$$

introduced in Sect. 2, Eq. 5. Given the vector spaces  $X_1, \dots, X_N$  as defined in Sect. 3.1, an arbitrary signal  $w : \mathbb{X}_1 \times \dots \times \mathbb{X}_N \rightarrow \mathbb{R}$ , we associate a tensor  $\hat{D} : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$

with the residual  $R := D(\zeta_1, \dots, \zeta_N)w$  of the difference equation. Note that  $R : \mathbb{X}_1 \times \dots \times \mathbb{X}_N \rightarrow \mathbb{R}$  and its associated tensor  $\hat{D}$  can be represented with respect to the standard bases for  $X_1, \dots, X_N$  as follows

$$\hat{D} = \sum_{\ell_1} \dots \sum_{\ell_N} \hat{d}_{\ell_1 \dots \ell_N} e_1^{(\ell_1)} \otimes \dots \otimes e_N^{(\ell_N)} \tag{32}$$

where, given a point in the domain of the difference equation  $(x_{\ell_1}, \dots, x_{\ell_N})$  the coefficients  $\hat{d}_{\ell_1 \dots \ell_N}$  are defined as

$$\hat{d}_{\ell_1 \dots \ell_N} = R(x_{\ell_1}, \dots, x_{\ell_N}) = [D(\zeta_1, \dots, \zeta_N)w](x_{\ell_1}, \dots, x_{\ell_N}). \tag{33}$$

Given sets of POD basis functions  $\{\varphi_n^{(k_n)}\}_{k_n=1}^{r_n}$  a Galerkin projection of the discrete time model  $D(\zeta_1, \dots, \zeta_N)w = 0$  is defined as

$$\left\langle \varphi_1^{(k_1)}, \dots, \varphi_{N-1}^{(k_{N-1})}, D(\zeta_1, \dots, \zeta_N)w \right\rangle_{N-1} \dots \Big|_1 = 0 \tag{34}$$

for  $k_n = 1, \dots, r_n$  and  $n = 1, \dots, N - 1$ . Here, the expression of nested inner products should be interpreted as follows. Firstly,  $R = D(\zeta_1, \dots, \zeta_N)w$  is to be associated with the tensor  $\hat{D} : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$  as above so that  $\hat{D}_{N-1} := \langle \varphi_{N-1}, \hat{D} \rangle$  becomes a linear functional  $\hat{D}_{N-1} : X_1 \times \dots \times X_{N-2} \times X_N \rightarrow \mathbb{R}$ , etc. Note that the  $N$ th independent variable is not projected. Throughout we assume this independent variable corresponds to time. Equation 34 defines a collection  $\mathcal{D}_G$  of tensors  $\hat{D}_G : X_N \rightarrow \mathbb{R}$  defined by

$$\begin{aligned} \hat{D}_G &:= \hat{D}(\varphi_1^{(k_1)}, \dots, \varphi_{N-1}^{(k_{N-1})}, \cdot) \\ &= \sum_{\ell_1} \dots \sum_{\ell_N} \hat{d}_{\ell_1 \dots \ell_N} \left\langle e_1^{(\ell_1)}, \varphi_1^{(k_1)} \right\rangle \dots \left\langle e_{N-1}^{(\ell_{N-1})}, \varphi_{N-1}^{(k_{N-1})} \right\rangle \left\langle e_N^{(\ell_N)}, \cdot \right\rangle \end{aligned} \tag{35}$$

for  $1 \leq k_n \leq r_n$ , see Lemma 1 in the appendix for background. Now,  $\mathcal{D}_G$  is defined by

$$\mathcal{D}_G = \{ \hat{D}_G : X_N \rightarrow \mathbb{R} \mid 1 \leq k_n \leq r_n, n = 1, \dots, N - 1 \}. \tag{36}$$

Equation 35 can be simplified by defining

$$b_{k_1 \dots k_{N-1} \ell_N} = \sum_{\ell_1} \dots \sum_{\ell_{N-1}} \hat{d}_{\ell_1 \dots \ell_N} \left\langle e_1^{(\ell_1)}, \varphi_1^{(k_1)} \right\rangle \dots \left\langle e_{N-1}^{(\ell_{N-1})}, \varphi_{N-1}^{(k_{N-1})} \right\rangle. \tag{37}$$

This gives

$$\hat{D}_G(\cdot) = \sum_{\ell_N} b_{k_1 \dots k_{N-1} \ell_N} \left\langle e_N^{(\ell_N)}, \cdot \right\rangle \tag{38}$$

Now, we are in a position to define a reduced order model. Given a time instance  $t = t_{k_n}$ , the reduced model is given by the following equations

$$D_G(t_{k_n}) = b_{k_1 \dots k_n} = 0 \tag{39}$$

for  $1 \leq k_n \leq r_n, n = 1, \dots, N - 1$ . Given the order of the reduced model,  $r = (r_1, \dots, r_{N-1})$ , the spectral expansion used for  $w(x_1, \dots, x_N)$  is given by

$$w_r(x_1, \dots, x_N) = \sum_{k_1} \dots \sum_{k_{N-1}} a_{k_1 \dots k_{N-1}}(x_N) \varphi_1^{(k_1)}(x_1) \otimes \dots \otimes \varphi_{N-1}^{(k_{N-1})}(x_{N-1}). \tag{40}$$

Trajectories of the reduced model are thus formed by the coefficients  $a_{k_1 \dots k_{N-1}}$  that satisfy the residuals (39) for all  $t_{k_n}$ .

4.2 Model reduction of a heat transfer process

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

$$0 = -\rho c_p \frac{\partial w}{\partial t} + \kappa_x \frac{\partial^2 w}{\partial x^2} + \kappa_y \frac{\partial^2 w}{\partial y^2}. \tag{41}$$

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.

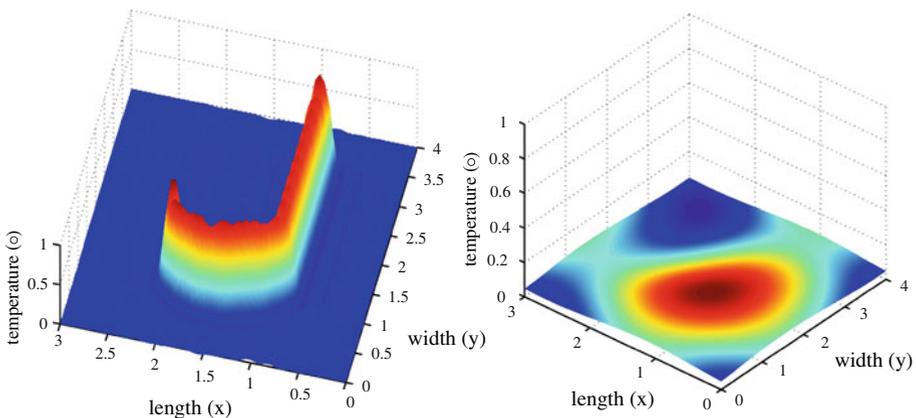
Solutions of the reduced model are then given by

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

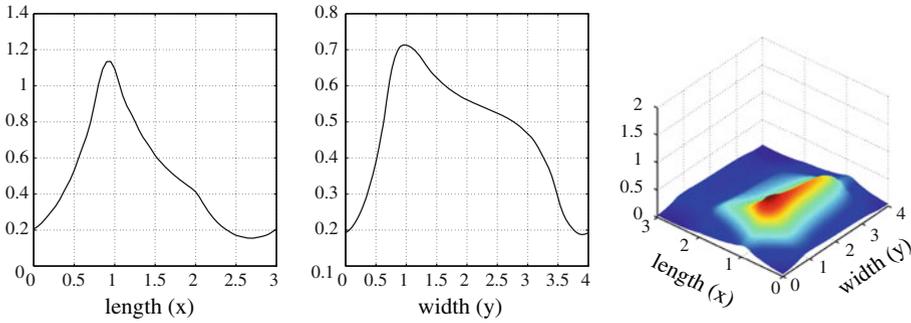
$$0 = -\rho c_p \dot{A} + \kappa_x F A + \kappa_y A P. \tag{42}$$

**Table 4** PDE Parameter Values

Parameter	Value	Unit
$\rho C_p$	5	$\frac{J}{m^3 \cdot K}$
$\kappa_x$	0.5	$\frac{W}{m \cdot K}$
$\kappa_y$	0.5	$\frac{W}{m \cdot K}$
$L_x$	3	m
$L_y$	4	m
$T_f$	3.6	s
$\Delta_x$	0.05	m
$\Delta_y$	0.05	m
$\Delta_t$	0.05	s



**Fig. 3** First and final time slices of the FE solution of (41)



**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

$$\begin{aligned} \rho c_p \dot{a}_{\ell_1 \ell_2}(t) = & \kappa_x \sum_{k_1=1}^{r_1} a_{k_1 \ell_2}(t) \left\langle \ddot{\varphi}_1^{(k_1)}(x), \varphi_1^{(\ell_1)}(x) \right\rangle \\ & + \kappa_y \sum_{k_2=1}^{r_2} a_{\ell_1 k_2}(t) \left\langle \ddot{\varphi}_2^{(k_2)}(y), \varphi_2^{(\ell_2)}(y) \right\rangle \end{aligned} \quad (43)$$

for  $1 \leq \ell_1 \leq r_1$  and  $1 \leq \ell_2 \leq r_2$ . A FE solution of (41) has been computed with physical and discretization parameters as given in Table 4. 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.

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 Sect. 3.2 are shown in Fig. 4.

**Table 5** Reduced model simulation error results, basis functions were computed using TSVD (left) and dedicated construction (right)

$r$	$\frac{\ W - W_r\ _F}{\ W\ _F}$	$\frac{\ W - W_r^d\ _F}{\ W\ _F}$
(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

The simulation time of the FE implementation is 17.22 s, the reduced models have a simulation time of approximately 0.35 s. Table 5 gives the simulation error of the reduced model for different model orders. The reduced models were given the same initial condition as the one 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 5 using a dedicated construction to compute basis functions does not give a more accurate reduced model for this example.

### 5 Numerical computation of tensor SVD

In this section we propose an algorithm for the computation of the singular value decomposition of an arbitrary order- $N$  tensor  $W \in \mathcal{T}_N$ , as defined in Definition 1. The algorithm is based on the fixed point properties of a contractive mapping  $G$  that is iterated in a power-type algorithm to compute the singular vectors of order  $k$  and the corresponding singular value.

Let  $W \in \mathcal{T}_N$  be a tensor defined on the Cartesian product  $X_1 \times \dots \times X_N$  where  $\dim(X_n) = L_n$ . The operator  $\text{col}$  stacks all its arguments in a vector, i.e.,  $\text{col}(a, b)$  is the vector  $(a \ b)^\top$ . Let  $L = \sum_{n=1}^N L_n$  and define the mapping  $G_W : \mathbb{R}^{L+1} \rightarrow \mathbb{R}^{L+1}$  by

$$G_W(x) := \begin{pmatrix} \frac{1}{\sigma} \nabla_1 W\left(\frac{x_1}{\|x_1\|}, \dots, \frac{x_N}{\|x_N\|}\right) \\ \vdots \\ \frac{1}{\sigma} \nabla_N W\left(\frac{x_1}{\|x_1\|}, \dots, \frac{x_N}{\|x_N\|}\right) \\ W\left(\frac{x_1}{\|x_1\|}, \dots, \frac{x_N}{\|x_N\|}\right) \end{pmatrix} \tag{44}$$

Here,  $x = \text{col}(x_1, \dots, x_N, \sigma)$  and

$$\nabla_n W = \text{col}\left(\frac{\partial W}{\partial x_n^{(1)}}, \dots, \frac{\partial W}{\partial x_n^{(L_n)}}\right)$$

is the  $n$ -mode gradient of  $W$ . Then  $G_W$  is well defined provided that  $x_n \neq 0$  for any  $n$  and  $\sigma \neq 0$ .

The following TSVD algorithm is based on a power iteration of the map  $G_W$  and proves a rather efficient tool for the computation of singular values, singular vectors and the singular value core tensor of  $W$ . The algorithm computes the  $k$ th order singular values and singular vectors for consecutive values of  $k$ .

Step 0 (Initialization) Set tolerance level  $\varepsilon_{\text{tol}} > 0$ , order  $k = 1$ , and  $W_k = W$ .

Step 1 Select for  $n = 1, \dots, N$  random elements  $x_n^0 \in X_n$  and  $\sigma^0$  with  $\|x_n^0\|_n = 1$  and  $0 < \sigma^0 < 1$ . Set  $x^0 := \text{col}(x_1^0, \dots, x_N^0, \sigma^0)$ .

Step 2 Define  $G_{W_k}$  according to (44) with  $W = W_k$  and iterate

$$x^{i+1} = G_{W_k}(x^i), \quad i = 0, 1, 2, \dots, i^* \tag{45}$$

where  $i^*$  is the smallest integer for which  $\|x^{i^*} - x^{i^*-1}\| < \varepsilon_{\text{tol}}$ .

Step 3 Write  $x^{i^*} = \text{col}(x_1^*, \dots, x_N^*, \sigma^*)$  and define, for  $n = 1, \dots, N$ ,

$$\begin{aligned} \sigma_k &= \sigma^*, \quad \varphi_n^{(k)} = x_n^*, \\ \Phi_n^{(k)} &= \left( \varphi_n^{(1)} \ \dots \ \varphi_n^{(k)} \right), \\ Q_n^{(k)} &= I - \Phi_n^{(k)} [\Phi_n^{(k)}]^\top. \end{aligned}$$

Step 4 Define the tensor  $W_{k+1} \in \mathcal{T}_N$  by

$$W_{k+1}(x_1, \dots, x_N) := W(Q_1^{(k)}x_1, \dots, Q_N^{(k)}x_N)$$

and set  $k$  to  $k + 1$ .

Step 5 Repeat Step 1, Step 2, Step 3, Step 4 until  $k = K = \min_n \text{modrank}(W)$ .

Step 6 For every  $n$  for which  $K < L_n$  complement  $\Phi_n^{(K)}$  to a unitary square matrix  $\Phi_n^{(L_n)}$ .

Step 7 (Output)

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

with  $w_{\ell_1 \dots \ell_N} = W(\varphi_1^{(\ell_1)}, \dots, \varphi_N^{(\ell_N)})$ .

The algorithm is easy to implement and has shown satisfactory performance.

To investigate convergence properties of the TSVD algorithm, let  $G_W^p$  denote the  $p$ th power of the operator  $G_W$ , i.e., the  $p$ th iterate of  $x^{i+1} = G_W x^i$  with initial condition  $x^0$  in (45) satisfies  $x^p = G_W^p x^0$ . The following theorem generalizes a result in Weiland and van Belzen (2010) and shows convergence of the above sequential series of iterations to the exact singular vectors and singular values of the tensor  $W$ .

**Theorem 6** Suppose that  $G_W : \mathbb{R}^{L+1} \rightarrow \mathbb{R}^{L+1}$  maps a closed subset  $\mathcal{D} \subset \mathbb{R}^{L+1}$  into itself and that

$$\|G_W^p x - G_W^p y\| \leq \alpha_p \|x - y\|, \quad \text{for all } x, y \in \mathcal{D}, \quad p = 1, 2, \dots \tag{46}$$

where  $\beta = \sum_{p=1}^{\infty} \alpha_p < \infty$ . Then for every  $k = 1, \dots, K$ , with  $K = \min_n \text{modrank}(W)$ , the operator  $G_{W_k}$  has a unique fixed point  $x^* \in \mathcal{D}$  (depending on  $k$ ) and the iteration (45) converges to  $x^*$  as  $i \rightarrow \infty$ . Moreover, every iterate establishes the error estimate

$$\|x^i - x^*\| \leq \beta \|x^i - x^{i-1}\|, \quad i = 1, 2, \dots$$

and the components  $\varphi_n^{(k)}$ ,  $n = 1, \dots, N$  and  $\sigma_k$  of the fixed point  $x^*$  are extremal values of the optimizations (11) (if  $k = 1$ ) and (12) (for  $1 < k \leq K$ ).

The proof of the above theorem is an application of Theorem 12.1.1 in Rheinboldt (1970) combined with the observation that the inequality (46) holds with  $G_W$  replaced by  $G_{W_k}$  with  $k > 1$  whenever (46) for  $G_W = G_{W_1}$ . In particular, this observation makes the convergence rate  $\beta$  independent of  $k$ .

In practice it is not trivial to explicitly verify whether  $G_W$  satisfies (46) An interesting special case of Theorem 6 applies to tensors  $W$  for which  $G_W$  maps a closed subset  $\mathcal{D} \subset \mathbb{R}^{L+1}$  into itself and is contractive on  $\mathcal{D}$  in the sense that there exists  $\alpha < 1$  such that  $\|G_W x - G_W y\| \leq \alpha \|x - y\|$  for all  $x, y \in \mathcal{D}$ . In that case, the result of Theorem 6 simplifies to the contraction mapping theorem for nonlinear operators. Specifically, if  $G_W$  is contractive, (46) holds with  $\alpha_p = \alpha^p$  and  $\beta = \frac{\alpha}{1-\alpha}$  defines the convergence rate. This means that under the contractivity condition of  $G_W$ , the sequence (45) in step 2 of the TSVD algorithm converges to the unique fixed point of  $G_{W_k}$  in  $\mathcal{D}$  whenever  $x^0 \in \mathcal{D}$ . More refinements of convergence conditions go in the direction of transforming  $G_W$  into  $G'_W := T G_W T^{-1}$  where a suitable homeomorphism  $T : \mathbb{R}^{L+1} \rightarrow \mathbb{R}^{L+1}$  is chosen so as to make  $G'_W$  contractive, or, alternatively, to study iterated contractions of the form  $\|G_W(G_W x) - G_W x\| \leq \alpha \|G_W x - x\|$  where  $\alpha < 1$  and  $x \in \mathcal{D}$ . We refer to Rheinboldt (1970) for more details.

Theorem 6 promises that whenever the algorithm converges, it converges to an extremal solution to the optimization problems (11) and (12) that define the singular value and singular vectors of order  $k$ . Here, by ‘extremal solutions’ we mean that the fixed points of  $G_{W_k}$  satisfy the *first order necessary conditions* for the optimal solution of the maximization problems formulated in (12). Solutions to the optimization problems (11) and (12) satisfy these conditions but we can not guarantee that the iterated map (45) converges to a fixed point  $x^*$  of  $G_{W_k}$  that also satisfies the *sufficient conditions* for the optima. This means that if  $G_{W_k}$  is contractive, the algorithm converges to a fixed point  $x^* = \text{col}(\varphi_1^{(k)}, \dots, \varphi_N^{(k)}, \sigma_k)$  where the  $N$ -tuple  $(\varphi_1^{(k)}, \dots, \varphi_N^{(k)}) \in \mathcal{S}^{(k)}$  and where the gradient of the cost function  $|W(x_1, \dots, x_N)|$  vanishes in  $(\varphi_1^{(k)}, \dots, \varphi_N^{(k)})$ .

*Remark 2* A numerical algorithm for the computation of a dedicated singular value decomposition requires a minor change to the TSVD algorithm. Indeed, if  $W \in \mathcal{T}_N$  with  $X' = X_1 \times \dots \times X_i$  and  $X'' = X_{i+1} \times \dots \times X_N$ . The dedicated singular value decomposition in Definition 3 is numerically calculated from the TSVD algorithm in which the definition of  $Q_n^{(k)}$  in step 3 is replaced by

$$Q_n^{(k)} := \begin{cases} I - \Phi_n^{(k)}[\Phi_n^{(k)}]^\top & 1 \leq n \leq i \\ I & i + 1 \leq n \leq N \end{cases}.$$

### 6 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.

### Proofs

*Proof of Theorem 1.* If eigenvalues of  $\Phi$  are absolute summable,  $\Phi$  is self-adjoint and nuclear. This means that it admits a representation  $\Phi = \sum_{n=1}^N \lambda_n \langle \psi_n, \cdot \rangle \psi_n$  where  $1 \leq N \leq \infty$ , the eigenvalues  $\lambda_n$  are positive, non-increasingly ordered and summable, and the eigenfunctions  $\{\psi_n, n = 1, \dots, N\}$  are orthonormal in  $\mathcal{H}$ . Moreover, for any orthonormal basis  $\{\varphi_n, n \in \mathbb{I}\}$  of  $\mathcal{H}$  we have

$$\begin{aligned}
 J(\varphi_1, \dots, \varphi_r) &= \int_{\mathbb{X}''} \langle w - w_r, w - w_r \rangle dx'' \\
 &= \int_{\mathbb{X}''} \left\langle \sum_{n>r} \langle w(\cdot, x''), \varphi_n \rangle \varphi_n, \sum_{n>r} \langle w(\cdot, x''), \varphi_n \rangle \varphi_n \right\rangle dx'' \\
 &= \int_{\mathbb{X}''} \sum_{n>r} \langle w(\cdot, x''), \varphi_n \rangle \cdot \langle w(\cdot, x''), \varphi_n \rangle dx'' \\
 &= \sum_{n>r} \int_{\mathbb{X}''} \langle w(\cdot, x''), \varphi_n \rangle \cdot \langle w(\cdot, x''), \varphi_n \rangle dx'' \\
 &= \sum_{n>r} \langle \varphi_n, \Phi \varphi_n \rangle.
 \end{aligned}$$

Now first suppose that  $\varphi_n = \psi_n$  for  $n = 1, \dots, N$ . Then,  $\{\varphi_n, n = 1, \dots, N\}$  is an orthonormal set of eigenfunctions of  $\Phi$  and  $J(\varphi_1, \dots, \varphi_r) = \sum_{n>r} \lambda_n$  is finite and minimal for all  $r$  by the monotonicity of the sequence  $\lambda_n$ . Hence,  $\{\varphi_n, n = 1, \dots, r\}$  is a POD basis of order  $r$  for any  $r$ . Second, for any POD basis  $\{\varphi_n, n \in \mathbb{I}\}$  the above expression for the error implies that

$$J(\varphi_1, \dots, \varphi_r) = \sum_{n>r} \left\langle \varphi_n, \sum_{m=1}^N \lambda_m \langle \psi_m, \varphi_n \rangle \psi_m \right\rangle = \sum_{n>r} \sum_{m=1}^N \lambda_m \langle \varphi_n, \psi_m \rangle^2$$

which is minimal for all  $r$  only if  $\langle \varphi_n, \psi_m \rangle = \delta_{n,m}$  for all integers  $n, m$  between 1 and  $N$ . But then it is immediate that  $\{\varphi_n, n = 1, \dots, N\}$  is also a set of orthonormal eigenvectors of  $\Phi$ . □

For the next proof we need the following lemma

**Lemma 1** *Let  $W \in \mathcal{T}_N, W : X_1 \times \dots \times X_N \rightarrow \mathbb{R}$ , with  $X_n$  inner product spaces, possibly infinite dimensional, and  $x_n \in X_n$  for  $n = 1, \dots, N$ . Then*

1.

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

2.

$$W(x_1, \dots, x_N) = \langle x_N, \dots \langle x_2, \langle x_1, W \rangle_1 \rangle_2 \dots \rangle_N.$$

*Proof of Lemma 1.* 1. Let  $\{\xi_n^{(\ell_n)}\}_{\ell_n=1}^\infty$  be an orthonormal basis for  $X_n, n = 1, \dots, N$ .  $W$  can be represented with respect to these bases as  $W = \sum_{\ell_1} \dots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} \xi_1^{(\ell_1)} \otimes \dots \otimes \xi_N^{(\ell_N)}$ . The tensor evaluation can be written as  $W(x_1, \dots, x_N) = \sum_{\ell_1} \dots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} \langle x_1, \xi_1^{\ell_1} \rangle \dots \langle x_N, \xi_N^{\ell_N} \rangle$ . Let  $U := x_1 \otimes \dots \otimes x_N$ .  $U$  can be represented as  $U = \sum_{\ell_1} \dots \sum_{\ell_N} u_{\ell_1 \dots \ell_N} \xi_1^{(\ell_1)} \otimes \dots \otimes \xi_N^{(\ell_N)}$  with  $u_{\ell_1 \dots \ell_N} = \prod_{i=1}^N \langle x_i, \xi_i^{(\ell_i)} \rangle_i$ . Then,

$$\begin{aligned}
 \langle W, U \rangle &= \sum_{k_1} \dots \sum_{k_N} \sum_{\ell_1} \dots \sum_{\ell_N} w_{k_1 \dots k_N} u_{\ell_1 \dots \ell_N} \\
 &\quad \cdot \underbrace{\langle \xi_1^{k_1}, \xi_1^{\ell_1} \rangle}_{0 \text{ unless } k_1 = \ell_1} \dots \langle \xi_N^{k_N}, \xi_N^{\ell_N} \rangle
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\ell_1} \cdots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} u_{\ell_1 \dots \ell_N} \\
 &= \sum_{\ell_1} \cdots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} \langle \xi_1^{\ell_1}, x_1 \rangle \cdots \langle \xi_N^{\ell_N}, x_N \rangle
 \end{aligned}$$

which is the tensor evaluation.

2. To prove the second statement, we first show that  $\langle x_1, W(\cdot, v_2, \dots, v_N) \rangle_1 = W(x_1, v_2, \dots, v_N)$  for some  $v_n \in X_n, n = 2, \dots, N$ . Let  $\{\xi_n^{(\ell_n)}\}_{\ell_n=1}^\infty$  be an orthonormal basis for  $X_n, n = 1, \dots, N$ .  $W$  can be represented with respect to these bases as  $W = \sum_{\ell_1} \cdots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} \xi_1^{(\ell_1)} \otimes \cdots \otimes \xi_N^{(\ell_N)}$ . Then

$$W(x_1, v_2, \dots, v_N) = \sum_{\ell_1} \cdots \sum_{\ell_N} w_{\ell_1 \dots \ell_N} \langle \xi_1^{(\ell_1)}, x_1 \rangle \prod_{k=2}^N \langle \xi_k^{(\ell_k)}, v_k \rangle.$$

On the other hand, we can write  $x_1$  as  $x_1 = \sum_{k=1}^\infty \langle x_1, \xi_1^{(k)} \rangle \xi_1^{(k)}$ . Then

$$\begin{aligned}
 \langle x_1, W(\cdot, v_2, \dots, v_N) \rangle_1 &= \sum_{k_1} \langle x_1, \xi_1^{(k_1)} \rangle_1 W(\xi_1^{(k_1)}, v_2, \dots, v_N) \\
 &= \sum_{k_1} \langle x_1, \xi_1^{(k_1)} \rangle_1 \sum_{\ell_2} \cdots \sum_{\ell_N} w_{k_1 \ell_2 \dots \ell_N} \prod_{k=2}^N \langle \xi_k^{(\ell_k)}, v_k \rangle \\
 &= W(x_1, v_2, \dots, v_N)
 \end{aligned}$$

Thus, we have that  $\langle x_1, W(\cdot, v_2, \dots, v_N) \rangle_1 = W(x_1, v_2, \dots, v_N)$ . Since tensors are multi-linear functionals, this completes the proof. □

*Proof of Theorem 2.* 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  $\lambda$  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 (11) 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 of Theorem 5.*

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 of Theorem 4*

1. Since extremal solutions to the optimization problems (17) and (18) are guaranteed to exist for any tensor  $W \in \mathcal{T}_N$ , also the dedicated representation is guaranteed to exist.
2. Using Lemma 1 we have

$$\left\langle U_{\ell_1 \dots \ell_N}, U_{\ell'_1 \dots \ell'_N} \right\rangle = \prod_{n=1}^i \left\langle \psi_n^{(\ell_1)}, \psi_n^{(\ell'_n)} \right\rangle \prod_{n=i+1}^N \left\langle e_n^{(\ell_n)}, e_n^{(\ell'_n)} \right\rangle.$$

Therefore, the inner product between unequal rank-one tensors is zero whenever one of the inner products on the right-hand side is zero. Since the bases are orthonormal, all rank-one tensors are orthogonal unless  $\ell_n = \ell'_n$  for  $1 \leq n \leq N$ .

3. This is by construction
4. Let  $L = L_1 + \dots + L_N$  and associate with the optimization problem (17) the Lagrangian  $L_1 : \mathbb{R}^{L+N} \rightarrow \mathbb{R}$  by setting

$$L_1(x, \lambda) := W(x_1, \dots, x_N) + \sum_{n=1}^N \frac{1}{2} \lambda_n (1 - \langle x_n, x_n \rangle).$$

It has already been argued that an  $N$ -tuple  $x^{(1)} = (\psi_1^{(1)}, \dots, \psi_N^{(1)})$  exists that attains the maximum in (17). From the theory of variational analysis Bertsekas (1982), Fletcher (1981), one then infers the existence of an  $N$ -tuple  $\lambda^{(1)} = (\lambda_1^{(1)}, \dots, \lambda_N^{(1)})$  of Lagrange multipliers such that

$$\nabla L_1^{(1)}, \lambda^{(1)} = 0, \tag{47}$$

where  $\nabla L_1$  denotes the gradient of  $L_1$ . The  $n$ -mode Fréchet derivative  $\partial_n W(x_1, \dots, x_N)$  of  $W$  at the point  $(x_1, \dots, x_N)$  is an order-1 tensor (a linear functional) that maps  $X_n$  to  $\mathbb{R}$  and satisfies

$$\partial_n W(x_1, \dots, x_N) = W(x_1, \dots, x_{n-1}, \cdot, x_{n+1}, \dots, x_N)$$

where the ‘dot’ is at the  $n$ th spot. By the multi-linearity of the tensor,  $\partial_n W(x_1, \dots, x_N)$  is independent of  $x_n \in X_n$ . Hence, rewriting (47) for each independent modal direction gives that  $x^{(1)}, \lambda^{(1)}$  satisfies, for  $n = 1, \dots, N$ ,

$$W(\psi_1^{(1)}, \dots, \psi_{n-1}^{(1)}, \cdot, \psi_{n+1}^{(1)}, \dots, \psi_N^{(1)}) = \lambda_n^{(1)} \left\langle \cdot, \psi_n^{(1)} \right\rangle, \tag{48a}$$

$$\|\psi_n^{(1)}\| = 1. \tag{48b}$$

(48a) implies that for each  $n = 1, \dots, N$ ,

$$W(\psi_1^{(1)}, \dots, \psi_{n-1}^{(1)}, \xi_n, \psi_{n+1}^{(1)}, \dots, \psi_N^{(1)}) = 0 \quad \text{whenever} \quad \left\langle \xi_n, \psi_n^{(1)} \right\rangle = 0.$$

In a similar manner, for  $k > 1$  we associate with the optimization problem (18) the Lagrangian  $L_k : \mathbb{R}^{L+N+i(k-1)} \rightarrow \mathbb{R}$  defined by

$$L_k(x, \lambda, \mu) = W(x_1, \dots, x_n) + \sum_{n=1}^N \frac{1}{2} \lambda_n (1 - \langle x_n, x_n \rangle) + \sum_{n=1}^i \langle g_n(x_n), \mu_n \rangle.$$

where  $x_n \in X_n, \lambda_n \in \mathbb{R}, \mu_n \in \mathbb{R}^{k-1}$  and  $g_n : X_n \rightarrow \mathbb{R}^{k-1}$  is given by

$$g_n(\xi_n) := \begin{pmatrix} \langle \xi_n, \psi_n^{(1)} \rangle \\ \vdots \\ \langle \xi_n, \psi_n^{(k-1)} \rangle \end{pmatrix}.$$

Again, there exist  $N$ -tuples  $x^{(k)}, \lambda^{(k)}$  and  $\mu^{(k)}$  that satisfy the stationarity condition

$$\nabla L_k(x^{(k)}, \lambda^{(k)}, \mu^{(k)}) = 0. \tag{49}$$

Rewriting (49) for each modal direction gives, for  $n = 1, \dots, i$ , that

$$W(\psi_1^{(k)}, \dots, \psi_{n-1}^{(k)}, \cdot, \psi_{n+1}^{(k)}, \dots, \psi_N^{(k)}) = \lambda_n^{(k)} \langle \cdot, \psi_n^{(k)} \rangle + \langle g_n(\cdot), \mu_n^{(k)} \rangle, \tag{50a}$$

$$\|\psi_n^{(k)}\| = 1, \tag{50b}$$

$$g_n(\psi_n^{(k)}) = 0. \tag{50c}$$

Now suppose, again for  $n = 1, \dots, i$ , that  $\xi \perp \text{span}\{\psi_n^{(1)}, \dots, \psi_n^{(k)}\}$ . Substituting  $\xi$  for the dotted argument in (50a) gives that

$$W(\psi_1^{(k)}, \dots, \xi, \dots, \psi_i^{(k)}, \psi_{i+1}^{(\ell_{i+1})}, \dots, \psi_N^{(\ell_N)}) = 0 \tag{51}$$

for all  $\xi \perp \text{span}\{\psi_n^{(1)}, \dots, \psi_n^{(k)}\}$ , where  $\xi$  is at the  $n$ th spot, with  $1 \leq n \leq i$ .

□

*Proof of Theorem 3* Without loss of generality let  $p = 1$ . Then  $W_r^* = W|_{\mathcal{M}_1^{(k)} \times X_2 \times \dots \times X_N}$  with  $\mathcal{M}_1^{(k)} = \text{span}\{\varphi_1^{(1)}, \dots, \varphi_1^{(k)}\}$ . Then

$$W = W|_{\mathcal{M}_1^{(k)} \times X_2 \times \dots \times X_N} + W|_{\mathcal{M}_1^{(k)\perp} \times X_2 \times \dots \times X_N} - W|_{(\mathcal{M}_1^{(k)} \times X_2 \times \dots \times X_N) \cap (\mathcal{M}_1^{(k)\perp} \times X_2 \times \dots \times X_N)}.$$

Since

$$(\mathcal{M}_1^{(k)} \times X_2 \times \dots \times X_N) \cap (\mathcal{M}_1^{(k)\perp} \times X_2 \times \dots \times X_N)$$

is equal to

$$(\mathcal{M}_1^{(k)} \cap \mathcal{M}_1^{(k)\perp}) \times (X_2 \cap X_2) \times \dots \times (X_N \cap X_N)$$

and  $W|_{\emptyset \times X_2 \times \dots \times X_N} = 0$ , we infer that

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

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  $\|E_{k+1}\| \leq \|E_k\|$ . □

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Bertsekas, D. P. (1982). *Constrained optimization and Lagrange multiplier methods*. New York: Academic Press.
- Beylkin, G., & Mohlenkamp, M. J. (2005). Algorithms for numerical analysis in high dimensions. *SIAM Journal of Scientific Computing*, 26(6):2133–2159.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3), 283–319.
- De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multi-linear singular value decomposition. *SIAM Journal of Matrix Analysis and Applications*, 21(4), 1253–1278.
- De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). On the best rank-1 and rank  $(r_1, r_2, \dots, r_n)$  approximation of higher order tensors. *SIAM Journal of Matrix Analysis and Applications*, 21(4), 1324–1342.
- de Silva, V., & Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal of Matrix Analysis and Applications: Special Issue on Tensor Decompositions and Applications*, 30(3), 1084–1127.
- Eldén, L., & Savas, B. (2009). A Newton-Grassmann method for computing the best multi-linear rank- $(r_1, r_2, r_3)$  approximation of a tensor. *Siam Journal of Matrix Analysis and Applications*, 31(2), 248–271.
- Fletcher, R. (1981). *Practical methods of optimization 2 constrained optimization*. London: Wiley.
- Fujimoto, K., & Ono, S. (2008). On balanced realization and finite-dimensional approximation for infinite-dimensional nonlinear systems. Proceedings of the 47th IEEE CDC.
- Glover, K., et al. (1988). Realisation and approximation of linear infinite-dimensional systems with error bounds. *SIAM Journal of Control and Optimization*, 26(4), 863–898.
- Hackbusch, W. (2008). *Use of tensor formats in elliptic eigenvalue problems*, preprint
- Ishteva, M., Absil, P.-A., van Huffel, S., & De Lathauwer, L. (2010). Tucker compression and local optima. Chemometrics and intelligent laboratory systems, Accepted.
- Jolliffe, I. T. (1986). *Principal component analysis*. Berlin: Springer.
- Khoromskij, B. N., & Khoromskaia, V. (2009). Multigrid accelerated tensor approximation of function related multidimensional arrays. *SIAM J. Scientific Computing*, 31(4), 3002–3026.
- Kirby, M. (2001). *Geometric data analysis, an empirical approach to dimensionality reduction and the study of patterns*. New York: John Wiley.
- Kolda, T. G. (2001). Orthogonal tensor decompositions. *SIAM Journal of Matrix Analysis and Applications*, 23(1), 243–255.
- Kolda, T. G. (2003). A counterexample to the possibility of an extension of the Eckart-Young low-rank approximation theorem for the orthogonal rank tensor decomposition. *SIAM Journal of Matrix Analysis and Applications*, 24(3), 762–767.
- Kunisch, K., & Volkwein, S. (2002). Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. *SIAM Journal of Numerical Analysis*, 40, 492–515.
- Ortega, J. M., & Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*. New York/London: Academic Press.
- Oseledets, I. V., & Tyrtshnikov, E. E. (2009). Breaking the curse of dimensionality, or how to use SVD in many dimensions. *Siam Journal of Scientific Computing*, 31(5), 3744–3759.
- Oseledets, I. V., & Tyrtshnikov, E. E. (2010). TT-cross approximation for multidimensional arrays. *Linear Algebra and Its Applications*, 432, 70–88.
- Thomee, V. (1997). *Galerkin finite element methods for parabolic problems*. Springer, Berlin, Computational Mathematics 25.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3), 279–311.
- van Belzen, F., Weiland, S., & de Graaf, J. (2007). Singular value decompositions and low rank approximations of multi-linear functionals. In: *Proceedings of 46th IEEE CDC*.
- Weiland, S., & van Belzen, F. (2010). Singular value decompositions and low rank approximations of tensors. *IEEE Transactions on Signal Processing*, 58(3), 1171–1182.

## Author Biographies



**F. van Belzen** received both her Bachelors (2005) and Masters (2006) degree at the Control Systems Group Department of Electrical Engineering at Eindhoven University of Technology. She received her Masters degree with honors and is currently working towards her Ph.D. degree in the same group. Her research interests include model reduction of large-scale dynamical systems, general systems theory, tensor analysis and robust control.



**S. Weiland** is Professor at the Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology. He received both his MSc. (1986) and Ph.D. (1991) degrees in mathematics from the University of Groningen in the Netherlands. He was a postdoctoral research associate at the Department of Electrical Engineering and Computer Engineering, Rice University, Houston, USA from 1991 to 1992. Since 1992 he has been affiliated to Eindhoven University of Technology. His research interests are the general theory of systems and control, robust control, model approximation, modeling and control of hybrid systems, identification, and model predictive control. He was Associate Editor of the IEEE Transactions on Automatic Control from 1995 to 1999, of the European Journal of Control from 1999 to 2003, of the International Journal of Robust and Nonlinear Control from 2001 to 2004 and Associate Editor for Automatica from 2003 till 2006. He is currently board member of the Dutch Institute of Systems and Control.